

THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Understanding and Supporting Users in Visual Network Exploration

Mashael Hamad AlKadi

Doctor of Philosophy

Institute for Language, Cognition and Computation
School of Informatics
University of Edinburgh

2024

Abstract

Network visualization tools are used in numerous domains to explore data. Various network visualizations have been designed to explore data from different perspectives. However, few resources investigate how analysts create and interact with the visualizations to explore networks in the wild. By analysts, we denote users varying in their background and level of expertise. Being in the wild signifies that analysts work outside controlled settings where no pre-designed tasks are set. This thesis focuses on studying how to understand and support analysts in the process of network visual exploration. In such a process, analysts need to learn about the concepts, the tool(s), the processes, the visualization(s) and interactions, and the workflow. They also need to ensure that they can apply what they have learned accumulatively on their data toward their goal(s). Thus, to understand and support analysts, we have applied and collected data through mixed methods: interaction logging, mini-questionnaire, and visualization's state annotation, running an intensive 6-week course, designing an analytical dashboard, and implementing a coaching program. We ran our studies using the Vistorian a web-based tool that offers 4 types of interactive network visualizations.

This multimethod research led to the following contributions. Interaction logging allowed identifying 4 types of users based on their tool usage and advancement in the visual exploration process: demo users, data strugglers, single-session and multi-session explorers. To examine user types further, we designed a utility to capture and annotate visualizations' states, which we call bookmarks. We identified eight barriers that might cause analysts to struggle through the visual exploration process. We designed an analytical dashboard by specifying Key Performance Indicators (KPIs) and analyzing interaction logs accordingly. Those KPIs informed the assessment of the tool, the visualizations, the help resources, and the users' exploration. To support analysts, we designed a self-regulated guide for network visual exploration, which we call a roadmap. The roadmap describes step-by-step the processes of network visual exploration and associated activities. We also described 16 exploration strategies analysts follow, classified into three categories. We found 4 distinguished analyst groups based on how they aim to explore whether with/without research goals and/or data, which we call roadmap pathways. We evaluated the roadmap through a coaching program and found that it plays a crucial role in teaching networks visual exploration. Those findings have implications for the network's visual exploration through enhancing the design of its tools and associated educational efforts, mitigating barriers within, and supporting various user types.

Lay Summary

This research is about understanding how people use special visualization tools to help them see and understand complex networks of information. Imagine these networks as big webs of connected dots and lines, like a social network with friends and connections. These tools are like maps that show you the relationships between these dots and lines. They're used in many fields, like science and business, to make sense of this complex information.

But we wanted to know how people use these tools in the real world on their own, whether they have a specific goal to accomplish or just to learn the concept or a tool. So, we studied different types of users, some with more experience and some with less, to see how they use these network tools in everyday situations. To understand this, we used various methods like tracking how people click and interact with these tools, taking notes on important moments, teaching a course on how to use them, creating a dashboard to show important information, and making a step-by-step guide to using these tools effectively. In the end, our research is all about making these network tools easier to use for everyone. It's like helping people read and understand maps but for complex networks of information instead of roads and cities.

Acknowledgements

In the name of God ALLAH, the Most Gracious, the Most Merciful

Praise to Allah, who has guided us to this; and we would never have been guided if Allah had not guided us.

Firstly, I would like to thank my supervisors, Benjamin Bach, Uta Hinrichs, and Dragan Gasevic, for their unconditional support and guidance throughout my studies. Their advice and support over the past four years have been invaluable. I am truly blessed to have had you as my supervisors. I also want to thank Maria Wolters for her precious advice and feedback on both my studies and motherhood; they helped.

To my mother, Munira AlBassam, words will never be enough to express my deepest gratitude and love for always being there and believing in me. Thank you. My mother passed away before witnessing the fulfillment of my PhD requirements. May Allah shower her with mercy and reward her abundantly.

To my husband and companion, Yasir Alguwaifli, I love you, and we shall treasure our moments in Edinburgh, both studying for our PhDs and being parents for life.

To my sweet little children, Mohammed and Reem, I love you, and thank you for being patient while Mama was away for some time.

To my sisters Nada, Deya, Jawaher, Nourah, Hana, and Munira Adel, I love you. Without your support, this wouldn't have been possible. I cannot thank you enough.

To my father, Hamad ALKadi, and brother, Khalid AlKadi, thank you for your love and support.

To my best friends Maha, Wadha, and Razan, I cherish every word and the support you have provided me.

To my nephews, nieces, and my whole family, I love you, and thank you.

To Mohammed, Fahad, Nawaf, Abdulazeez, Hamad, Saad, Ibrahim, and Faisal.

To Noura, Moudi, Munira, Leen, Dina, and Lamees.

To my extended family: Uncle Mohammed, Auntie Munira, and Lubna.

To Auntie Lulu and Reem, may Allah have mercy on your souls. We love you and miss you.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

The work presented in chapter 4 was published in IEEE Transactions on Visualization and Computer Graphics 2022 as ”Understanding Barriers to Network Exploration with Visualization: A Report from the Trenches” by Mashael AlKadi (Primary Investigator), Vanessa Serrano (Co-author), James Scott-Brown (Co-author), Catherine Plaisant (Co-author), Jean-Daniel Fekete (Co-author), Uta Hinrichs (Co-supervisor) and Benjamin Bach (Primary Supervisor). As the primary investigator, I collected and analyzed interaction logs with the Vistorian tool, conducted interviews, administered surveys, and made session observations. I also prepared course content, led weekly sessions, provided technical support, and collaborated with my supervisors to thematically analyze interview data.

chapter 5 was published in IEEE Transactions on Visualization and Computer Graphics (2023 IEEE Visualization and Visual Analytics (VIS)) as ”Show Me My Users: A Dashboard Visualizing User Interaction Logs” by Jinrui Wang (co-author handled the design of dashboard pages), Mashael AlKadi, and Benjamin Bach (Primary Supervisor). I analyzed the collected interaction logs to derive metrics for evaluating the performance of the visualization tool, examined their impact on improving the design of visualization tools, and contributed to the design of the dashboard.

chapter 6 is under submission as ”Roadmaps, Pathways, and Activities: Helping Analysts Understand, Plan, and Personalize their Visual Network Exploration” by Mashael AlKadi (Primary Investigator), Magdalena Boucher (co-author and assisted in digitalizing roadmap design), Wolfgang Aigner (co-author), Uta Hinrichs (Co-supervisor) and Benjamin Bach (Primary Supervisor). As the primary investigator, I designed the roadmap and prepared each step’s content including videos, slides, some examples, and strategies description cards. I ran the coaching program, collected data, and analyzed it from various resources such as observations from coaching sessions and 1-to-1 exit interviews with the participants.

Please note that I have included an authorship statement at the beginning of each chapter to clearly differentiate my contributions from those of my colleagues. I use the pronoun ”I” to describe my individual work and ”we” to refer to any joint or collaborative efforts.

(*Mashael Hamad AlKadi*)

Table of Contents

1 Introduction **1**

1.1 Overview 1

1.2 Research Questions 3

1.3 Contributions 4

1.4 Thesis Organization & Outline 5

1.5 Publications 6

1.6 Definitions & Concepts 7

2 Literature Review **9**

2.1 Network Visualization Concept and Types 9

2.2 Network Visualization Tools 12

2.2.1 Network Modeling for Visual Exploration Purposes 13

2.2.2 Data Wrangling in Networks Visualization Tools 14

2.3 Evaluating Networks Visualization Tools 15

2.3.1 Approaches and Metrics for Network Visualization Tools Evaluation . . 17

2.3.2 The Visual Exploration Process as the Evaluation Focus 19

2.3.3 Insight-based Evaluation 20

2.3.4 Evaluating by Interaction Logs Analysis 21

2.3.5 Learning Analytics in Interactive Visualizations 22

2.4 Mental Model & Knowledge Generation in Visual Analytics 24

2.5 Network Visualization Literacy, Education and Guidance 27

2.5.1 Networks Visualization Literacy 27

2.5.2 Guidance through Visualization Creation and Exploration 30

2.5.3 Designing Educational Guides for Visualization Purposes 32

2.5.4 Educational Activities for Visualization Purposes 36

2.5.5 Supporting Self-Regulated Learning in Educational Efforts 38

3 The Methodology **41**

3.1 The Vistorian: The Network Visualization Tool 44

3.1.1 Visualizations and Interactions . . . 44
3.1.2 The Vistorian Visualizations and Data Encoding . . . 48
3.1.3 Network Schema Specification Requirements . . . 48
3.1.4 Network Creation, Data Import Interface & Wizard . . . 49
3.1.5 Types of Help & Support Resources Provided in the Vistorian . . . 51
3.1.6 The Support of Temporal and Geographical Data Exploration . . . 53
3.2 First Approach: Logging Analysts' Interactions . . . 53
3.2.1 Interaction Log Events Description . . . 55
3.2.2 Interaction Log Events Categories . . . 56
3.2.3 Participants' Demographics, Data Collection and Analysis . . . 61
3.3 Second Approach: Mini-Questionnaire and Network Visualization State Annotation . . . 63
3.3.1 How the Bookmarks Work and Support Visual Exploration . . . 63
3.3.2 Participants' Demographics, Data Collection and Analysis . . . 67
3.4 Third Approach: Course . . . 68
3.4.1 Networks Visual Exploration Course Content . . . 68
3.4.2 Participants' Demographics, Data Collection and Analysis . . . 69
3.5 Fourth Approach: Analytical Dashboard . . . 70
3.6 Fifth Approach: Coaching Program & Roadmap . . . 71
3.6.1 Designing A Self-Regulated Guide for Networks Visual Exploration . . . 72
3.6.2 Participants' Demographics, Data Collection and Analysis . . . 73

4 Understanding Barriers to Network Exploration with Visualization: A Report from the Trenches 75
4.1 Introduction . . . 76
4.2 Background . . . 78
4.3 Network Visualization with The Vistorian . . . 80
4.4 Study 1: Interaction Logging & Workshops . . . 81
4.5 Study 2: Network Exploration Course . . . 82
4.5.1 Data Import Wizard . . . 83
4.5.2 Course Outline . . . 83
4.5.3 Participants . . . 84
4.5.4 Data Collection & Analysis . . . 85
4.6 Participants' Goals & Course Journeys . . . 86
4.6.1 Initial Participant Goals . . . 86
4.6.2 Participant Course Journeys . . . 87
4.7 Observations . . . 88

4.7.1 Defining Network Structures . . . 88
4.7.2 Iterating on the First Visualization . . . 92
4.7.3 Visual Exploration . . . 92
4.8 Steps in Network Exploration . . . 93
4.9 Barriers to Network Exploration . . . 94
4.10 Discussion . . . 98
4.11 Conclusions . . . 100

5 Show Me My Users: A Dashboard Visualizing User Interaction Logs 103
5.1 Introduction . . . 104
5.2 Background and Related Work . . . 105
5.3 Analysis . . . 106
5.4 Dashboard Design . . . 108
5.5 Findings from Dashboard in Use . . . 112
5.6 Discussion & Future Work . . . 113

6 Roadmaps, Pathways, and Activities: Helping Analysts Understand, Plan, and Personalize their Visual Network Exploration 115
6.1 Introduction . . . 116
6.2 Background . . . 118
6.2.1 Guidance for Understanding Visualization . . . 118
6.2.2 Activities and Educational Material . . . 119
6.2.3 Network Visualization Literacy and Training . . . 120
6.3 Strategies & Pathways . . . 121
6.3.1 Exploration Strategies . . . 121
6.3.2 Pathways . . . 124
6.4 Resources: Roadmap, Materials, & Activities . . . 126
6.4.1 Roadmap Design process . . . 126
6.4.2 Roadmap Design . . . 128
6.4.3 Exploration Steps and Activities . . . 128
6.4.4 Design of Learning Resources . . . 134
6.5 Coaching Program Study . . . 136
6.5.1 Coaching Program & Coaching Sessions . . . 136
6.5.2 Coaching Program Participants . . . 137
6.5.3 Data collection & Analysis . . . 138
6.6 Findings . . . 138
6.6.1 Coaching Sessions and Personal Success Stories . . . 139

6.6.2 Roadmaps and their Contribution to Visual Network Exploration 139
6.6.3 Strategies' Impact on Visual Exploration 140
6.6.4 Design of Activities & Learning Resources to support self-regulated Learning 141
6.6.5 The Role of Logging Observations and Reflecting on the Exploration Process 142
6.6.6 Validation of Strategies and Trust of Visual Exploration in Networks . . 143
6.7 Discussion 143
6.8 Conclusion 147

7 Discussion, Future Work & Conclusion 149
7.1 RQ(1): What do analysts struggle with when exploring visualizations? 149
7.2 RQ(2): How to capture analysts' exploration in the wild? 150
7.2.1 Metrics to Evaluate Visual Exploration Process 150
7.2.2 Metrics to Evaluate Networks Visualization Tools 152
7.3 RQ(3): How to enhance visualization tools to support visual exploration? . . . 153
7.4 RQ(4): How to characterize the visual exploration process? 154
7.5 RQ(5): How to prompt self-regulated learning of visual exploration of networks? 155
7.6 Reflection on the Findings in Visualization Research 158
7.6.1 Generalization of Findings on Visualization Tools 158
7.6.2 Views and Recommendations on the Evaluation Methods for Visualization Researchers 162
7.6.3 Toward Explainable Guidance in Networks Visual Exploration 166
7.7 Future Work 168

A Supporting Material of Barriers Network Visual Exploration Study 171
A.1 One-to-One Interview Questions: Reflection on the Course and the Vistorian Tool 171
A.1.1 General Feedback on the Course 171
A.1.2 Individual Features of the Vistorian 172
A.1.3 Use of the Vistorian in Comparison to Other Tools 173
A.2 Participants Profiles 173
A.3 The Story of a Journey 175
A.4 6-Weeks Course Outline 176

B Analytics Dashboard Pages 179
B.1 Overview Page 179
B.2 Visualizations Page 180

B.3 User Exploration Page 181

C Resources of Networks Visual Exploration Roadmap 187

C.1 What is Networks Visual Exploration? 187

C.2 Concept Map Sketching 196

C.2.1 Understanding your Data and Sketching Concept Maps (Starting from Data) 196

C.2.2 Understanding your Data and Sketching Concept Maps 214

C.3 The Selection and Description of Networks Visual Exploration Strategies . . . 227

C.3.1 Data-Driven Exploration Strategies 227

C.3.2 Open-Ended Exploration Strategies 230

C.3.3 Task-Driven Exploration Strategies 233

C.4 Network Schema Creation 240

C.5 Which visualization to read and How? 257

C.6 Logging & Validating your Observations 294

C.7 Data Transformation and Wrangling 305

C.8 Visual Exploration Walkthroughs 310

C.8.1 Networks Visual Exploration Roadmap for Data Collectors 310

C.8.2 Networks Visual Exploration Roadmap for Concept & Tool Learners . 328

C.8.3 Extra Practice Example(Students Dataset) 341

D Coaching Program Feedback Questions 355

E Earlier Sketches of Roadmap 359

F Designing Educational Material for Networks Visual Exploration 361

F.1 How to Design a Roadmap for Visual Exploration Purposes? 362

F.2 How can we support the Knowledge Generation Model when Designing Network Visualization Tools and Guides? 363

F.3 Bloom's Taxonomy and Knowledge Acquisition in Networks Visualization . . 364

F.4 How do our Findings fit with the Knowledge Generation Model? 367

Bibliography 371

Chapter 1

Introduction

1.1 Overview

Network visualizations play a major role in networks analysis [150], information exploration [271, 65], discovery [241], and explanation [51]. Network visualizations are highly used in various domains (e.g., social sciences, biology, business) dealing with various data sizes and scales [202]. It allows analysts to investigate different aspects of their data through the visual representation of a network. A network is a set of connected data elements that are known as a set of nodes and edges. Networks allow having additional data describing nodes and/or links such as geography, time, types, directionality, and others, which makes them intrinsically hard to understand. Thus, network visualizations are designed to enable a better understanding of data based on the type of data available and the researcher's goal(s).

Network visualizations have evolved through time into various types, offering different aspects of the data to be investigated. In the beginning, networks were explored using node-link diagrams. This enabled exploring aspects related to connectivity, such as node degrees, central nodes, layouts, and others. Then, other visualizations were designed, such as adjacency matrices, to allow further pattern discovery (e.g., clusters), especially for dense networks. Further aspects of networks arise due to the dynamic nature of networks [31, 222]. For example, timeline visualization allows exploring the dynamic changes of a network through time. On the other hand, large scale data represented a challenge for researchers on how to visualize and explore. Thus, methods were proposed to have a clearer view of the data. For example, how to declutter views based on the network dimension [207].

To facilitate network visualization and exploration, tools were developed to visualize networks in general (e.g., Gephi [20]) or for specific purposes (e.g., Cytoscape [240] for visualizing molecular interaction networks). In addition, software packages were designed (e.g., NodeXL [247]), and libraries were written (e.g., R, Python NetworkX [115]). Some of those tools produce static visualizations, whereas some offer interactive visualizations (e.g., Vistorian [10]). However, it is neither an easy nor a straightforward task to decide on which tool to use in order to address a specific research goal or to learn and master a tool [42]. Each tool offers its own exploration experience based on its features, its visualizations, and the tool's workflow. Thus, researchers aimed to understand how analysts work and cope with such tools.

Network visualizations can be used to explain and explore data. Explaining data means selecting network visualization(s) to represent certain analytics or facts. Visual network exploration, on the other hand, means the ability to investigate research goals and data through visualizations. Visual data exploration holds its own challenges [236]. That is, initially, analysts do not know what they are looking for yet [259], and are not able to determine which parts of the data are relevant to the task to be accomplished [236]. Those challenges upscale once we discuss large graphs, where the analysts need to decide what to represent to understand the data better. For tool developers, it is challenging to understand what makes some tools more common, usable, and supportive of the exploration process than others. As a result, the exploration process of networks has been investigated, focusing on various aspects such as: exploration tasks (e.g., data wrangling), data domains (e.g., health), exploration goals (e.g., temporal changes), and/or certain user types (e.g., novices).

The process of visual data exploration empowers analysts with the ability to question more about their data. This includes examining the data model to understand how elements are interconnected, assessing whether the data is dense or sparse, identifying the types of data values and outliers, determining what kinds of questions the data can answer, and uncovering potential insights. The ability to answer such questions depends on how data is represented, visualized, and explored as a network. That is, creating different network schemes (e.g., selecting different nodes and links) and visualizations (e.g., adjacency matrix, network over map) from the same data set can address different research questions and goals. Also, tools should support analysts by creating additional views (e.g., coordinated views [65]) and interactive features (e.g., ego-network exploration with X-ray mode [122]). In addition, analysts, when needed, should be supported with guidance to accomplish various exploration tasks. Schulz et al. [236] define guidance as *"to actively guide the user, be it in creating visual representations for different types of tasks or in navigating to different regions of interest"*. They also define it as *"the goal is to incentivize users more subtly to carry out steps that lead to promising results"*. The success of visual exploration of networks is determined by the analyst's ability to create suitable

network schemes and visualizations, capture and interpret visual patterns, tweak data to support exploration goals, ask the proper questions, and verify the results [264]. But how can we understand and examine such a process?

Our knowledge about how such tools are being used outside controlled and lab settings [65] is very limited. For example, how do analysts use such tools to explore their own data? what kind of obstacles do they face? which features do they use mostly? And how to support analysts with proper guidance? Most studies were conducted in controlled studies with preset tasks and/or on specific data sets. They tracked participants' actions through various types of data collection approaches (e.g., video recordings and application events logs). This thesis focuses on understanding how analysts explore their data in *the wild*. By analysts in the wild, we mean users coming from different backgrounds (e.g., domains) and levels of expertise (no/limited expertise), who are exploring their own data using network visualization tools outside controlled settings (e.g., labs). We focus on investigating different users needs and challenges (chapter 4, chapter 5), finding how to capture what they are doing (chapter 3), understanding the process they follow and the decisions they make (chapter 5, chapter 6) until they reach satisfying results (chapter 7).

To accomplish the goals of this thesis, we ran our studies following the methodology described in (chapter 3). We conducted our studies using the Vistorian, an online platform for interactive network visualizations (section 3.1). The Vistorian is *selected* as the tool for our studies for the following reasons : (1) it is an open-source accessible web application that does not require any installation, (2) it does not require any coding at any stage, and (3) it supports the network visual exploration process through its interactive visualizations, views, and features. Those reasons are essential to study network visual exploration as a process as they: (a) isolate any prerequisites that might hinder the analysts from using the tool (e.g., requires coding skills), (b) accessible for both in-person and remote settings, (c) allows study conductors to support remote users needs (e.g., through logs),(d) easy to use with no/simple training, and (e) offers adaptive and flexible ways to explore data. A full description of the Vistorian can be found in section 3.1.

1.2 Research Questions

To the best of our knowledge, our studies are the first to trace the analysts' usage outside controlled settings while they use their own data. Hence, this thesis addresses and aims to answer the following research questions:

- RQ(1): What do analysts struggle with when they use visualization tools to explore their network? **[Barriers]** (chapter 4)

- RQ(2): How do we capture what analysts are doing through their exploration process in the wild? **[Methods]** (chapter 3)
- RQ(3): How can we enhance visualization tools design to support analysts in their visual exploration of networks? **[Interventions]** (Dashboard chapter 5 & Roadmap chapter 6)
- RQ(4): How can we characterize this exploration process as a matter of interactions and strategies followed by analysts? **[Strategies]** (chapter 6)
- RQ(5): How can we teach or prompt self-regulated learning of visual exploration of networks through the tool? **[Education]** (chapter 6)

Through this thesis, we examined several *methods RQ(2)* aiming toward a better understanding of the analysts' visual exploration process of networks. Initial methods (interaction logging & 6-week course) enabled us to identify *barriers RQ(1)* analysts face in such a process as well as user types. An analytical dashboard presented a *intervention RQ(3)* that better understood analysts' needs. Methods such as the coaching program and course *RQ(2)* enabled us to identify *strategies RQ(4)* and pathways analysts follow in their exploration process. We used our findings to formalize the network visual exploration roadmap as an *educational RQ(5)* guide and *intervention RQ(3)*. The roadmap steps present an *educational RQ(5)* approach to support self-regulated learning of networks visual exploration.

1.3 Contributions

This thesis contributes to the literature which focuses on studying visual exploration of networks. The outcomes of this thesis can be summarised as follows:

- Identification of network visualization tools **user types** (chapter 4): demo users, data strugglers, single-session explorers, and multi-session explorers. The identification process relied on data acquisition and the user return rate to the tool.
- Identification of **barriers** (chapter 4) analysts face in the visual exploration of networks process of their data.
- Defining learning analytics in the scope of network visual exploration in the form of evaluation **metrics and Key Performance Indicators (KPIs)**. And how this led to outline the **dashboard design pipeline** (chapter 5).
- The description and classification of **strategies** analysts follow in network visual exploration (chapter 6).

- The design of **network visual exploration roadmap** (chapter 6). This includes the **steps** analysts follow and the **pathways** that distinguish them.

- The **methods** allows capturing analysts' exploration process and how they complement each other (chapter 3).

The studies and findings of this thesis have contributed to understanding how Vistorian users are using the tool, what usability issues they face, and how the tool design can be enhanced to support visualization (e.g., change the data upload process within the Vistorian).

1.4 Thesis Organization & Outline

This thesis contributes to the literature that focuses on studying network visualization tools usage, including the following:

- In chapter 2, we review *background and literature* around network visualizations, network visualization tools, and how to understand analysts' exploration process via them. We also review the knowledge generation model, how analysts make sense of data, and how guidance can support analysts in their network visual exploration. In addition, we describe to which extent learning analytics have been used in such a context.

- In chapter 3, we outline the *methodology* followed in all of our studies and how one has led to another. This includes the decisions taken on the level of the tool selection (The Vistorian), the study outline, and the material prepared in each study.

- In chapter 4, we report on our earlier findings based on our first two methods (interaction logging and a 6-week course). Firstly, the interaction logging method allowed for the identification of two major processes where analysts face obstacles, which are the process of data formation and the process of visually exploring network visualization. Thus, a 6-week course on network visual exploration has been conducted to investigate more about those obstacles. Through this course, we presented techniques on how to form network data to support the visual exploration process and how to explore different network visualizations visually. As a result, we describe the nature of challenges users face through their visual exploration of networks framed as *barriers*. We also describe the **user types** we found in visual exploration of networks.

- In chapter 5, we report on how we used the interaction logs collected through our studies to understand better analysts' exploration process and their needs based on user types. User types identified in our second study (chapter 4) helped us in studying users' needs

in-depth rather than treating all users as a single user type. As a result, we were able to identify metrics and formulate KPIs to measure users, usage tools, and visualizations. Eventually, a dashboard was co-designed with Jinrui Wang to reflect a live demonstration of analysis and evaluation metrics.

- In chapter 6, we report on the design of networks' visual exploration roadmap. This includes the discussion of the steps the analysts follow in their visual exploration process. The strategies users choose and follow in their visual exploration process of networks. In addition, the different paths that distinguish users' exploration experiences. And what sort of material and activities support analysts in their exploration process.

- In chapter 7, we discuss the outcomes of all studies and how they contribute to our understanding of the users' exploration process. In addition, we describe limitations and future work in the visual exploration process.

1.5 Publications

The outcomes and contributions of the studies and analysis we conducted in this thesis are submitted and/or published as follows:

- Mashael AlKadi, Magdalena Boucher, Wolfgang Aigner, Uta Hinrichs, and Benjamin Bach "Roadmaps, Pathways, and Activities: Helping Analysts Understand, Plan, and Personalize their Visual Network Exploration", IEEE Transactions on Visualization and Computer Graphics (Under Submission)

- Magdalena Boucher, Mashael AlKadi, Benjamin Bach, and Wolfgang Aigner "Exploring Instructional Comics for Data Visualization Education", IEEE Transactions on Visualization and Computer Graphics (Under Submission)

- Jinrui Wang, Mashael AlKadi, and Benjamin Bach (VIS 2023) "Show Me My Users: A Dashboard Visualizing User Interaction Logs", IEEE Transactions on Visualization and Computer Graphics.

- Mashael AlKadi, Vanessa Serrano, James Scott-Brown, Catherine Plaisant, Jean-Daniel Fekete, Uta Hinrichs, and Benjamin Bach (VIS 2022) "Understanding Barriers to Network Exploration with Visualization: A Report from the Trenches", IEEE Transactions on Visualization and Computer Graphics, vol 20, 907-917.

1.6 Definitions & Concepts

Network Visual Exploration is the process of visually exploring data as a network through one or more network visualizations. The exploration process includes looking into various aspects such as connectivity, layout, density, dynamic changes, and others.

Analysts is the terminology we use to describe tool users with/out data sets and/or research questions and goals.

Novice analysts, when mentioned, are network visualization tool users who have *no or limited experience* in creating network visualizations and exploring them.

Using a tool in **the wild** when mentioned indicates that the study was conducted outside controlled settings (e.g., labs). In such an environment, participants or tool users are not contacted, and no specific tasks are set or given. People participate in such studies voluntarily.

The Vistorian (`https://vistorian.net/`) is a state-of-art network visualization tool. It is an open-source tool that offers four different network visualizations in addition to a coordinated view. Further description of the tool can be found in (section 3.1).

A **Network Schema** describes the network nodes, links, and additional information. A network to be created needs a schema that at least contains two attributes which are: node1 (source node) and node2 (target node) where one or more links exist between them. Additional attributes can describe nodes and/or links[1].

[1] P.S. Attributes specification - basic or additional - can vary from one tool to another.

Chapter 2

Literature Review

In this chapter, we discuss the background of network visualizations, the network visualization tools, and how they are used in visually exploring networks. We start by describing the main concepts related to the creation and interpretation of network visualizations (section 2.1). Then, we describe network visualization tools and what they provide analysts with (section 2.2). In (section 2.3), we discuss how network visualization tools are studied and evaluated. In (section 2.4), we discuss the knowledge generation and mental model behind the visual exploration process. That is, such concepts form the backbone of how network visualization tools and educational guides can support the visual exploration of networks. So, in (section 2.5), we discuss the background of network visualization literacy, the nature of guidance available to support the visual exploration process, and how to support it further through educational guides.

2.1 Network Visualization Concept and Types

A network is a set of nodes(actors) and links(relationships) that represent how data elements are connected. Network visualizations *"translate network data into a visual representation of some combination of the actors, relationships, clusters, and data attributes"* [282]. A **network schema** describes the network nodes, links, and additional information (e.g., link/node type, link weight). A network schema needs at least two attributes, which are node1 (source node) and node2 (target node), where one or more links exist between them. The network schema specification is crucial for analysis and exploration purposes. Different schemes of the same data set can answer different questions and offer different insights. The network schema can be created based on a *node table* or a *link table*. A **node table** is a table where each row describes a node with all possible links represented as columns. A **link table** is a table where each row describes a single link between two nodes. A node can be described by several attributes such as its label/name, node's icon, node's location, and node's type. A link can be described by

various attributes, such as the nodes it connects, the link type, the link weight, the date the link originated, the node's location, and others. However, tools vary in the number and type of attributes supported in their tools. The analysis and exploration of networks are applied in various domains such as healthcare, marketing, history, social sciences, and cyber-security. Several tasks are associated with multivariate network analysis and exploration [208]. Tasks in network analysis are classified into four categories, which are structure-based tasks (e.g., find clusters that are regions of highly connected nodes within a network[156]), attribute-based tasks (e.g., find nodes with specific properties), browsing tasks (e.g., follow the path), and estimation tasks (e.g., overview trend) [156, 208].

Networks can be analyzed and explored considering several dimensions [58] such as pattern and knowledge discovery, scalability, information integration, and visualization. The *pattern and knowledge discovery* discusses the capacity of algorithms, methods, and techniques to extract knowledge. *Scalability* discusses the limit and capacity of algorithms, frameworks, and methods and the data size they can work with. *Information integration* discusses the extraction and fusion of data from different sources and types. The *visualization* discusses how and what we can show through visualizing and exploring data, which is our focus in this thesis. For example, Figure 2.1 shows an image of major visual patterns a network node-link diagram provides, such as isolated (disconnected) nodes, paths (from one node to another), and clusters. Node-link diagrams present the layout of a network's data in matters of connectivity. Different layouts can be used to create a node-link diagram, such as force-directed layouts [99], which simulates physical forces between nodes. Other layouts include hierarchical layout [8, 184], which divides the graph into layers revealing the parent-child relationships, and orthogonal layouts [90] where links are plotted only along vertical and horizontal axes. Thus, different layouts can reveal the underlying structure of networks in various ways. For example, clusters in a force-directed layout, parent-child relationships in a hierarchical layout, cyclic patterns in a circular layout, and nodes' spatial distribution in a grid layout. Each layout has its strengths and limitations, and the choice of layout depends on the specific characteristics of the data and the insights that need to be conveyed.

There are many types of *network visualizations*. Some visualizations are more common such as node-link diagrams [23, 127] and adjacency matrix [121, 139] (see 1st and 2nd visualizations from left Figure 2.2). Both node-link diagrams and adjacency matrices emphasize link structure and connectivity. However, adjacency matrix [30] are good in visualizing clusters, whereas node-link diagram supports representing paths and layout. In addition, other visualizations have been developed to represent and explore different aspects of networks. For example, timeline visualization was designed for temporal exploration and examining network changes through time (see 3rd visualization from left Figure 2.2). For example, how clusters form or connect and

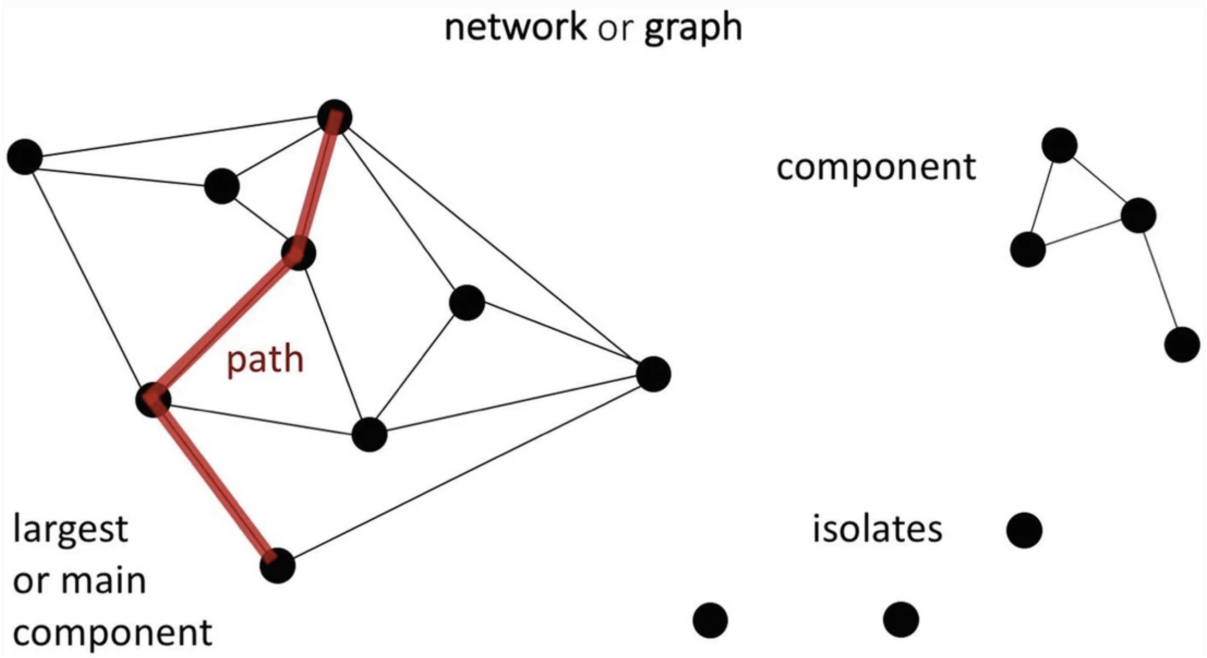


Figure 2.1: Basic Graph/Network Concepts and Network-related Terminology [282]

how central nodes change over time. Network over-map visualizations illustrate the geographical distribution of nodes [233] (see 4th visualization from left Figure 2.2). Different efforts are made to emphasize specific aspects of a graph. For example, PivotGraph by Wattenberg [271] focuses on the relationship between node attributes and connections rather than the global graph topology. Another example is hive plots [151] that were designed to represent a rationalized drawing of networks. Graphs can be explored in different forms, ranging from social networks to Markov chains [271]. This helps simplify multidimensional comparisons and answer quantitative questions (e.g., which node type has more connections?).

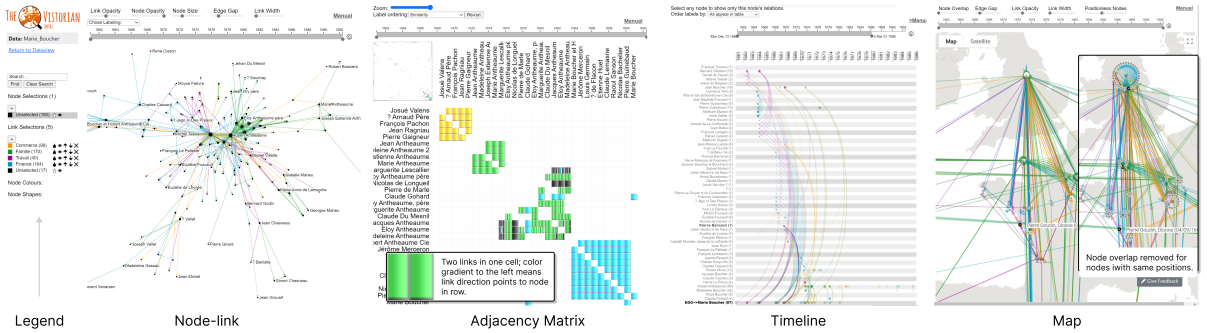


Figure 2.2: Example of a Network Visualizations

Network Modeling is the process of creating the mathematical, computational, and visual representation of the network's structure, data, and characteristics. [270, 99]. The network model is strongly influenced by the data structure underlying it. Thus, several network models have been proposed, including simple networks (e.g., co-occurrence networks) and complex networks

(e.g., multilayer networks [144], multivariate networks [165], knowledge graphs [128]). Simple networks allow for the investigation of specific and direct research questions. Whereas complex networks require an in-depth analysis to investigate various variables in the network. Various methods were designed to model networks to reflect their underlying nature. For example, social network analysis (SNA) is a method to model the social relationships between a set of entities into a network [139, 197, 125]. Bipartite networks to model relationships between two types of nodes. Multivariate networks to represent nodes and links associated with various properties [271]. Furthermore, network models are continuously studied to support analysis from various perspectives such as: *weighted networks* to reflect the importance of links/nodes, *multiplex networks* to represent mixed relationships, and *dynamic networks* to represent the changing nature of network components.

Networks can be analyzed and explored from different aspects [58] such as structural analysis [270] that includes investigating nodes' centrality in a network, connectivity, and interactions(e.g., communities detection), network topology and layout (e.g., sparse vs. dense). For temporal analysis, networks can be explored for their dynamic changes through time. For spatial analysis, networks can be explored over maps. Networks can be analyzed as a whole or based on a specific node's network, called *ego-networks*. Ego networks are special forms of networks that highlight the close neighbors of a specific node. Thus, networks not only can be analyzed based on their nodes and links but also on a large number of variables and aspects - if they exist -.

2.2 Network Visualization Tools

As we have reviewed the main concepts in network visualizations, we discuss network visualization tools and how they comply with and support the exploration of networks. Due to the nature of network data, visualization tools face a set of challenges across different interaction paradigms [227, 253, 124]. Those challenges [177] include the variance of data types in a network (e.g., geographical-data, temporal), the network type (e.g., multivariate), the different networks that can be extracted and formed from the same data (e.g., projected network), the size of the network and the impact of the data field and domain. All challenges collectively contribute to and reflect the complexity and requirements of the network's data. In addition, networks require having different visualizations to examine data from different perspectives. There are dozens of network visualization tools available, each working to offer more features to meet the needs of analysts [42].

Several tools have been developed to visualize and explore networks in *different forms* such as: *standalone applications* (e.g., Gephi [20], Honeycomb [259], PivotGraph [271], Pajek [21],

Krackplot [148], Palladio [131, 73], Cytoscape [240] and UCINET [225]), *applications extensions* (e.g., NodeXL [115, 248]), *web-applications* (e.g., The Vistorian [17], *graph databases* (e.g., Neo4j [153, 94]) or *libraries* (e.g., D3 [182], JUNG [193]). Network visualization tools have evolved through time from offering only static visualization to generating dynamic interactive visualizations. Tools' interactivity is provided through interactive features, such as zooming, filtering, highlighting, grouping nodes, switching between views, etc. For example, by interacting with the time-slider feature, the analyst can change and examine the level of data displayed based on time [259].

Network visualization tools vary in the experience, *analysis and exploration capabilities* they offer to users, and in their learn-ability level. For example, graph databases allow querying and visualizing networks [256]. Some tools allow analysts to explore networks by offering different views and features that represent metrics of social network analysis (e.g., Perer and Shneiderman's work [197]). For example, Gephi [20] is a cross-platform interactive application that allows adjusting the graph's layout, size, appearance, and other functionalities. Another example is Tableau, a standalone application that allows analysts to choose values and place them on the graph from a created database. Tableau is based on VizQL [116], which is a language for query, analysis, and visualization. Thus, Tableau extends its capabilities by supporting analysts with functionalities such as sort and filter values [123]. Some tools support visualizing large graphs, such as Pajek [82] and Netminer [78]. Those tools support the exploration process with several functionalities, such as extracting sub-graphs, choosing analysis algorithms, and various visualizations. Thus, the power of network visualization tools increases by revealing the network's characteristics, allowing several networks to be derived from the original network data, and offering harmonized views to support exploring networks from different angles.

2.2.1 Network Modeling for Visual Exploration Purposes

One of the network visualization tasks is to model the network data. Data modeling for network visualization or what we can call *network modeling* is a challenging task due to some factors contributing to its complexity. Those factors include choosing nodes and links, specifying multi-condition links and multilayer networks, and deciding on the optimal data structure that serves the goal of the network's visual exploration process. Here, we mention some of the tools that offer network modeling. For example, Orion [124] presents a unified model and workflow language for interactive modeling, transformation, and visualization of network data. Orion allows for defining, filtering, aggregating, and statistically analyzing networks based on proposing possible multiple network models and linking conditions between user-selected nodes. Ploceus [163] is a tool that enables analysts to extract network models and automate the generation of networks from tabular data.

The first step in network visualization tools is to create the network model by specifying its schema. So, how do the tools support such a task? Some tools allow analysts to create a network through visual drawing rather than defining its schema. For example, Ploceus [163] and Gilbert and Auber's [104] graph database allows analysts to create multiple network schemes through drag-and-drop of attributes as nodes. Graphiti [253] allows analysts to examine possible network options using a demonstration-based interaction technique. That is, by offering various network models by demonstrating a subset of the expected network to the tool to simulate. VISAGE [203] allows users to retrieve different sub-graphs based on textual query. Some tools, such as Data Wrangler [138], recommend the generalization of analysts' edits to reach the desired network model. However, most of those tools have not been evaluated in the wild. Also, they do not offer analysts the means to compare and select among different network models.

The workflow of network visualization tools is usually driven by analysts' needs. However, workflows in network visualization tools vary due to the different designs and infrastructures they have. The design of tools workflows impacts not only the usability-related issues of the network creation process but also how it can be interpreted and explored by the analyst.

2.2.2 Data Wrangling in Networks Visualization Tools

One key feature in network modeling and visualization tools is to offer data wrangling operations. Data transformation or wrangling is the process of changing data representation by applying specific methods and techniques. For example, *network abstraction* is a technique used in describing a network's data to reduce the network size in large graphs [259]. That is by applying different techniques such as sampling and clustering. Making a decision about the level of data represented and how abstraction is applied can be challenging due to the varying considerations involved.

Thus, network data wrangling and modeling tools are at the early stages where they only offer data wrangling partially, such as Graphiti [253], D-Dupe [37], and Ploceus [163]. D-Dupe [37], for example, is a tool that assists in finding and dealing with duplicates. Ploceus [163] provides analysts with both data wrangling and network visualization on the same page. That allows analysts to examine any changes in their network schema and model directly. On the one hand, this confirms what Bartem et al. [18] found, that some participants with non-numerical data use other tools to check their tabular data. On the other hand, it shows that specific domains require certain types of data wrangling operations, which have not been investigated thoroughly in network visualizations yet [140].

Specific tools for Data wrangling offer data transformation to clean and prepare data. For example, Origraph [35] and Mure.js [34] support visual data wrangling of networks. Kandel et al. [138] offered data cleaning and integration through programming-by-demonstration. There

are other tools that offer advanced mining algorithms accompanying the data manipulation process, such as KNIME [146]. Overall, *combining data wrangling and network modeling in network visualization tools* will push toward a better analysis and visual exploration experience, especially in the wild.

2.3 Evaluating Networks Visualization Tools

Network visualization, modeling, and data wrangling have been offered through various tools. Network visualization tools differ in their features, workflows, and functionalities. Thus, there are several studies investigating and comparing between those differences. Well-designed tools are known, for example, to have the error-free performance of common tasks and to provide creative intuitive interfaces [244]. Visualization tools are evaluated through several methods and settings, including controlled user study, classroom-based evaluation, crowd-sourced evaluation, a meta-review of related literature, and in the wild [95]. However, most evaluation methods require specific tasks to be designed or set, except for *in the wild*. Evaluating tools in the wild simulates real-world scenarios where analysts explore their data in their own way and rely on their skills to reach their goals. Firat et al. [95] compared the "in the wild" method to other approaches as Figure 2.3 shows. We can see that the "in the wild" method is distinguished by having the ability to capture observations from a medium distance without being in a strictly controlled environment. Statistically, they found few studies conducted in the wild: only 4 papers out of 34 [95]. Among these 4 papers, 3 targeted adults. Two papers assessed the reading and interpretation of visualizations, while the third one evaluated the visualization construction process. The fourth one by Peppler et al. [196] targeted children's literacy with reading and constructing visualizations. Their findings address design practices to aid data visualization literacy in museums.

Evaluation method and characteristics	Physical distance	Proximity of the observations	Control over the environment	Number of participants
In the wild	Medium	Medium	Medium	30–400
Controlled user Study	Close	Close	High	10–180
Classroom-based user study	Close	Medium	Medium	10–50
Crowdsourcing study	Far	Distant	Less	30+
Literature review	N/A	N/A	N/A	N/A

Figure 2.3: A summary of the evaluation methods and different characteristics of the classification categories by Firat et al. [95]

Supporting the use of visualization tools *in the wild* and specifically network visualizations require further investigation [64]. Chamberlain et al. [64] describe "in the wild" methods as *"observing and recording what people do and how this changes over suitable periods of time"*. And Roger et al. [221] define the goal of "in the wild" approaches as *"understanding how technology is and can be used in the everyday/real world, in order to gain new insights about how to engage people/communities in various activities, how peoples lives are impacted by a specific technology, and what people do when encountering a new technology in a given setting"*. Thus, they discuss case studies, methodologies, and practical and ethical considerations involved in-the-wild research. The authors argue that such type of research allows for the evaluation of new technologies and interactions as they are integrated into everyday life, potentially leading to more innovative design solutions.

To *evaluate network visualization* ***tools***, we need to investigate several aspects as follows. Borner et al. [44] studied the familiarity of visualizations among adult and young museum visitors. They found that most visitors were able to interpret basic visualizations only. In terms of visualizations' readability, they found charts to be the easiest to read, followed by maps and graphs. Also, they found that network layouts were the most difficult to read. Wojton et al. [274] have proposed a simplicity-familiarity matrix as a platform as Figure 2.4 shows for designing interpretive materials for museum visitors. This matrix represents their research-driven model for situating complex data visualizations in the context of exhibition design. Wojton et al. [274] matrix uses the two continua of simplicity and familiarity to classify designed visualizations. Study participants were given visualizations, were asked to interpret them, and then design similar ones. Participants needed, on average, 150% longer time to understand and construct network visualization in comparison to other simple and familiar visualizations (e.g., bar charts, maps). Thus, they have classified *networks to belong to quadrant D being complex and unfamiliar*. The third study was by D'Ignazio and Bhargava [87], who proposed *teaching methods* focused on studying feminist theory to address inequalities via three case studies. In their second case study, they examined how participants use networks as one of the visualizations that help in "asking questions". In addition, in their third case study, they used networks in the "convince me" activity. That is by using network visualizations to explain and reason their arguments.

Networks ***visualizations*** specifically were investigated in limited number of studies. Firas et al. [95], through their literature survey of 35 papers and 45 publications in 2022, found only 3 papers that studied network visualization. Two of those mentioned earlier are studies in the wild (Borner et al. [44], Wojton et al. [274]). The third paper by Rodrigues et al. [220] aimed to report on the nature of the initial questions asked by a group of novice undergraduate and postgraduate students when they encounter visualizations for the first time. This classroom-

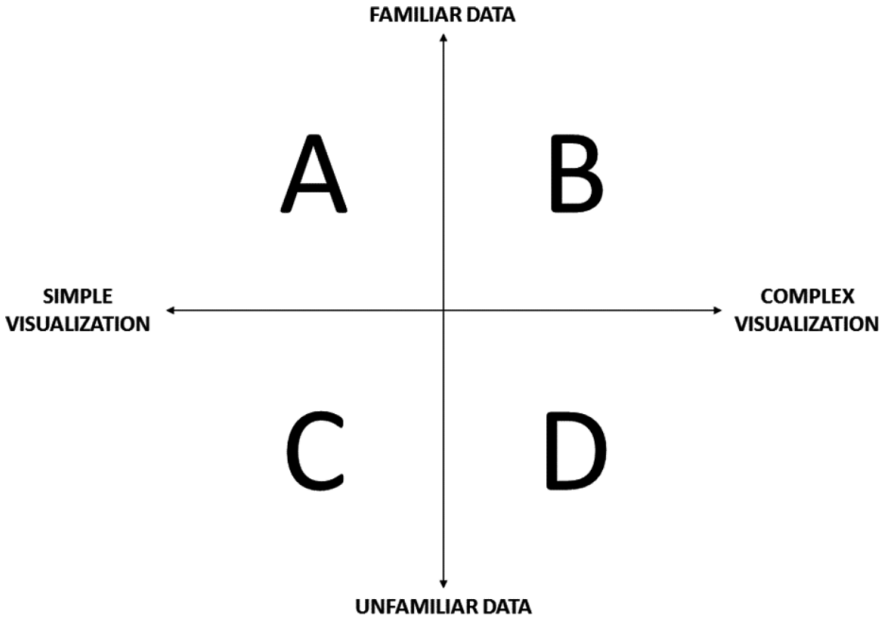


Figure 2.4: Data visualization quadrant map [274]

based study investigated networks as one of the 20 visualizations presented to the students. The study analyzed the common mistakes in the questions asked by students when interpreting data visualized. Overall, novice [56] users have a different pace in capturing and interpreting visualizations than practitioners. Thus, when assessing tools and designing activities, it's crucial to account for variations in users' literacy levels in network visualization. This includes their familiarity with the data, such as data type and collection method, and the specific tasks they aim to accomplish.

2.3.1 Approaches and Metrics for Network Visualization Tools Evaluation

Studying network visualization tools can follow different **evaluation approaches**. Some studies have followed a single approach, whereas others have combined multiple hybrid approaches [114, 244] such as "Multi-dimensional In-depth Long-term Case Studies" (MILC) [244, 243]. In this MILC study [243], researchers have assessed their tool based on (1) usage and (2) the expert users' success in achieving their goals. The user's *performance and interface efficacy* was measured and evaluated *multi-dimensionally* through logging, surveys, observations, and interviews. Different individual approaches aimed to understand analysts' usage and evaluate visualization tools through task-based evaluation [5], insights-based evaluation [114, 189, 206, 66], and interaction logs analysis [114].

Task-based evaluation targets assessing visualization tools through the predefined task(s) and measuring specific pre-designed metrics. For example, Ghoniem et al. [103] have used

task-based evaluation to compare the *suitability of a visualization* (e.g., node-link diagram vs. adjacency matrix) for specific tasks based on the network size. Additional studies have used **other methods** such as examining a tool's paradigm impact on how visualizations are created and explored. In the context of network visualization tools, a paradigm refers to a fundamental framework or approach that guides the design, functionality, and usage of the tool [127]. For example, Saket et al. [227] studied visualization-by-demonstration as an approach to construct visualization and their impact on the visual exploration process. Overall, all of those approaches aim to understand more about analysts and how they visually explore their data. However, in the context of this thesis, I do not aim to evaluate visualizations based on pre-designed tasks; rather, I aim toward evaluation in the wild.

Several studies investigated and evaluated visualization tools **based on their scope**. For example, a number of them studied *exploration as a process* in general [28, 238, 22, 40], some investigated *visualizations-related aspects* (e.g., visualization techniques [42]). In contrast, some have studied exploration from a domain-based or application-specific point of view [1, 166, 19, 166]. However, a limited number of studies have investigated network visualization tools *from the visual exploration perspective as a whole*. Some studies focused on how specific techniques promote visual exploration, such as [254]. Other studies focused on the evolving complexity of networks as in interdependent networks and how it opens opportunities and proposes challenges in visual exploration [6]. Furthermore, several studies discussed how network data modeling is a challenge in visual exploration [19, 124, 253], and some proposed new ways to model networks (e.g., visualization-by-demonstration) [227]. Block et al. [40] studied how interactivity affects the mental model of data processing over time. For example, Bokhare et al. [42] have compared using two network visualization tools of different paradigms.

Evaluation metrics have evolved over time based on the study design such as: tools' features alignment with the study goals [42] (e.g., datatype of graph elements, type of graph domain), **study settings** (e.g., controlled environment vs. in the wild), **data collection methods** from a single source or multiple-sources such as MILC studies [244], **duration to proficiency** that is how to take a user to transit from beginner to a practitioner, and the ability to move from *training* to *application and analysis* to *changing strategies*, **study tasks' nature** as being explanatory vs. exploratory, **scalability or the size of the graphs** [259], **the nature of the data** being structured, semi-structured, or unstructured, **network's visual appearance** [42], **the data readiness for visualization** meaning to what extent data was wrangled and prepared, **the tool's interactive features** such temporal analysis, **the representation of where values are missing** [259], **algorithmic metrics** or what Ham et al. [259] describe as *pluggable metrics* represent how indicators distinguish between algorithms for specific tasks such as clustering and layout design. These evolving metrics underscore the importance of a comprehensive approach

to evaluation, ensuring that visualization tools are assessed holistically based on the chosen study context and data characteristics.

Evaluation metrics of controlled environments, such as task accuracy and task completion time, are not applicable or precise in cases such as visual exploration in the wild. So, what sort of metrics can be used in the wild to evaluate visual analysis systems? And how can we measure the analyst's advancement in the visual exploration process? Overall, each visualization tool offers a varying set of opportunities and challenges that will have its own impact on the analysts' exploration journey toward achieving their goals, especially from the perspective of network modeling that was not investigated in-depth.

2.3.2 The Visual Exploration Process as the Evaluation Focus

To evaluate network visualization tools from the aspect of the visual exploration process, we need to understand more about the process's nature. The visual exploration of data is a process formed from several steps leading toward a specific goal or reaching insights. Facilitating this process has been *used as a metric to evaluate visualizations* [228], the design of visualization tools, and their provided interactions. Insights have been defined as *"an individual observation about the data by the participant, a unit of discovery."* [228]. Insights have their own nature of being complex, subjective, and qualitative [189]. The visual exploration enables users to find unexpected results, provoke new insights, and look into data from different perspectives [244]. The visual exploration process can be applied as a top-down approach (e.g., hypothesis-driven [162]) or a bottom-up approach (e.g., open-ended exploration [25]). Thus, understanding the nuanced nature of the visual exploration process is a crucial prerequisite to having a better evaluation of network visualization tools.

Analysts' strategies in visual exploration have been investigated as how they reach insight(s) [114, 162, 206, 213, 229], task-driven exploration [5], or observing analysts working with their own data [212]. To achieve a certain goal or objective, analysts follow a set of steps (similarly called sub-tasks or tactics) [107] such as data manipulation [52], view and query specification [123, 252], getting an overview [252], or detect patterns [277]. Tools support visual exploration through utilities such as search, export, collaboration, and history tracking [244]. On the other hand, exploring data holds various challenges, such as the need to use a variety of tools in some cases, transform and manipulate data, make hypotheses, look for patterns, and others [205, 244]. Analysts can follow different ways to explore the same dataset, but which of those is the form of the best strategy? Guo et al. [114] describe the sequence of essential analyst-application interactions to accomplish a specific visual exploration task as *critical path* [114]. So, which strategies do analyst adopt while exploring data and how can they balance between systematic and flexible exploration [197].

2.3.3 Insight-based Evaluation

Insights-based evaluation aims to understand how analysts arrive and generate insights [66, 189]. In insight-based tools evaluation studies [114], researchers aim to find how to quantify and describe the insights' characteristics [277, 66, 189, 45] and how they differ across applications and tools. In addition, what sort of evaluation questions each study answers, and what tools are best suited for specific data sets and/or research questions? Some insight-based evaluation questions were hard to address, such as *"How do users use the system to arrive at insights?"* and *"Which interface design factors are potentially hindering insights generation?"* from Guo et al. [114]. Due to the difficulty of precisely knowing when and how analysts arrived at insights, efforts were made to *trace insights' arrival* through recording analysis sessions and coding actions along with interaction log histories. The outcome of insights tracing is identifying which analysis strategies were followed by analysts in lab settings. Those were defined by what actions were taken, in which sequence, and what interface components and features were preferred more commonly [114]. Some of those studies focused more on what insights visualization tools offer [229]. However, understanding how analysts arrive at insights in predesigned task evaluation studies might be easier than in the wild. That is, additional work might be required from the analyst, such as preparing the data, thinking about how to visualize the data, and answering their research question through it. Thus, insight-based tool evaluation is not straightforward, especially in the wild.

Assessing insights cannot be done based on traditional evaluating metrics such as task duration and completion time. Measuring how analysts arrive at insights is based on two steps [114], which are: (1) *Abstraction*: the process of identifying and classifying analyst's actions based on chosen application interactions and (2) *Pattern Extraction/Discovery*: the process of identifying the actions' pattern leading to insight. A similar process is applied when collecting data from participants' think-aloud sessions and recorded videos of their exploration sessions. However, what differs is the level of information to be extracted and the level of understanding based on the type of data collected. For example, interaction logs of the exploration process by themselves might not provide detailed information in comparison to combining them with a think-aloud session. Whereas the insights-based tool's evaluation depends on evaluating to what extent the tool can help in revealing new insights and re-thinking the data, assumptions, and hypothesis [244, 205]. That is, analysts need to examine their data from different perspectives several times. Therefore, various factors must be considered that impact the insight-arrival process, including those influenced by the visualization tool's capabilities, the nature and structure of the data, and the specific research goals and questions being addressed.

2.3.4 Evaluating by Interaction Logs Analysis

Another method to study visualization tools is through *logging* analysts' interactions with the tool as this enables researchers to understand how analysts explore their data [212, 255, 239]. That is by extracting visual tasks such as gaining an overview, finding outliers, extracting trends, comparing values, and others [183]. However, such tasks differ based on the nature of the data, the visualization, and the data domain. Serrano Molinero et al. proposed to analyze logs based on users' interactions with the Vistorian, an online visualization tool [239]. Eltayeby and Dou [91] discussed the analysis of logs from different sources and using different tools (e.g., Google Analytics). Gotz et al. [106] analyzed interaction logs and characterized them based on a hierarchical classification of interactions: from atomic low-level to high-level semantic events such as actions, sub-tasks, and tasks. Some studies have combined logs with other methods to investigate the challenges analysts face through multi-dimensional long-term case studies methodology [205, 244, 198]. Blascheck et al. [38] complemented both interaction logs and eye-movement data to extract user behavior through spatiotemporal patterns. Furthermore, user exploration behavior has been investigated using Markov models to infer more about users' navigation and decisions [24, 214].

Actions, steps, sub-tasks, or terminals are all terms used to describe the semantics of the atomic low-level operations (e.g., create network schema). Different efforts have analyzed interaction logs based on various conceptualizations of users interactions such as typologies [52, 89], frameworks, models [226], conceptual spaces [217, 235], general visualization taxonomies [242, 156], taxonomies for networks visualizations specifically [275, 142] and dynamic networks [28]. Thus, different studies have used the outcome of such efforts to conceptualize user interactions for various purposes, including understanding users' intentions and objectives [106], finding analytical gaps in visualization tools [5], and generating suggestions for users [80]. Brown et al. [55] classified user interactions into 3 different categories, which are: (1) state-based, (2) event-based, and (3) sequence-based. They used those categories to calculate metrics, measure user performance, and learn more about personal traits. Other efforts were conducted to find metrics excluding usability issues (e.g., technical errors) and time [33]. There were efforts such as Gatham et al. [101] grammar to bridge the gap between high-level visualization taxonomies and low-level user interactions. That is to allow the generalizability of such taxonomies and extend their applicability. However, such taxonomies need further customization for complex visualization, such as network visual exploration.

Patterns represent a higher level of data abstraction aiming to represent interactions in the context of the user's objective or strategy. Rae et al. [212] investigated patterns by analyzing interaction logs within an enterprise setting, focusing on field observations. They have identified 4 common objectives or what they describe as the highest analytical goals, whereas actions

are the smallest atomic interactions based on Gotz and Zhou taxonomy [107]. They have used such classification to analyze logs and extract patterns, which is defined as *a series of actions performed repeatedly by the user to achieve their objective* [212, 106] as Figure 2.5 shows. That is, low-level system interactions are grouped into a sequence of actions. Then, each series of actions represents a task. And finally, on the highest level, the analytical reasoning (e.g., goal or objective) interprets why such interactions occurred. The identified objectives are: (1) check for anomalies or impact of a change, (2) investigate and reason unexpected patterns, (3) determine similar known patterns to take action, and (4) refine the definition of a new metric. Guo et al. [114] identified patterns based on the user's objectives, which are orienting, locating, sampling, and elaborating. In addition, they identified 3 additional recurring visual patterns (Fanout, Plug and play, and Sidestep) in addition to the earlier ones described (Drilldown [106], Sawtooth [106, 114, 204], and Side-by-side [276]). Gotz et al. [106] discussed two additional patterns, which are scan and flip. Finally, Rae et al. [212] framed 5 emergent themes that describe the analysis context, which might impact how logs are analyzed and understood (e.g., real-time collaboration). However, we still lack a systematic framework that can process and interpret various users' interactions with visualization tools on a higher semantic level. In other words, we need a method to transform interactions logged from different tools or visualizations into a sequence of readable exploration actions that reflect high-level analytical operations.

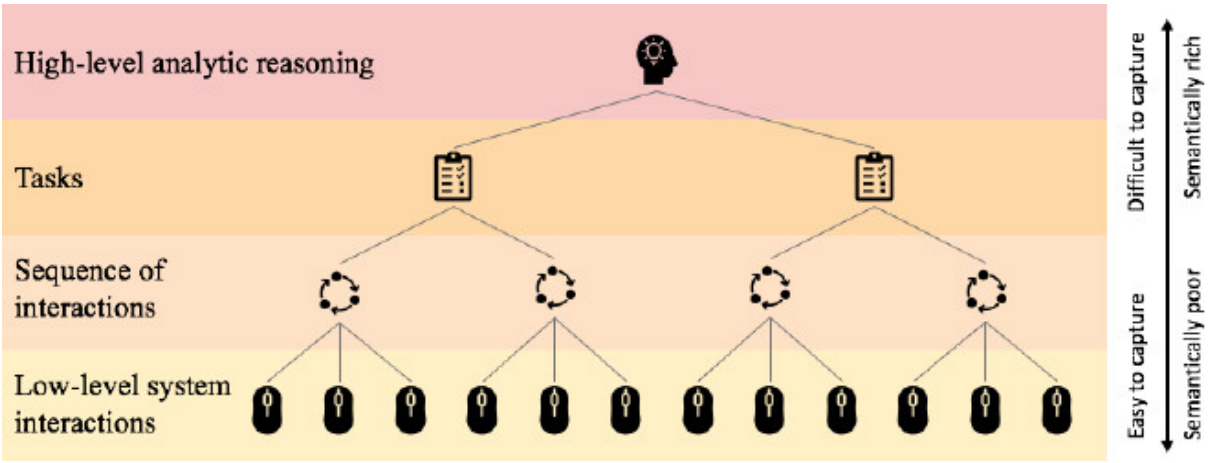


Figure 2.5: Mapping of Low-level System Interactions into High-level Reasoning (e.g., user objectives and strategies) [107]

2.3.5 Learning Analytics in Interactive Visualizations

Users' interactions with visualization tools along with other data sources (e.g., interviews, video recordings, think-aloud sessions, observations) represent the raw data of visualization tools' actual usage. Thus, various metrics were set and designed to evaluate tools and visualizations

from several aspects. Once metrics are computationally evaluated, the resulting analytics will provide insights about the examined target (e.g., tool, visualization). Analytics can be used for self-monitoring, monitoring others, or administrative monitoring [237]. Analytics differ based on the domain and context in which they are applied. For example, learning analytics represent the process of collecting, analyzing, and reporting on learners within an educational context and how students learn [155, 246]. For example, Li et al. [160] have analyzed students' interaction logs with videos offered in Massive Open Online Courses(MOOC). This has enabled them to find which parts are most of interest and measure their engagement with the content (e.g., returning to a specific moment of the video or replying to certain parts). In the context of the tool's evaluation, usage analytics can represent evaluating users' engagement, task accomplishments, and others. Thus, metrics should be carefully selected based on the goal of evaluation. The resulting analytics can be used and/or represented in various forms (e.g., dashboards, taking actions, feeding machine learning models). A learning dashboard is defined as *"a single display that aggregates different indicators about learner(s), learning process(es), and/or learning context(s) into one or multiple visualizations"* by Schwendimann et al. [237]. Such dashboards can be in one of the following 3 types described by Verbert et al. [260, 261] which are dashboards: (1) for traditional face-to-face sessions, (2) for face-to-face group work, and (3) for reflection, awareness, behavior change and sensemaking in online and blended learning. The third case applies to the context of this thesis of exploring in the wild.

In the context of network visual exploration, we aim to find which metrics can reflect and measure the visual exploration using network visualization tools. Learning analytics in network visual exploration can describe several aspects, such as how analysts create networks, explore data, and choose visualizations. However, designing learning analytics to measure the success of a network's visual exploration is a bit tricky to do for two reasons. Firstly, network visual exploration has no fail/pass marker as most educational assessments have (e.g., test's pass mark, assignments). For example, there is no tangible value that can be used to classify the exploration session based on it. Secondly, the content in each exploration session varies based on the user, the data, and the visualizations being explored. In other words, there are no distinct indicators pinpointing how the analysts are progressing through their visual exploration process. Thus, those two points represent a challenge in determining learning analytics to represent the network visual exploration.

Schwendimann et al. [237] represented metrics and indicators that measure the learning process from different aspects as Figure 2.6 shows. Each indicator is described by the questions it aims to answer, along with illustrating examples. By comparing those indicators in the learning and teaching context to the visual exploration context using visualization tools, we can say that:

- the first 3 indicators are applicable in the context of our studies, which are: *learner-related*

Indicator group name	Question the indicators aim to answer	Examples of indicators in the group
Learner-related	Who are the learners?	Prior education, Competences, Age, Prior courses they took, University entrance grade
Action-related	What do they do while learning?	Number of page visits, Number of file downloads, Time spent on tasks, Login time, Timeline
Content-related	What is the content involved in their learning?	Sentiment of the messages, Topics covered and omitted in the report, Number of concepts and links in a concept map
Result-related	What is the result of their learning?	Average grade, Grades distribution in a group
Context-related	In which context does the learning take place?	Location of learners around a tabletop, Placement in a classroom, Geographical location
Social-related	How they interacts with others during learning?	Network showing communication with others in a group forum, Direction of interaction in a group around a tabletop

Figure 2.6: Overview of Indicator Types Presented in the Systematic Literature Review by Schwendimann et al. [237]

(information describing the learner), *action-related* (information describing the actions taken by the learner), and *content-related* (information describing the content the learner interacts with).

- The fourth indicator, which is *result-related* (outcome of learners activities), as discussed above, is a bit tricky and hard to measure in the visual exploration process. Thus, we investigate and seek further metrics and indicators to measure the visual exploration process.
- The fifth indicator, which is *context-related* (context where learning takes place), is applicable in the places where training or exploration takes place (e.g., course, in the wild)
- The sixth indicator, which is *social-related* (how learners interact with each other), is not applicable in the cases of visualization tools that do not offer collaboration between tool users and learning happening in the wild.

Thus, it is important to keep the selection process of such indicators and metrics in mind when evaluating visualization tools in the wild. In visual exploration of networks, we need to enrich the assessment of the indicators by learning more about who the explorers are, what they do while exploring, what is the content involved in their exploration, what is the result of their exploration, and in which content it takes place in the wild.

2.4 Mental Model & Knowledge Generation in Visual Analytics

To understand, support, and guide analysts in their visual exploration process of networks, we need to understand the mental model behind data visualization and how knowledge is generated. Ham et al. [259] describe the understanding of how different factors impact the design of a network's visualizations and views within a single mental model as a *daunting task*. As the

analyst explores different visualizations and views, zooming between levels of abstraction and trying to link their observations and findings. Various **mental models** have been developed to understand and investigate how analysts reach insights. Those models have been used for different goals such as evaluating design changes and predicting their outcomes [110], benchmarking visualization tasks [105], and classifying users with different task strategies [55]. In addition, one main role of those models is to reason about the actions taken by users while visually analyzing and exploring their data. The sensemaking process of the visual analysis process included abstraction and task categorization [208], interactions chunking [114], and building taxonomies for characterizing users based on their analysis tasks [101] and how they reach insights [277, 156, 142]. Taxonomies are developed and used to describe different tasks such as data analysis tasks [4, 178], dynamic networks [28], and visual exploration process [275]. By leveraging these mental models and taxonomies, we can enhance the design, support, and even the evaluation of data visualization tools to better facilitate analysts' insights and understanding.

The Knowledge generation model illustrated in Figure 2.7 by Sacha et al. [226] consists of two parts. The left part represents the visual analytics system consisting of data, visualizations, and models. The right part represents the human knowledge generation process. The latter process consists of exploration, verification, and knowledge generation loops. By interacting with visual analytics systems, analysts can explore data by capturing findings and taking actions based on them. Verifying their findings, forming insights, validating hypotheses, and generating knowledge based on this whole iterative process.

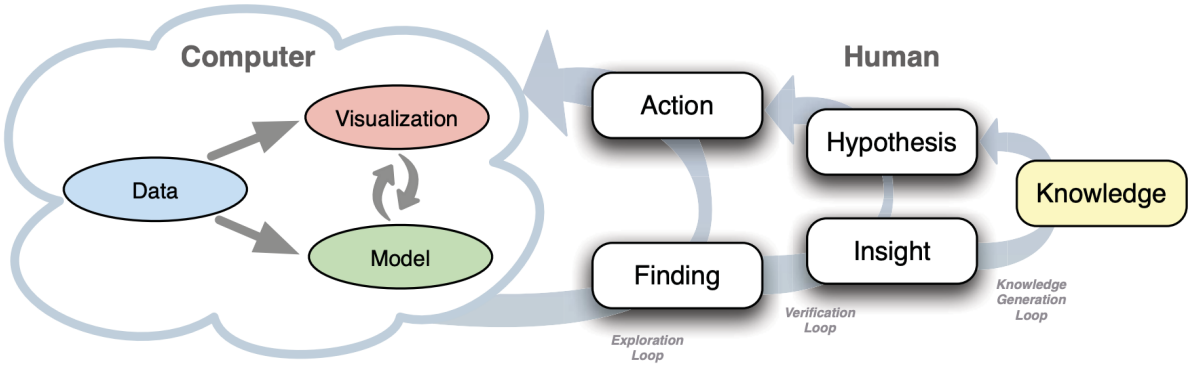


Figure 2.7: Knowledge Generation Model for Visual Analytics [226]

Analysts working and engaging with visualizations perform tasks ranging from low-level (as in data foraging) to high-level tasks (as in sensemaking) [60]. Klein et al. [145] define sensemaking as *"the process of fitting information into a frame, which is a mental structure that organizes the data"*. Frames are the result of expertise in a domain or process. Pirolli and Card [204] described that the sensemaking process includes several tasks such as: *"rep-*

resentation of the information in a schema that aids analysis", *"the development of insight through the manipulation of the representation"*, and *"the creation of some knowledge product or direct action based on the insight"*. Pirolli and Card [204] describe that the sensemaking *process consists of two loops*: the foraging loop and the sensemaking loop. The *foraging loop* contains processes that seek information, search and filter it, and read and extract it. Whereas the *sensemaking loop* is *"the iterative development of the mental model (conceptualization) from the schema that best fits the evidence"* [204]. It is crucial to acknowledge that the progression in the sensemaking process may vary between a user encountering the visualization for the first time and one already acquainted with it. Freedman and Shah [97] discussed the sensemaking process in their work toward "Knowledge-based Graph Comprehension". They state that users vary in how they comprehend graphs based on several factors, such as their prior technical and domain knowledge, skills (e.g., scientific reasoning), and biases [97]. Freedman and Shah [97] say "During the initial processing, people chunk the visual features in the graphs. Nevertheless, prior knowledge guides the processing of visual features.". Specifically, the quantity or proportion of loops engaged in the process, whether they pertain to sensemaking or foraging, may vary depending on the user's level of familiarity. Thus, investigating the visual exploration process is crucial to leverage many of the problems related to the sensemaking process, such as generating alternative hypotheses and revisiting the confirmation bias.

Amar and Stasko [5] introduced *analytic gaps* to represent the problems and challenges analysts face when using visualization systems to learn and perform high-level analytical tasks such as sensemaking and decision-making. For example, supporting the process of decision-making faces 3 challenges: (1) limited affordances of analytical operations provided by visualization tools, (2) predetermined representations of data of a specific domain or problem, and (3) decline of determinism in decision-making (e.g., as in the case of uncertainty in data) [5]. Thus, Amar and Stasko [5] identified 2 main analytical gaps, and for each gap, they identified several knowledge tasks to measure and ensure that such gaps do not exist. The identified gaps are: (1) rational gap *"between perceiving a relationship and expressing confidence in the correctness and utility of that relationship"* and (2) worldview gap *" representing the gap between what is shown to a user and what actually needs to be shown to draw a representational conclusion for making a decision"*. Thus, it is important to measure and overcome such analytical gaps that impact the findings' validity.

Information processing can be driven by a top-down process (from theory to data) or vice-versa bottom-up (from data to theory). Pirolli and Card [204] recommend that both top-down and bottom-up processes can be invoked in an opportunistic mix. The bottom-up process includes searching and filtering data, reading and extracting from data, schematizing information (deciding on how to represent information), and building a case or theory by proving

or disconfirming a hypothesis with evidence. Top-down processes encompass re-evaluating data representations, seeking guidance or support, searching for evidence (such as specific patterns), identifying relationships between data entities, and gathering information based on hypothesized assumptions. To effectively facilitate these processes, tools must provide interactive features and robust support mechanisms to assist analysts in transitioning between different visualizations and views, such as dynamic labels [259]. On another level, guidance shall support processes such as data wrangling, selecting and creating visualizations, the exploration process, the verification of findings, and knowledge generation [63]. Thus, the knowledge generation model [226] should aim toward enhancing its flow by allowing the characterization of users' behavior and their analysis objectives. This emphasizes how studying the mental model is an important process as it involves the human, the data, the visualization concepts, and the tool's design and its features.

2.5 Network Visualization Literacy, Education and Guidance

In this section, we discuss how to learn and teach network visualization literacy. The learning and teaching of network visualization tools to address the needs of building a mental model of data visualized involves challenges on different levels [135]. One challenge is how to teach and learn network visualization concepts, such as how to read network visualization, how to explore them, and how to evaluate their representation. This is known as "network visualization literacy". Boy et al. [48] described visualization literacy by saying *"a concept is generally understood as the ability to confidently create and interpret visual representations of data"*. Borner et al. [44] explained it further as *"the ability to make meaning from and interpret patterns, trends, and correlations in visual representations of data"*. Bristor and Drake [54] add to the definition by the *"ability to evaluate visual messages"*. But what is the state of network visualizations' literacy? To find out, we will first discuss network visualization literacy in subsection 2.5.1. Following that, we will explore how to provide guidance to assist analysts in creating and interpreting visualizations in subsection 2.5.2. Next, we will examine the design of educational guides for visualization purposes in subsection 2.5.3, and review the types of educational activities used to support guidance in visualization literacy in subsection 2.5.4. Finally, subsection 2.5.5 will conclude this section by discussing how to support self-regulated learning in visual exploration of networks.

2.5.1 Networks Visualization Literacy

By defining visualization literacy, we can look into network visualization literacy that extends such basic definitions. In the context of networks, Zoss et al. [282] defined the literacy of network visualization as *"the ability to read, interpret, and visualize different types of networks"*.

Zoss et al. [282] discuss several topics around network visualization literacy, such as its role in teaching and learning, how to evaluate network visualization literacy, and how to support the literacy of network visualizations. Thus, to understand how to teach network visualization, evaluate its literacy and support it, we need to identify the skills and topics required to acquire the knowledge. Zoss et al. [282] classify and describe the literacy topics and aspects in network visualizations as Representational Literacy, Topological Literacy, and Metaphoric Literacy, which we describe in the following.

The *representational literacy* [282] focuses on answering the question of *"do individuals understand how network data are converted into visuals?"*. Sayama et al. [231] have articulated a set of seven essential concepts and core ideas [231] that are must-have concepts for a network-literate person. The core concepts are: (1) networks are everywhere, (2) networks describe how things connect and interact, (3) networks can help reveal patterns, (4) visualization can help provide an understanding of networks, (5) today's computer technology allows studying real-world networks, (6) networks help in comparing a wide variety of systems, and (7) the structure of a network can influence its state and vice versa. Some of these concepts constitute the minimal knowledge a network-literate person should have (e.g., networks describe how things connect and interact). In contrast, others represent a proficient level of literacy in network visualizations (e.g., the structure of a network can influence its state and vice versa). Each concept of those is subsumed by several core ideas that illustrate the concept in practice and can be found here in detail. Those concepts and ideas form the cornerstone for building educational material about networks.

The *metaphoric literacy* [282] addresses the question of *"how intuitive is the arrangement of nodes and links in network visualization?"*. This includes the analyst's ability to understand the metaphoric properties and aesthetics of the network visualizations. Such as the the distance-similarity metaphor that denotes that close nodes are more similar than nodes apart in distance. Another example is how shorter links represent stronger links than longer ones. So, why is metaphoric literacy important? Zoss et al. [282] explain by saying *"Novices without a sophisticated understanding of layout algorithms will likely make judgments based on the length of lines."* These examples are only a glimpse of the wide range of metaphors a visualization can hold, varying from one network representation to another. Thus, misinterpretations of aesthetics in a network visualization may lead to incorrect findings.

Thirdly, the *topological literacy* [282] addresses the question of *"do individuals understand basic network properties when reading network visualizations?"*. This literacy enables analysts to extract the *structural* properties and relationships within a network that describe how nodes and edges are arranged and interconnected. For instance, the node degree and the clustering coefficient are an example of the statistical properties describing a network's topology. Zoss et

al. [282] highlighted that this type of literacy *may be affected by the data's domain* by saying: *"depending on the field of the study, different parts of a network dataset may be considered especially important."* For example, in some domains, studying clusters is more important, whereas in other domains, identifying highly influential nodes (e.g., nodes with high betweenness centrality score) is more important. The topological literacy depends on many factors [282] including:

- the analyst's prior training with both network visualization and data structuring,
- the selection of the network visualization layout with an understanding of the underlying algorithm,
- the additional design decisions the analyst takes (e.g., selection of node colors),
- the distinct properties of the selected network dataset (e.g., number of links, size of the network),
- and the selected analytical measure of the network topological properties.

Measuring and evaluating network visualization literacy is a very important process. Zoss et al. [282] emphasize that measuring and evaluating should be similar to how individuals interpret and create such visualizations in daily life tasks. Zoss et al. [282] promoted the importance of having a *mix of formal and informal education* to guide analysts and improve network visualizations' literacy. Zoss et al. mentioned 3 existing mediums, which are: (1) museums such as The Nature of Networks exhibit at the New York Hall of Science, (2) research programs in schools such as NetSci program for high school students, and (3) online courses such as Information Visualization Massive Open Online Courses(MOOC) course at Indiana University. Museums represent an informal learning environment, whereas research programs and online courses represent formal learning environments. Informal approaches and learning environments are known for the small amount of knowledge transferred. Thus, formal approaches or the mix of formal and informal approaches are crucial for teaching *novices* network visualizations as an instrument to analyze and communicate their findings. Zoss et al. [282] explains by *"significant scaffolding is required for effective learning and knowledge transfer"*. They also recommend improving network visualization literacy by reviewing relevant research and experiences with teaching and learning with network visualization.

In addition to the challenge of how to teach basic literacy (section 2.5), network visualizations face another challenge of how to teach the process of network creation. Creating a network involves making multiple decisions, such as identifying possible nodes and links, deciding on the network schema that helps to address the research goal, and even thinking about which visualization will help identify patterns in the selected network. In addition, deciding upon the suitability of a visualization tool(s) for the creation and exploration processes involves understanding its features, provided visualizations, and whether it can limit or inspire the

exploration process. Thus, teaching network visualizations has been and continues to be a challenge that requires more structured and formalized guidance [68, 14]. Novice users often struggle to understand new visualization techniques without proper introductory materials. And some analysts struggle to decide upon which visualization to create or explore.

2.5.2 Guidance through Visualization Creation and Exploration

As we explored in subsection 2.5.1 the network visualization literacy, we focus here on how we can transform this knowledge and skill set into formalized guides. And how can guides meet the diverse needs of analysts, from novices to practitioners, and accommodate different data types? It's important to note that limited resources are available for teaching network visualizations, especially in the wild [133, 59, 136, 135]. Thus, in this part, we aim to investigate how network visualization can be taught and learned in a systematic way that offers clear, informative guidance. So, what does guidance mean in the context of network visualizations? And how do tools and educational guides support it? Guidance in visual analytics and data exploration has been investigated and developed theoretically to support users and tool developers [72, 62, 236]. Guidance is defined as *the active process of addressing knowledge gaps that hinder the analytical progress* [201, 62]. So, how can we overcome knowledge gaps, address them, and in which means.

Guidance can be offered in different means (e.g., computer-assisted, printed) and different forms (e.g., videos, written tutorials). Guides in visual analytics can be created for *general* purposes, providing overall information or outlining a process (such as an onboarding procedure), or they can be *customized* and tailored for specific tasks, such as preparing data files. However, creating custom guides requires knowing the users and analysts' *intent*, which is not always captured [250]. *Intent or analysis goal* can be either explicitly stated by the user or captured by the system such as *speculative guidance* mechanism [249]. Speculative guidance aims to capture the user's intent by suggesting different hints and capturing feedback from the user to eliminate undesired options. Creating guidance material to support various analytical and exploration tasks requires knowledge to fill the gap between [201] what the analysts want to do and how to do it. The guidance process should be designed toward a goal, which is described by Engels [92] as a *goal state*. That is, task compositions should be proposed to achieve that goal. And this is not a straightforward process, which requires additional efforts offer better support.

Guides characterization frameworks [250, 61] were mostly developed for guidance designers. Guidance can be provided as a system initiative -inter/action- only or as a mixed-initiative between the user and the system [75, 62, 251]. Sperrle et al. [251] describe guidance as *"a continuous process of action-reaction pairs"*. However, we should consider providing such process is impacted by the medium of delivery, such as guidance provided through software

versus a written guide. Guides for information visualization follow either a top-down or a bottom-up teaching method. Guides can be classified under *skill-training, knowledge acquisition or development, or both of them.* There are several efforts on designing guides [61, 250].

Describing guidance is a crucial step toward guide design. Schulz et al. [236] proposed a *schema* to characterize guidance from 4 perspectives which are: (1) *context* which specify the level of knowledge the user has, (2) *domain* where the guidance shall be provided, (3) *target* defines the guidance objective, which can be a situation or an entity based on the guidance domain, and (4) *degree* which describes the level of guidance to be grasped. The **guidance context** describes the user's knowledge or expertise and classifies it into 4 situations, which are: 1) *zero knowledge* where neither the goal nor the path to achieving it is known, 2) *goal is known* where the knowledge about the goal exists but without knowing how to reach it, 3) *the path is known* to get somewhere but there is no knowledge about the goal, and (4) *full knowledge* where both goal and path are known. In each of those cases, a different guidance to be given or advised for optimality.

Schulz et al. [236] have described success in guidance by reaching a goal or progressing **guidance domain** can be on several levels, including data (e.g., semi-automatic selection of data subsets of interest), visualizations (e.g., suggesting visualization techniques and parameterizations), infrastructure (e.g., software as tools, hardware as displays) and users (e.g., assigning tasks to users). The **guidance target** can be direct (guidance toward a specific point), indirect (guidance toward several points that share a state or condition such as query-by-example), or inverse indirect (guidance toward all points that are different from a specific point such as discovering the unexpected).

Finally, the **guidance degree** [236] specifies the level of guidance provided, which are orienteering (little guidance, much freedom), steering (medium guidance, medium freedom), storytelling (much guidance, little freedom), and annotated animation (full guidance, no freedom). Schulz et al. [236] describe the *effectiveness of guidance* solution by its level of freedom and how it adapts based on the context. Sperrle at al. [250] designed a practice-oriented model to minimize the gap between the guidance design and implementation. That is, by automating the generation of guidance based on its specifications. Sperrle et al. [250] distinguishes between *guidance degree* and *guidance level* as follows. The guidance degree describes to what extent a system supports mixed-initiative guidance. Whereas guidance level describes, *"the semantic complexity of a given suggestion"*. Cenedea et al. [61] describe 3 *degrees of guidance*, which are orienting, directing, and prescribing. Orienting means highlighting through visual cues, directing is about providing a ranked list of suggestions, and prescribing is about the system selecting the optimal suggestion.

This section discussed how to offer guidance to support network visualization literacy with the required skill set. And how to transform knowledge into formalized guides that cater to various analysts' needs, from novices to experienced practitioners, across different data types. Designing effective guidance should ensure the capture of the user's intent, which can be explicitly stated by the user or inferred by the user's actions. The design of guidance materials should bridge the gap between analysts' goals and their approaches and be presented in different forms. Guidance in information visualization can follow top-down or bottom-up teaching methods and be classified into skill training, knowledge acquisition, or both. A well-structured guidance schema includes context, domain, target, and degree, which determine the level of guidance needed based on the user's knowledge and the analytical tasks. Effective guidance balances the level of freedom and adaptability to the user's context, ensuring optimal analytical progress. Thus, given the limited resources on teaching network visualizations in real-world settings, in this thesis, I aim to develop a systematic teaching and learning approach that focuses on guidance in the wild. The provided guidance should address knowledge gaps hindering analytical progress and be applicable through various means and forms.

2.5.3 Designing Educational Guides for Visualization Purposes

Before designing guides, we need to understand the types of learning. Learning types are either active versus passive [98]. Active learning means actively participating in a corresponding dialog to proceed within a lesson. Meanwhile, passive learning indicates that the learner only receives information with no participation from the participant side. Teaching methods can be either bottom-up or top-down [279]. The top-down teaching method is a deductive approach that starts by giving an overview through a high-level and abstract level about the whole concept/process and then discussing details. On the other hand, the bottom-up approach is an inductive approach that discusses details and smaller learning units and builds upon them until the whole concept is delivered. Studies have shown that the learning process of data visualization and engaging with it can be supported by managing its complexity and how it is taught [147]. This is done by assisting students in focusing on the big picture and concrete meaning rather than the abstract representation when learning to use data analytical tools.

Understanding the goal of the analysis task plays a major role in designing educational guides for visualization purposes. Sperrle et al. [250] described the characterization of interactions between the system and the user based on the user's intent and analysis goal as **guidance dynamics**. Guidance can be either to *support* by providing suggestions or *correct* by adjusting the analysis behavior. Andrienko et al. [7] identified general guidance goals, which are to inform, reduce cognitive load, train, mitigate bias, engage, and verify conclusions. Also, Andrienko et al. [7] stated the existence of various levels of guidance, starting from the smallest efforts, such

as low-level operations, to high-level, such as building models. Ceneda et al. framework [61] focused on "when and how" guidance is provided. However, [250] proposed an abstraction of guidance level as Figure 2.8 illustrates. It shows how the semantic complexity increases as analysts advance from tool onboarding, where system functionality is explained, to open-ended guidance, where basic practices in visual exploration are explained, to goal-oriented guidance, where support to accomplish specific tasks is provided. To automate such guidance, further challenges will faced, such as adapting to context, timing actions, capturing user feedback, and determining when a suggestion or action is applicable.

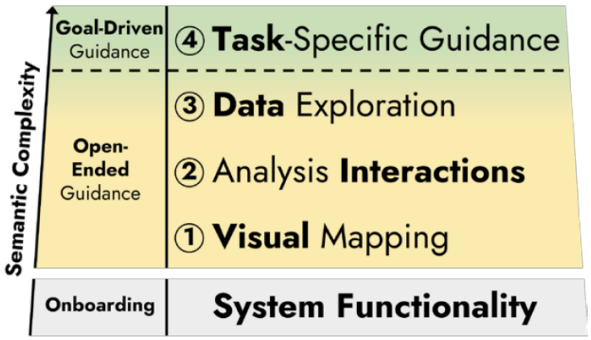
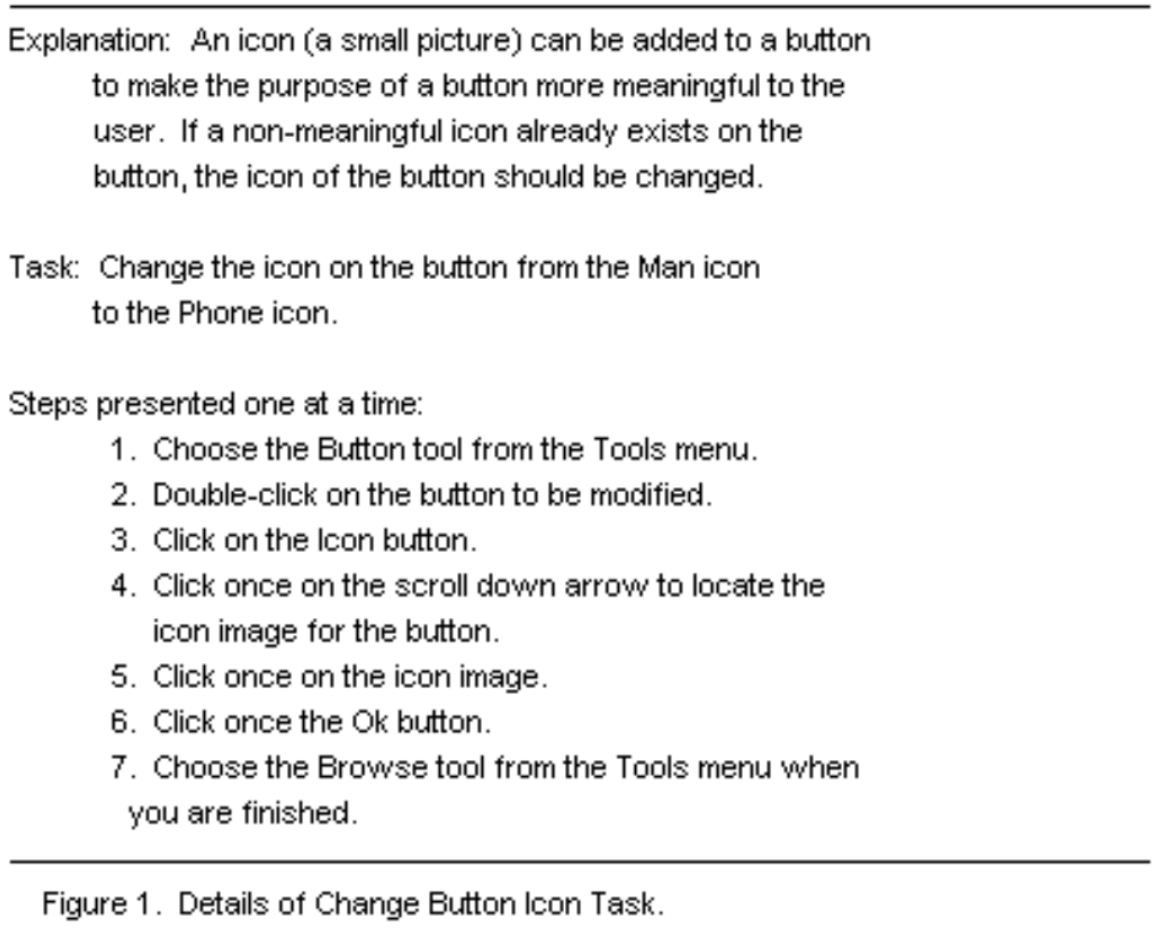


Figure 2.8: Four Levels of Guidance by Sperrle et al. [250]

The design of skill-training guides in visualization differs in its nature based on the targeted skill. Tanahashi et al. [257] created online introductory guides for information visualization. Novice users were introduced to practice learning segments, followed by testing segments on similar visualizations. They have used active learning along with the top-down teaching method as they found it has the best improvement on testing segments or evaluating participants. Their guides were presented in the form of text-plus-question [257]. There are additional forms of guidance besides describing strategies and tasks. For example, Dabek et al. [80] provide suggestions in the form of exploration interactions recommendations based on learning from previous users' interactions. Also, Gotz et al . [106] use similar techniques to guide the visualization building. Thus, training on a skill can often be described in a detailed, step-by-step manner where applicable. However, this is not always the case, as multiple processes may overlap, making it difficult to clearly delineate where one step or process ends and the next one begins.

Knowledge-acquisition tutorials focus, as Tanahashi et al. [257] describe it as *"the development of the theoretical or practical understanding of a subject rather than the cultivation of a skill"*. For example, designing guides to introduce non-familiar visualizations [158]. Tanahashi et al. [257] designed guides where contained each a visualization taught and presentedz through 2-4 slides. The first two slides contained text describing the example visualization, and the last two, if present, each contain an exercise question. The first slide describes the visualization

technique in general and how information is encoded. The second slide explains the example visualization more specifically (types of interactions supported and type of information visible). Participants were asked to choose among options to answer an exercise rather than an open-ended question. The exploration process differs from the testing process as participants learn more as they explore (not to remove the in-test learning) [257]. Tanahashi et al. [257] found that guides that includes tasks such as questions positively affect the learner's performance. Designing guides is also impacted by incorporating visualizations and tool features such as different sorts of interactions [257].


Explanation: An icon (a small picture) can be added to a button to make the purpose of a button more meaningful to the user. If a non-meaningful icon already exists on the button, the icon of the button should be changed.

Task: Change the icon on the button from the Man icon to the Phone icon.

Steps presented one at a time:

1. Choose the Button tool from the Tools menu.
2. Double-click on the button to be modified.
3. Click on the Icon button.
4. Click once on the scroll down arrow to locate the icon image for the button.
5. Click once on the icon image.
6. Click once the Ok button.
7. Choose the Browse tool from the Tools menu when you are finished.

Figure 1. Details of Change Button Icon Task.

Figure 2.9: Hyper Cards describing Tasks [118]

Different efforts are made to design guides that describe strategies and tasks. For example, Harrison [118] used what they call "Hyper Cards"Figure 2.9 to describe evaluation tasks. Each hypercard contained 3 parts: an explanation, a task, and steps to perform it. Sperrle et al. [250] developed the Lotse framework for guidance in visual analytics. This framework consists of 4 steps, which are: (1) design the guidance strategy, (2) translate it into guidance specification, (3) generate guidance based on specification using Lotse engine, and (4) implement or adapt by the front-end. Sperrle et al. described their guidance specification using low-level as in Figure 2.10.

Lotse framework [250] describes how systems can assist during visual exploration through abstracting and reasoning existing approaches. Lotse framework [250] specify that each section starts with an *analysis goal*. Analysts shall break and structure their approach toward a goal into tasks and implement them sequentially to reach that goal.

Basic Info	**Strategy Name**: _____ **Description**: _____ _____
Goal	**Guidance Goal**: ○ Support ○ Correct **Level**: ○ Onboarding ○ Open-Ended ○ Goal-Driven **Dynamic**: ○ Learn ○ Teach ○ Other _____ **Knowledge Gap Type**: ☐ Path to Solution ☐ Optimal Solution **Knowledge Gap Domain**: ☐ Data ☐ Tasks ☐ VA Methods **Details**: _____
Action	**Content**: _____ _____ **Timing**: ○ Static ○ Contextualized on: _____ **Adaptation**: ○ No ○ Yes: _____
Visual ization	**Degree**: ○ Orienting ○ Directing ○ Prescribing **Placement**: ○ In-situ ○ Ex-situ **Visual Elements**: _____

```
type:
  action
metadata:
  description: Suggest alternative month
  degree: orienting
  action_id: timeslider
generate_suggestion_content:
  args: [state]
  load: return (state.month + 1, 'Move', '[...]')
threshold: .10
condition:
  type: function
  args: [state]
  load: |
    month = state.get_current_month()
    all_data = len(month)
    een = len([d for d in month if d.hovered])
    kreturn seen / all_data > self.threshold
is_applicable:
  args: [state, delta]
  load: return self.condition(state) \
              and not self.suggested
accept:
  args: [suggestion, state, delta]
  load: |
    state.month = suggestion.suggestion.event.value
    self.threshold *= .95
    self.suggested = False
reject:
  args: [suggestion, state, delta]
  load: |
    self.threshold = min(0.9, self.threshold + 0.05)
    self.suggested = False
```

Figure 2.10: Guidance Template and Guidance Specification by Lotse Framework [250]

The visual exploration of networks or even data in general contains several stages, such as ideation, visualization, and exploration. However, such a process follows an iterative non-linear behavior as students in a visualization course [149] reported (see Figure 2.11). Students differ in the start point, path, number, and sequence of processes they follow in their visualization journey. Thus, a general educational guide should be designed to give an overview of the whole process. Furthermore, activities shall be placed in one of the design thinking stages as inspiration, ideation, and implementation as proposed by Palacin-Silva et al. [195].

In this subsection, we have reviewed the different efforts in educational guides' design for

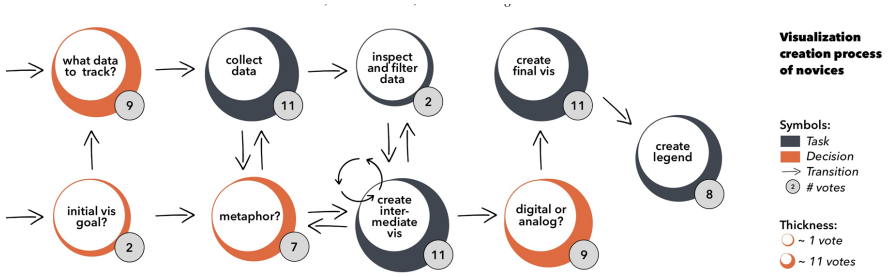


Figure 2.11: Common Approach of Ideation and Visualization by Course Students [149]

visualization purposes. Thus, this thesis focuses on both knowledge acquisition and skill training. I aim to design a guide that forms the ground for learning visual exploration of networks and even evaluating network visualization tools. Such a guide should contribute to this thesis's research questions about providing an intervention, supporting self-regulated learning, and describing strategies in visual network exploration. The guide should adopt an active learning mode, allowing learners to interact, choose, and customize their exploration journey. The guide should offer the learner guidance using a top-down teaching method, offering an overview of the process, the possible exploration options, and different resources that explain each step. This enables reducing the exploration process's complexity by explaining step-by-step how to move from data to findings. The guide should also offer means to capture the user's intent and how to enable a better outcome of the exploration process. It should also accommodate the analysts' differences as much as possible, such as their expertise, domains, learning pace, and their understanding of data. The design of the guide and supplemental material should adapt to the skills and knowledge introduced and the knowledge taught. The design of the whole guidance process should simulate and reflect the real-world scenarios of exploration. Aiming to allow analysts to think and learn about their data and goals from various perspectives. The guide and the supplemental material should inform, guide, and provide feedback/reflection on the user's choices based on each stage's need for the exploration process. Such design guidelines shall support analysts toward a better learning and exploration process in the wild.

2.5.4 Educational Activities for Visualization Purposes

As discussed in subsection 2.5.1, the content of educational activities associated with educational efforts in network visualization should establish the core concepts described by Zoss et al. [282]. Activities can teach or assess one or more skills based on their granularity and goal (e.g., creating visualization [108]). For example, Imre et al. [135] focused on *explaining visualization layouts* and how different representations of a single network impact its understanding. Although this looks like a delivery of knowledge, it can be transformed into an activity that helps the analysts understand and practice what they have learned. However, research on the quality and prevalence

of skills required to interpret and construct *network visualizations is still limited to a number of skills* [282]. For example, Ghoniem et al. [103] have investigated the readability of node-link diagrams and matrix-based representations by designing several *readability* tasks. Another study was by He and Adar [120], who designed the "VizItCards" toolkit and used it in a card-driven workshop for graduate InfoVis classes. They aimed to investigate how information visualization education can be improved using new practices such as toolkit cards. The toolkit has classified visualization tasks into several categories such as: *domain tasks* related to field needs, *abstract tasks* related to how and what visualization is created, and *encoding tasks* related to how data within a visualization is represented. Network visualizations were one of the visualization types He and Adar [120] experimented with the toolkit. Participants were asked to find key features for generalizations as an example of an abstract task. By testing the toolkit, the researchers found that design activities related to abstract tasks might be challenging for the participants due to some tasks' vagueness and time-consuming nature. In addition, they found that separating abstract tasks from domain/encoding tasks is hard. We can see that most of the studies focus on a few tasks in network visualization (e.g., readability and constructing a simple network). However, the tasks do not cover other real-world tasks in network visual exploration, such as choosing the correct network schema for a specific goal. Thus, in network visualizations, further activities should be designed to assist analysts in their visual network exploration.

Furthermore, activities' tasks in information visualizations can aim toward different objectives. For example, Brehmer and Munzner [52] classified visualization tasks based on their purpose: why, how, and what tasks. They why tasks can be high-level (e.g., discover), mid-level (e.g., locate, browse), and low-level (e.g., identify, compare). Brehmer and Munzner [52] used such classification to create a topology of abstract tasks in visualization, aiming toward a general description that is not domain-specific. Some tasks require interactions, whereas others are more about information acquisition. Thus, in information visualization education, the success of an activity outcome relies primarily on the task description [52]. The task description varies based on its objective and should address the task's purpose, method, and rationale. As an example, McKenna et al. [57] have created worksheets for novices on how to implement four main visualization design activities: understand, ideate, make, and deploy. Each activity worksheet included the description of concrete steps leading toward specific goals. The worksheets also included the assessment of their effectiveness along with their strengths and limitations. Also, Byrd et al. [57] designed a series of activity worksheets to guide novice instructors and students through the data visualization process. Each worksheet focused on an individual step in the data visualization process. Their worksheets covered various stages such as acquire, parse, mine, filter and represent, critique, refine, and interact. Such worksheets played an active learning role in practicing various visualization activities [120]. Thus, it is crucial that any activity or task

description be comprehensive and walks analysts step-by-step toward the objective. Also, it should ensure that learners are actively involved in the learning process and can make decisions on different exploration levels.

Lastly, evaluating an activity/task is crucial in achieving its objectives and desirable outcomes. The evaluation process can differ based on the objectives and scope of the activity or task. Some objectives can be high-level, such as in the VizItCards [120] toolkit learning objectives (e.g., design, ideate, collaborate, apply, and synthesize). High-level objectives might be hard to evaluate and require breaking tasks into smaller ones. Thus, He and Adar [120] to evaluate the outcomes of a task, they set a heuristic metric, which they call the expressiveness-effectiveness score (EES). This enabled them to evaluate if the task was doable in the domain the participant chose. The researchers validated the outcome by giving a score ranging from 0 (meaning information cannot be expressed) to 4 (meaning information can be expressed easily). Thus, good practice in activity design should encourage thinking about data from different perspectives rather than limiting it to a narrow range of representations. Also, it should develop metrics to evaluate an activity's effectiveness and its success in achieving its goals and objectives. In addition, it should focus on training learners through practicing the same activity with different possible inputs/outputs, exposing them to think about their data from different perspectives, such as deconstructing a visualization rather than building one.

2.5.5 Supporting Self-Regulated Learning in Educational Efforts

One of the key strategies to a successful learning process is self-regulated learning [273]. Self-regulated learning involves the ability to regulate their learning productively. Supporting self-regulated learning generally relies theoretically on metacognition in learning theory. Metacognition differs from cognition. That is, cognition describes the cognitive skills necessary to perform a task. Whereas metacognition describes the awareness, the understanding of one's own thought processes, how one learns, and how a task is performed [100]. It plays an important role in learning, including how we capture, process, and apply information. Metacognition contains two main components: metacognitive knowledge and metacognitive regulation [234]. Metacognitive knowledge describes the level of awareness one has about one's own cognitive process, how one learns, and what techniques are most effective for learning. Metacognitive knowledge includes:

- Self-knowledge: this is where a person understands their own learning processes along with their strengths and weaknesses.
- Task knowledge: is when a person understands the nature of the task and the demands it places on the learner.

- Strategy knowledge: is the knowledge about different learning strategies and techniques and when to use them.

Metacognitive Regulation [234] regulates one's cognitive processes by planning, monitoring, and evaluating. It includes the following processes:

- Planning: knowing and deciding how to approach a learning task by setting goals and selecting strategies.
- Monitoring: the learning process and the performance by tracking one's understanding through suitable activities (e.g., asking oneself questions to assess understanding).
- Evaluating: through assessing the performance of one's progress and how effective the strategy is after completing a task. This reflection contributes to the future application of strategies chosen and tasks applied.

Metacognition is crucial to raise learners' self-awareness and enable them to take control of their learning process [119]. That is by assessing their level of understanding and identifying which areas need more work and focus. Thus, designing education materials and guides that promote self-regulated learning and foster metacognitive skills should encourage learners to reflect on their learning process, ask questions about their understanding, and discuss strategies for learning and problem-solving. Dunlosky et al. [85] described how designing self-regulated material is achieved by *"explaining how new information is related to known information, or explaining steps taken during problem-solving"*.

Supporting self-regulated learning in network visual exploration plays a crucial role in designing guides and transforming informal and unstructured learning into structured learning. As the analyst will be using guides to create and explore network visualization, they should be able to apply instructions and navigate through guides solely. There is limited research on how to design effective guides for network visualizations. Thus, to offer an effective learning experience, we investigate how to design a guide that supports a self-regulated active learning experience.

In summary, there are few studies focused on measuring and evaluating network visualization literacy[95, 48, 44, 54]. Network literacy can be taught and learned in one or more forms: representational literacy, topological literacy, and metaphoric literacy[282]. Thus, this thesis investigates, in particular, the network's visual exploration in the wild. It aims to investigate the barriers analysts face in visual network exploration (RQ1). And to do so, I have reviewed studies in information visualization and how they investigated understanding and/or supporting the visual exploration process using different methodologies such as: analyzing interaction logs [80, 106], characterizing guidance[62], creating guidance frameworks[250, 61], building

typologies[201] and designing knowledge model[251, 7]. However, most of those efforts are from the design perspective, focusing on tools, visualizations, and interactions rather than the exploration process as a whole. Thus, this thesis aims to find how to capture what the analysts are exploring in their networks in the wild (RQ2). That is, by examining the applicability of some of those methodologies in the wild (e.g., logs analysis) and by looking into new means to evaluate tools and understand analysts' needs. Such investigation shall help us understand how to characterize analysts' strategies in the network visual exploration process (RQ4).

Furthermore, to support analysts, I reviewed current literature on how guidance is provided to investigate how to enhance the design of network visualization tools (RQ3) and support education in information visualization (RQ5). I found that guides can be designed for knowledge acquisition or training skills. Guidance in network visual exploration can be provided through different forms and mediums. The guidance provided should be characterized in the matter of context, degree, domain, and target [250, 61]. In this thesis, I investigate what sort of guidance is needed for educational purposes to support network visual exploration in the wild. This thesis aims towards guidance that supports the visual network exploration as a process and guides analysts step by step. Thus, guidance must be a process that is user-centered and should allow analysts to choose how to be guided and where. In addition, most of the efforts reviewed focus on a specific step from the exploration process, such as choosing a visualization [106] and interacting and exploring [80]. This thesis aims to provide guidance in each step, giving a general overview of the process and detailed information when needed (e.g., smaller operations). The desired guidance should be based on the understanding of the current literature state and current analysts' needs that tools lack. In the following section (chapter 3), I will discuss the methodology I will follow in my thesis and the outline of how studies are conducted.

Chapter 3

The Methodology

As discussed in the literature (chapter 2), there are limited studies on how analysts explore network visualizations in the wild. Studies were mostly done within controlled settings (e.g., inside labs or pre-designed tasks). This has taught us little about how analysts use visualization tools to explore their own data toward their goals. Thus, this thesis investigates the barriers analysts face and struggle with through their visual network exploration (RQ1:chapter 4). In addition, the thesis investigates how we can capture what analysts are doing through their exploration process in the wild (RQ2:chapter 3). The thesis aims to find a new approach to evaluate the exploration process and the analysts' advancement. Also, to find to what extent visualization tools support the network's visual exploration. The findings should contribute to designing solutions or interventions that can enhance the design of visualization tools to support analysts in their visual exploration of networks (RQ3:Dashboard chapter 5 & Roadmap chapter 6). In addition, this thesis investigates how we can characterize strategies followed by analysts through the exploration process (RQ4:chapter 6). And finally, how can we teach or prompt self-regulated learning of visual exploration of networks through the tool (RQ5:chapter 6,chapter 7).

Thus, to address the thesis research questions, this thesis examines five methods that complement each other and assist in understanding the analyst's exploration process. By *analysts*, I mean tool users with/out datasets and/or research questions who aim to explore networks visually for various goals (e.g., learn a tool). I first describe the Vistorian as the network visualization tool adopted in this thesis's studies (section 3.1). As Figure 3.1 shows, each method was applied according to chronological order and accompanying sessions of various types (e.g., tutorials, interviews). In each section of this chapter, I describe one of the approaches, its effectiveness, pros and cons, and how it led to the following approach - if any - as Table 3.1 outlines.

The first method was *logging and analyzing analysts' interactions* due to its ease and because it allows us to reach people remotely (section 3.2). To enhance the logging process and understand more about analysts' goals, we offered *networks state capturing and annotation* tool

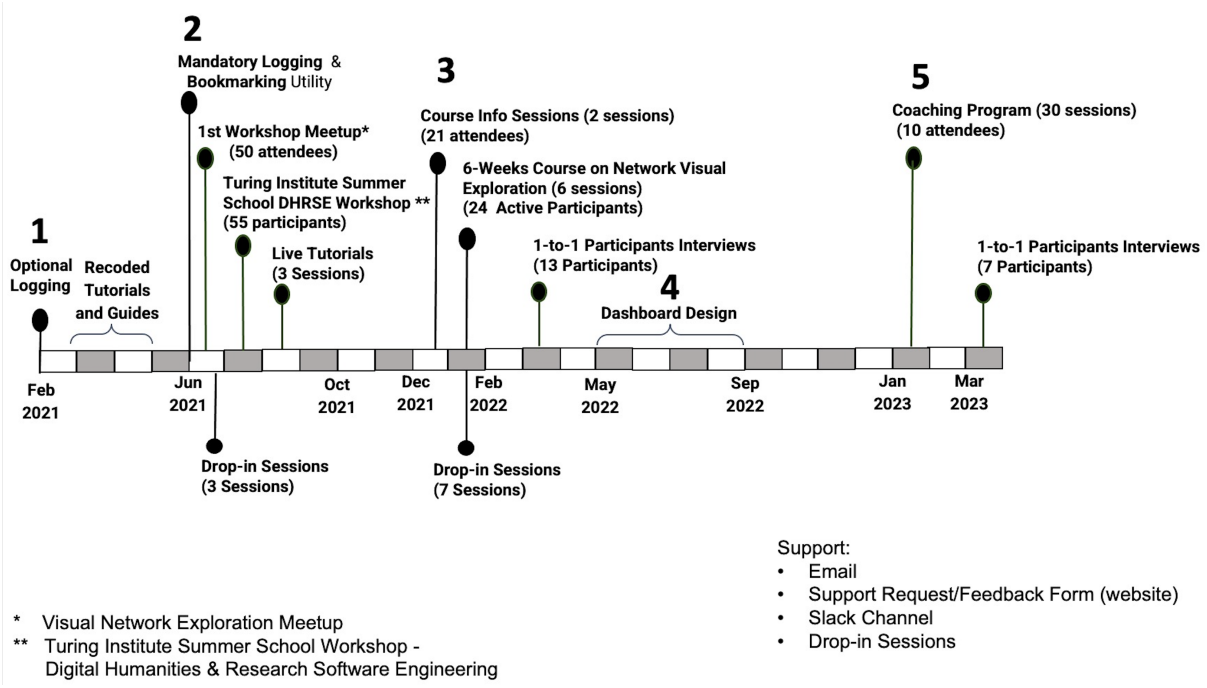


Figure 3.1: Approaches and Data Collection through our Studies Timeline

(section 3.3). This approach allows analysts to capture which patterns caught their attention and reason about it, which we call bookmarking. Although bookmarking showed analysts' interest in such a feature, the logging process did not allow us to know more about the analyst's intentions, why certain actions are taken, and what hinders them through their exploration. Thus, we ran our 6 weeks to work with analysts closely and learn more (section 3.4). The 6-week course enabled us to learn more about analysts' needs and some of the barriers they face. Based on it, we decided to read more into our logged events and extract metrics to describe the exploration process through an analytical dashboard (section 3.5). Finally, we used our findings from the previous 4 methods to create a step-by-step guide to accompany analysts through their exploration process, which we call a roadmap (section 3.6). We also ran a coaching program to observe closely how analysts use the roadmap, what assists them, and how such a guide can promote self-regulated learning. For each method, I start by describing the data collection process, the data analysis process, the participants' demographics, and the recruitment process. I conclude each method's subsection with an overview of the advantages and disadvantages of each it, explaining why one method led to the following one.

Approach	Goal	Main Outcomes (O) and Limitations (L)	Discussion Chapter
section 3.2: **Logging Interactions**	Initial analysis of the analysts' visual exploration process	**O:** Logs categories, Users types and where users get hindered **L:** Cannot reason analysts' decisions	chapter 4
section 3.3: **Bookmarking**	Capture analyst's intention	**O:** Analysts' interest in bookmarks and their usefulness (capturing, annotating and exchanging bookmarks) **L:** Limited understanding of analysts' exploration decisions	chapter 4
section 3.4: **Intensive Course**	Understand problems analysts face in (1) network creation (2) visual exploration of networks	**O:** Offers an in-depth understanding of users' needs and profiles **L:** Cannot be applied on a large number of analysts	chapter 4
section 3.5: **Analytical Dashboard**	In-depth analysis of logs to support cross-application understanding of analysts' usage	**O:** Offers live readings of tool usage by different user types, visualizations, and resources **L:** Doesn't explain some users' decisions in their exploration	chapter 5
section 3.6: **Coaching Program**	Understand how to support analysts of visual exploration of networks in (1) supporting self-regulated learning (2) offering systematic, detailed guides	**O:** Provides a self-regulated architecture that supports the exploration process and provides an understanding of it **L:** To get the best of it, strategies and steps should be automated	chapter 6

Table 3.1: Approaches in Capturing and Analyzing Analysts' Visual Exploration of Networks

3.1 The Vistorian: The Network Visualization Tool

To accomplish the goals of this thesis and investigate its research questions, I adopted a single network visualization tool *The Vistorian* (`https://vistorian.net/`). The Vistorian is a state-of-art and open-source network visualization web-application with four interactive visualizations: node-link diagram (Figure 3.2), adjacency matrix (Figure 3.3), timeline [81] (Figure 3.4), and a map with overlaid nodes and links (Figure 3.5). In addition, a coordinated or a side-by-side view of a node-link diagram and an adjacency matrix offers a synchronized exploration experience (Figure 3.6). The Vistorian was conceived as a means to provide, pilot, and evaluate complementary interactive visualizations for temporal, geographic, and multivariate network visualizations [17]. It supports visual encodings for link direction, link weight, node and link type, as well as multiple links. It also supports time for nodes and links. In subsection 3.1.1, I describe the visualizations provided by the Vistorian and their interactive features. In subsection 3.1.3, I describe how to create a network using the Vistorian and what its network schema can support. In subsection 3.1.4, I describe the workflow in the Vistorian toward creating a visualization.

3.1.1 Visualizations and Interactions

Each visualization has a legend showing node and link types as colors and a search box for text search. The coloring and visibility of nodes and links can be determined through the legend. If the network has a temporal component, a time range-slider is shown with each visualization to specify the visibility of elements present in a specified period. Individually, each visualization has a set of specific visual encodings and parameters that can be defined by a user.

The **node-link diagram** (Figure 3.2) uses the force-directed [99] layout implemented by D3 [46] and provides small sliders to set node size and opacity, link width and opacity, and the distance between links (lines) connecting the same two nodes (a value of 0), visually overlays all those links, collapsing them into a single line. In that case, increasing link transparency allows comparing the number of links across node pairs. The node-link layout stays fixed when filtering by a time range, helping to preserve the user's mental model. When a node is clicked, the ego network of that node is viewed, where other non-linked nodes will be dimmed using the same layout [46].

The **adjacency matrix**(Figure 3.3) allows for pan and zoom, keeping the positions of node labels fixed at the sides of the screen. A cell encodes link type (color), link weight (opacity), multiple links (splitting the cell into vertical segments of equal size), and link direction (a gradient from white to color pointing up encodes a link pointing at the node in the column, a gradient from white to color going left to right points to the node in the respective row). The adjacency matrix allows ordering nodes in the matrix by the following algorithms: similarity,

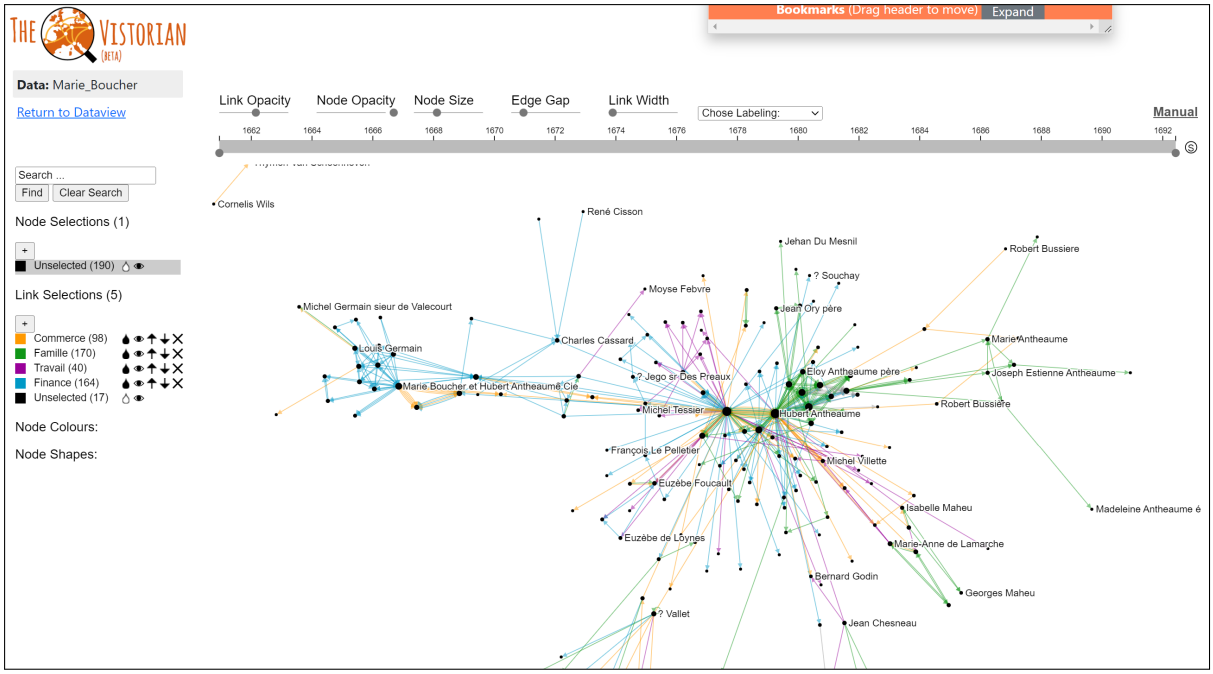


Figure 3.2: **Node-link Diagram** Visualization in the Vistorian

alphanumerical, reverse alphanumerical, and node degree.

The **timeline visualization** shows nodes along a vertical axis and links between nodes as arcs along a horizontal time axis (Figure 3.4). A user can scroll down the list of nodes or zoom into a time period through the time range slider. Clicking a node keeps only this node's neighbors visible, effectively compressing the view.

The **Map visualization**(Figure 3.5) overlays a node-link diagram onto a geographic map (Google Map API). Nodes with the same geographic position can be shown in a circle layout to make all their links visible. The circle radius can be specified by a small radius, allowing the complete collapse of all nodes into a single visual node or a circle that shows all nodes without overlap.

The **coordinated-view** (side-by-side)((Figure 3.6)) of node-link diagram and adjacency matrix. Interactions in such a view are coordinated through brushing and linking. It allows analysts to explore networks by complementing their findings and interpretations of provided visualizations.

Each visualization in the Vistorian runs in its browser tab, allowing tabs to be arranged in a split-view or across multiple screens to access more than one visualization simultaneously. Interactions are synchronized across visualizations, such as highlighting (brushing-and-linking), filtering time through a time-slider, and filtering the visibility of node and link types; this is implemented by using the browser's LocalStorage to share state.



Figure 3.3: **Adjacency Matrix** Visualization in the Vistorian

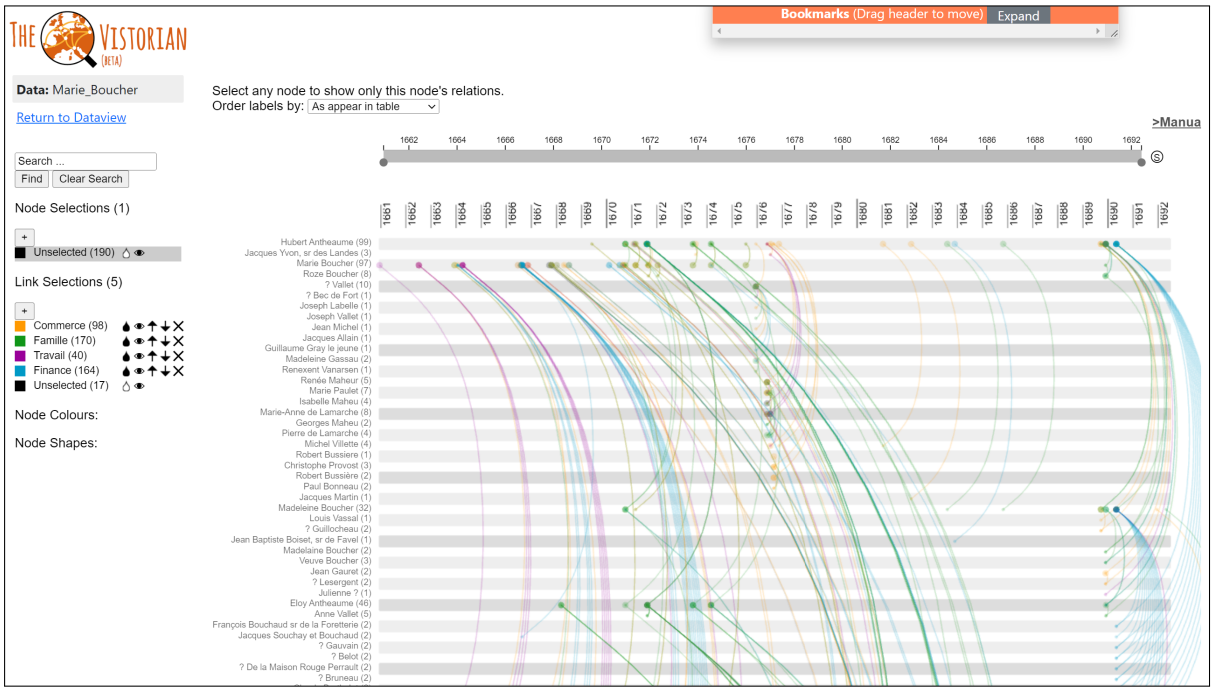


Figure 3.4: **Timeline** in the Vistorian

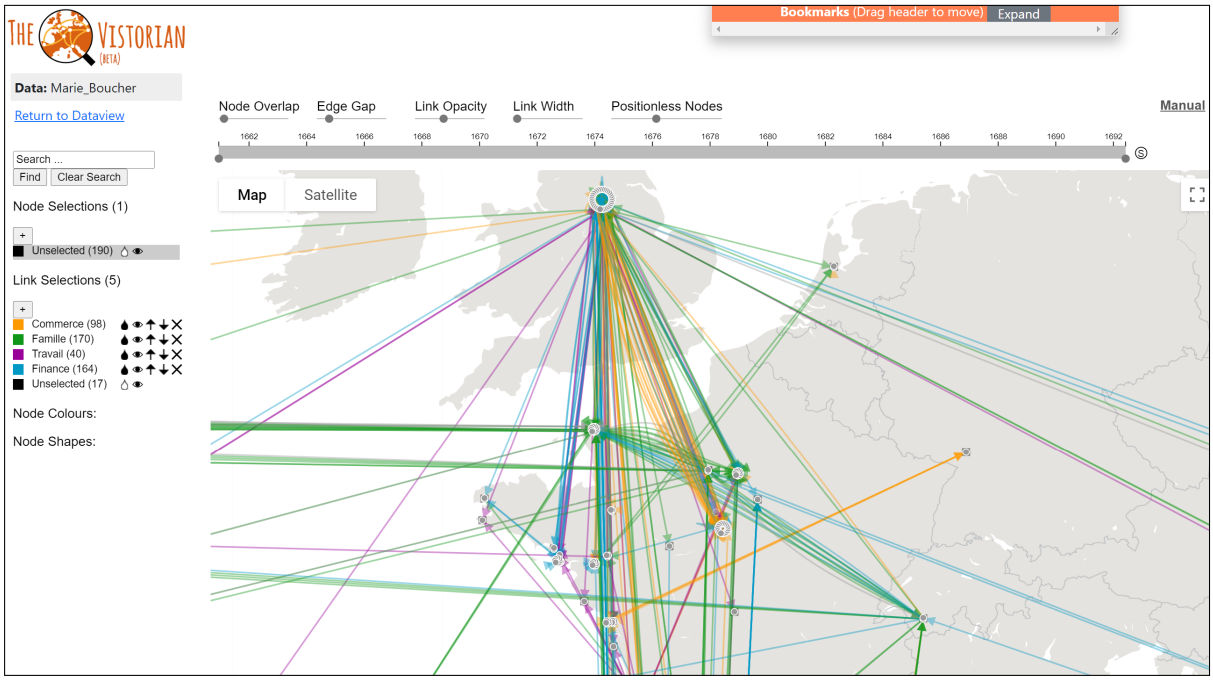


Figure 3.5: **Network over Map** in the Vistorian: map with overlaid links and edges

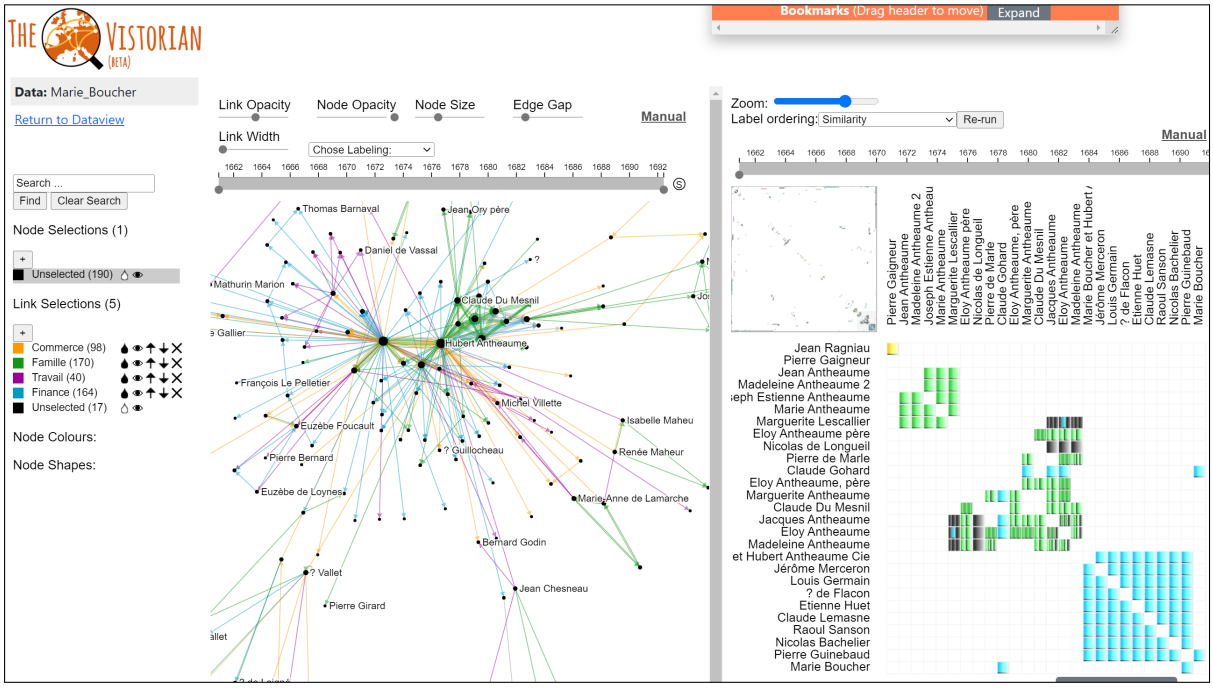


Figure 3.6: The Vistorian **Coordinated View** of node-link diagram and adjacency matrix

3.1.2 The Vistorian Visualizations and Data Encoding

Coloring schema: used for node/links are assigned by the Vistorian. That is, once links/nodes types are specified by the analyst during the network schema setup, the Vistorian assigns colors automatically to each type. Analysts are able to change colors by clicking on the color box next to the link/node and choosing from the color-picker palette.

Directionality is represented in node-link diagrams, timelines, and maps using directed arrows. The adjacency matrix represents directionality as a gradient square directed from the source to the target nodes.

3.1.3 Network Schema Specification Requirements

As a first step toward creating any network visualization (subsection 3.1.1), a *network schema* should be specified. A **Network Schema** is the data structure that describes the network's nodes and links along with any additional information. A network schema should include at least two attributes, which are node1 (source node) and node2 (target node), and how they are linked. The additional attributes can describe nodes and/or links further[1]. For nodes, in the Vistorian, we can specify the node type (String) and/or location (coordinates/location address). For links, we can specify link type (String), weight (Integer), time/date, ID (positive Integer), and link direction (True/False). The network schema specification is crucial for analysis and exploration purposes. Different schemes of the same data set can answer different research questions and offer different insights about the data.

The Vistorian supports creating network schema from tabular data format as well as common network file formats (e.g., GEDCOM, PAJEK [82], or GraphML (XML)). The network schema can be created using a *node table* or a *link table*. The *network schema* supports the following attributes for each data table type. In the *node table, each row* describes a node along with all of its links as relations. No link weight or location can be described in a node table, as each row might have multiple links. The ***node table*** supports the following specification:

- **Node Label** (*required*) : representing the node name
- **Nodes Relation(s)**: this one or more attribute represents the nodes linked to the specified node as a relation/link. At least one relation is required.
- **Node Location**: represent the geo-location of each node.

[1]. P.S. Attributes specification - basic or additional - can vary from one tool to another.

In the link table, each row describes a single link between any two nodes. This can also include the description of the link weight, type, direction, ID, and nodes' locations. The ***link table*** supports the following attributes:

- **Node 1/Source Node** (*required*): first node in a link or a source node if directed.
- **Node 2/Target Node** (*required*): second node in a link or a target node if directed.
- **Link Type**: the type(s) of links connecting any two nodes.
- **Link Weight**: an integer value representing the weight of a link specified by the analyst.
- **Link ID**: a positive integer value representing the ID of a link specified by the analyst. This is used in cases where a specific link value changes over time, for example (e.g., link weight).
- **Node 1/Source Node Location**: the geo-location of the first node (e.g. city, country)
- **Node 2/Target Node Location**: the geo-location of the second node in a link.
- **Timestamp/Date:** date/time related to each link (e.g., Contract date). The Vistorian supports a list of short/full-time/date formats.

The *link table supports additional attributes* including:

- **Latitude/longitude coordinates**: The analyst is able to upload the value of the latitude/longitude coordinates to be used as nodes' locations instead of the coordinates auto-extracted by the Vistorian using the MapTiler API [168].
- **Node Types**: The Vistorian supports specifying node types by uploading node types data files.

3.1.4 Network Creation, Data Import Interface & Wizard

Initially, the Vistorian relied on a single page for data import and network schema specification. Through this page, analysts upload their data in table format. There are two types of tables inspired by NodeXL [247]. As Figure 3.7 shows, the analysts upload their data file as a node table or a link table. Then the analyst specifies the network schema (e.g., source node, target node, links types,..etc.) as Figure 3.8 shows. Like NodeXL [247], the analyst can specify the network schema by selecting a role (Source Node, Target Node, Link Weight, etc.) for each column using drop-down menus. The possible options to specify for a network schema in the Vistorian are described in subsection 3.1.3.

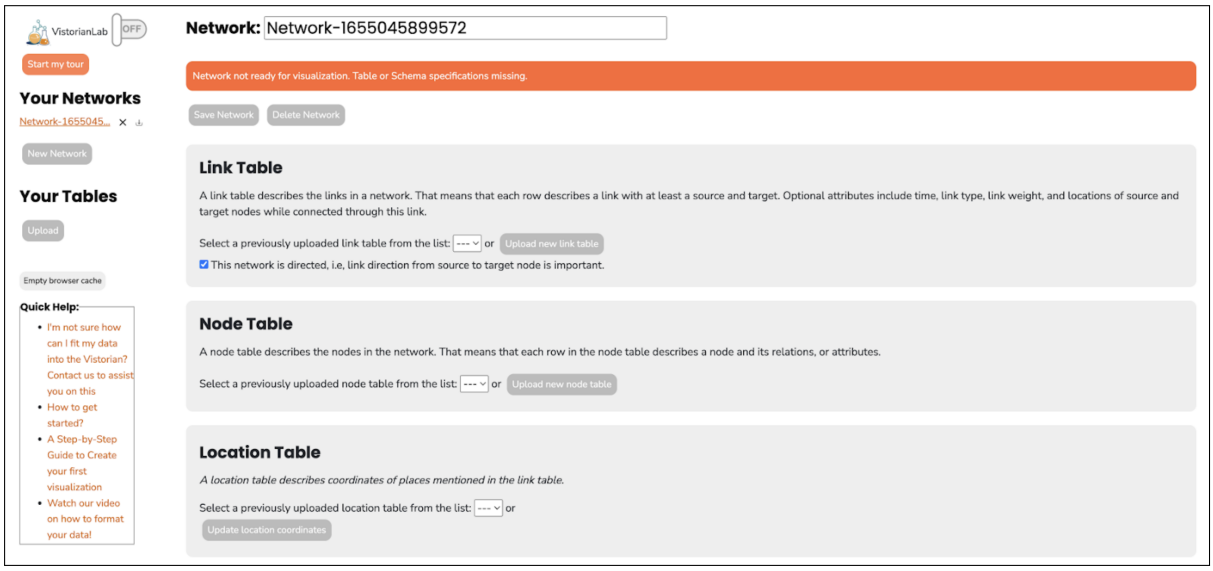


Figure 3.7: Vistorian Previous Single-Page Network Schema Setup (*Uploading Files*)

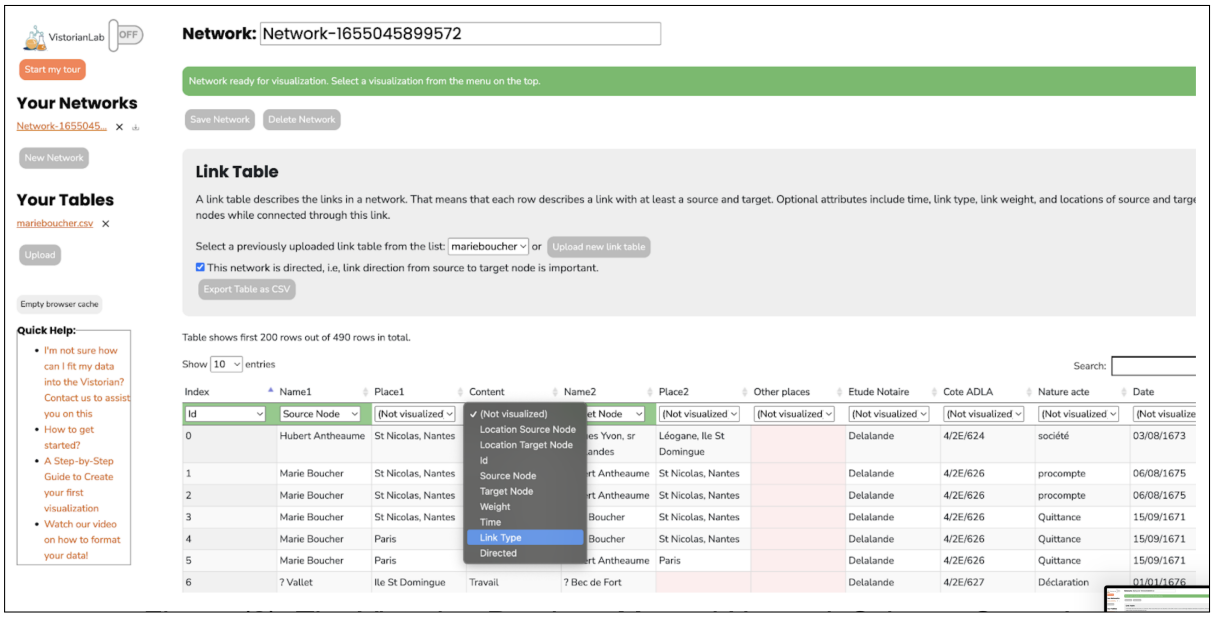


Figure 3.8: Vistorian Previous Network Schema Setup Interface (*Specifying Schema*)

After several months of logging analysts' interactions and workshops, logs showed that some analysts could not decide which table format suits their data to create a network, as discussed in section 4.4. We utilized our findings on potential analyst struggles during exploration to enhance the user experience and design of the Vistorian. Thus, to assist analysts and guide them through network creation and formatting, the Vistorian transformed its network creation interface into a wizard (see Figure 3.9). The wizard walks analysts step-by-step until they create their desired network. This helps analysts to understand the impact of their decisions at each step. The wizard's steps as Figure 3.9 shows are described here:

1. Enters the network name,
2. Choose the data file format to be uploaded (e.g., tables, GEDCOM, Pajeck, etc.)
3. Choose the type of table: a link table or a node table
4. Upload table and specify the schema for that table (Required: node1 and node2, Optional: e.g., link weight, time)
5. the wizard asks for optional data such as location table and node types table. In this step, the user can upload an additional file that contains latitude and longitude geo-coordinates of locations if location names were specified in step (4). This is an optional step as The Vistorian can automatically extract geolocation coordinates using geo API [168].
6. Network visualizations are ready! Choose a visualization by clicking on it, and it will be launched in a new tab.

Both the 'single-page' and 'wizard' import methods saved data in the browser's LocalStorage, avoiding the need to upload private data to a server. Analysts can share imported networks by exporting them into a proprietary `.vistorian` format.

3.1.5 Types of Help & Support Resources Provided in the Vistorian

The Vistorian offers a variety of help resources that deliver different content through different mediums. This includes the following: **Data Examples** are examples of datasets formatted to be used to create networks (e.g., node/link tables), start exploring it, and learn more about the tool/concept. **Data Formatting** is a detailed reference on how to prepare data for network creation and exploration purposes. **Demo** is a live interactive example of network visualizations of demo data. This enables the user to try the tool and interact with it directly. This removes the burden of preparing data and creating a network in order to try the tool. **Network Manual** is a detailed reference about creating networks using the tool and how to address different needs. **Step-by-Step Video** is a video guide on how to perform specific tasks using the tool. **Tutorials** are textual step-by-step guides on how to perform specific tasks using the tool. **Visualization Examples** are different examples of network visualizations and how to read them. This provides users with further insights about how to analyze data.

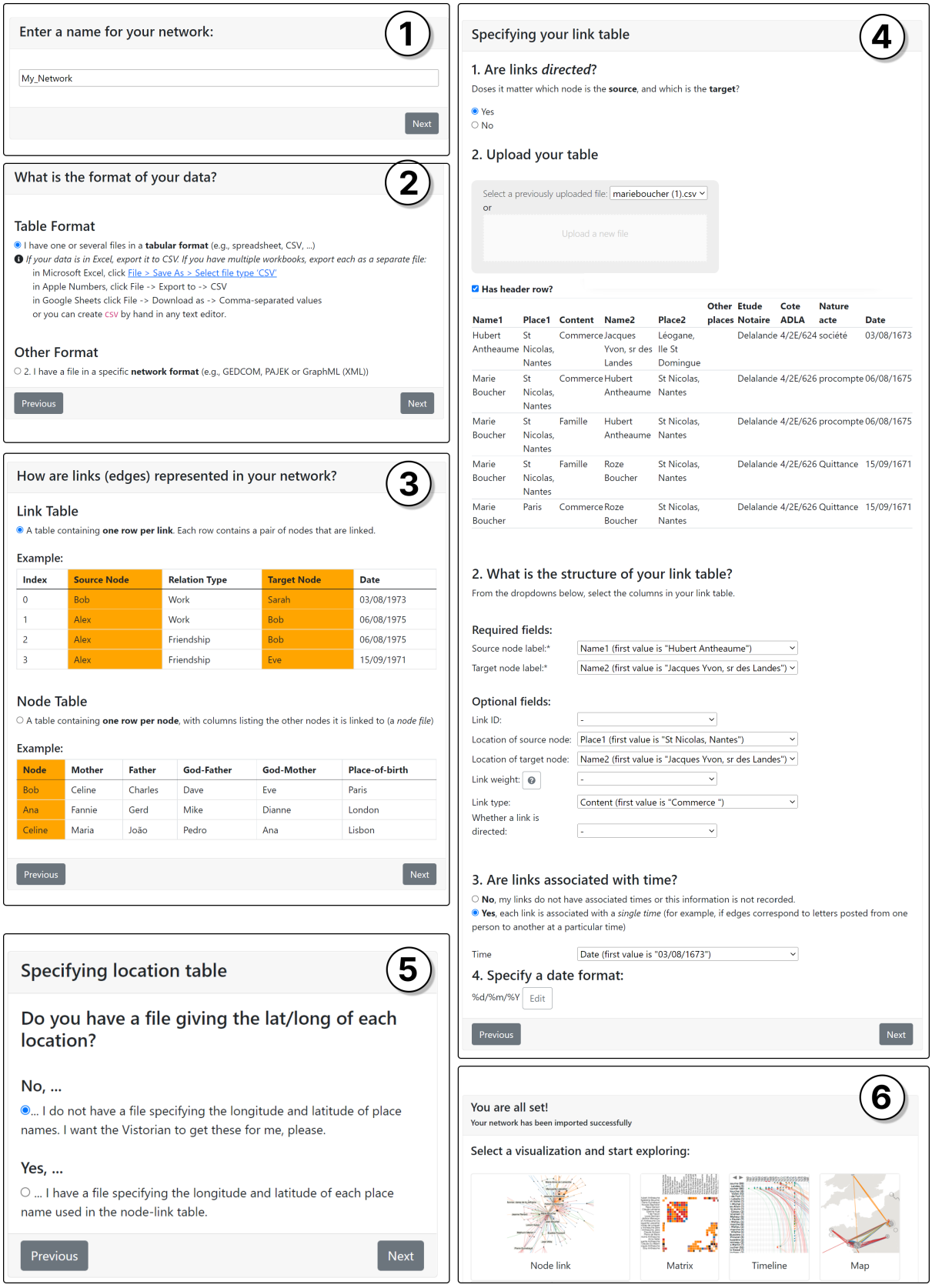


Figure 3.9: The Vistorian New *Step-by-Step Wizard* of Network Schema Setup

3.1.6 The Support of Temporal and Geographical Data Exploration

The Vistorian supports exploring different data types such as **geographical and temporal data** through its *network schema setup, the type of visualizations and features* it provides. For example, dynamic changes can be explored *temporally using a visualization* as in the timeline visualization, where it shows when and how nodes connect through time. Another way to examine temporal changes is through using *interactive features* such as *the time-slider feature*. It allows exploring how networks dynamically change from different perspectives across visualizations over time. For example, how clusters change in an adjacency matrix through time or how nodes' distribution changes over the map. The Vistorian supports temporal data in any format (e.g., full/short date/time), which is important in supporting visual exploration.

In addition, the Vistorian supports visualizing networks that contain *geo-locations* on maps. This includes, for example, providing cities and countries where the Vistorian will automatically extract their coordinates using the MapTiler API [168]. Or an analyst can provide their own table with coordinates mapped to places mentioned in the link/node table. Deciding upon how to use nodes' or links' locations is a decision to be taken based on the goal of the exploration or the research question to answer. This holds many challenges, such as how to manage the scale between displaying nodes on a global scale side-by-side to nodes condensed in small areas. Locations in the link table can be expressed in place names, such as `Paris`, and are resolved by The Vistorian using the MapTiler API [168]. To the analysts, the challenge was not how to create a network through the Vistorian; it was prior to that what to select as a node, link, or any network-related value. This is not limited to the Vistorian itself but to any visualization tool. The matter of enabling analysts to choose the suitable network schema from a possible set of network specifications.

3.2 First Approach: Logging Analysts' Interactions

Analysts' interactions with a visualization tool are influenced by their intentions, but these interactions may not always accurately reflect their true objectives. For instance, if the analyst randomly clicks buttons due to unfamiliarity with the tool or if the tool itself is inadequate or poorly designed, the recorded interactions might not accurately represent the analyst's actual intentions. Thus, as a first step toward capturing what analysts do in the wild, I described possible analysts' interaction events with the tool and their data attributes (subsection 3.2.1). Then, I classified all possible interactions with the Vistorian into one of the top-level categories based on their purpose (subsection 3.2.2). Here, I describe how interaction events are processed for data collection and analysis purposes as Figure 3.10 illustrates:

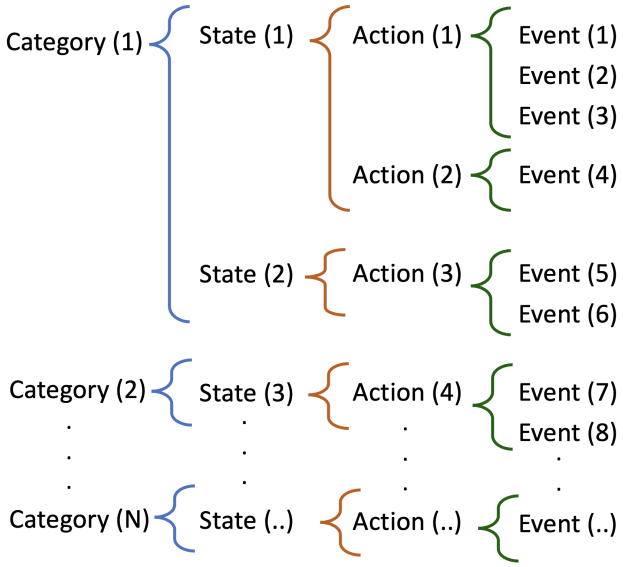


Figure 3.10: Hierarchy of Interactions Logs Classification

- An analyst can perform different sorts of interactions with the tool; we call each interaction *an event*. For example, selecting a data attribute to specify a network schema or creating a visualization.

- Each *event* or group of events represents *an action* performed by an analyst. The action describes what the analyst actually did at a higher level. For example, in Figure 3.11, uploading a link table is an event, and uploading a node table is a different event. However, both events represent the action of *uploading a data file* by the user to create a network schema.

- Then, all *actions* are grouped based on what *state* the actions' belongs to in the exploration process. Each state represents a specific step toward creating and exploring a visualization. For example, uploading a data file for the first time is an action. Selecting one of the previously uploaded files is a different action. However, both actions represent the *import data* state.

- Finally, all *states* are grouped and classified into one of the top-level *categories*, each *category* represents a different phase of the exploration process. For example, importing data, exporting data, and shaping data as in Figure 3.11 are states that belong to a single category that is called *data management*.

This hierarchical classification ensures that each interaction is appropriately labeled as subsection 3.2.1 shows, facilitating detailed analysis and understanding of analyst behavior as subsubsection 3.2.3.3 explains. This structured approach not only aids in accurately capturing

user interactions but also enhances the overall data analysis process. This classification is essential to describe user interactions needed to support process mining [102, 169]. Process mining is a data analysis technique used to extract insights and infer valuable knowledge from event logs to visualize and analyze the flow of processes [102, 169]. Process mining allows for uncovering typical user behavior patterns, identifying process bottlenecks, and detecting discrepancies or variations between actual and intended processes. [102]. In other words, we use it to understand more about the users' intentions, what they are trying to do, and how far they have progressed in the exploration process.

3.2.1 Interaction Log Events Description

The following basic information describes each interaction event:

- **Event Name**: is the name of the atomic unit of events triggered by an analyst interaction with the tool. The name of each event describes its purpose.

- **Event ID**: a unique identification code assigned to each distinct event. The following convention forms each event ID: starts with three letters describing the top category the event belongs to (e.g., dat denotes to the data management category), followed by an underscored (_), and ends with a unique number to distinguish events under the same top-category. Some codes may contain an additional two letters to distinguish further between events. For example, in Figure 3.11 (dat_2_LK) means that the link table was uploaded.

- **Action**: describes the goal of each interaction event. Several different events can belong to the same action as they describe their purpose. For example, in Figure 3.11, all events of (dat_2_*) means that the user has uploaded a data file of any of the possible types (e.g., link table, node table, location table).

- **State**: provides a higher-level of description about which phase if the visual exploration of networks the user is in. States provide insights into users' intentions and their objectives. This understanding can be achieved through process mining and by identifying the sequence of states navigated by the analyst. For example, in Figure 3.11, any changes in specifying network schema in any form are represented as the data shaping state.

- **Category**: is the high-level category describing the user's interaction. For example, in Figure 3.11, various (dat_*) events describe interactions related to managing data (e.g., uploading data or specifying schema).

Some interactions are described by additional information for analysis purposes (e.g., process-mining), such as:

- **Preceding Event ID**: if provided, it describes the value of the event that should/might precede the currently described event. This field *is optional and depends on the design of the visualization tool*. That is, determining what log types that precede another is a tool-dependent process. For example, in Figure 3.11 before specifying a network schema attribute (dat_11), a data file should be uploaded (dat_2_*) or a pre-uploaded file should be selected (dat_3_*).

- **Event Attribute**: describe additional values of the event. For example, in Figure 3.12, this attribute describes the type of the visualization being created (ID: vis_1). This means it can be any of the possible visualizations or views offered by the tool (e.g., node-link, timeline, coordinated view, .. etc).

3.2.2 Interaction Log Events Categories

Log event types depend on the visualization tool and possible analyst-application interactions. Such as interactions with the interface include data processing, visualization creation, and exploration features.**The logged events were classified to the following *seven main categories***:

- **Data-Management Events Logs** include logged analyst-application interactions related to how analysts create, specify, and update their network schema as Figure 3.11 shows. For example, events of creating a new network, uploading data tables, updating table entries, specifying (adding, removing, changing) network schema, or exporting data. Those logs are important in understanding what sort of data analysts usually use, how analysts create their network schema, and how they change it over time.

- **Visualization & Interaction Events Logs** include logged events related to the creation or interaction with any of the visualizations as Figure 3.12 shows. For example, creating a visualization, setting the time range slider, filtering nodes and links, panning, zooming, highlighting nodes and links, etc.

- **Help & Support Events Logs** are those related to engagement with help resources provided as Figure 3.13 shows. For example, manuals, demo data, guide videos, and others.

- **Interest & Communication Events Logs** related to analysts approaching the developers to report a technical bug, provide feedback, and request a consultation as Figure 3.14 shows.

Category	State	Action	Event ID	Event Name (Atomic Interaction)
Data Management (dat)	Import Data	Create Network	dat_1	New Network Created
		Upload Data File	dat_2_DT	Upload new data table
			dat_2_LK	Upload new link table
			dat_2_NT	Upload new node table
			dat_2_LO	Upload new locations coordinates table
		Select Data	dat_3_LK	Select previously uploaded link table
			dat_3_NT	Select previously uploaded node table
			dat_3_LO	Select previously uploaded location table
	Delete Data	Delete Data	dat_5	Delete a Network
		Clear Cache	dat_6	Empty Browser Cache
	Export Data	Export Data	dat_7	Download network data file
			dat_8	Export CSV file
	Create/Update Network	Update Data	dat_9	Change a Network Name
			dat_10	Update Table's Data Cell
		Set Network Schema	dat_11	Specify Schema Data Attribute
			dat_12	Set Network as Directed
			dat_13	Update location coordinates
			dat_14	Extract locations from node and/or link tables
	Overview Data	Overview	dat_15	Search data tables
			dat_16	Sort data table by Header

Figure 3.11: Data-Management Events Logs: shows log types related to network creation, specification, and amendment.

- **Bookmarks Events Logs** relate to creating, updating, and using bookmarks and annotating them as Figure 3.15 shows. Bookmarks enable users to capture network visualization states, add their comments on the captured state, and share them with others by exporting them. Bookmarks are described in section 3.3 and in chapter 4.

- **Technical-Problems and Usability-Related Events Logs** tracked the list of technical errors raised during users usage of the Vistorian and supported debugging and improvement process (Figure 3.16).

- **Study & Activity-related Events Logs** related to the activity and engagement with tool. Such logs include events not included in other categories, as Figure 3.17 shows. For example, user actions include navigating between pages, user activity status, sharing feedback, support-request, etc.

Category	State	Action	Event ID	Event Name (Atomic Interaction)
Visualization (Vis)	Explore Visualization	Change Network Appearance	vis_4	Change Link Opacity Value (Node-link/Map)
			vis_5	Change Node Opacity Value (Node-link)
			vis_6	Change Node Size Value (Node-link)
			vis_7	Change Edge Gap Value (Node-link/Map)
			vis_8	Change Link Width Value (Node-link/Map)
			vis_38	Change Node Overlap Value (Map)
			vis_40	Change Opacity of Positionless Nodes (Map)
		Explore	vis_9	Change Labeling Type (Node-link)
			vis_10	Change of Search Term
			vis_11	Find in Visualization
			vis_12	Clear Search
			vis_13	Add Node/link
			vis_14	Show/Hide Node/Link
			vis_15	Show/hide Node Group Color (dimming)
			vis_16	Change a Node Group Color
			vis_17	Delete a Node Group
			vis_22	Show/hide Link Group Color (dimming)
			vis_23	Change a Link Group Color
			vis_24	Delete a Link Group
			vis_27	Change (Min Value) of Time-Slider
			vis_43	Change Time-Slider (Min and Max Values)
			vis_28	Change (Max Value) of Time-Slider
			vis_29	Click a Node
			vis_30	Hover over a Node
			vis_31	Click a Link
			vis_32	Hover over a Link
			vis_33	Remove All Data Filters
			vis_34	Change Zoom-Slider Value
			vis_35	Select Ordering Algorithm (Adjacency Matrix)
			vis_36	Re-Run Sorting button
			vis_37	Label Display Selection (Timeline)
			vis_42	Switch to Map Satellite View (Map)

Figure 3.12: Visualizations & Interactions Logs Events: shows logs related to creating and interacting with visualizations.

Category	*State*	Action	Event ID	Event Name (Atomic Interaction)
Help (hlp)	Fetch Help	Search Resources	hlp_1	Open Network Exploration Manual
			hlp_2	Open Data Formatting Reference
			hlp_4	Browse Visualization Example(s)
			hlp_5	Download Example Datasets
			hlp_6	Open Interactive Demos
			hlp_7	Open Visualizations Explanation
		View Tutorial	hlp_3	Watch Video Guides
			hlp_8	Open Step-by-step Tutorials

Figure 3.13: Help & Support Events Logs: show log types related to user support and guidance.

Category	*State*	Action	Event ID	Event Name (Atomic Interaction)
Intrest & Communication (com)	Communicate	Communication	com_1	Check GitHub Page (Development & Code)
			com_2	Contact Team
			com_3	Request a Team Consultiation
			com_4	Report Technical Bug

Figure 3.14: Interest & Communication Events Logs: shows logged interactions related to visualizations.

Category	*State*	Action	Event ID	Event Name (Atomic Interaction)
Bookmarking Tool	Bookmark State	Note-taking	bkm_1	Add New Bookmark
		Note-taking	bkm_2	Update Bookmark
		Note-taking	bkm_3	Specify BookmarkType
		Note-taking	bkm_7	Select/Filter Bookmarks view option
		Note-taking	bkm_8	Show/hide Bookmark window
		Export	bkm_9	Export Bookmarks file
		Explore	bkm_10	Import Bookmarks file
		Explore	bkm_5	Restore Bookmark state

Figure 3.15: Bookmarks Events Logs: shows logs interactions related to network visualizations state capture and restore.

Category	*State*	Action	Event ID	Event Name (Atomic Interaction)
Troubleshooting Issues and Problems (err)	**Report Issue**	Error Captured	err	**Capture Technical Problem Details** *(e.g. problem in :uploading a data file, creating a network, fetching a node location)*

Figure 3.16: Technical-Problems and Usability-Related Events Logs: include troubleshooting and usability enhancement events.

Category	*State*	Action	Event ID	Event Name (Atomic Interaction)
Study & Activity-related (log)	**Navigate**	Start	log_1	Land for the first time on any Vistorian page
		Switch View	log_2	Visit a specifc Visitorian Page
			log_3	Return to My Session Page (all networks)
			log_4	Return to Data View
		Extract Session ID	log_11	Reterive Session ID
	Consent	Managing Consent	log_5	Display Consent
			log_7	Download Consent
			log_9	Accept Consent
			log_10	Disable Logging
			log_11	Show/Hide SessionID
	Change Activity-Status	Interruption	log_12	Close Webpage (End Activity)
			log_13	No Activity Detected
		Switch View	1og_17	Focus on a Page
			log_18	Change Page (Page on Blur)
	Share Feedback	User-Initiated Feedback	log_14	Open Feedback Form
			log_15	Submit Feedback
			log_16	Cancel Feedback Submittion
		System-Initiated Feedback	log_17	User Submit Feedback of (No-Activity Prompt Pop-up Form)

Figure 3.17: Study & Activity-related Events Logs: shows logs related to the user's activity, navigation, and other study-related events.

3.2.3 Participants' Demographics, Data Collection and Analysis

This subsection discusses how the participants were recruited, their demographics, which data was collected, and how it was analyzed.

3.2.3.1 Participants' Demographics and Recruitment

An open email invitation was distributed to mailing lists, including those of the School of Informatics and the Humanities and Social Sciences Center at the University of Edinburgh, as well as to groups of analysts interested in data visualization, exploration, and analysis. Analysts were introduced to the tool and how to use it through workshops. In addition, help and training resources were offered through the Vistorian web page. Support and drop-in sessions were offered to assist analysts with various needs, such as data preparation, formatting, and troubleshooting issues. **Participants Demographics**— Users' interactions with the Vistorian were logged anonymously at the initial step, without any knowledge of their background or goals.

3.2.3.2 Data Collection

To understand how analysts use the tool, I ran several public workshops to introduce attendees to how to explore multivariate network visualization using the Vistorian and how to prepare their data. For example, one workshop was an online DataVis Meetup with approximately 25 participants. Another workshop was part of an online Digital Humanities Summer School with around 55 participants and others. Workshop participants were introduced to the Vistorian demos, step-by-step tutorials, handouts, and exercises (e.g., sketching networks). Using this method, I logged interaction events using the Intertrace JS library [93] and collected audience questions from the workshops and drop-in sessions. I have collected interaction logs over ten months as described in the logging process earlier in subsection 3.2.1 and subsection 3.2.2, and analyzed them as described in subsubsection 3.2.3.3. According to the Informatics Research Ethics Process, RT number 2019/20353, the collection process was done as part of the certified study.

3.2.3.3 Data Analysis: Extracting Exploration Steps from Interactions Logs

Analysts go through different states in their visual exploration process. Each state represents a specific step toward creating and exploring a visualization. Thus, we started by analyzing which steps and paths analysts follow and how they transition from one state to another. An analyst needs to pass through the minimum number of states for us to consider their interactions as *an exploration process* through our logs analysis. Those essential transitions are highlighted with

solid lines in Figure 3.18, which includes importing data, creating a network schema, creating a visualization, and exploring it. Additional states (dashed rectangle in Figure 3.18) can be visited and vary from analyst to analyst, such as fetch help, overview data, etc. States are colored based on their category, as described in subsection 3.2.2. The results of the logs analysis are discussed in chapter 4.

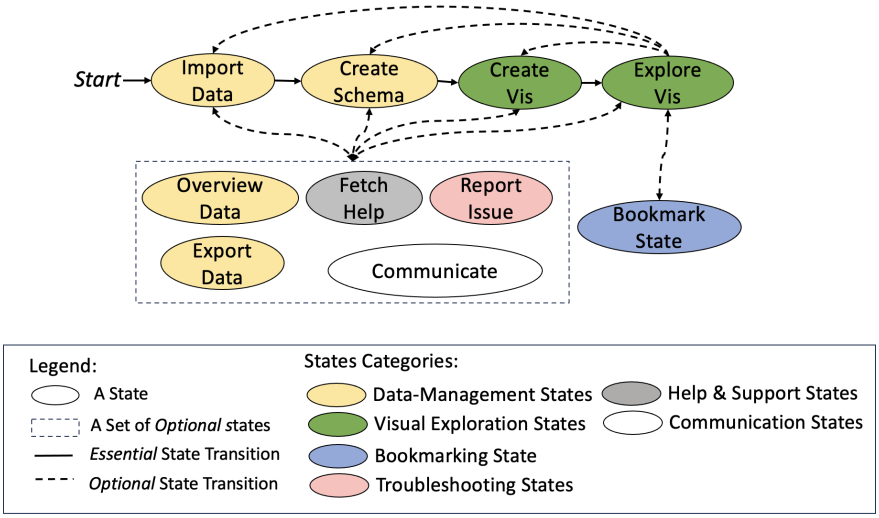


Figure 3.18: Process-Mining of Network Visual Exploration: through analyzing analyst's transition from one state to another.

Logging analysts' interactions has both benefits and drawbacks. While the interactions logging process provides valuable statistical information on tool usage and insights into analysts' behavior, it does not explain the difficulties analysts face during exploration. Moreover, logs cannot clarify why an analyst made certain decisions or specific interactions. To address this, I added a network state-capturing utility to capture analysts' goals through mini-questionnaire. This enables analysts to provide context and explain their interactions as I explain in the following section 3.3. Additionally, in collaboration with my supervisor, Benjamin Bach, we designed an intensive course to capture analysts' goals and investigate their needs alongside the logs, as I will describe in section 3.4. The questions and observations from the workshops have provided valuable insights that informed the design of the network visual exploration course. Although the logging approach laid the foundation for understanding some analysts' needs, it had limitations. Therefore, I transitioned to a method that allows analysts to annotate their visualization states and work closely with analysts through an intensive course.

3.3 Second Approach: Mini-Questionnaire and Network Visualization State Annotation

Logging and analyzing analysts' interactions cannot inform us about analysts' intentions and how they explore networks. Thus, we provided analysts with a mini-questionnaire to capture analysts' goals and understand the reasons behind their actions. The mini-questionnaire is provided as a part of this utility that I have added to the Vistorian to allow analysts to capture any network visualization state, annotate it, and save it as a *bookmark*. Bookmarks are data structures describing a specific network state that contain details about the network, such as time range, highlighted nodes, visible elements, colors, zoom level, pan position, etc. A network's state shows how it looks under a specific data filter or condition (e.g., date period, link type). In addition, the network state describes the network appearance based on chosen visual encodings (e.g., allowing nodes to overlap or how links look). Each bookmark features a timestamp, a title, and a description entered by the analyst. In the context of this research, bookmarks serve two objectives: (a) capturing users' journey [244, 41], and (b) prompting users to provide information about their overall task using a mini-questionnaire [239]. I extracted a set of possible user intentions from the initial mini-questionnaire used in the previous study [239]. Then, I refined and expanded the list of options through several iterations, ultimately allowing analysts to annotate their captured states. It was shown when a user created a new bookmark (Figure 3.19-left) and asked them what they were currently doing: *Analyze Data*, *Learn* (The Vistorian), *Demo to others*, *Test & Explore*, *Discuss Findings*, *Report an Issue*, and *Other, which prompted the analyst to specify their own goals using plain text*. So, how can such a utility support the network visual exploration process? This is what will be discussed next.

3.3.1 How the Bookmarks Work and Support Visual Exploration

The bookmark utility supports analysts in annotating their findings and reviewing their exploration history. Analysts can capture, annotate, and restore different states of their network(s). As shown in Figure 3.19 on the right, the bookmarks' timeline, and at the bottom, the utility offers means of exporting and importing bookmarks and other features (e.g., troubleshooting). The analyst can create a bookmark by clicking on the *Add New* button, which captures the network visualization state and shows a pop-up mini-questionnaire window as the one on the left side of Figure 3.19. Then the analyst can specify the following values: the bookmark label, own notes, the goal of the bookmark - how they characterize it-, and what sort of analysis they are performing. Each bookmark created is stored in chronological order and displayed as a timeline of bookmarks. The timeline's importance comes from its ability to show how findings were reached chronologically and logically (e.g., what was explored and how). A bookmark state can

be restored by clicking on its top-right restore button.

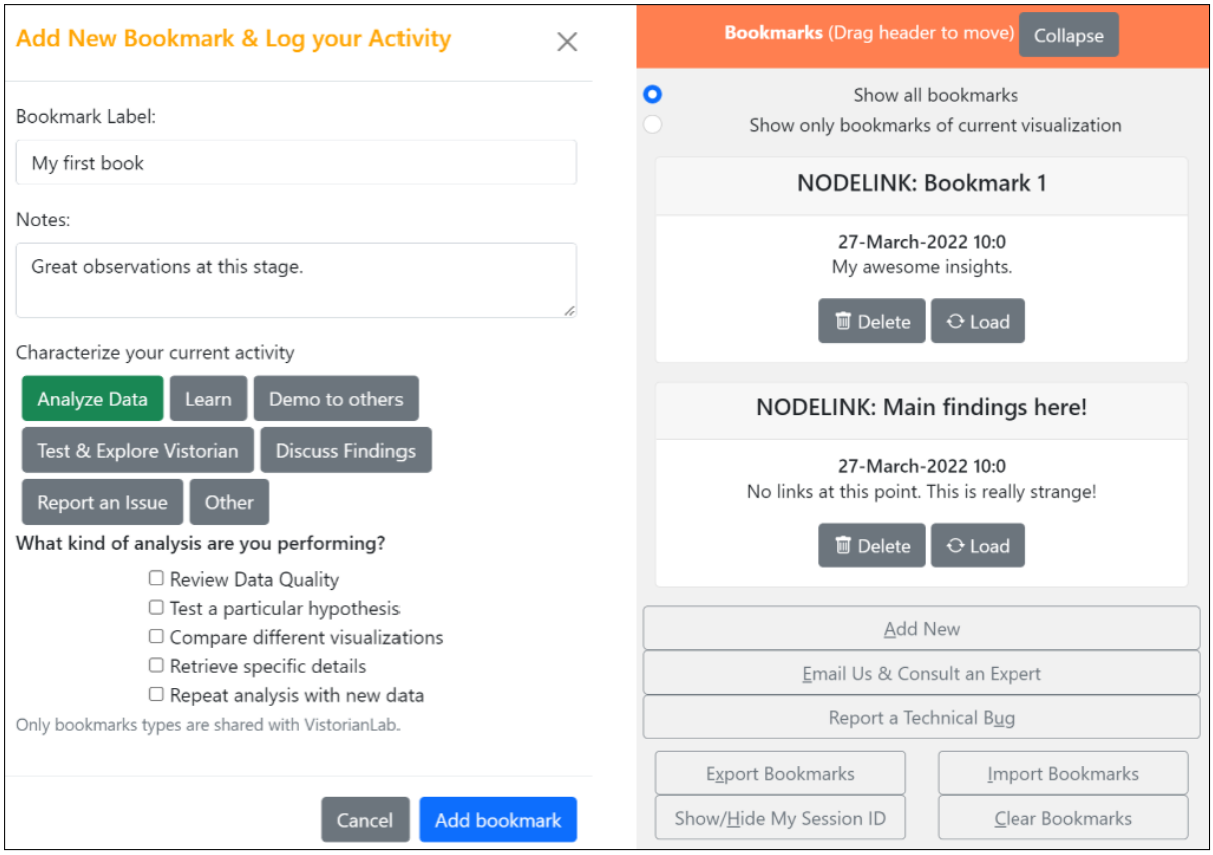


Figure 3.19: The Vistorian Bookmark Utility: the image on the (right) shows the bookmarks timeline from latest (top) to earliest (bottom) along with other features. The image on the (left) shows how analysts can describe a bookmark once created.

Figures from (3.20-3.23) demonstrate how bookmarks are used. In Figure 3.20, the timeline on the right shows the existing bookmarks from different visualizations (e.g., node-link diagram and adjacency matrix). It also shows the initial state of the demo dataset used. Then, in Figure 3.21, the restored bookmark on the right is expanded, and the state is restored on the left. We can see the filtered time range and its impact on the displayed links and nodes. Additional states are restored as (Figure 3.22, Figure 3.23) show and reflected on the network state. In chapter 4, I describe how bookmarks were used in combination with interaction logs (section 3.2) and the 6-week course (section 3.4) to understand more about the exploration process.



Figure 3.20: Screenshot(#1) showing a visualized dataset with its initial state on the left and the bookmarks timeline on the right.

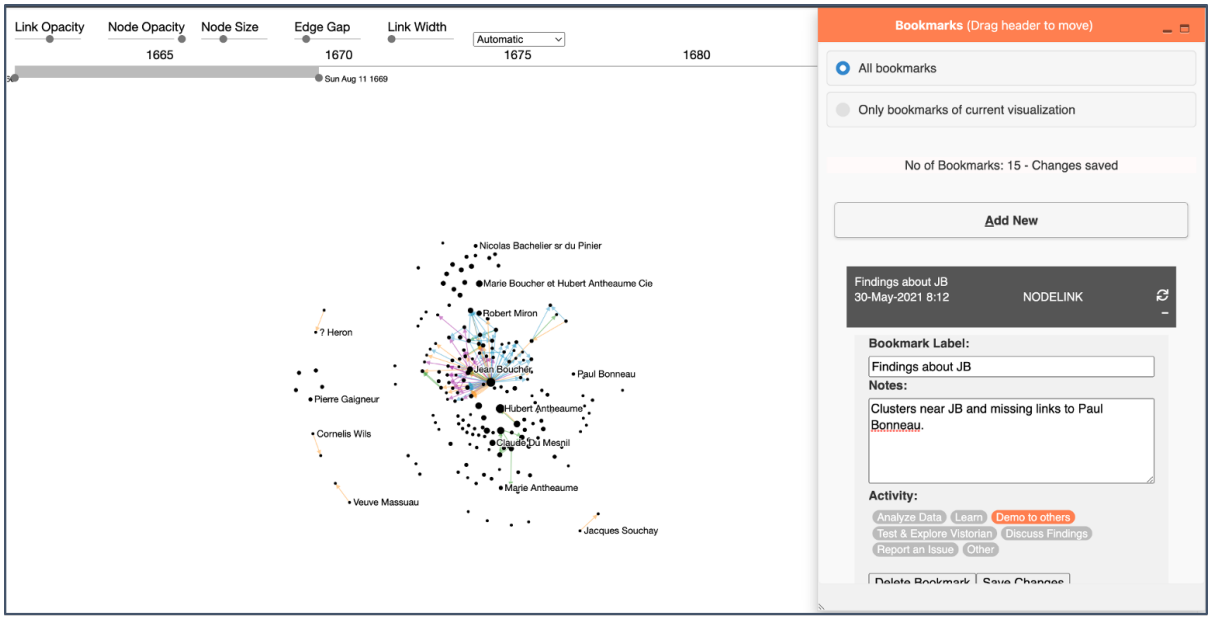


Figure 3.21: Screenshot(#2) showing a restored bookmark state on the left (filtered timeline) along with bookmark details on the right (e.g., notes, analysis goal)



Figure 3.22: Screenshot(#3) showing a second bookmark and its network state. It shows a different period.

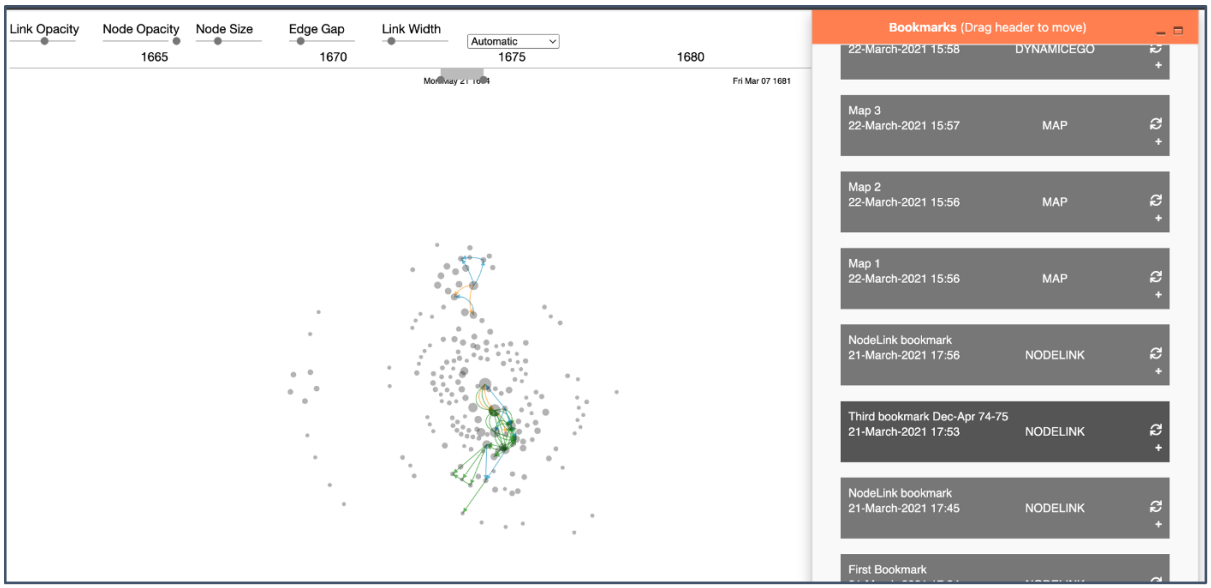


Figure 3.23: Screenshot(#4) shows a timeline of different bookmarks captured on the right from different network visualizations. On the left,t we can see a different period affecting the network state along with tweaked features

3.3.2 Participants' Demographics, Data Collection and Analysis

This subsection discusses how the participants were recruited, their demographics, which data was collected, and how it was analyzed.

Participants Demographics and Recruitment– Similar to logs collection and analysis, an open invitation was widely distributed through mailing lists, encouraging interested analysts to attend our workshops and use the tool (see subsection 3.2.3). Since the logging process was anonymous, no background or demographic data about the analysts was collected.

The **Data collection process** combined interaction logs with additional data entries detailing the analysts' self-declared goals and objectives. The goals and objectives were captured when the analyst specified them once a bookmark was created. The **Data Analysis**— process included (1) analyzing interaction logs as explained in section 3.2, and (2) exploration objectives and goals declared by analysts once a bookmark is created. Goals and objectives were extracted from the logs and analyzed based on when and where a bookmark was created, on which visualization, and for what reason. However, as this included anonymous logging, no background or demographics of the analysts were captured. The results of the data analysis process are discussed in chapter 4.

Using the state-capturing and annotating approach (aka bookmarking) to capture and understand analysts' visual exploration has its own advantages and disadvantages. As the bookmarking method allows analysts to capture visualization's states of interest, annotate, and restore them, it offers us the following advantages. Firstly, it allows analysts to convey their goals and objectives from the bookmarks created. Secondly, it allows analysts to track their exploration history and how findings were reached. Thirdly, it supports co-exploration experiences where analysts can share their findings through bookmarks with their colleagues. And finally, it helps researchers to understand more about the usage of the visualizations, their interactive features, and how they contribute to visual exploration of networks. On the other hand, the drawback of using such a method is that analysts rarely share their real exploration goal or reason for creating a bookmark (e.g., a random reason is provided). And even if the goal was declared, it could not be analyzed in-depth without capturing the context, such as the *nature of the dataset* (e.g., why data with date/time wasn't explored), and the *why the captured state was of interest* (e.g., the layout of the visualization was sparse or dense). Thus, we decided on our third approach, working closely with the analysts through an intensive course section 3.4.

3.4 Third Approach: Course

Neither interactions logging and analysis section 3.2, nor the state-capturing and annotation section 3.3 approaches provided insights about analysts' goals and intentions. Thus, to overcome their limitations, I designed with my supervisor a 6-week course to capture analysts' intentions, understand how they work with data, and explore it visually. The course walks analysts through the exploration process step-by-step. The course content covers two parts (3 sessions each): (1) data preparation for network visualizations and (2) network visual exploration. The course was designed to enable participants to achieve the following **goals**: (1) Structure their network data and prepare it for visualization using the Vistorian tool, (2) Define the objectives of their exploration process and identify what they aim to learn about their network data through visualizations, and (3) Learn and utilize a range of network visualizations through theoretical instruction and hands-on practice to explore their data.

3.4.1 Networks Visual Exploration Course Content

The course included weekly sessions covering the following topics: Network Data Preparation (2 sessions), Data Shaping Techniques and Challenges (1 session), and Reading, Interpreting, and Exploring Network Visualizations (3 sessions). Each session provided tutorials and hands-on exercises to learn the process of visual exploration of networks. The course's outline can be found at section A.4. The hands-on exercises provided analysts with various **activities** to understand the exploration process, such as how to*sketch concept map, form different networks from the same dataset, create node/link tables, and read and interpret different network visualizations*.

In addition to the weekly tutorial sessions through 6 weeks, the course provided the analysts with the following. Firstly, offer various types of **resources** such as recorded videos, cheat sheets, additional practice examples, and reading material. Also, it provided analysts with **support** at any time through drop-in sessions and on-demand help (e.g., Python scripts). In addition, it discussed **topics of the analysts' interest** such as how to extract and create a network from qualitative data (e.g., interviews), creating networks out of different data types (e.g., videos), dealing with semi/unstructured data (e.g., social media accounts), how to employ network features to describe data (e.g., how to use link type), how to use different data types in supporting the visual exploration of networks (e.g., geo-locations), working on data consistency, and others. The details and details of the course and its outcomes are discussed in detail in section 4.5.

3.4.2 Participants' Demographics, Data Collection and Analysis

This subsection discusses how the participants were recruited, their demographics, which data was collected, and how it was analyzed.

Recruitment— An open invitation was sent to recruit analysts interested in learning and exploring their data using network visualizations. Participants volunteered to opt in and could withdraw from the study at any time. No background was required to enroll in the course. Any participant, with or without their own data, was welcome to join to learn how to explore data through network visualization using the Vistorian. **Participants Demographics**— We had 24 active participants out of 36 registered analysts from different backgrounds, such as humanities, natural sciences, business and management, health and medicine, education, computer sciences, engineering, history, social sciences, and law. Their level of experience in data visualization and exploration varied as follows. *First-timers* (around 50%) are participants who the course was their first chance to explore a network. Participants with *limited expertise* (around 33.3%) as they only explored 1-2 networks. And *Advanced participants* (around 16.7%) who have explored 3+ networks previously. Fifty percent of the participants joined the course without a specific research question to investigate. At the same time, 27.8% had their research questions, and 22.2% were in different analytical stages, such as data collection or preparation. Participants' profiles are described and provided in section A.2 for reference.

Data collection— Through the course, I have collected observations from the leading weekly 2-hour course sessions, drop-in sessions, and on-demand technical support meetings. Also, I collected weekly feedback survey responses, interaction logs, and answers to 1-to-1 exit interview questions (see section A.1) from the participants who agreed. **Data Analysis**— Interaction logs were analyzed quantitatively using Python scripts. Qualitative data (surveys, observations, and feedback) was analyzed using thematic analysis. That is, by coding and tagging the interviews and classifying them using Nvivo, done by me (primary investigator) and Uta Hinrichs (co-supervisor). The results of the data analysis process are discussed in chapter 4).

The logs section 3.2 provided quantitative analysis, whereas the state-capturing and annotation section 3.3 and the course provided qualitative analysis. The course resulted in identifying barriers in visual exploration of networks and their implication on the design of resources and visualization tools. The course outcomes are discussed in detail in chapter 4. Using intensive courses to understand the visual exploration of networks has the following **pros**: (a) Allows in-depth understanding of analysts' needs, goals, and struggles through discussions (qualitative data), and (b) Provides a rich knowledge exchange environment through group discussions from various backgrounds. However, running courses to understand and investigate the visual exploration of networks has the following **drawbacks**:

- Coordinating with analysts from diverse backgrounds (varying levels of expertise, domains, and exploration goals) necessitates accommodating their different needs simultaneously.

- Courses are mostly conducted as instructor-led rather than self-regulated. Consequently, analysts with varying levels of expertise cannot progress at their own pace and require the presence of an instructor to assist them as needed, unlike self-regulated guides.

- Courses do not facilitate the simultaneous investigation of many analysts, unlike log collection and analysis in the wild, which can be efficiently conducted using scripts and log analysis software.

Due to such limitations from the methods applied and based on their findings simultaneously, I decided to provide a self-regulated guide (section 3.6). The guide should inform and allow analysts to explore network visualizations based on their needs and at their own pace. At the same time, I decided with my supervisor, Ben, to use the collected logs to offer a live dashboard that will inform the tool's use as discussed in section 3.5.

3.5 Fourth Approach: Analytical Dashboard

Events logs hold detailed information about analysts' usage and the visual exploration of networks. Through our research, we aim to understand the visual exploration process further to enhance the visualization tools and educational guides. However, due to its nature, extracting metrics to describe the process of network visual exploration is a bit tricky (see subsection 2.3.5). How can we decide on the success of an analyst's visual exploration process? How can we describe it? And how do we measure it? Thus, in this approach, we aimed to find more about the visual exploration of networks by analyzing the logs from three different perspectives, which are: (1) the tool, (2) the visualizations, and (3) the user. Those three perspectives are the pillars network visual exploration stands on and which we aim to understand further.

To understand the visual exploration of networks process better, we need to have tangible values that reflect what analysts do. The values we aim to have should also help evaluate the visualization tools. To do so, learning analytics play a significant role here as it allows extracting evaluation metrics that enable a better understanding of the tool's usage in network visual exploration. The learning analytics design and calculation are applied to the logs I collected from the logging process (section 3.2) for a period that lasted over 28+ months. The design process took around three months. The resulting metrics and key performance indicators (KPIs) are represented through a dashboard. The dashboard informs the decision-making toward better visualizations and tool design.

Participants Recruitment and Demographics— No specific recruitment has been done to collect KPIs extraction and calculation logs. The dashboard and KPIs were designed as group work as follows. I handled the metrics calculation and logs analysis in the design of the analytical dashboard, whereas my colleague Jinrui Wang handled the design of the dashboard visualizations. Benjamin Bach, our supervisor, discussed with us and evaluated the KPIs and the dashboard design over three months through several iterations. No demographics about the analysts who used the tool and generated the log events were captured as anonymous logging.

Data Collection and Analysis— We used the logs collected from the logging process section 3.2 over 28+ months. Those logs were collected where analysts were using the tool on their own in the wild and in predesigned activities (e.g., workshops and courses). The logs were analyzed based on the KPIs we decided and set to evaluate the three perspectives mentioned earlier: the tool, the visualizations, and the user. The data analysis is detailed in section 5.3. The analysis also included how to design a dashboard to represent meaningful metrics to evaluate the exploration process and the tool based on it. This resulted in a dashboard of 3 pages to represent overall statistics about the tool usage, the visualization usage, and the user exploration process based on their different needs. The dashboard design process is detailed in section 5.4.

The outcomes of this approach are evaluation metrics, an analytical dashboard, and a dashboard design pipeline for network visual exploration. The outcomes are detailed in (chapter 5). However, the findings are limited to the type and amount of captured log events and what they can tell. In addition, dashboards don't allow reasoning and investigating further specific findings without complementing the analysis process with qualitative data (e.g., interviews with analysts). So, how can we use our dashboard readings to enhance the analyst's visual exploration process by using the tool, learning, and applying the concepts? Thus, my supervisor and I decided to move to our fifth approach (section 3.6) to understand how the metrics we extracted assist in creating better guides for visual network exploration.

3.6 Fifth Approach: Coaching Program & Roadmap

Our previous approaches offered us a better understanding of the network visual exploration. This includes understanding some of the barriers analysts face, user types, metrics, and KPIs needed to describe the visual exploration process. So, how can we employ our understanding to support a better exploration and self-regulated learning of the visual exploration process? In other words, how can we transform the learning process to be *analysts-centered* rather than facilitator-centered learning? To do so, I used our earlier findings to create a network visual exploration roadmap. This roadmap shall guide analysts through their exploration process based on their goals, data, and strategy. It provides analysts with detailed steps and associated activities.

Each step is defined and described through (1) its goals, (2) how to apply, (3) practice examples, and (4) supporting material. To test the roadmap, I ran a coaching program to evaluate how much it helps analysts in their exploration process independently(*self-regulated*). The **outcomes** of this study are: *the roadmap, exploration steps, steps' activities, the supporting material, exploration strategies, and their description.* In the following subsection, I discuss how I designed the roadmap to support self-regulated learning and exploration.

3.6.1 Designing A Self-Regulated Guide for Networks Visual Exploration

As discussed in subsection 2.5.5 to support **self-regulated learning and navigation** through the visual exploration of networks, I designed, along with my supervisors' feedback the roadmap with the following characteristics in mind:

- **Creating a Resource-Rich Environment**: where different supporting material is created and provided. The provided resources vary in their delivery medium (e.g., videos, textual, drawings) and content (e.g., tutorials, examples, demos). The need for various resources was found through our earlier studies (chapter 4) and was emphasized by our findings from our dashboard study (section 3.5,chapter 5).

- **Setting Clear Learning Objectives**: inspired by SMART [192, 74] goals in education, the learning objectives of the roadmap are designed to be: *specific* for each pathway, *measurable* through the outputs of each step, *achievable* as each step breaks down its process into one or more activities the analysts select from based on their needs, and *relevant* through the dataset selection, proposed activities, and exploration strategies. We identified four learning goals based on the analyst's data and exploration goal: learning tools and concepts, data collection, free-form exploration, and goal-oriented exploration, as I will explain later in subsection 6.3.2.

- **Fostering Autonomy**: analysts should have ownership of their learning and exploration journey through the roadmap. They can choose their preference, such as datasets and resources. They can choose what to apply (e.g., activities), how many times to apply it (*iterations*), and how to apply it (e.g., *strategies*). This ownership of their exploration journey is customized by their needs, background, and preferences.

- **Providing Guidance and Support**: the roadmap helps in offering credible sources, suggesting exploration strategies and visualization to explore, and providing guidance on the expected workflow (input, output, navigation direction).

- **Promoting Self-Reflection or Self-Evaluation** Contrary to the educational context, in network visual exploration, we promote self-reflection rather than self-assessment. As

self-assessment is more about having a straightforward, clear answer (e.g., correct and wrong answers), the exploration is more about what we are answering and how we can best answer it. Reaching polished findings in network visual exploration depends on several network schema creation, exploration, and reflection iterations.

- **Emphasizing Critical Thinking**: through several aspects, such as understanding how sketching different concept maps can answer different questions. As well as how those concept maps are reflected in the generated schema.

- **Dialogue-based Exploration and Navigation**: where the visual exploration process is guided through decision points (e.g., yes/no questions). Those decision points help the analyst to decide on how to initiate the visual exploration of what visualization to start exploring, which additional visualization(s) to explore, and how to proceed (helps in overcoming barriers B7-B8 that will be discussed in section 4.9.

The full design of the roadmap is detailed in section 6.4, including the steps and associated activities. The design of the supporting material is detailed in subsection 6.4.4. In addition, the exploration strategies extraction and description is detailed in subsection 6.3.1. I conducted the study through a coaching program where analysts were provided with the roadmap along with supporting resources. The study's outcomes and results are detailed in section 6.6.

3.6.2 Participants' Demographics, Data Collection and Analysis

This subsection discusses how the participants were recruited, their demographics, which data was collected, and how it was analyzed.

Recruitment—- An open invitation was sent to recruit analysts interested in exploring their own data and/or learning the visual exploration of networks. The invitation described that analysts will use the roadmap as their guide and that a coaching program (see section 6.5) will accompany their exploration process. That is by providing analysts with kick-off sessions, follow-up sessions as needed, and an exit session. No prior experience was required. **Participants' Demographics—** 10 analysts participated in our program. Participants' network exploration varied: 4 analysts had explored 3+ networks, 2 analysts had explored 1-2 networks, and 4 analysts had no previous experience in network exploration. Participants had backgrounds in education, computer sciences, mathematics, engineering, history, social sciences, literature, and humanities. Four out of the ten analysts (40%) have explored three or more networks, two out of the ten analysts (20%) have explored 1-2 networks, and Four out of the ten analysts (40%) did not explore any networks previously. Most of the analysts' previous experience was with Python visualization libraries and MS Excel, and few had experience with other tools such as Tableau.

Eighty percent of the ten analysts brought their own data to work with. Detailed information about the analysts' objectives and goals is provided in subsection 6.5.2.

Data Collection— Through surveys (e.g., registration, feedback), observations from sessions, recorded video meetings, sketches of concept maps by the analysts provided for discussion, and exit 1-to-1 meetings. **Data Analysis**— Video and audio data from the coaching sessions and interviews were transcribed and subsequently analyzed by me using a thematic analysis approach [49, 113]. The analysis concentrated on participants' feedback and reflections on the roadmap design and content, learning resources, and the visual exploration process. This included their goals, the nature of their data, the activities they engaged in, their experiences with these activities, and their feedback on iterations of the roadmap design. In addition, the analysis included observations from the sessions and comments from the analysts.

The roadmap design for network visual exploration emphasizes the importance of understanding how analysts use visualization tools to explore data. Roadmaps have their own advantages and limitations. The roadmap offers the following advantages: supports the self-regulated design of both visualization tools and educational efforts, offers a more systematic and clear approach to how to explore data visually through steps, pathways, and strategies, addresses different needs of user types and pathways to overcome barriers, and allows the description of strategies in the visual exploration process. This offers the opportunity to build automated solutions to assist analysts in their exploration process. However, the roadmap may be *limited* to certain aspects, such as it requires detailed capturing of strategies and how they can be applied visually, requires a high number of participants to provide a comprehensive understanding of various users' needs, and needs to ensure that the roadmap design should not limit how analysts think about their data (to be neutral). In the following chapters (chapter 4, chapter 5, and chapter 6), I will discuss in detail each approach along with their findings.

Chapter 4

Understanding Barriers to Network Exploration with Visualization: A Report from the Trenches

Chapter Overview— This chapter reports on an in-depth study that investigates barriers to network exploration with visualizations. Network visualization tools are becoming increasingly popular, but little is known about how analysts plan and engage in the visual exploration of network data—which exploration strategies they employ, and how they prepare their data, define questions, and decide on visual mappings. The study involved a series of workshops, interaction logging, and observations from a 6-week network exploration course. The findings shed light on the stages that define analysts' approaches to network visualization and barriers experienced by some analysts during their network visualization processes. These barriers mainly appear before using a specific tool and include defining exploration goals, identifying relevant network structures and abstractions, or creating appropriate visual mappings for their network data. The findings inform future work in visualization education and analyst-centered network visualization tool design.

Authorship Statement— This chapter has been published, as noted in section 1.5, with authorship details clarified in the declaration at the beginning of this thesis. I declare that the work in this chapter was mostly done by the student Mashael AlKadi. By work, I mean and include the following: logging user interactions with the Vistorian visualization tool, designing the course material, running the course weekly sessions, supporting course participants through email and drop-in sessions, collecting data from event logs, surveys, observations, videos, interviews with course participants, and analyzing collected data. Before running the course, I ran several workshops on the Vistorian with my supervisor Benjamin Bach. Benjamin also supervised the weekly sessions and the entire study, while also reviewing the course content, data

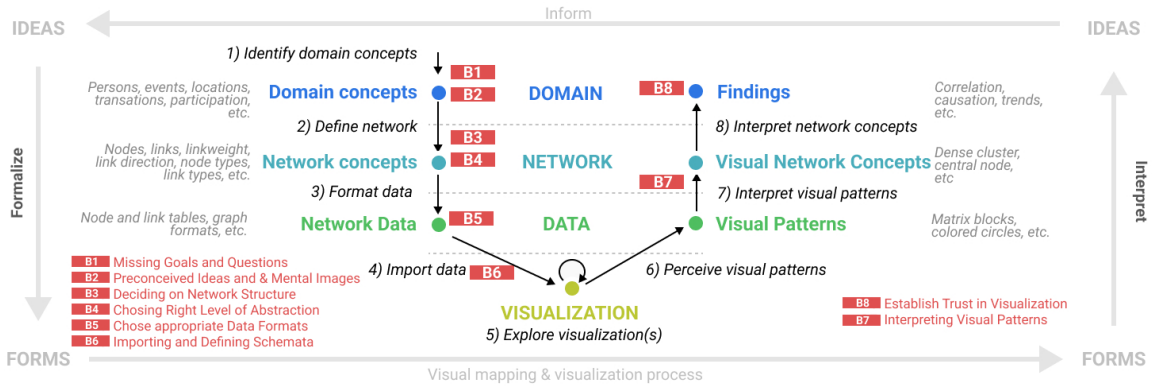


Figure 4.1: Barriers (red) that may be encountered during visual network exploration process, while translating domain concepts (ideas) into network structures and visualizations (forms) and back into findings.

collection, and analysis processes. Uta Hinrichs (co-supervisor) assisted in the thematic analysis and the analysis process in general. I have generated the content of this chapter. All paper authors (Mashael AlKadi, Vanessa Serrano, James Scott-Brown, Catherine Plaisant, Jean-Daniel Fekete, Uta Hinrichs, and Benjamin Bach) contributed to the discussion, organization, and polishing of the paper published out of this chapter in VIS 2022 under the title of "Understanding Barriers to Network Exploration with Visualization: A Report from the Trenches".

4.1 Introduction

As explained in chapter 2, network analysis is becoming an established methodology across many disciplines, such as biology, the social sciences, and the humanities. While the specific datasets are diverse and heterogeneous with respect to how they have been captured and what research questions they aim to address, they have in common that the analyst thought of their data as a *network*—i.e., as nodes and links. By applying methods from network analysis, such as calculating metrics or producing a visualization, the analyst aims to learn something about these nodes and their relationships. Hence, network analysis is a *tool* and a *lens* to model and interrogate data. It comes with assumptions, such as that a network is a meaningful data representation. And, it involves multiple steps in defining nodes and links, formatting data, applying visualization, and interpreting these visualizations correctly.

Especially for novice analysts, performing these steps and making the respective decisions can pose problems. Formerly, and over several years, my supervisor Benjamin Bach ran workshops, tutorials, talks, and interdisciplinary collaborations on network visualization and built our own network visualization tool, the Vistorian [17, 10]. Following that, I have ran several workshops on how to explore networks visually, and how, using the Vistorian. During this

time, we experienced first-hand some of the problems, challenges, and misconceptions that can hamper engagement with network analysis and visual network exploration. Many of the issues we encountered were very common, such as identifying the meaningful entities (nodes, links, attributes, time, geographical locations) from the initial dataset—e.g., a corpus of documents—that should represent nodes and links to create a network that is meaningful to answering a specific research question. Other issues we observed were about using visualizations other than node-link diagrams to understand multivariate, dynamic, and geographic networks or properly formatting a dataset for import into a tool. While some of these issues are echoed in studies on expert data workers [179, 36, 140, 18] or the context of network visualization literacy [44, 231, 200], a comprehensive account of the problems novices face when engaging with network exploration is missing. Whereas multivariate and quantitative data can be represented by familiar visualizations (e.g., bar charts or line charts), networks pose specific challenges due to their qualitative (i.e., structural) character, the combination with other data types, and the abundance of specialized visualization techniques for temporal [27], geographic [233], and multivariate [185] networks.

In this chapter, I present an in-depth study into the barriers some analysts face when preparing their data for visual exploration and using an interactive visual network exploration tool. For simplicity, I relied in my study on a single tool, the Vistorian, that we consider a state-of-the-art network visualization tool comparable to Gephi [20], NodeXL [247, 248], or Palladio [131, 73]. The Vistorian has been specifically developed to provide an easy entry point to network visualizations (node-link, adjacency matrix, timeline, map) and interactive exploration for data analysts without a technical background (section 4.3).In a first log study (Study 1, section 4.4), I tracked users of the Vistorian over several months to understand how they use visualizations and interactions in their day-to-day work. To complement these anonymous data with qualitative data about potential barriers, I designed with my supervisor a 6-week network exploration course. We delivered this to 36 analysts while closely monitoring their progress, providing personal advice, and engaging in individual interviews (Study 2, section 4.5). The course was open to anybody and included analysts from computer science, the social sciences, and business. It introduced the basic concepts of networks, for example, visualizations for multivariate, temporal, and geographical networks, hands-on activities for sketching and formatting data, and hands-on tutorials with the Vistorian. While many of our participants succeeded in visual exploration and did not encounter major problems, our observations reveal barriers, experienced by some participants. For example, preconceived ideas and specific mental images about one's network blocked analysts from thinking creatively about their research questions and network structures. We also found a barrier in defining (multiple) possible networks to match different research questions and abstract from a domain problem into a network problem and back [180, 48].

Mitigating these barriers has implications for future tools and research. In summary, I report in this chapter on:

- the user profiles we identified based on anonymous interaction logs across 56 weeks (section 4.4),
- observations (section 4.7) from a 6-week course (section 4.5) on visual network exploration, including
- a set of detailed participant profiles and observations about goals and strategies toward network visualization and exploration (section 4.6),
- 8 barriers to engaging with network exploration (section 4.9)
- implications for tool design and training (section 5.6).

4.2 Background

In this section, I provide a recap of the literature discussed earlier in chapter 2. I examine the current state of network visual exploration, including their support mechanisms, study methodologies, and evaluation techniques. Additionally, I review network visualization literacy and training, focusing on the ease of acquiring the necessary skills.

Tools for Visual Network Exploration—Numerous tools enable analysts to produce network visualizations without writing any code: Gephi [20], SoNIA [32], Palladio [131, 73], NodeXL [247, 248], Visone [26], NetworkWorkbench [190], SYF [199], Paejk [82], UCINET [225], or NetMiner [78, 130]. These tools provide node-link diagrams with some degree of interactivity, maps with overlaid node-link diagram; or animation support for temporal networks. These tools load data in a range of formats, such as GraphML, GEDCOM, GEFX, etc. [223], some of which are not human-readable or manually editable. NodeXL asks users to upload data as *node tables*, which contain information about nodes, and *link tables*, which describe links between these nodes. Users then specify the meaning of each column and perform transformations such as link projection. That said, data-wrangling tasks such as re-formatting data and correcting errors in 'dirty' datasets can be a significant portion of data-analysis work [137]. Dedicated tools can be used for data wrangling, for example, Ploceus [163], Orion [124], or Origraph [35]. Our network visualization tool, the Vistorian, is representative of these tools in that it creates interactive node-link diagrams visualizing link weight, direction, and type, as well as node types, and can represent temporal and geographical data. It has been developed and iterated over with many domain collaborators and allows us to log interaction data during in-the-wild and day-to-day use. We were inspired by NodeXL's table format, which is easy to create and edit in common spreadsheet applications. Also, the Vistorian extends existing tools by providing alternative visualization techniques such as adjacency matrices (e.g.,

for dense networks), a timeline and time-sliders for temporal networks, a map for geographic networks, and coordinated linked views.

Studying Network Visualizations and Tools—Many studies [278] have empirically evaluated the value of specific visual representations for networks (e.g., [103, 186, 191, 215, 129, 232]), their respective visual affordances (e.g., [9, 210, 262, 3]), interaction techniques (e.g., [15]) or the combination of multiple coordinated views for node-link and matrix diagrams [126, 65]. Due to the controlled nature of these studies and the questions they pursue, they have often relied on tasks described as *low-level* by common task taxonomies [156, 142, 275]. Most of these studies focus on quantitative records to describe task-completion time and task-error rate. Still, alternative measures such as learnability, user satisfaction [205], and engagement (e.g., [132, 167]) can be highly important.

To achieve greater ecological validity and investigate tasks that are higher-level and potentially more sophisticated studies have used more complex methodologies and data collection mechanisms, such as studying analysts using design-studies [238] or involving experts and their data in qualitative studies [199]. Still, many of these studies focus on developing, understanding, and improving a specific visualization tool within a particular context. The Multidimensional, In-depth, Long-term Case (MILC) approach [244] streamlines the process of in-depth evaluation of visualization tools in the wild and suggests using interviews, training, improving the tool, interaction logging [91] and ethnographic observations, collecting user annotations through logbooks and micro-entries to help tracking analysts thoughts and discoveries [41], as well as entry and exit questionnaires. In our pilot study (section 4.4), I collected user logs over several months, supported by short mini-questionnaires as described by Serrano et al.[239] and enhanced it by allowing analysts to provide further context. The mini-questionnaires prompted users to provide feedback on using our tool to complement the interaction logs [112]. In our second study (section 4.5), I follow more closely the MILC methodology but significantly increase participant numbers compared to prior studies (usually 5 [244]) to approx. 24 in our case.

Network Visualization Literacy and Training—Through brainstorming with scientists, teachers, and students, Sayama *et al.* [231] define seven *"essential concepts about networks [...] that every person in the 21 century [should] know by the time he/she finishes secondary education"*. The concepts include *"networks describe how things connect and interact"*, *"networks help reveal patterns"*, and *"visualizations can help provide an understanding of networks"*. However high-level these concepts seem, studies on teaching network science in high schools and public exhibitions [77, 282, 44] revealed how little is generally known about the basic concepts of networks, their analysis, and visualization in the wider population. For example, when shown images of node-link diagrams, visitors to science museums often

struggled to explain how these images would be interpreted [44].

Courses and explicit training for network analysis [77] have been used in the context of MILCs such as Bonsignore et al.s [43], and Hansen et al. [117] to evaluate NodeXL and learn about teaching network exploration concepts [43, 117]. Bonsignore et al.s [43] and Hansen et al. [117] studies introduced graduate students (21 and 15 respectively) to Social Network Analysis and NodeXL and tracked them over 5 weeks. Students used demo datasets and smaller self-chosen datasets, described as *discovery learning* and *exploratory learning* in the paper. Our course (section 4.5) focused on practitioners working on their own datasets, a stark qualitative difference from prior courses and studies. Moreover, instead of evaluating our tool, the Vistorian, I aim to understand better the challenges analysts face during the network exploration process.

4.3 Network Visualization with The Vistorian

A full description of the Vistorian visualization tool is provided in section 3.1. It has been selected to assist us in understanding analysts' usage and needs in real-world scenarios.

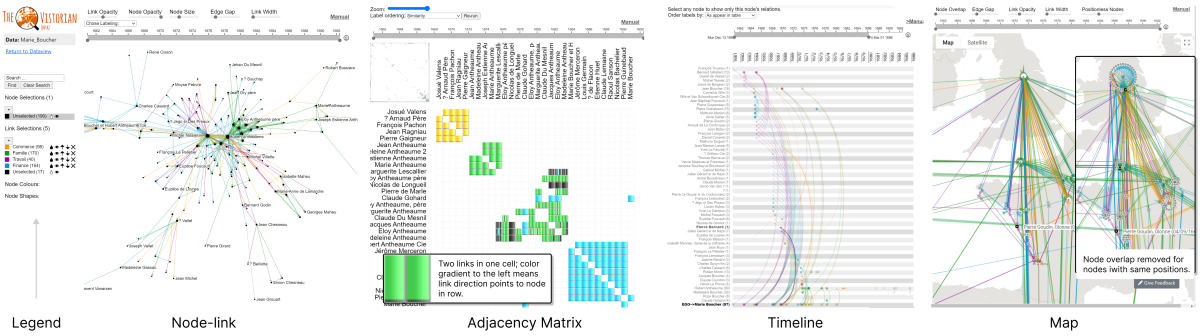


Figure 4.2: **Visualizations in the Vistorian:** node-link diagram, adjacency matrix, timeline, and map with overlaid links and edges. Each visualization runs in its own tab and is coordinated through brushing and linking.

Studying Real-World Usage of the Vistorian—While designed for analysts in the social sciences and humanities, it addresses a potentially broad audience. It has been designed and optimized with collaborators in these areas since 2014. It supports exploration processes that can easily take weeks, months, or years in which an analyst iteratively obtains new data, refines their research questions and explores their dataset from different perspectives. In such scenarios, an analyst's understanding of the dataset constantly evolves alongside their analysis goals, questions, and methods. Data exploration almost always requires learning and mastering new methodological tools, including visual exploration. Understanding these processes and how to facilitate them is the main aim of our research. Our studies of real-world usage of the Vistorian consist of two phases that build upon each other, as detailed in the following two sections.

4.4 Study 1: Interaction Logging & Workshops

The first phase of our research focused on anonymously logging interactions and conducting workshops to gain a general understanding of *how real-world analysts make use of the Vistorian* and *how to introduce analysts of various backgrounds to network visualization*.

Interaction Logging— I anonymously tracked all interactions with the Vistorian. Over 10 months, I collected data on how analysts upload data, specify network schema, create and interact with a visualization, and request help. A pop-up informed people about the logging taking place, linked to information for study participants, and obtained their consent. This procedure was approved by our University Ethics Board. Logging started from the moment a user landed on the Vistorian website and continued until they closed all related browser pages. Logs were collected using the *Intertrace* JS library [93]. To facilitate the analysis of our interaction logs, I classified all log events, based on categories I defined in advance. The main categories (the list of all categories is included in subsection 3.2.2) included *Data-Management* events (e.g., interactions such as creating a new network, uploading data tables, updating table entries, specifying table schema), *Visualization & Interaction* events (e.g., any interactions with visualizations), *Support & Communication* events (e.g., engagement with the Vistorian help resources such as manuals, demo data, and videos), and *Bookmark* events (e.g., interactions related to the creation, update, and usage of bookmarks and annotations).

Workshops—While still developing the Vistorian platform, we ran a number of public workshops that varied in length from half a day to three days. Notable here are the most recent (online) workshops that we ran in parallel to starting to log interactions with the platform—one ran in the context of an online DataVis Meetup (approx. 25 participants) and one at an online Digital Humanities Summer School (approx. 55 participants). These workshops included lectures on multivariate network visualization, sketching networks and discussing appropriate visualizations, demos of the Vistorian, and hands-on tutorials that guided participants through the process of importing their own or demo data to the platform. Data exploration using the Vistorian was facilitated at the workshops, but also beyond through drop-in sessions and a dedicated Slack channel where people could ask questions and seek support for their data explorations. While these workshops did not include structured data collection, they provided valuable insights that contextualized our analysis of interaction logs, and informed the network visualization course design (section 4.5).

Findings from Interaction Logs & Workshops— Over the course of 10 months we logged 534 individual users sessions which we analyzed using the interaction categories described above, looking for interaction patterns (i.e., interaction times, what parts of the Vistorian platform people engaged with and in what order, and the number of returning users). Based on our analysis, I identified four types of usage patterns: ***Demo-users*** ($298/534 \approx 55\%$) solely played

with the demo dataset or looked into help resources, without exploring their own data using the Vistorian platform. ***Data-strugglers*** ($71/534 \approx 13\%$) did try to explore their own datasets using the Vistorian platform, but struggled with shaping their data and creating their networks. These users sometimes returned to the platform several times, but without managing to successfully create a network visualization.

Single-session explorers ($93/534 \approx 17\%$) managed to upload their own data and created a network visualization. They then spent a short time (less than 2 min.) exploring their visualization and did not return.

Multi-session explorers ($72/534 \approx 13\%$) uploaded their own data, created (one or more) network visualizations and explored these over the course of multiple sessions that each lasted a minimum of 2 minutes. Multiple short sessions were more common than long continuous sessions.

Our results confirm observations we made during earlier workshops and demos. However, the mini-questionnaires did not help generate detailed insights about users' strategies and their problems in using the tool. We wanted to learn more about people's network visualization and exploration strategies and inherent issues that seem to go beyond potential usability problems as many people managed to upload and visualize their data using the Vistorian. To address these questions from a more qualitative data perspective, we designed the course for Study 2 (section 4.5).

4.5 Study 2: Network Exploration Course

The target audience of our network visualization course were analysts without a technical background and/or no or little experience in network visualization and exploration. We advertised the workshop widely, targeting analysts across the sciences, arts and humanities, and at different stages of their analysis process. We aimed to teach different aspects of network exploration—from data preparation to visual mapping to visual network exploration. We utilized the Vistorian as the sole network visualization and exploration tool in our course *(1)* to avoid overwhelming participants by introducing them to multiple network visualization tools within a relatively short amount of time, and *(2)* because the Vistorian supports a large variety of network visualization and exploration techniques, making it accessible even to analysts without a technical background as described in section 3.1. Based on our findings from Study 1, as I analyzed the users' interaction event logs, I have identified some users who face issues when creating a network. Some of the issues were related to how the analyst's data files were prepared. Thus, I have added interactive and written guides to the Vistorian to assist the users. Also, I communicated what users face to my supervisor, Benjamin Bach, and discussed it with the developer James

Scott-Brown, who developed the import wizard and some features. Below, I describe these changes to the Vistorian, the course design, and the methods of data collection and analysis.

4.5.1 Data Import Wizard

We replaced the 'manual' process of setting up a network schema with a comprehensive wizard to support analysts importing the correctly formatted data and defining meaningful schemata. The wizard (see subsection 3.1.4 for the figure and full description) has the following steps: *(1)* network name, *(2)* data format (tables, GEDCOM, Pajek, etc.), *(3)* choosing between a link or a node table, *(4)* uploading a table and specify a schema for that table, *(5)* specifying a location and/or a node table with node types (optional), and *(6)* launching a visualization in a new tab.

4.5.2 Course Outline

I have designed with my supervisor a 6-week network visualization course that covered both data preparation and visual network exploration. From a teaching perspective, the goals of the course were to help participants (1) Define the goals of their data exploration, (2) Become familiar with a range of network visualization techniques through theory and hands-on use, and (3) Use different types of interactive visualizations to explore their own data. From a research perspective, we aimed to learn more about analyst approaches and strategies when using network visualization to explore their data and potential barriers inherent in these processes.

Following the recommendation by Batch *et al.* [22] to increase the use of visualizations in data exploration by providing education, the course was based on problem-based learning. We encouraged participants to bring their own data and related questions/issues/problems to the course so they could directly apply lessons learned in a real-world context relevant to them. However, we also provided demo datasets to those who did not (yet) have their own data.

Each week consisted of a two-hour session combining online live lectures (recorded and shared online) and hands-on activities. Some sessions set homework for the following week to encourage participants to practice and identify the areas that they needed help and support with. We offered weekly one-hour drop-in sessions during which course participants could ask questions about the lecture material or raise issues they had encountered. We also moderated a Slack channel that allowed participants to discuss course-related questions and problems with each other and with the course instructors. We describe the topics and activities covered in the individual course sessions below.

Weeks 1 & 2: Network Data Preparation introduced participants to the topic of network visualization and the goals of the course. We covered the basic terminology around network visualization (e.g., introducing the concept of a network, nodes, and links) as well as questions

that can be answered using network visualization and exploration. In *Activity 1*, participants prepared a sketch of the potential network(s) that could be constructed from their data and thought about potential questions that these networks could help answer. In *Activity 2*, participants identified attributes in their data that could serve as nodes and links in their data. In *Activity 3*, participants sketched small networks based on the nodes and links identified in Activity 2. In *Activity 4*, participants formalized these sketches into a concept map by starting to define nodes and links more formally. Finally, in *Activity 5*, participants specified their node and/or link tables. We had initially planned for a single week of network data preparation, but during Week 1, we realized that participants struggled with (1) mapping data attributes to nodes and links to structure their network and (2) formatting their data into a network table. During Week 2, we therefore dedicated more time to re-iterating and finishing the activities introduced in Week 1.

Week 3: Data Shaping Techniques & Challenges introduced techniques for ensuring consistency in data, reducing data size (e.g., filtering nodes by a degree or node type, or filtering edges by time), or aggregating data. We provided respective Python scripts and assisted participants with applying them to their data. Activities included loading the finalized data tables into the Vistorian using the import wizard.

Week 4: Network Exploration Using Node-Link Diagrams introduced node-link diagrams and related interactive features.We presented different exploration tasks [156] in the form of questions, e.g., *What to observe about network structure?*, *What does that mean for the data?*, or to explore a network's change through time interactively.

Week 5: Network Exploration Using Adjacency Matrices introduced adjacency matrices. We explained the visual encodings and how the node ordering aims to optimize for visual patterns in the matrix.

Week 6: Timeline, Maps, and Coordinated Views introduced timeline visualization and the geographic network visualization techniques, as well as the coordinated view feature in the Vistorian.

Again, throughout the course, participants were guided to prepare, visualize, and explore their own data using the different network visualization features provided by the Vistorian. We also encouraged participants to share and discuss insights that they had gained and challenges they had encountered. Our course website [11] provides more details on the course and related materials.

4.5.3 Participants

36 participants from a wide variety of backgrounds, including history, social sciences, business & management, health & medicine, law & politics, natural sciences and technology, education, and event management, registered for the course. In a pre-course questionnaire, 18 (50%) participants

stated that they had no experience with network exploration, 12 (33.3%) had explored one or two network visualizations prior to the course, and 6 (16.6%) participants stated experience with three or more network visualizations prior to the course. Due to the nature of the course, which was organized during the COVID-19 period and allowed participants to attend sessions synchronously online or watch recorded lectures asynchronously, we cannot report how many participants attended all sessions of the course. Of those who registered, 24 attended the first session synchronously, with subsequent course sessions being attended by 14 participants on average. Our findings are based on data collected across all participants, regardless of their pattern of course attendance.

4.5.4 Data Collection & Analysis

Our findings are based on the following data collected across the course.

Questionnaires. We asked all participants to fill out questionnaires prior to and after the course to capture their starting points and expectations when entering and their insights and reflections after the course. We received completed pre-questionnaires from 36 participants and exit questionnaires from 8 participants.

Interviews. We invited all 36 course participants to a semi-structured interview where we asked them about their initial motivation to take the course, their lessons learned, to what extent they managed to create visualizations with the Vistorian, what strategies they applied, and what they found the most challenging about the course and the visualization process. To learn more about the Vistorian, we also included questions about which features participants most appreciated, which ones they found most difficult to use, what additional features they would like to see added, and how they would overall describe their workflow with the Vistorian (See section A.1). 13 participants followed our invitation to be interviewed. All interviews were automatically transcribed prior to their analysis.

Course Recordings & Online Discussions. All course sessions were recorded, and we took notes of comments, questions, and discussions that drove the tutorial and drop-in sessions. We also captured inquiries and discussions on the course's Slack channel, where participants would ask questions about the course content and ask for help on their network visualization processes and challenges. Course recordings were not fully transcribed but reviewed to complement our written notes.

Our analysis across this variety of data largely follows a thematic analysis approach [50, 113] where I qualitatively coded participants' statements, and the emerging coding was then discussed and iteratively refined between a team of three co-authors who are myself Mashael, Benjamin Bach and Uta Hinrichs. In the first qualitative coding pass, we captured themes visible in our data top-down, driven by our research questions and questions highlighted in our questionnaires

and interviews. In a second coding pass, we allowed themes to emerge directly from our data collected. Our thematic analysis started already during the course, where we reflected on observations noted as part of the course and drop-in sessions. Emergent themes (e.g., observations around approaches to data preparation, how participants defined their first network, and approaches to visual exploration) were refined after the course in light of participants' statements from the exit questionnaires and interviews.

Our findings from Study 2 are structured into four sections. I first describe participants' goals and types of course journeys (section 4.6). Based on this, I describe the range of observations about network preparation, visualization, and exploration phases that participants went through (section 4.7, Figure 4.1). From these observations, we derived an 8-step model of network exploration, which helps to formalize and contextualize participants' exploration processes (from data preparation to deriving findings, section 4.8) as well as the barriers that can hamper these. Finally, I report on the barriers that some participants faced as part of these processes (section 4.9).

4.6 Participants' Goals & Course Journeys

I start the findings by characterizing participants' starting points when entering the course, including their types of goals and questions (G1–G3) with respect to their data and analysis. We then provide a high-level overview of participants' learning journeys (J1–J3) through the course, illustrated by 14 representative participant profiles. Participants' profiles are described in detail here section A.2. For the remainder of the chapter, we refer to individual participants by the topical focus of their data (e.g., *P-Business*).

4.6.1 Initial Participant Goals

Statements from the pre-course questionnaires show that participants' motivations for engaging in network visualization varied in terms of the specificity of goals, exploration questions and the data at hand.

G1: Learning About Network Visualization To Inform Data Collection. Approximately 22.2% (8/36) of our course participants were still in the process of collecting data when entering the course or planning to start with the collection. Their main goal was to learn about network visualization and analysis in order to inform their data collection processes and questions of potential interest: *"I am preparing for network data collection but would like to see what can be analyzed before deciding on types of network data to collect."*[*P-SupplyChain*].

G2: Exploring Data to Help Define Questions. 50% (18/36) of our participants already had data, but no specific exploration questions prior to the course; Their goal was to *visualize*

and explore their data to help generate questions (e.g., *P-Dance*,*P-Laureates*), check if network visualization is applicable to their data(e.g., *P-RocksStudies*) or aimed to explore the quality of their data (e.g., *P-Documents*).

G3: Specific Questions To Inform Exploration. Approximately 27.8% (10/36) of our participants had defined questions about their data that they wanted to explore, for example, how a particular network had changed over time (e.g., *P-Business*), understanding the structure of a network (e.g., *P-Criminology*, *P-Business*), or who talks to whom in a social network (e.g., *P-Carers*, *P-ElderlyCare*).

4.6.2 Participant Course Journeys

While all active participants (24/36) were able to create at least one network, they followed a range of course journeys, influenced by their data and initial goals. In order to anchor our findings on network exploration strategies and potentially hampering barriers, I illustrate the different types of journeys based on selected participant profiles.

J1: Vistorian Exploration using Demo Data. Participants who did not bring their own data to the course explored the features of the Vistorian based on the demo datasets provided, to learn about network visualization (G1).

For example, *P-UrbanStudies* an engineer with a social sciences background and experience in network visualization was *"interested in learning about the timeline and geographic map networks, plotting adjacency matrices, and learning what Vistorian is about."*

J2: Assisted Network Exploration. 2/24 participants required assistance in creating their first visualization (e.g *P-Carers* health sector; no network visualization experience). They struggled to define a network structure relevant to their questions. However, by the end of the course, they were able to create and explore network visualizations using their own data. Other participants managed to create a first visualization on their own but required assistance with refining their network structure to align with their goals and questions. For example, *P-Letters* (history; some experience in network visualization) focused on data on a collection of historical documents and letters. Their goal was to understand the structure of the network spanned by documents and correspondents (G3). They—with the help of the course instructors—iterated on their initial network visualization leading to an adequate network visualization that enable interrogating the data according to these goals.*"before I explored my network, I supposed that the institution [X] was the most influential actor in my network, but it was [Y]."*[*P-Letters*].

J3: Independent Network Exploration. At least 5/24 participants actively and without assistance explored their data from the different perspectives offered by the Vistorian. For example, *P-Business*(G3) (computer science, social sciences, humanities; experience in network visualizations but not with the Vistorian) wanted to explore how funding rounds of commercial

businesses assist in the growth of industries and business networks, and how the people behind businesses are linked. To explore these questions, they created and iteratively refined a number of temporal and geographic network visualizations and adjacency matrices. *P-RocksStudies* (G2) (Natural Sciences; some experience in network visualization) had a dataset describing studies and methodologies about a specific phenomenon related to rocks. They were able to create visualization but found it did not contribute to their research. *P-Documents*(G2) started with no research question in mind but used network visualization to figure out the best way to link nodes that might reveal interesting patterns in their data.

4.7 Observations

Based on our observations throughout the course and our analysis of the post-course questionnaires and interviews, we identified different aspects that influenced (1) how participants went about defining network structures, (2) how they iterated on their visualizations, and (3) how they approached their visual explorations of created networks. We describe our observations (**O1**–**O10**) below.

4.7.1 Defining Network Structures

The first part of the course (Weeks 1–3) focused on helping participants prepare and shape their data to create a first network visualization. Some participants with prior experience in network visualization were able to accomplish this easily (***J3***) and were able to skip ahead to the visual exploration phases. For participants with less experience, we scaffolded this process through lectures, activities (see subsection 4.5.2), and individual assistance (***J2***). Participants without their own data (***J1***) did not have this problem since data was already defined and formatted. Below, we report on the factors that influenced participants' process of creating a first network visualization and the challenges we observed.

4.7.1.1 Preconceived Ideas and Mental Images

Our conversations with participants throughout the course revealed that around 9 participants had an initial mental image of the network they wanted to visualize in the Vistorian. These mental images were only partially influenced by participants' initial questions and goals but, rather, by their 'intuitive' understanding of the nature of their data and pre-conceived ideas of what their network *should* look like and what entities (nodes) and relationships (links) it expressed.

👁 ***O1: Preconceived Ideas About Networks Can Blind Analysts to Alternative Approaches***. Around 2 participants thought of network visualizations predominantly as social networks showing people and their social relationships. Consequently, they initially defined nodes in their network as people and links as their respective relationships. It can blind participants to alternatives and hinder them from exploring other mappings of nodes and links that could express different information about their data or support more specific goals. This reflects findings about *latent abstractions* [36], i.e., participants not having clear abstractions of their data, such as networks, tables, or timelines. For example, *P-Carers* collected data about social relations between 80 patients and their carers from interviews. Initially, *P-Carers* wanted to create an ego-network centered around each patient. However, discussion and initial network visualizations revealed that the patients were not actually that relevant to the network, as *P-Carers* was interested in the *types* of carers and their relations through the act of caring. In other words, the collection of individual ego networks had to be re-imagined as a network of relationships between the roles of carers. Re-thinking her network structure, *P-Carers* reformatted her data tables several times again, a process common to data workers [18]. Similarly, *P-Letters*'s data focused on *"two persons exchanging letters on several topics."* Similar to *P-Carers*, the participant imagined a network of authors (nodes) connected through multiple links, one for each letter. The initial network visualization—a two-node network with approx. 50 connections—was not useful. During the discussion (***J2***), we suggested representing topics mentioned in the letters as nodes beside persons and linking them through response type; this drastically changed the network structure and the questions that could be explored.

👁 ***O2: Analysts have Mental Images of their Networks.*** Besides preconceived ideas of what the nodes and links of a network should be, several participants had specific visual mental images of what their network would look like and what they would be able to observe: *"I'm envisioning one map [(network)] per person with an interconnected web surrounding all the people, things, artifacts, systems, which make up care in their everyday life."*[*P-ElderlyCare*]. Often, these images were disjointed from participants' goals and questions but were based on an intuitive understanding of their data. Similarly, *P-Documents* worked with labeled documents and wanted to create a network visualization based on the co-occurrence of labels for any two documents (nodes). The participant expected that this would result in a clear network structure. However, their initial visualization ended up being too cluttered to be readable: *"I assumed the visualization would have the answer ready, but what I found it needed multiple rounds of data formation and understanding the underlying aspects to achieve it"*. This effect is commonly observed in visualization design processes [266].

4.7.1.2 Defining and Formatting Networks

40% of our participants (in particular those with less specific questions or goals) were initially unsure about the aspects of their data that should form the nodes and links in their network(s). When asked about particular challenges they faced during the course, one participant highlighted the difficulty of defining the structure of their network data: *"deciding which and how information should be included prior to exploration [before having seen the network visualized]."*

During the course, we facilitated this abstraction process using sketching exercises that asked participants to envision possible nodes and links (subsection 4.5.2, Weeks 1&2). To a certain extent, this approach was successful, enabling active participants to create at least one visualization of their dataset. However, besides defining nodes and links in ways that led to networks that were difficult to explore or simply misaligned with participants' goals, we observed participants struggling to decide on particular aspects when it came to defining their network schema.

👁 ***O3: Alternative Mappings to Nodes and Links.*** Intuitively defining nodes and links often resulted in huge networks with over 10,000 nodes or links, turning most of the visualizations into colored hairballs. One strategy discussed in the course was to reduce network size by filtering by node or link type, filtering nodes based on their degrees only, or slicing the network over time (subsection 4.5.2, Week 3). Another common strategy is to transform the network through graph transformations such as projections or clustering. Link projection was especially new to participants and helped improve their network models in the case of multi-modal networks. For example, *P-Dance* tried to identify participants (nodes) who participated together in one or more events, whereas the initial data they had was who participated in which event. *P-Archaeo* tried to explore the movement of artifacts from one site to another over time and which artifacts were found at the same sites. Thus, we assisted them in creating the respective network projection. The resulting network enabled a meaningful cluster analysis.

P-Festivals's geographic network described links between countries and specific festival locations within the same single city. The respective network featured countries on the one side and street-level information in a single city on the other, making a good-scaled representation of those locations impossible. Through discussion, we explored three alternative networks. Network 1 focused on that single city, with nodes being festival locations and links showing that two locations hosted artists from the same country. Network 2 would keep countries as nodes but collapse all locations within a single country into a single node, resulting in a more readable map visualization. Network 3 would be the initial network but visualized as node-link or matrix only. Without discussing such aspects with the participants, some workarounds might be hard for them to get by their own.

O4: Dealing with Multiple Occurrences of the Same Data Instance. Several participants had data where the same instance of a data point occurred multiple times (e.g., because of contextual or temporal changes). These participants found it difficult to decide whether to represent each data point as a separate node or perform aggregation and represent all data points from the same entity as a single node. For example, *P-Carers* had multiple instances of the same family doctors in their data as these doctors were connected to several patients. Having different node instances representing each doctor made it hard to identify roles of high importance. They finally decided to use a single node to represent each carer's role, e.g., 'family doctor'.

Similarly, *P-Laureates*, working with a dataset of laureates across multiple years, aimed to find those who shared the same types of awards. They first decided to represent awards each year as individual nodes. However, it turned out that this offered a limited understanding of laureates across the years. In a subsequent iteration, *P-Laureates* therefore represented each award title (regardless of the year) as a single node. Creating one node per year and linking to the laureates in that year would help *P-Laureates* answer their question about when the same person had received the award twice.

O5: Defining Link Types & Weight. Another common issue was the definition of link types and how to map relevant data to link weight. One participant mentioned that *"deciding the type of link between the nodes [...] which aspect of the connection [...] is most relevant to answering the research question"* (anonymous) was the most challenging aspect of their network visualization process.

Particular problems arose from wanting to assign link weights based on non-quantitative attributes, such as rankings expressed through words (e.g, 'strong', 'weak', *P-Carers*) or different types of activities ('applied', 'interviewed', 'job-offered'). *P-Carers* initially used a *ranked* scale from 1–10 (1 representing the most important links). However, in the Vistorian, low numbers assigned to link weight result in faint lines, while link weights ranked higher are visually emphasized. They found that the distribution of values in their data was rather small, making it difficult to differentiate between important and less important links visually. While these issues can be fixed quickly and might not be specific to network visualizations, they provide insights into the small barriers that might hinder people from using a tool.

O6: Formatting Data into Node and Link Tables. In line with other network visualization tools, the Vistorian supports node and link tables as data formats. In line with previous work [117], many participants were initially confused about what type of table formatting to apply. Further confusion resulted from declaring links with changing weight over time, which required an individual row in the link table for each time point (and their weight), linked together by an ID column to identify identical links. Without IDs, the Vistorian would interpret each row as a different link. On the other side, this option could be useful to show all possible link weights

simultaneously, i.e., through individual links between the same two nodes. Most participants were able to determine the suitable format for their network.

4.7.2 Iterating on the First Visualization

As illustrated above in several examples, most participants iterated on their first visualization multiple times. We observed participants—both those with and without prior experience in network visualization—not only re-defining their nodes and links but also their data based on issues they encountered with their initial visualization.

👁 ***O7: Data-level Refinement.*** One common issue that participants encountered with their initial network visualization (e.g., *P-Festivals*, *P-Laureates*, and *P-Documents*) was that too many nodes or edges made the network appear as a hairball that was too cluttered to explore in a meaningful way. The primary distinction between ***O3**: Alternative Mappings to Nodes and Links* and this observation lies in their focus. **O3**emphasizes how the selection of nodes and links impacts the resulting network's structure and goal fulfillment. In essence, modifying nodes and links specifications can transform a cluttered network into a clearer, more refined structure. On the other hand, ***O7**: Data-level Refinement* centers on preventing a cluttered network by applying data transformation operations at the data level rather than at the network schema level. The course covered strategies such as slicing the network over time or filtering nodes and edges based on relevance or type. As illustrated above, several participants had defined network schema that resulted in duplicate nodes or links—either because of issues in the data or because of the very nature of the data. For example, *P-Festivals* found that the same venue recruited certain ensembles multiple times for different events. *P-Documents*'s data included labeled documents that included multiple co-occurrence links with other documents. During the course, we provided technical support (in the form of dedicated Python scripts) to help participants explore duplicate links and, based on their goals and questions, transform these meaningfully (e.g., aggregating data points to be represented in the form of link weight) (subsection 4.5.2, Week 3). This helped several participants reduce the size of their networks and highlight important patterns in their data without affecting the nature of the network.

4.7.3 Visual Exploration

👁 ***O8: Tool Features Can Inspire Exploration Strategies.*** Interactive or visual features that seem trivial to visualization designers and exploration experts can be game changers to novice analysts, such as time sliders or timeline visualizations (*"The time-slider provided an entirely different perspective on my data."*). They can provide important impulses for exploring data and inform analysis questions: *"I largely ignored the time element of my data up to this point. The*

timeline provided an entirely new dimension to my data and certainly changed my perspective.". Features such as the timeline assist in finding the story the analyst wants to tell.

Similarly, analysts appreciated the coordinated views showing node-link and matrix representation simultaneously. Participants liked how they enabled interpreting and comparing different visualizations of the same network at the same time and how matrix and node-link diagrams complemented each other (*"Even if I missed noticing that there are two groups in the node-link diagram, the matrix would show me that."*[*P-Criminology*]).

👁 ***O9: Unfamiliar Visual Encodings are not immediately understood.*** A well-known problem, yet most visualizations in the Vistorian were well-understood by course participants. Adjacency matrices caused some confusion in the beginning. While understanding the matrix ordering was easier than we thought, most questions arose from the visual encoding of cells in the matrix, particularly when cells used a multitude of visual encodings: multiple links (cell split into equally sized rectangles), link type (color), link weight (opacity), and link direction (gradient). After understanding the visual encoding, participants faced the challenge of interpreting their own data using these encodings. This observation holds for both the layout of nodes in a node-link diagram and the ordering in a matrix. Participants reported the need to develop trust in the visualizations and their visual observations.

👁 ***O10: Building Trust In Visualization.*** Participants often asked us about measuring the reliability of the observations in the visualization and building trust with respect to the visualization: *"How do I assess whether what I'm visualizing is meaningful, useful, reliable from the data?"*[*P-Carers*]. Moreover, the issue was about the general communication of findings and the scientific rigor in visualization, i.e., to understand *"reliability in network science"*[*P-Carers*] and *"where [visualization] fits on the spectrum of qualitative, quantitative and also mixed methods. Partly, because the course talked a bit about how analysts can always cut things multiple ways, and [then] need to know what is the meaning of [their findings]."*[*P-Carers*]. These questions had implications about the participants' own work and how much they could trust their own observations: *"[Hence,] how will I know that something matters or has meaning, rather than just it's a sort of artifact of having played around with it for a long time."*[*P-Carers*].

4.8 Steps in Network Exploration

Our observations point to specific barriers in the *process* of visual network exploration. To better contextualize these barriers, this section develops a model of 8 steps that we found in exploring networks. The model is informed by our observations (O1-O10) and the participant journeys (J1-J3). It is also inspired by observations in visualization literacy more generally, notably the need to translate abstract domain concepts into visual encodings and back, in order to solve

tasks with visualizations [48]. It also links to Norman's *Gulfs of Execution and Evaluation* for cognitive processes that accompany interactions in interactive systems [188]. Eventually, there are parallels to Munzner's Nested Model describing the increasing abstraction and formalization from domain concepts and problems into algorithms and systems [180].

In the first phase (left side in Figure 4.1), data is formalized from abstract *domain concepts* (i.e., ideas) into visual encodings (i.e., concrete forms ondisplay) in a network visualization. Then, based on this visualization, information is interpreted to inform findings (right side in Figure 4.1) leading back from forms in the visualization to findings in the domain. Throughout these two phases, we can describe 8 steps across an analysts' journey from *domain* to *network* to *data* to *visualization* and back. **1) Identify domain concepts** in the data (e.g., people, events, locations, transactions, participation in events). This is a necessary preparatory step; **2) Define a network** to turn domain concepts into network concepts, i.e., nodes, links, link weight, direction, and (if necessary) apply network transformations; **3) Format data** to turn the network concepts into a machine-readable format such as node and link tables; **4) Import data and specify data schema** to create mappings from the data format (i.e., columns in the case of the Vistorian) to semantics and visual variables in a visualization; **5) Interactively explore visualization(s)** through interaction and choosing among different visualizations; **6) Perceive visual patterns** in a visualization, understanding meaningful constellations of pixels, colors, etc; **7) Interpret visual patterns** to understand visual patterns as network concepts, i.e., 'a cluster' or 'a central node'; **8) Interpret network concepts**, i.e., clusters, paths, and central nodes in the respective domain, and understand their meaning for that domain and respective questions.

Depending on an analyst's expertise and the specific tool used, these steps can be less deliberate and more based on tacit knowledge. For example, an experienced analyst may directly format their network structure into a machine-readable format as they define it or re-work an existing node table to tailor it to a new question. The fact that data can be already formatted *before* the analyst decided on the network structures does not invalidate the model but simply means that a data format is required. However, the model is deliberately kept simple to illustrate the decisions involved in network exploration and to inform future support (tools and methods) for overcoming the barriers.

4.9 Barriers to Network Exploration

Of our 24 active participants, nearly all succeeded in creating at least one network visualization of their own data using the Vistorian. One participant's data (*P-RocksStudies*) turned out not to be useful to be mapped as a network. 11/24 participants used the Vistorian with little help, and 5 participants did not experience any barriers at all. Each of the other participants

experienced between 1-5 barriers each, with the specific barriers experienced greatly varying across participants. Through discussion and explanations, we were able to resolve most, if not all, of these barriers. In the following, we now describe 8 barriers (numbers in brackets indicate the number of participants facing this barrier) and discuss how we mitigated them in our course.

B1: Missing Goals & Questions (3)—Missing goals for explorations are not a barrier *per se*. A more open-ended approach can inspire creative perspectives on the data and novel angles of exploration. However, a lack of specific goals can become a barrier to all steps of network exploration, in particular to early tasks, i.e., Steps 1 and 2 in our model, where relevant data aspects have to be identified and translated into a network schema, and to Steps 4 and 5 where decisions about suitable network visualization techniques influence what types of patterns can be explored. In the worst case, a lack of goals and questions can lead to irrelevant findings or incorrect conclusions, e.g., via drill-down fallacies [157]). *Mitigation strategies*—Use constructive visualization methods [134] (e.g., sketching) to explore data in a low-cost and open-ended way that helps identify points of interest, goals, and questions. Show example network visualizations and explaining what they can reveal (**O8**), to stimulate creative thinking about network visualization and how it can be applied to facilitate sensemaking.

B2: Pre-conceived Ideas & Mental Images (5)—Like missing goals, we think that mental images (which can also be influenced by preconceived ideas about network visualization) about what the data might look like as a network (**O2**) are an essential part of visualization and exploration in that they can inform goals and starting points to network exploration. However, both preconceived ideas and mental images can turn into barriers. They can hamper tasks in Steps 2 & 3, leading to a mapping of domain concepts to network structures that may not match with analysts' goals and questions (**O3**). Even worse, preconceived ideas and mental images can hinder a creative exploration process that allows for new perspectives on the data. For example, they may lead to analysts dismissing new angles of exploration that may or may not include unfamiliar visual encodings (**O9**), to preventing analysts from exploring alternative mappings of domain concepts to nodes and links (see *P-Letters*). In the worst case, preconceived ideas and mental images can negatively influence the exploration and scrutiny of the data altogether, leading to confirmation bias, although we have not observed this in our course. *Mitigation strategies*—An early visualization where analysts can see their ideas realized can be a pivotal point to help iterate on early steps of the data formation process or explore alternative visualization techniques. Support the externalization (e.g., via sketching) and interrogation (e.g., via probing questions) of ideas and mental images early on in the network visualization process in a lightweight manner to help analysts align their visions with their goals and questions. Approaches that allow a critical reflection on data formation and design processes could facilitate this.

B3: Deciding on a Network Structure (4)—Defining network structures such as nodes, links, and link weight has a profound impact on all the consequent steps, as illustrated in Figure 4.1. Depending on an analyst's experience or domain, these mappings can be trivial. However, we found, in particular, participants with little experience in network visualization, struggled with deciding on a network structure (tasks related to Steps 2 & 3 of the network exploration process), which created a profound barrier and delayed subsequent processes. *Mitigation strategies*—Breaking down the process of defining a network structure into different activities can help overcome this barrier (e.g., identifying potential nodes and links, sketch potential outcome visualizations based on this, and formalize outcomes of these processes as concept maps; see our course activities in subsection 4.5.2). However, there are still open questions about how such activities can be integrated into network visualization tools.

B4: Choosing The Right Level of Abstraction (6)—The initial network visualizations of participants do not always yield insights because an intuitive mapping of domain concepts to nodes and links led to visually cluttered networks due to too many nodes or edges (or both). Participants struggled with introducing the right level of abstraction to create more meaningful networks. We define abstraction as the process of deducing a new network from the initial network that describes the entire dataset. This notion of abstraction is different from Bigelow et al. [36], which refers to data abstractions as tables, networks, timelines, or alike. *Mitigation Strategies*—In the course, we discussed options and gave participants access to dedicated Python scripts to facilitate network size reduction, as well as transformation and aggregation strategies. Tool support in this area exists (e.g., [35]), and, again, sketching can help identify goals and approaches to abstraction.

B5: Choosing the Appropriate Data Formats (2)—Choosing the appropriate (machine-readable) data format and transforming data accordingly (Step 3) can become a barrier through a range of issues (**O6**), e.g., (1) not understanding the data format required, (2) confusing different formatting options, and (3) inconsistent formatting. Besides the general benefits of tables for making data accessible [18], node and link tables are relatively straightforward to understand, are editable in common applications such as Excel, and are successfully used in other network visualization tools [247]. However, some participants struggled to decide whether a node table or a link table serves best for their data. *Mitigation strategies*—Provide illustrated guides for table formatting (e.g., [12]); promote individual hands-on activities to table formatting. At a tool-level, it is important to scaffold data import in a step-by-step manner, remind analysts how to interpret different tables, and provide side-by-side views of data tables and the resulting visualization (e.g., as already done in NodeXL [247]). Data validation and consistency checking can be done using different tools (e.g., Python).

B6: Importing Data and Defining Schemata (2)—A data schema creates an explicit mapping between the data and the visualization. For example, in an earlier version of the Vistorian, users uploaded a link table and then had to manually specify the role of each column. In addition, they could upload a node table and a location table, or click a button to retrieve geo-coordinates for locations mentioned in the tables. Observations from early workshops showed that planning and coordinating all these actions is a barrier for some analysts (e.g., *P-Migrants*). Our import wizard (subsection 4.5.1) addresses this issue, but can slow down the data import for advanced analysts. *Mitigation strategies*—Scaffold data import and mapping processes through wizards with detailed explanations while allowing advanced analysts to streamline the process.

B7: Interpreting Visual Patterns in Visualization (2)—Visual patterns in network visualizations need to be understood as network concepts before they can be properly interpreted (Steps 6 & 7). This includes not only low-level encodings of visual elements (e.g., through size or color), but their composition into more complex visual constructs. For example, blocks of cells in an adjacency matrix represent a densely connected cluster, while the same cluster may appear as a set of overlapping arcs in a time arc visualization. While we observed a steep learning curve in reading and interpreting adjacency matrices (**O9**), understanding the richness of network visualization can pose barriers. For example, it can be overwhelming to engage with a network visualization for the first time and to explore it for patterns and clusters—whether or not the technique is familiar. Also, to add complexity, interaction can lead to changes in visual patterns on the fly. To better understand visual patterns, participants reported on multiple coordinated views helping to interpret and verify a pattern visible in one visualization through another one (**O8**). *Mitigation strategies*—Document and illustrate network visualization techniques through demos and practical use cases; provide hands-on tutorials exploring different network visualizations and illustrate their strengths and limitations, e.g., through visualization cheat sheets [268]; support multiple coordinated views.

B8: Establish Trust in a Network Visualization (1)—Trust in visualization is a generic topic [173] and poses a potential barrier to network visualization in itself (**O10**). It can influence if analysts engage in a network visualization process in the first place, how findings are derived from the visualizations, and if a network visualization is deemed valid for evidence in (scientific) communication. One problem can be unfamiliar visualization (e.g., adjacency matrix) and their respective construction methods, such as ordering (in the case of matrices) or layouts in node-link diagrams. Another problem is understanding the *provenance*, including the many decisions along the network creation: conceptualization of nodes and links, network transformations, and filtered elements. For network visualization, there seems to be less of an established culture for how to create trustworthy visualizations in the first place and use them for scientific inquiry. *Mitigation strategies*—Provide sufficient explanation and documentation of the strengths and

limitations of individual visualization techniques (including interpretation pitfalls); explain the underlying algorithm, their assumptions, and possible artifacts they can introduce [268]; show examples of anti-patterns where visualization is misleading or introduces artifacts (e.g., different layout algorithms, clutter in node-link diagrams); use examples of visualizations to show limitations of analytic approaches to network science in terms of network metrics [69, 172] (e.g., an algorithm identifying clusters not visible in a visualization [70]); show examples of network visualization success stories and uses in journalism [47].

4.10 Discussion

We described 8 potential barriers that pose problems to *some* analysts during *some* of their exploration process. In our course, these barriers could all be solved through discussion rather than being built-in problems of specific tools or due to human incapacity. We think these barriers are due to inexperience and underestimating the decisions involved in sensemaking with data and visualization. Eventually, analysts will learn how to navigate the barriers, but if left on their own, our log study and discussions suggest that these barriers indeed block analysts from progressing. Like fire drills, these barriers highlight the importance of training and cautioning analysts upfront.

In summary, our findings also highlight the data preparation stages of network visualization as an important source for insights in themselves and how this process helps novice analysts learning: *"Getting the data prepared to analyze visually [. . .] helped me to realize things about the data that I did not know. It was less about the actual visualizations at the end, it was a lot more about looking at the types of structures we had and thinking about the process..."*[*P-Business*].

Implications for Education—Our work network visualization teaching approaches and related questions to be addressed in the future:

- **Work with Participants' Data.** Allowing participants to work with their own data was hugely important for the success of our course. At the same time, demo data sets are important in explaining and illustrating particular aspects of network visualization. Ideally, demo data should be relevant to the course audience and consistent across the course to allow for the comparison of techniques.

- **Focus on Understanding Goals.** Early course activities should focus on helping participants reflect on their own goals, which is also beneficial for course instructors to provide individual support. Pre-course questionnaires can be beneficial here.

- **Provide Activities to Identify & Mitigate Barriers.** We found that sketching activities provide an easy entry point for analysts to engage in the data preparation stages of network visualization. Additional research is needed to explore how to frame sketching activities in this context and also think beyond sketching.
- **Balance Independent Exploration & Guidance.** We tried to strike a balance between scaffolding participants' explorations and allowing them to come up with and explore their own strategies.
- **Define Clear Learning Goals and Essential Skills.** This is an area to explore more in the future. Our barriers can serve as waypoints and benchmarks for these learning goals.
- **Create Teaching Resources.** We need to start exploring additional resources that can facilitate teaching network visualization in ways that mitigate the barriers discussed in this paper. For example, success stories and galleries of successful network visualizations (beyond node-link diagrams) can help to show what is possible with network visualization and can motivate novices and expert analysts to consider network visualization as a method in the first place.
- **Help Establish Trust.** To build more trust in network visualizations, future research could look into explaining network visualization algorithms in concise and accessible ways (e.g., cheatsheets [268]).

Implications for Tool Design—Our work shows that the range of data and research goals that analysts bring to network visualization is vast. It is, therefore, sensible to start thinking about network visualization tools and teaching that consider the process of building visualization from the analyst's perspective and allow for different entry points in terms of exploration goals, data, visualization techniques, and exploration strategies. Below are some design considerations that can enhance the analyst's experience in visualizing and exploring data when designing tools as well as engaging with network analysts.

- Support the analyst's perception of the data form by offering a concept-demonstrating miniature (e.g., **concept map**) or diagram, an approach already taken by Origraph [35]. Changes and selections to this concept map can be reflected in the network.
- Encourage **incremental network building** and iterative exploration; this is similar to the idea of enabling 'progressive evaluation' in the Cognitive Dimensions of Notations framework [111].
- Provide **coordinated multiple views** using different types of visualization. This not only allows the analysts to take advantage of their complementary strengths but also allows

them to learn how to interpret an unfamiliar visualization by observing how it changes as they interact with a more familiar one.

- Support **creative approaches** to creating networks. For example, show possible networks derived from a given "seed" network through aggregation, projection, and filtering.
- **Explain visual patterns and encoding as part of the tool**, e.g., using automatic annotations suggesting explanations for specific patterns or miniature galleries of visual patterns found inside a dataset.
- **Automatically recommend or highlight** potentially insightful views and data patterns.

Methodology, Limitations, and Open Questions—In-the-wild research is inherently challenging [244, 43]. Our findings are situated in the specifics of our course and course participants, as well as the tool we used. Since analysts may have enrolled because they had particular conceptual challenges in network visualization, we cannot say how prominent our observed barriers are in other settings (e.g., among analysts with different backgrounds and training). Similarly, those participants who engaged in discussions and interviews might not be those who fail *silently* for reasons we were not able to capture. However, the consistency of our barriers across participants hints towards general issues. In the future, we need better methods to collect data about analysts in the wild, especially, to capture (thinking) processes where analysts are not using a computer where actions can be logged. Our experience with mini-questionnaires [239] and microentries [41] was not encouraging due to missing engagement by users. Participants' workflows were influenced by the Vistorian, its required data formats, import routines, and visualizations. Different tools may impact the existence and gravity of barriers. Still, many of our observations are independent of the particular tool, including the barriers during the *definition* phase (e.g., **B1**, **B2**, **B3**).

4.11 Conclusions

We have shown that some of the reasons for people having limited engagement with network visualization tools are barriers to preparing their data and planning their exploration, such as being unable to transform what they have in mind into compatible data structures. Throughout this chapter, I have described such issues in detail and discussed how the data formation process is an essential unit of the visualization tool's design and how the design of this unit should support the whole system/workflow. These findings *must not* be used to blame the analyst but rather shed light on the complex nature of understanding our world through data and visualization. Thus, how can we analyze the interaction logs in the context of user types, visualizations, and

analysts? And how can we describe the success of the network visual exploration process? This is what we shall discuss in our next study of dashboards and learning analytics to support networks' visual exploration. chapter 5.

Chapter 5

Show Me My Users: A Dashboard Visualizing User Interaction Logs

Chapter Overview— This chapter describes the design of a dashboard and analysis pipeline to monitor users of visualization tools in the wild. Our pipeline describes how to extract analysis key performance indicators (KPIs) from extensive log event data and a mix of user types. The resulting three-page dashboard displays live KPIs, helps researchers and developers understand visualization tool users, detect exploratory behaviors, plan educational interventions, and improve tool features. We propose this case study as a motivation to use the dashboard approach for more relaxed monitoring of users and for fostering supportive mindsets for visualization tools.

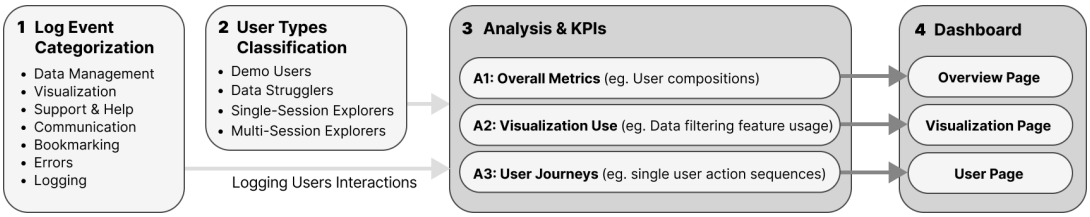


Figure 5.1: **Dashboard Design Pipeline of Visualization Tool Logs**: User interactions are logged and categorized by event types (1) to determine different stages of visual exploration. With user types classified (2), we are then able to apply analysis and calculations (3) according to our key performance indicators (KPIs). The dashboard (4) pages and contents are informed directly by the KPIs.

Authorship Statement— This chapter has been published, as noted in section 1.5, with authorship details clarified in the declaration at the beginning of this thesis. I declare that the work in this chapter was divided equally between myself Mashael AlKadi and Jinrui Wang. The work I handled was the data analysis of logs for metrics definition and calculation, whereas Jinrui handled the design of the dashboard interface. We both discussed with Benjamin Bach,

the supervisor, the metrics and the refinement of the visualizations and dashboard design. Jinrui Wang covered the design section details and I covered the data analysis section and content. And all the co-authors of this chapter: Jinrui Wang, Mashael AlKadi, and Benjamin Bach have contributed to the writing, discussion, organization, and polishing of the paper published in VIS 2023 under the title of "Show Me My Users: A Dashboard Visualizing User Interaction Logs".

5.1 Introduction

In the first two studies of this thesis (chapter 4), I was able to identify user types and barriers analysts face in network visual exploration. So, how can we use our findings to offer live analytics describing the tool's usage (e.g., visualizations and interactive features) to support developers in decision-making? In addition, how can we measure how successful the analysts are and advance in their network visual exploration? how can we measure and formalize the process from our collected interaction logs? In this chapter, I will focus on the analysis of interaction logs for a better evaluation of the visual exploration process. That is, understanding how people use a visualization tool in their daily work environment is key to improving tool features and providing users with adequate training. In particular, I am interested in understanding where people struggle to use a given visualization design or interactive features over time [14] and inform decisions about specific features, identifying users in need of help, developing new training methods, or planning for in-person interventions (e.g., workshops, drop-in sessions). This goal poses two distinct yet intertwined questions: *(Q1) What information to capture about tool users using the tool? and (Q2) How do we present these data to researchers and developers to support decision-making?*

Current methods to collect and analyze usage data across a diverse crowd of users (anonymous, geographically distributed, with different backgrounds, tasks, and data) over a longer period do not support these tasks well. For example, mini-questionnaires [239] that prompt users for written feedback or online check-in session [2] can respond to a user's specific questions but require active engagement by users, qualitative analysis, and do hardly scale. Interaction logs do capture users' interaction events with the tool, but their analysis usually requires bespoke analysis scripts [263, 101, 114, 212, 255] or time-intensive manual exploration [84].

In this chapter, we design a dashboard and analysis pipeline that provides an at-a-glance live view to monitor user types and their engagement with the visualization software. The approach is inspired by the use of analytics and dashboards in the field of Learning Analytics [96, 143]. However, unlike those approaches, which track students' progress throughout an often well-defined course structure, visualization systems are much more open, using multiple visualization designs and interactive features, loading individual data sets, and developing personal exploration

and analysis strategies. Hence, our pipeline (Figure 5.1) features a range of interaction events, their collection mechanism, classifying users into different types, calculating key performance indicators (KPIs) on the collected data, and designing an interactive visual dashboard with three pages. The resulting live overview allows us to understand general trends in tool use, identify the impact of changes to features, training resources, or educational activities, and plan for future interventions.

We base our design on a case study with *The Vistorian* (`vistorian.net`) [2, 17], as described earlier in section 3.1. The dashboard was developed over several rounds of iterations among the authors of this chapter, who at the same time take the role of the data analyst to inform their research around the Vistorian. In the following, we describe our analysis pipeline (section 5.3) and dashboard design (section 5.4), then discuss our work through findings from up to 2298 users using the Vistorian over 28 months (section 5.5). We close by discussing implications and lessons learned from the dashboard and the potential of monitoring user behavior in visualization tools live (section 5.6). A live version of the dashboard is online via `https://halcyonwjr.github.io/Vistorian_Dashboard/` with more details in the GitHub repository [1] and Appendix B.

5.2 Background and Related Work

In this section, I provide a recap of the literature provided and discussed fully in chapter 2. If literature has been read earlier, this section can be skipped to the analysis section (section 5.3).

5.2.0.0.1 The Vistorian Use Case The Vistorian is an open-source network visualization web application with five interactive visualizations: node-link diagram, adjacency matrix, timeline, map, and a coordinated view of node-link and matrix. It supports users in uploading data, creating networks, and exploring visualizations interactively. The full description of the Vistorian is discussed earlier in section 3.1. Through the Vistorian, we provided various help resources such as examples or visualization manuals. The description of different help resources provided is included in subsection 3.1.5. The Vistorian has been used in numerous workshops at scientific conferences, bespoke network visualization workshops, and summer schools [2]. Its main target audience are analysts without programming skills. While the authors of this paper have plenty of informal feedback from workshops about how people use the tool, we have so far no means to monitor progress during and outside of workshops.

[1] `https://github.com/HalcyonWjr/Vistorian_Dashboard`

5.2.0.0.2 Interaction Log Analysis Interaction logs capture single interaction events and are commonly used for webpages and software systems. In visualization research, they play a critical role in many studies to describe user intentions, behaviors[25], provenance[194], and workflows either in different levels of granularity [101] or as a sequence [114]. Many of these case studies are based on small data sets obtained from lab settings[101] with clear analysis tasks[39]. A common way to visualize user interaction events is through timelines [84] or diagrams (e.g., [280]). The field of Learning Analytics focuses on investigating users interactions with the goal to understand people's learning and progression through digital learning platforms [29, 171, 183], educational videos [160], and academic performance through courses and learning management systems [170, 171, 88, 152]. Our work builds on these approaches and extends them to the open nature of interactive multi-view visualization systems.Those systems lack explicit assessments of users' progress such as tests and assessments.

5.2.0.0.3 Dashboards design While dashboards are used across many applications, including learning analytics [170, 86], designing dashboards is not trivial. It requires contemplating the analysts' tasks and developing solutions that make the best use of the available screen space and visualization affordances. Sarikaya et al. [230] described example dashboards by their goals, purposes, audiences, visual features, and data semantics, among which an in-depth analytic dashboard requires careful choice of data, visual layout, as well as high demand on designer and user's visual literacy. More actionable, Bach et al. [13] described design patterns for dashboards, for example, layouts, visualization, or structure. Qu et al. [211] studied how designers balance visual consistency across multiple views, and Aparício developed a BI dashboard decision tree to support the design process [76]. Our design is based on this guidance and can generalize to visualization systems beyond our use case.

5.3 Analysis

The analysis pipeline consists of the following steps (Figure 5.1). Firstly, log events are used to distinguish various user interactions. The log categorizations (Figure 5.1-1) are defined on the basis of the common tasks in the visual exploration process. Then, we categorize four user types (Figure 5.1-2) based on key factors separating them in exploration stages. From logs and user types, we reach three analysis KPIs with detailed requirements (Figure 5.1-3). The calculation outcomes are used to feed the dashboard composition and layout design (Figure 5.1-4).

5.3.0.0.1 Log Events Types and Logs Collection I log 94 types of interaction events using the *Intertrace* JS library[93]. Events are grouped into seven categories: *Data-Management* (e.g.,

create a network), *Visualization & Interaction* (e.g., change visual encodings), *Support & Help* events (e.g., usage of help resources), *Communication* events (e.g., consult the team), *Bookmark* events (e.g., capture/annotate a network state), *Error-Tracking* events (e.g., report issues), and *Activity-Logs* events (e.g., share feedback). The full list of events is included in subsection 3.2.2. Once a user visits the Vistorian, a unique session ID is assigned to the user's interaction logs. The session ID will last with the user until the browser's cache is cleared or the browser page with incognito mode is closed. In such cases, a new session ID is assigned to the same user after revisiting the Vistorian page. I combine session files if they share the same IP address, and the new session was created within 1 minute or less of the last session's end time (the time to refresh the browser memory and continue using the Vistorian).

5.3.0.0.2 User Types and Log Calculations

Our pipeline features four user types (as reported in section 4.4) based on (i) whether they used their **own data** or demo data, (ii) whether they **created any own networks**, and (iii) whether they **returned for multiple visits**.

- **Demo user:** no own data, but only use the provided demo data.
- **Data strugglers:** users who tried by did not manage to create their own network from a data table.
- **Single-session explorers (SS_Explorers):** users with a single session using the Vistorian with their own data.
- **Multi-session explorers (MS_Explorers):** users who return for at least one additional session at least 20 minutes after their first session.

Combining log event types and user types in the logs analysis process assists in calculating various metrics. For example, exploration patterns can be classified based on their themes (e.g., temporal, geographical,..). Also, the tool's usability can be measured for each user type instead of just the overall usage. This enables a better understanding of the tool's design and how it supports different users needs. That can be reached by observing the variation in a number of user types over time, examining user decisions through exploration sessions, and assessing the impact of educational efforts (e.g., courses).

5.3.0.0.3 Requirements and KPIs

From log event categories and four types of users, we are mostly interested in unfolding KPIs from three aspects:

- **A1: General metrics** about number of users, time spent with the tool, change over time, types of users, etc.
- **A2: Visualization use** such as time spent with each visualization, trends over time, and use of interactive features.
- **A3: User journeys** for specific users and their use of the system.

The specific requirements are listed in (Table 5.1), which have been systematically identified in our previous studies on understanding users through (tedious manual) log analysis alongside interviews, workshops, and drop-in sessions [2].

Table 5.1: Analytical KPIs list covering three main aspects guiding dashboard design. Codes for each KPI (e.g., a1) refer to specific charts in the dashboard.

A1: Overview page (Figure 5.2)	
1	Number of total users (a1)
2	Trend of user visits over time (a2)
3	Distribution of session length by user type (a3)
4	Users returning rate (a4)
5	User visits changing with the external events (b)
6	Most frequently visited features (c1)
7	Use of help resources (c2)
8	Time spent by user type (c3)
A2: Visualization page (Figure 5.3(a))	
1	Time spent on each visualization per user (d2)
2	Number of users using a visualization (d1)
3	Use of interactive visualization features (d3, d4)
A3: User page (Figure 5.3(b))	
1	Interactions per individual (e2, e3)
2	Number of networks successfully created by each user
3	Exploration sequence for different user types? (e1)

5.4 Dashboard Design

The dashboard was designed over 3 months, featuring regular consultation among the co-authors of this chapter: myself, Jinrui Wang, and Benjamin Bach. We followed Bach et al.'s [13] six-stage design process for dashboards through identifying data abstractions, data board structure and pages, page layout, visualization design, screenspace optimization, and designing interaction.[2] Where appropriate, we refer to Bach et al.'s design patterns in *italic* formatting. Throughout the design process, I and the co-authors evaluated if the design answers the tasks in Table 5.1, and observed user patterns across charts and pages.

The three main goals (A1, A2, A3) informed a *multi-page structure*. The **Overview (A1)** page is the *hierarchical* leading page (Figure 5.2), followed by two pages about **visualization (A2)** (Figure 5.3:a) and **users journeys (A3)** (Figure 5.3:b). Each page is kept in a *screen-fit* size to provide all the information at a glance. Full screenshots of the dashboard can be found in Appendix B.

[2] `https://dashboarddesignpatterns.github.io/processguidelines.html`

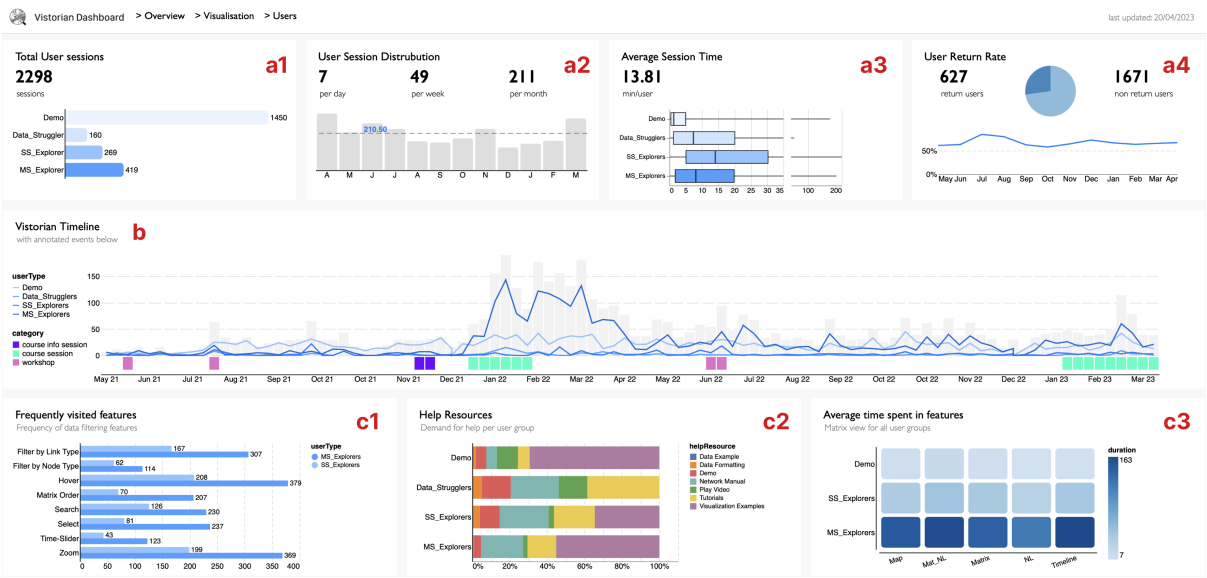


Figure 5.2: Page **Overview**: A navigation bar on top, three stratified rows to report overall trends and usage in the Vistorian.

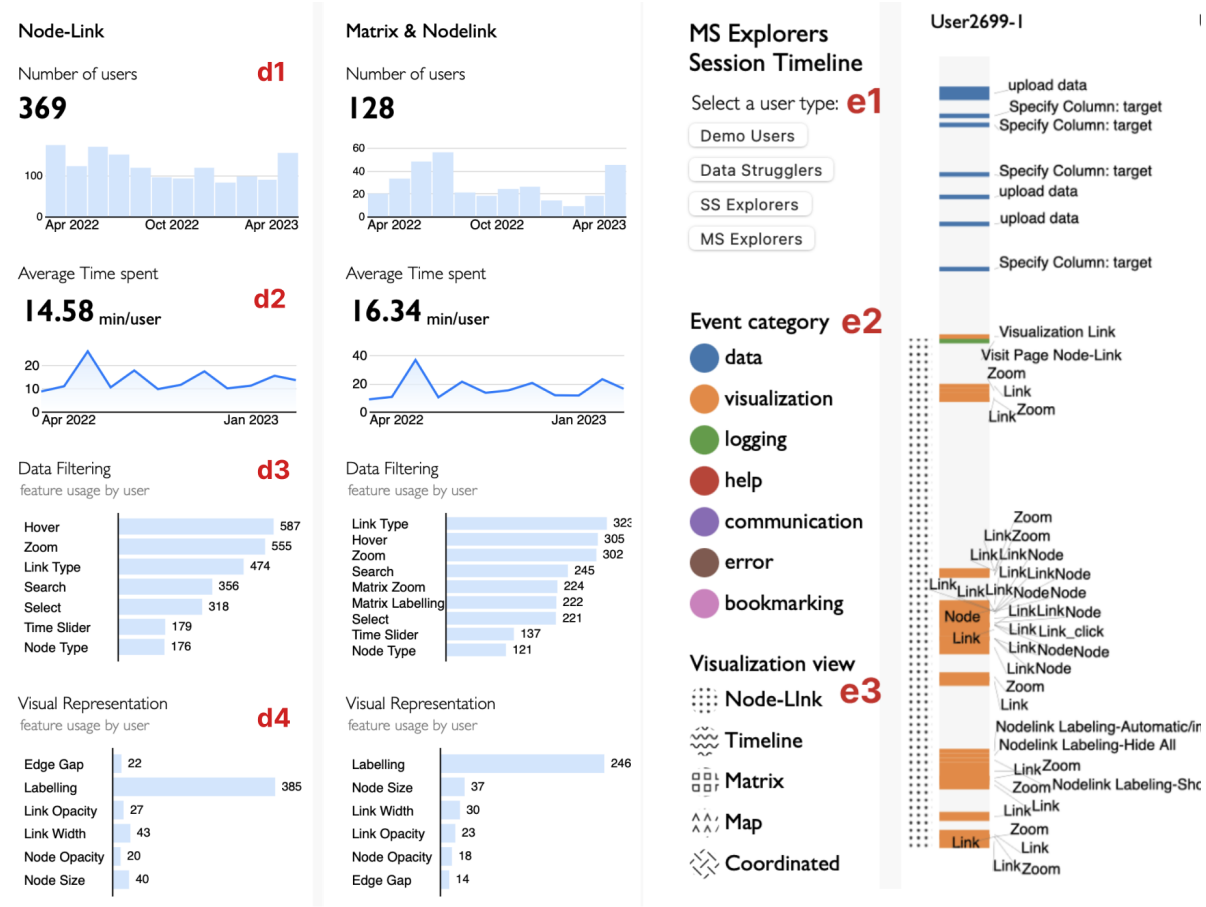
5.4.0.0.1 Overview page The overview page reports on the overall activities with the Vistorian. The *stratified layout* organizes three levels of information as Figure 5.2 shows: (a1-a4) user statistics, (b) temporal trends, and (c1-c3) key features usage.

To glimpse the initial profile of tool users, the first group of charts speaks of up-to-date facts about user composition (a1), monthly sessions (a2), session length (a3), and retention rates (a4), establishing the context for tool performance. Meanwhile, we are conducting research with the tool about visualization in practice, where we require simple, real-time access to how these efforts impact users, and to compare temporal trends of past efforts. The annotated timeline (b) accommodates this need, allowing us to account for workshops, and training courses in relation to the changes in users. Furthermore, we would like to learn what are the go-to resources for help and do different types of users show differences in feature preferences. This set of charts answers these questions from the popularity of interaction features (c1), supporting materials (c2), and users' interest in network visualization by session time (c3).

5.4.0.0.2 Visualization page The Visualization page uses a *table layout* (Figure 5.3:d) where each column depicts one visualization view, and each row measures shared metric across different views. The complete view of the visualization page (Figure 5.4) shows statistics for all of the Vistorian visualizations and views. Our previous study [2] revealed that users differ in view preferences and interaction techniques. Therefore, this page includes (d1) the number of users, (d2) time spent on each view, and (d3) popularity in interaction features (Table 5.1) to compare different visualizations in use.

We adopted the Gotz' and Wen's [107] taxonomy of user's intent to classify interactions.

The first group includes data filtering (d3), e.g., time slider, filter by node, and link type. The second group changes the visual representation (d4) of elements, i.e. label ordering techniques in the matrix view. The dashboard lists the frequency of events triggered to represent the usage popularity. These comparisons also encourage us to reflect on feature and user KPIs without distinguishing views, provoking further speculations into specific trends that are inconsistent with other views or the overall trend.



(a) Page Visualization (detail) (a) Page Users (detail)

Figure 5.3: (a) Page **Visualization**: a table layout showing data about each visualization. (b) Page **User**: individual interaction sequences color-coded by event categories.

5.4.0.0.3 User page Guided by the third aspect (A3), we designed a timeline to show users' action sequences. Previous studies summarized an 8-step visual exploration framework [2], where actions in each step are traced in our log list. Though the framework can guide the learning process, it will be heavily interrupted and tweaked in practice, where timelines provide an accurate representation. The user page in Figure 5.3:b adopts a *grouped layout* where each timeline shows interaction events for an individual user (Figure 5.3:b) showing only one timeline). Each interaction event is visualized as a horizontal dash on the gray timeline, color-coded by its category (e2) while the specific event name is shown as a label. Events that are

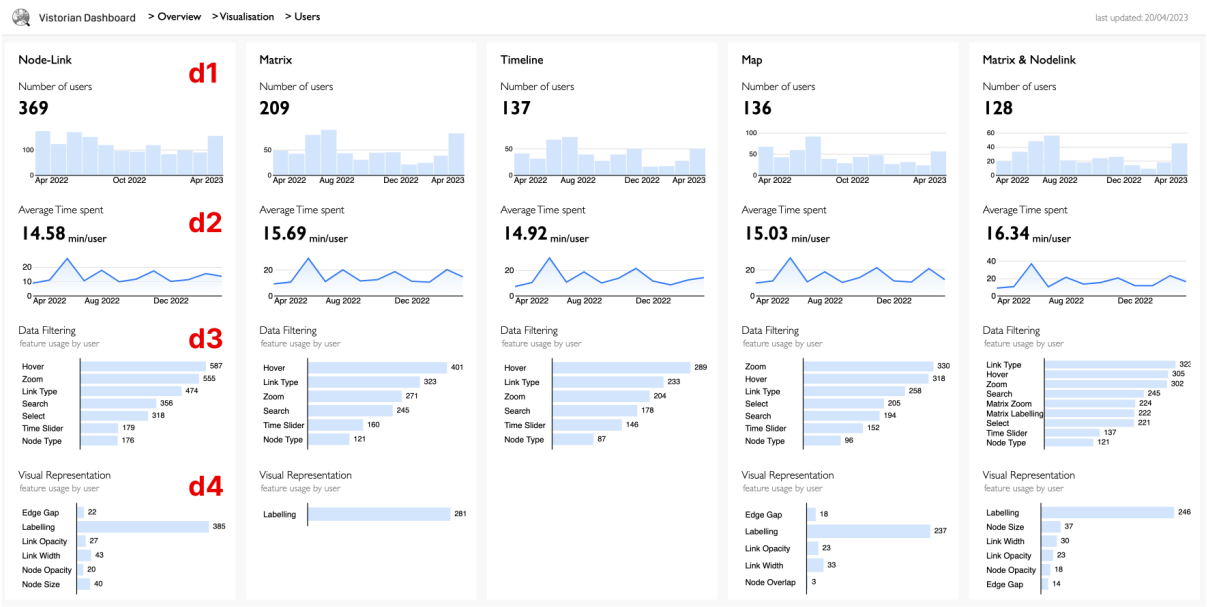


Figure 5.4: The Full Visualization Page: showing all visualizations' analytics along with coordinated-view.

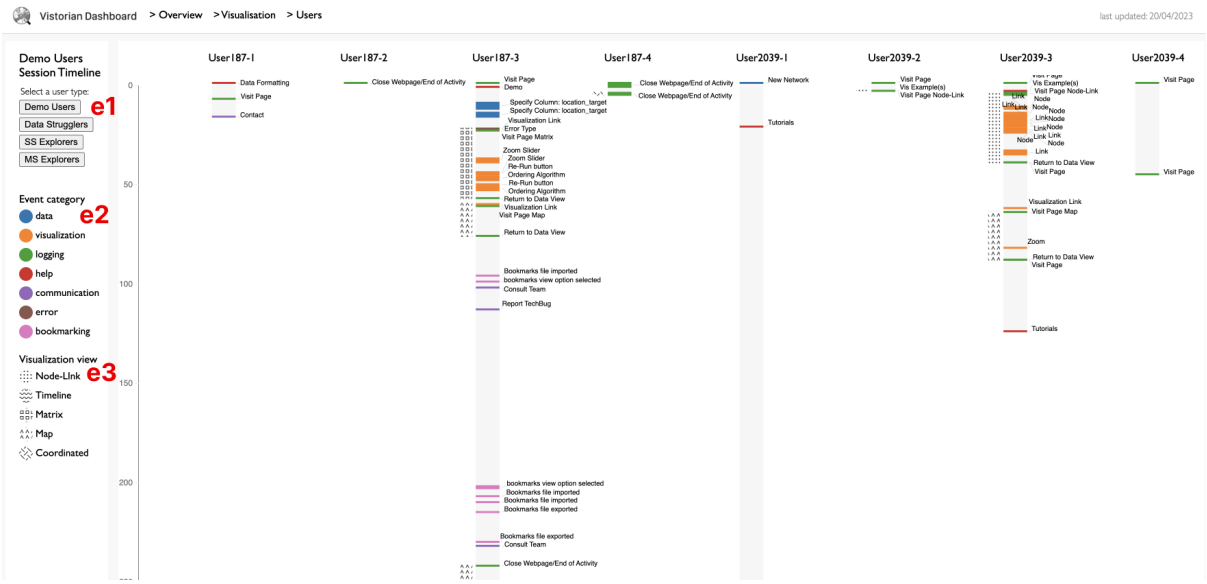


Figure 5.5: User Page: shows a timeline of each user's exploration sessions. Sessions can be filtered based on user type, visualization type, and event type.

triggered continuously within a one-second (e.g., continuous hovering of nodes while browsing labels) interval are combined into a block on the timeline and measured using the first to the last event timestamp. Adjacent to each session timeline are sidebars indicating the user's current visualization view through different textures (e3). Interactive filtering can be used to show and hide individual types of events (e2). The full view of the page is found in Figure 5.5. Additional scenarios are discussed in the appendix (section B.3).

5.5 Findings from Dashboard in Use

Over 28 months, we collected logs from 2298 users. Besides users using the Vistorian at home, the time period included a 6-week course on network visualization and the Vistorian,[3] a 2-month coaching program with individual drop-in sessions,[4] and four workshops stretching a few hours each. Below, we report on both findings from the logs and how these findings influenced decisions around the design of tool features and educational efforts.

Understanding user trends: just over 50% of all users were demo users and just about 5% (160) were data strugglers (Figure 5.2:a1). About 25% of users returned to the Vistorian for additional sessions (Figure 5.2:a4). Those returning (MS_Explorers) spend an average of 15 minutes on their sessions (Figure 5.2:a3), while demo users spend only about 5 min with the tool in each session. From the timeline (Figure 5.2:b), we can see a constant yet slightly increasing number of users (gray bars) with an increase in sessions during the 6-week course (cyan bars under the timeline). During the same period, we can observe a robust, increasing trend in the matrix and node-link coordinated view (Figure 5.3), justifying our educational effort. In (Figure 5.2:c2), we can see that most users use visualization examples (purple), except data strugglers who rely heavily on tutorials (yellow), and videos (green). The more experienced a user becomes (MS_Explorer), the less they rely on demos (red) and data formatting (orange) help. The analysis of help resource usage leads us to create more help videos and provide additional example datasets, particularly to support those who are struggling with data. Likewise, we used the individual timelines (Figure 5.3:b) to track users and prepare them for our encounters in the drop-in sessions.

Understanding individual exploration: Analysis of features' usage provides insights about the *exploratory behavior* per user type and raises further research questions as in (Figure 5.2:c1). For example, on average, users take about two minutes to upload data and create a network. Those taking longer keep creating networks without exploration, which requires our attention. We also observed in the individual timelines that most users start by using node-link diagrams and then explore the other visualizations. Currently, we are working on guidelines for self-directed learning to guide users through visual network exploration. These observations could lead to new KPIs (e.g., flag users that upload more than 3 data sets without exploration) to inform educators. Likewise, these observations can lead to new hypotheses and questions about, for example, why single-session users explore fewer ordering algorithms in the adjacency matrix, why they tend to examine temporal changes in their networks to a lesser extent, or whether this is related to a user's tasks (e.g., no interest in temporal exploration) or their skills (e.g., lack of awareness of temporal exploration).

[3] `https://vistorian.github.io/courses`
[4] `https://vistorian.github.io/upcoming_courses`

Improving tool feature: Based on the data about data strugglers and our observations from the workshops, we designed and implemented a more comprehensive data uploading wizard with a linear set of configuration steps and detailed explanations at each step [2]. In addition, further coordinated views can be created based on users' common exploration patterns. Currently, we are in the process of adding more help resources, improving interaction affordances, and increasing the range of visualization techniques.

5.6 Discussion & Future Work

This chapter presented a dashboard design and analysis pipeline for visualizing user interactions with an interactive visualization tool. Our observations show how monitoring users over a period of time can inform decisions for interventions and further analysis. They also call for **more 'casual' monitoring of users**, whether the respective tools are research prototypes or mature tools. By *casual*, we mean the *constant, exhaustive, and up-to-date display of interaction log data*. This approach is different from dedicated log analyses performed with either specific questions in mind, focusing on specific aspects of a tool, or not running alongside tool development and education. Our user types can be seeded as personas, further informing help resources for particular audiences.

I believe our dashboard design and KPIs are **generalizable to many other interactive visualization applications**. Adaptions would require extending or limiting the types of events and setting different parameters and thresholds for our user categorization. New KPIs could be added and more widgets and pages designed. For example, we can imagine pages dedicated to interaction features, specific user groups, or specific workshops. We can also imagine a general protocol and architecture for visualization tool logging and visualization with dashboards. Such a framework could also try to feed on **additional data sources**, such as textual notes from users [40], call information, email exchange, or workshop attendances. However, qualitative data requires bespoke analysis and careful evaluation to determine the extent to which the derived KPIs really reflect users' concerns and progress. We need more research to see if KPIs represent causation rather than correlation.

In particular, we want to encourage the creation of specific **KPIs that can capture learning or failure in visualization use**, such as an increase in the frequency of using visualization and the frequency of accessing help resources. Good KPIs must be able to capture such specific developments, but they are naturally harder to define and require proper evaluation. Such KPIs could feed into **automatic recommender systems personalizing help and support** to remote users. This will help address issues like intrusiveness and contextual inaccuracy typically associated with recommender systems. Ultimately, we might be able to create very individual

user profiles and help individualize and tailor (e.g., for specific groups of users) features and education. We could finally be able to track learning, i.e., how individual users learn visualization interfaces and visualization knowledge. Currently, we still know too little about our users or describe them based on assumptions and informal evidence (e.g., novices, domain experts, data scientists, etc.) [56, 141].

However, it is important to **create a 'carer mindset'** with principles that care about our tools, their deployment and use, their continuous development, and the education to help people master new tools and interfaces. The visualization community is full of stellar examples of care mindsets, e.g., for tools such as D3, Vega, or Eventflow [84]. I believe our approach is a step towards building care mindsets for visualization tools and helping increase the impact of data visualization. So, how can we make use of our findings about the tool's usage, analysts' needs, and the visual exploration process in teaching and learning networks visual exploration? And how can we design guides to support the self-regulated learning of analysts in the wild? This is what we will discuss in our next chapter.

Chapter 6

Roadmaps, Pathways, and Activities: Helping Analysts Understand, Plan, and Personalize their Visual Network Exploration

Chapter Overview— This chapter develops a visual roadmap—akin to flow diagrams —- alongside activities that help analysts with their network exploration. The process of exploring data can involve many steps, such as formulating goals, collecting data, learning visualizations, formatting data, choosing a visualization, performing interactive exploration, and evaluating observations. Yet, our prior research has identified a number of barriers that can hamper this process. One reason for these barriers is the cost or lack of bespoke guidance on how to get started and how to navigate the steps and decisions involved in an informed and potentially non-linear and iterative way. To address this problem, we develop a visual roadmap and associated activities that, for the first time, give a holistic overview of the steps, artifacts, and decisions involved in the exploration process. The map and activities are based on interaction logs, workshops, interviews, as well as 1-1 coaching sessions with 10 analysts over 2 months. Either individually or together with tutors and peers, analysts can identify their individual starting points and negotiate personal pathways. Our coaching sessions and interviews show that the roadmap supports analysts to think more deliberately about their data and analysis and fosters learning and understanding of network exploration. Our roadmap can be considered as a foundation for future work on designing and evaluating self-guided learning and associated activities, learning resources, and tools.

Authorship Statement— This chapter has been published, as noted in section 1.5, with authorship details clarified in the declaration at the beginning of this thesis. I declare that the

work in this chapter was mostly done by the student Mashael AlKadi. By work, I mean and include the following: designing the roadmap and supporting material such as videos, slides, and Python scripts. In addition, I ran all of the coaching program sessions and interviews, collected data from logs, observations, sessions, and discussions, and analyzed data. Magdalena Boucher assisted in polishing the design and aesthetics of the roadmap and the design of data comics, one of the support materials offered to participants. Benjamin Bach supervised the design and analysis process and the entire study. Uta Hinrichs (co-supervisor) supervised the study discussions as well. I have generated the content of this chapter. All paper authors (Mashael AlKadi, Magdalena Boucher, Wolfgang Aigner, Uta Hinrichs, and Benjamin Bach) contributed to the discussion, organization, and polishing of the paper that is still under submission with the title of "Roadmaps, Pathways, and Activities: Helping Analysts Understand, Plan, and Personalize their Visual Network Exploration".

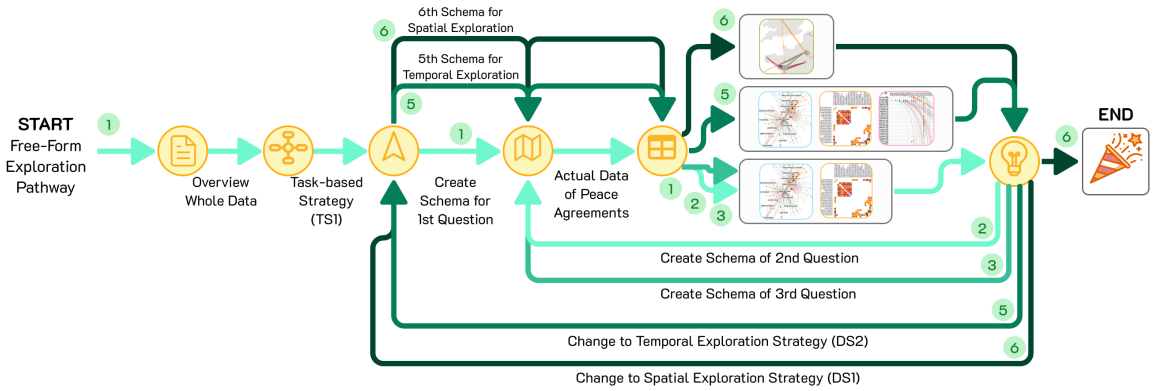


Figure 6.1: **Example of a network exploration trajectory**. This figure shows an exploration trajectory of an analyst on a free-form exploration pathway. The analyst navigates through the roadmap's steps over several iterations as the numbers indicate. Through different iterations analyst changes network schemes based on a chosen strategy.

6.1 Introduction

In the earlier chapters, we discussed several barriers analysts face in visual network exploration along with user types (section 4.9). In addition, we were able to define metrics and KPIs to measure the tool's usage and analysts' needs (e.g., help resources usage) (chapter 5). So, how can we use our findings to inform the design of educational guides for network visual exploration? And how can we design such guides to enable analysts' self-regulated learning in the wild? In this chapter, I focus on how we can assist analysts in overcoming some of the barriers by offering a step-by-step guide. And how can we understand and characterize what

sort of steps or strategies analysts follow to explore their data? Thus, in this chapter, I report on how we investigate further how the design of a roadmap for network visual exploration can assist analysts in exploring their networks.

To make data visualization more widely used, we need to promote ways to overcome barriers to utilizing the respective techniques and tools. For instance, my recent study, as detailed in chapter 4 and published as Alkadi et al. [2], has identified several potential obstacles to data exploration by examining analysts as they prepare for and conduct visual network exploration. These barriers point to issues and misconceptions when formulating research questions and possible exploration goals, defining subsets of the data for analysis, choosing relevant visualizations, and conducting an iterative exploration process. The level of complexity of these challenges can vary depending on the data model and structure, the size of the data, and the type of data being visualized. Most importantly, many of these barriers are concerned with high-level thinking and understanding of how to perform data analysis and exploration and are echoing findings in other areas of visual exploration [71, 154, 5].

One way to overcome the respective barriers is to provide better guidance and training to individual analysts. By *analysts*, I mean any person who aims to explore and analyze their data visually in any setting and from different backgrounds (e.g., level of expertise, domain). Such as users using visualization tools outside controlled settings (e.g., in the wild, remote settings). Such guidance needs to be *individual* and *context-specific* by considering an analyst's data, research questions, and personal data literacy. Or, as one interviewed analyst put it: "*I seem to have a black hole between understanding principles and examples and working out what I can do/apply to my* **own** *data!"* [2]. However, besides group workshops (section 4.4, section 4.5), and other forms of formal education, there are little to no means available that engage an analyst alongside their analysis in an informal and personalized way. Thus, to address the rest of this thesis's research questions on how can make use of the barriers (chapter 4), methods (chapter 3), and metrics (chapter 5) extracted to support visual network exploration in the matter of tool's design and learning process. Hence, in this chapter, I ask and focus on the following questions: *R1: How can we characterize the visual network exploration process(es) that individual analysts perform?* This would help to understand steps, crucial decisions in an exploration process, and individual differences. Following this, we can then ask *R2: How can we facilitate the exploration in a flexible way tailored to analysts' diverse individual goals and strategies?* Those two questions, R1 and R2, align and address this thesis's research questions RQ4 on strategies characterization and RQ5 on network visual exploration education, respectively. Please refer to section 1.2 for all of the thesis research questions.

This study addresses these two questions, R1 and R2, through an in-depth analysis of analysts' workflows as well as describing a set of materials and activities to guide analysts. This research

directly builds on our previous work (chapter 4) on identifying and understanding analysts' barriers as published in [2] and provides yet unpublished evidence on analysts' exploration process. I used the analysis of our previous study chapter 4 to inform the design of the roadmap and then evaluate it through the coaching program. Based on this analysis, I identified common exploration *strategies* as well as four groups of analysts (R1). Figure 6.1 presents an example of an individual analyst's exploration process, showing the steps, outcomes (yellow icons), and activities (arrows) this analyst performed. Strategies and analyst groups then informed the creation of *pathways*, which are common processes that we can suggest to individual analyst groups. A visual *roadmap* (Figure 6.4, similar to Figure 6.1) maps out the ensemble of all *steps*, *decisions*, *activities* and their respective *learning material*, *pathways*, and *strategies* that we either identified in our analysis or created to support R2. Akin to flow diagrams for clinical guidelines that guide physicians in standard treatment pathways, our roadmap aims to help analysts (and educators alike) overview, identify, and discuss possible pathways for their analysis.

To evaluate and iteratively improve our roadmap, activities, and materials, I ran individual 1:1 coaching sessions with 10 analysts over a period of 2 months. Our findings show strong support for our material and personalized approach and point to future work on guiding and training. In summary, the contributions of this work comprise:

- 16 strategies for network exploration (subsection 6.3.1).
- Four pathways informing common strategies and steps (subsection 6.3.2),
- a visual roadmap illustrating steps, decisions, learning resources, and activities involved in exploration (section 6.4,Figure 6.4),
- observations from individual coaching sessions (section 6.6).

6.2 Background

As detailed literature has been discussed in chapter 2, this section briefly summarizes related work around the following: guidelines for understanding visualization, activities, and educational material, as well as network visualization literacy and our own prior work.

6.2.1 Guidance for Understanding Visualization

Visualization design guidelines have some traditions to help inform good design practices for visualization [83, 181, 281, 269, 176]. However, apart from individual approaches to explain visualization techniques [224, 268], guidelines on how to *teach* the process of creating, designing, and exploring data through different types of data visualizations are not investigated

in-depth. Many studies[174, 83, 200] introduce guidelines on how to explore data, but it is unclear how they differ from one visualization type to another. Different visualization literacy assessment studies [48, 159, 44] show the importance of literacy assessments to determine gaps in the design of educational content. Some visualization literacy assessment tools show that high scores on assessment tests cannot be directly translated to the ability to "comprehend data and execute analytic tasks with visualization tools"[159].

The importance of having guidelines also emerges from mirroring the learning process of concept or hands-on skills [281]. Visualization tools working with complex data structures deal with data as 'black boxes' [5], where analysts feed their data to the tool and obtain visualizations without a way of understanding how they were created in the background. This results in what has been called *analytical gaps* [5] such as the gap between representation and analysis. This is described as "the gap between what is being shown and what actually needs to be shown to draw a straightforward representational conclusion for making a decision" [5]. Guidelines have been developed to guide analysts in—again—creating visualizations through concrete steps [174, 175]. Those steps have been differently represented over studies. For example, steps are presented in the form of worksheets that break down design activities into several actions (e.g., understand, ideate, make, and deploy) [174]. Other guidelines are designed based on the audience type and instructional focus [281], or based on visualizations query tasks (e.g., identify, compare, and summarize) [257]. To the best of our knowledge, there are no detailed accounts and guidelines to support analysts in exploration tasks around network data.

6.2.2 Activities and Educational Material

The development of educational material and activities for data visualization design has been investigated in several studies aiming to support learners in different educational settings and mediums (e.g., toolkit) [219, 53, 267, 218]. A different range of analytical skills and knowledge are required to process data and create suitable data visualizations [159]. Hence, some activities focus on supporting the discovery process in both theory and tools, such as CLUE model [109], which stands for Capture, Label, Understand, and Explain. However, developing educational material dedicated to the learning cycle or part of it has been partially investigated. For example, Tanahashi et al. [257] focused on designing *introductory* materials for information visualization, whereas Roberts et. al. [219] focused on creating a framework that walks students through the visualization creation process in general. Other studies focused on developing educational material based on different aspects such as visualization tasks and teaching methods [52] and learning styles [85] such as active learning [98], where participants are involved in activities through the learning process. Again, all of these materials focus on learning design, but we require materials and activities to help exploration.

6.2.3 Network Visualization Literacy and Training

Zoss et al. [282] have defined *network visualization literacy* as the ability to read, interpret, and visualize different types of networks. A number of studies have investigated network visualization literacy from various perspectives, such as how to evaluate network visualization literacy or how to support it in education [135, 282]. The training and teaching of network visualization concepts have been investigated in different educational and instructional environments such as courses [77], public exhibitions and museums [44], as well as brainstorming sessions on networks literacy [231]. Each study targeted different types of participants such as students, museum visitors, and professionals. Other studies have focused on how to support analysts navigating through networks and how to overcome some of the obstacles faced while interpreting networks [127, 272, 216]. However, few studies focused on how to teach network visualization concepts and support analysts' learning and exploration processes by offering different learning resources [59, 136, 135]. Studies suggest that a mix of different forms of education are suitable to support the learning process [281, 282, 95]. Our investigation complements these existing studies by providing a new set of learning materials and activities specifically for network data.

As part of our previous work [2], I logged and analyzed interactions with the Vistorian as discussed in section 4.4, a web-based network visualization tool, over the course of 10 months. The Vistorian offers 4 network visualizations, which are node-link diagrams, adjacency matrix, timeline, and network over a map, in addition to a coordinated view (see section 3.1). We also conducted a 6-week network exploration course with 24 analysts section 4.5. As part of that course, participants worked with the Vistorian on analyzing either their own data or a provided demo data set. The findings of our previous work are based a high-level analysis of our collected data; interaction logs, questionnaires, semi-structured interviews, challenges and course journeys, as well as recordings from the course sessions. Our study sheds light on processes around (1) defining a network at a data level, (2) defining and iterating on network visualizations, and (3) approaches to visual network exploration. Most importantly, 8 barriers that can hamper network exploration were identified: *B1: missing goals and questions, B2: preconceived ideas and mental images, B3: deciding on network structure, B4: choosing the right level of abstraction, B5: chose appropriate data formats, B6: importing and defining data schemes, B7: establish trust in visualization*, and *B8: interpreting visual patterns*. While our previous work provided a first high-level overview of real-world approaches to network exploration, it emphasized the importance of this thesis's research questions *RQ4 about charactarizing strategies* and *RQ5 discussing the literacy of networks visual exploration*.

6.3 Strategies & Pathways

To address R1 (*How can we characterize the visual network exploration process(es) that individual analysts perform?*), I started a fresh, in-depth analysis of interaction logs and interviews from our previous work (section 4.4,section 4.5). As part of this, we analyzed log data from more than 534 individual user sessions. This analysis was complemented with a second analysis pass (thematic analysis [49, 113]) on course participants' interviews, coding for comments on their strategies. In this process, we identified a set of individual *strategies* that captured analysts' exploration process. In the context of this research, we define a **strategy** as *a set of specific, linked steps that an analyst follows to identify and understand patterns based on the associated exploration goal(s).* We describe it as a set of individual steps that could be easily followed by another analyst later (section 6.4, Figure 6.2). The strategy impacts the decisions taken at various steps as follows: (1) how nodes and links are identified in the data, (2) the network schema structure — specifying how network is created: what are the network's nodes, links, and additional description —, (3) the visualizations chosen to explore, (4) how to read the visual patterns in the context of the selected goal, and finally (5) how to trust the findings based on all of the taken decisions. By visual patterns, we mean recurring and identifiable arrangements of visual elements within a network visualization.

6.3.1 Exploration Strategies

Overall, we identified 16 exploration strategies that we grouped based on the motivation for choosing a strategy as (Figure 6.3) shows. Each strategy is described in detail using a description card as (Figure 6.2) shows. Each description card includes the following information about the strategy: a description of the strategy and what it aims for, the type of network visualizations that support the strategy - if applicable-, a proposed set of questions to reflect on, the nature of data the strategy can be applied on, and the steps to apply the strategy. The strategy description card is written in non-technical language with two goals: (1) to explain the steps and (2) to ignite ideas on how to think about data. The language used in describing the strategy's steps is similar to pseudocode, where algorithms are explained in an informal and self-explanatory manner, describing actions and conditions. The complete set of strategies and detailed descriptions for each can be found in the appendix (section C.3). Now, I discuss the main groups of strategies classified by their goal.

6.3.1.0.1 Data-Driven Strategies (3) aim at exploring data partially or fully informed by a network's characteristics—e.g., temporal, spatial, or dense. Example strategies include: **Temporal exploration** in which an analyst examines the temporal changes from different

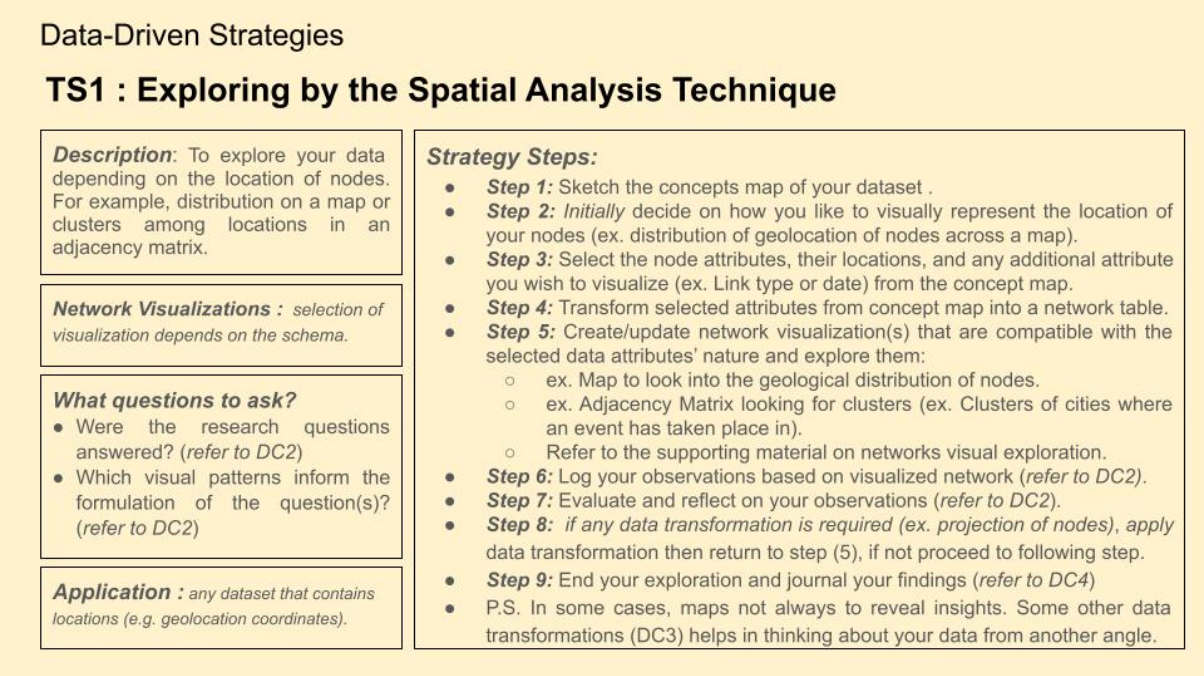


Figure 6.2: **Exploration Strategies Description Card Example.** A strategy description card includes the definition of the strategy goal, the strategy steps, the application of the strategy, and other information.

perspectives (e.g., temporal change, link frequency over time). For example, an analyst can examine temporal changes in the locations of peace agreements on a map or investigate peace agreements that took place in similar years by similar actors. **Spatial exploration** focuses on the spatial distribution of nodes (e.g., geo-locations or abstract locations of nodes). A third strategy focuses on the **structure** of the networks, e.g., exploring clusters and central nodes).

6.3.1.0.2 Task-Driven Strategies (10) aim at exploring data with a specific task in mind. As part of this strategy, observations are evaluated against the initial task and an exploration may be changed or refined accordingly. For example, in **Comparing & Contrasting**, an analyst compares multiple network visualizations for the same data set or from different data sets that share the same network structure (e.g., two differently labeled data sets). The **Content-extraction** strategy explores a network of features describing the content of an object (e.g., images, videos, social network posts). For example, for a medical image, features are extracted to describe the image and then linked to a disease or to similar images. Other task-driven strategies include *answer research question*, *find a pattern*, or *detect anomaly*.

6.3.1.0.3 Open-Ended Strategies (3) are applied to explore data if neither features in the data (data-driven strategies) nor tasks (task-driven strategies) are motivating enough. Possible strategies include experimentation with the network schema or network visualization technique

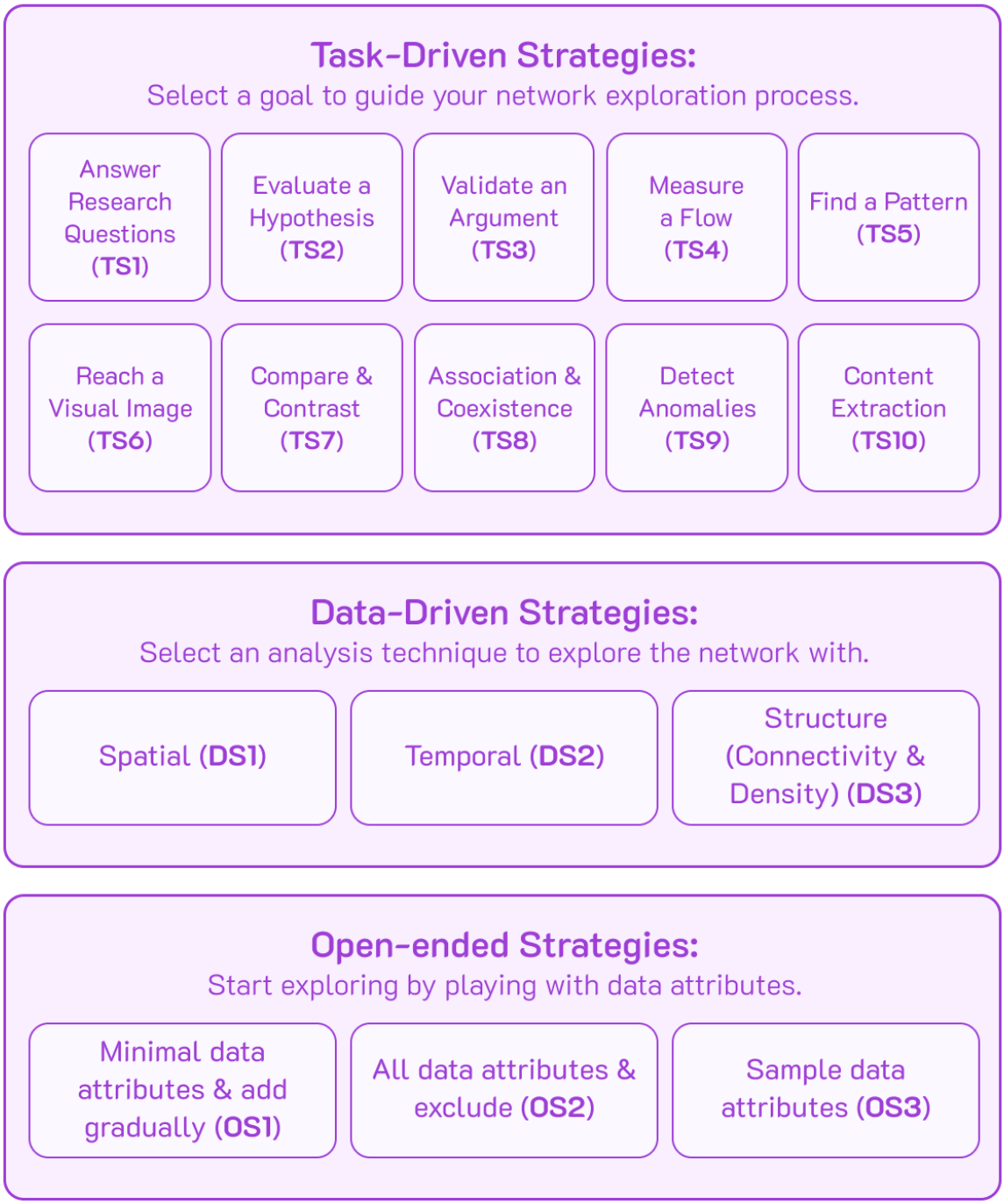


Figure 6.3: **Exploration Strategies Per Group.** Each group contains a set of identified strategies and more can be added in the future.

in search of identifying interesting visual pattern(s) or insights. **Gradual exploration** suggests starting by exploring the minimal number of data attributes (e.g., only network topology or node type) and then iteratively adding additional attributes in order to derive additional insights. For example, the analysts start by exploring visualizations of a network of only links between nodes

in the first iteration. Then, in each following iteration, add/change a describing data attribute (e.g., time, location, link types). In contrast, the **Explore and eliminate** strategy suggests starting with *all* data fields visualized and then iteratively excluding data attributes, informed by insights from the initial overview. For example, attributes may be excluded to avoid clutter or because the initial overview has low relevance to the analysis goals.

I will discuss an example of how to apply a strategy card on a dataset. For example, let us select *Spatial Analysis Technique* card and apply its strategy on a dataset of 'cars accidents'. Then, to find where most accidents happen, we shall do the following: (1) to decide if we are interested in representing nodes' location as geolocation used in a map *or* to use locations as nodes themselves for clustering purposes. (2) Then, based on the decision taken, we create the network schema in the first case, it will contain at least (e.g., node1:car1 Model, car1 location, node2: car2 model, car2 location), whereas in the second case, we should have at least (e.g., node1: Car model, node 2: accident location). (3) Then we choose a visualization: in the first case, we will go with a map, whereas in the latter case, we will go with an adjacency matrix or node-link diagram. (4) Explore and read visual patterns: in the first case, we will look into how accidents are distributed geographically over the map, whereas in the second, we will look into clustering patterns. Some visualization (e.g., adjacency matrix) serves better if the network is dense. (5) Finally, based on the strategy and its impact on each step, we can decide if the observations represent a better understanding of the data or if we should explore using another strategy. We can use the questions in the description card to reflect on our exploration and findings. So, how do analysts differ in their exploration based on their data and exploration goals? This is what I will discuss next in what we call pathways.

6.3.2 Pathways

While identifying strategies and during interactions with participants during the 6-week network course (section 4.5), we also noted four groups of analysts. Each group could be identified by whether the analysts had a clear exploration *goal* (yes, no) and whether they had access to their own data (yes, no). By goal, I refer to any high-level research question or hypothesis an analyst aims to study through the respective data. Table 6.1 shows the four resulting groups, which I describe in the following. Each analyst from our previous study (section 4.5) fell into exactly one of these groups, with consequences on their approach to exploration and their respective strategies (R1). Those four groups have contributed to the design of the roadmap. To inform our second research question R2 (*How can we facilitate the exploration in a way that is flexible and tailored towards analysts' diverse individual goals and strategies?*), we define the sequence of shared exploration processes per group as *pathways*. We found through our analysis that analysts from each pathway share common needs among them as well as some practices. For

Data / Exploration Goals	**No Goal**	**Have Goal**
No Data	Learning about Concepts and Tools	Informing Data Collection
Have Data	Free-form Exploration	Goal-oriented Exploration

Table 6.1: The Four Pathways of Visual Exploration Roadmap

example, how can we test the suitability of an aimed exploration strategy for analysts from Pathway 2 (Informing Data Collection), where no data exists yet? A pathway is similar to strategies in that it prescribes specific steps along the exploration for guidance. However, while strategies prescribe how to explore data (Section 6.3.1) toward a goal, pathways prescribe how each of the four groups explores data. Thus, pathways are characterized by routes that *i)* start out by identifying if an analyst has access to data and has an exploration goal and *ii)* can include specific strategies while generally allowing for any strategy to be employed. The four pathways for our four groups of analysts are:

6.3.2.0.1 Pathway 1: Learning about Concepts & Tools (no data, no goals) Analysts in this group had no access to data yet and had no specific exploration goal in mind. They were mainly interested in learning about network visualization tools and strategies to see if this could be a helpful method for the future and to understand what questions and data could be explored. In this pathway, analysts would either be given demo data, or they should use data that they were familiar with in order to better understand how patterns in the data would be visualized in different network visualizations. Explorations to learn about concepts & tools were not driven by a specific goal. Therefore, strategies such as open-ended and data-driven strategies would serve better (subsection 6.3.1). By trying different strategies with a given data set, analysts can gain an understanding of how to create networks and format data tables, what types of visualization are there to explore, and which observations each visualization can yield.

6.3.2.0.2 Pathway 2: Informing Data Collection (goals, no data) Analysts in this group did not have any ready-to-go data set to explore (similar to Pathway 1), but had an idea about goals to address through visual network exploration. Like those in Pathway 1 (Learning concepts and tools), they wanted to inform any future analysis and associated data collection: what

information can I visualize in a network and what data shall I collect?

6.3.2.0.3 Pathway 3: Free-Form Exploration (data, no goals) Analysts in this group had their own data but had no specific exploration goal in mind yet. These analysts engaged in free-form exploration mostly choosing open strategies. To inform strategies, analysts would start by describing their data. Then, analysts are open to choosing any strategy from any of the strategies in (subsection 6.3.1).

6.3.2.0.4 Pathway 4: Goal-Oriented Exploration (data and goals) Analysts in this group had both their own data and a specific goal for their exploration. These analysts started their exploration by selecting a strategy based on the type of data (data-driven strategies) or informed by their research goal (task-driven strategies). For example, to find specific patterns (task-driven strategy) to understand the network structure (goal). This promotes pathways analysts to start by picking one strategy, examine it, and then engage with any of the other strategies and experiment with it.

6.4 Resources: Roadmap, Materials, & Activities

To communicate strategies and pathways to future analysts (R2), I developed *i)* a visual roadmap showing all possible steps and pathways for overview. To support analysts in their analysis, we also developed *ii)* activities to help analysts transition from one step in the analysis to another one, and *iii)* a set of material explaining concepts such as data formatting and visualization techniques. Visualization techniques in the context of networks refer to methods used to represent network data visually. Such techniques aim to provide insights into aspects related to the structure, relationships, and properties of the network. All these resources were designed to complement each other and to be as self-explanatory as possible.

6.4.1 Roadmap Design process

Our design process went through several conceptual and visual iterations (see appendix Appendix E for earlier versions). In each iteration, we aimed to make pathways and their entry points (i.e., data & goals), potential steps, and directions visible and to facilitate a self-regulated learning process. The idea was that, at any point, analysts should know where they are currently at and be able to identify the next steps toward their goal (even if this may change as part of their exploration process).

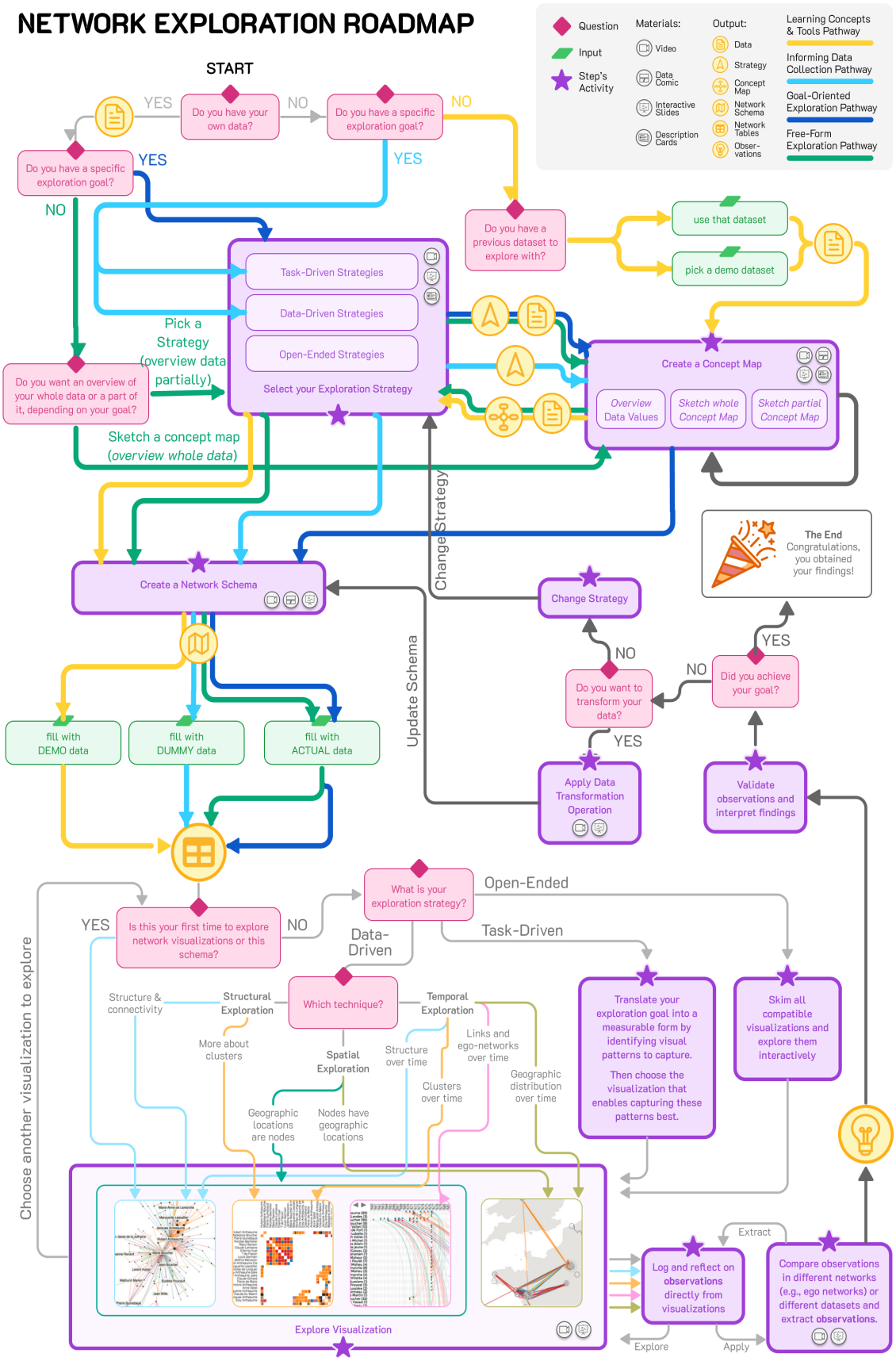


Figure 6.4: **Visual Network Exploration Roadmap: From Data to Findings**

6.4.2 Roadmap Design

Our final version of the roadmap is shown in Figure 6.4. The roadmap illustrates steps (purple rectangles with star symbols) that produce certain outputs (yellow circles with icons), questions and respective decisions (red rectangles with diamond symbols), pathways (colored arrows), strategies, intended outcomes (goals), and materials (grey circles with icons). Green boxes indicate steps in which a *data input* is required from the analyst (either example data or an analyst's data).

Selecting an appropriate learning style is important when designing material for visual exploration. We designed the roadmap to have an active learning style with a top-down teaching method. The effectiveness of using such a roadmap is by emphasizing the analyst's involvement, interaction, and activeness within the exploration process. For example, through the roadmap and supportive material design, we used questions (e.g., decision points, strategies description cards) to enhance user engagement and understanding as recommended by Tanahashi et al.[257]. It is important as well to break down the process into steps to *reduce the cognitive load*, *mitigate bias*, and *higher levels of user engagement* [7].

The exploration process starts at the `START` label. The first action is a question asking about the available data. In both cases of the answer (yes or no), the next question asks for a goal (yes or no). These two first questions are identical to those mentioned in subsection 6.3.2 and serve to identify one of the four pathways for each analyst. Consequently, pathways are color-coded in the roadmap.

Based on the chosen pathway, some analysts start with an activity to *create a concept map* (Learn Concept & tools pathway). Or, they will first select a strategy *strategy selection step* (Informing Data Collection pathway, Goal-oriented exploration pathway), whereas analysts from (Free-form exploration pathway) can start from either activity: *create concept maps or select a strategy*. To have a detailed understanding of the roadmap, the remainder of this section describes the individual network exploration steps, their activities, and their relevance for particular pathways.

6.4.3 Exploration Steps and Activities

Each step in the roadmap (purple rectangles) describes one or more activities that help an analyst to produce a certain output (yellow icons): a *dataset*, a selected *strategy*, a (concept map) *sketch*, a network *data schema*, a network *data table*, or an *observation* made from a visualization. Each step's material and detailed instructions are provided in Appendix C.

6.4.3.0.1 Step: Create a Concept Map Concept maps are diagrams that show concepts, their properties, and relationships between these concepts akin to Entity-Relationship diagrams: concepts represent anything of importance in the data set (e.g., people, time, steps, attributes) and relationships designate any type of relationship between these concepts. A concept map aims to help analysts understand their data and inform research questions. To help create concept maps, we designed three activities: *1) overview data values*, *2) sketch whole concept map*, and *3) sketch partial concept map*. An analyst could pick as many activities as they wish. While many steps have multiple activities, we limit ourselves to one representative activity per step and point to *Appendix C* for the complete description of the 7 activities.

Activity: Draw a complete concept map. This activity guides analysts in sketching *all* possible concepts and their relationships in the data. The activity is described in several steps, such as identifying concepts, drawing concepts, listing possible research questions, etc. This activity produces one (or more if repeated) concept maps. Analysts can create their data starting from either:

- **Data**: such as a formatted dataset or qualitative data (e.g., interviews). Analysts start *describing all data* using a concept map. Then, they generate possible question(s) their concept map can answer.

- **a Research Goal**: such as comparing two datasets' network layouts or examining a flow in a network. Analysts gradually build their concept map, starting from a single research goal. Then, they add more elements to their concept map as they go. They aim to create an initial concept map to address their research goal.

Symbols used in sketching concept mapping play a crucial role in understanding and expressing the underlying structure of networks. Each symbol represents the semantics and structure of proposed nodes and links. Thus, for network visual exploration and through our work with analysts, we have described the following symbols for concept mapping:

Symbols used in sketching concept maps are not important only for representing nodes and links. It exceeds that to represent *how the network data can be interpreted*, *how network schemes can be created*, and *how findings can be derived*. Using the symbols Figure 6.5, analysts can think about their data and what they represent as following:

- Does the element we are representing describe a node? (a concept attribute)
- Is the element describing a link? and reflecting its changing nature (a relationship attribute)
- Does the element represent a missing value that you expect to have? (A missing concept)
- Does the element form a partial part of a compound value (e.g., the city name as part of the address) (a sub-attribute)
- Is the value to be calculated or derived from other data attributes? (derived attribute)

Visual Network Exploration

Symbols of Concept Map Sketching

Symbol	Name	Description	Can be used as (network table)
	Concept	Represents any concept (ex.entities, events, persons)	Nodes
	Directed Relationship	Relationship that is an associations between or among concepts where direction is important	Directed Link
	Undirected Relationship	Relationship that is an associations between or among concepts where direction isn't important.	Undirected Link
	Concept Attribute	An attribute that describes a concept (ex. Name of a person, specialization of an employee)	Node / Node Type
	Relationship Attribute	An attribute that describes a relationship between concepts (ex. date of when a letter was sent).	Link Type / Link Weight / Link ID

Visual Network Exploration

Symbols of Concept Map Sketching

Symbol	Name	Description	Can be used as (network table)
	Multi-Relation	A relationship that connects 3 or more node entities (ex. Ternary relation)	Link types derived from any of the participating nodes
?	A missing Concept	A node that represents a concept that you expect to have or need but do not no about it for now at least.	Absent Node
	A sub-attribute	An attribute that can be partially considered part from another attribute such as composite attribute.	Node Type / Node
	Derived Attribute	An additional attribute that you can create by deriving its values from 1+ attributes (ex.computed	Node/ Link Type, Link Weight

Figure 6.5: Concept Mapping Symbols for Networks Visual Exploration

Providing a semi-/systematic approach to applying each step along the roadmap removes the ambiguity about how to apply it. That is, having general guidelines to sketch concept maps and then transform them into network schemes assists analysts in deciding what best suits their goals and data. Different symbols can be used for different data models (other than networks).

6.4.3.0.2 Step: Select an Exploration Strategy. This step asks analysts to select an exploration strategy (see section 6.3). For some pathways (*goal-oriented exploration* (dark blue) and *informing data collection* (turquoise)), this step can be a starting point; other pathways will pass through this at a certain point of the exploration process.

Activity: Strategy Description Cards. We facilitated the selection of exploration strate-

gies through dedicated *strategy description cards* (Figure 6.2) which provide (1) a summary description of the strategy's goal, (2) questions to guide the exploration process, (3) data-related requirements to apply the strategy, and finally, (4) a step-by-step guide to applying the strategy.

6.4.3.0.3 Step: Create a Network Schema. This step includes the process of transforming the concept map into a network schema and applying the suitable data based on the pathway. The selected strategy (see above) influences how a network schema is created for the network data tables.

Activity: Network Schema Activity. Our previous work showed that this step, in particular, can form a barrier for analysts (Barriers 5–6 in section 4.9); creating an appropriate mapping from the concept map to a network schema and the appropriate data structure (e.g., node or link table) is not trivial. To facilitate this process, we provide guidelines to create and fill in their data table based on (1) their selected strategies and (2) their pathway. This activity can be repeated as analysts navigate through the roadmap. It aims to assist in: *i)* creating different schemes based on the analyst's exploration strategies and concept maps, *ii)* easing decisions on the type of data table (node and/or link table), and *iii)* offering hints on which tool features could be used for interactive exploration based on the schema they have created. The complete material of the activity can be found in the appendix (section C.4).

The analyst can emphasize the importance of specific attributes that can help reveal patterns, such as the type of nodes/links, derived attributes (e.g., link weight), and others. Derived attributes represent an attribute whose values are driven from another attribute (ex., Calculated formulas, generated strings, ..). Such attributes are important in network visual exploration to represent the importance of some links (e.g., link weight) and nodes (e.g., age) or participate in the data transformation and wrangling (e.g., reducing the size of a network). For example, node/link types could be grouped through attribute categorization to reflect the hierarchy in node/link types. This can help reveal patterns by breaking down data attributes (e.g., educational background broken into majors) or, vice versa, grouping them into higher levels. In addition, this activity informs the decision to go with the whole composite value of an attribute (e.g., full address) or one of its values (e.g., city only). The outcome of such activity enables the analyst not only to create a schema but also to understand more about their data and exploration process.

6.4.3.0.4 Step: Explore Visualizations. This step involves the process of deciding which visualization to create and explore based on the strategy and data type, as well as using interactions (e.g., filtering) to look for interesting patterns.

Activity: Visualization Exploration Activity. Our previous work has shown that deciding on the type of network visualization technique to serve a particular goal and interpreting visual

patterns in visualization is not trivial (Barrier 7 in section 4.9)). To facilitate this, we designed a dedicated activity that provides guidance on what network visualization technique to choose and how to explore the resulting visualization based on a set of questions:

- *Is it the analyst's first time to explore networks?* and *Is it the first analysis with a given network?*: For beginners, it's recommended to start with node-link diagrams since they represent the base concept, can give a first impression about the data, and eventually motivate other visualizations (e.g., matrices), questions, and strategies.
- *What is/are the selected strategy(ies)?*: Some strategies directly point to a specific network visualization technique (e.g., adjacency matrices can be particularly suitable for identifying clusters). For other strategies, a range of or even multiple visualization techniques might be suitable. For example, to identify flows through a network, node-link diagrams can be used in combination with an adjacency matrix). We help analysts further narrow down options by pointing out the following considerations via guiding questions: (1) the chosen network schema and (2) any further analytical operations necessary (e.g., comparing different networks).

6.4.3.0.5 Step: Log Observations. As analysts explore various visualizations, we encourage keeping track of observations and decisions that have accompanied the network exploration process.

Activity: Log Observation. In order to facilitate this, we have designed an activity where analyst are invited to document their decisions taken at the network schema creation level. Then, for each explored visualization, analysts log their observations based on the visual patterns discovered (e.g., the central nodes in a node-link diagram or nodes that share links in an adjacency matrix). Each visualization has its own list of visual patterns that analysts are encouraged to look for and to reflect on. Logging observations, as the analysts explore, helps them to reflect on their decisions taken at any step and how this helped to reveal further insights.

6.4.3.0.6 Step: Validate Observations & Interpret Findings. On the basis of the observation logs, this step encourages the analyst to reflect upon, validate observations [265], and interpret their observations—if possible in an iterative process, if findings are not satisfactory or new questions/goals have been raised as part of the process.

Activity: Storyboarding In this activity, analysts are asked to validate their overall understanding through their observations. Analysts start by taking screenshots of all of their network visualizations, place them on any storyboard (e.g., Miro `https://miro.com/`), list their observations on each, and then look into how different visualizations can collectively

(1) offer a better interpretation, (2) identify conflicts in observations, or (3) confirm a previous finding. Further interpretations can be found and drawn from the observations.

6.4.3.0.7 Apply Data Transformation. This step offers different operations to transform the data into different formats to enable different network visualization techniques and exploration methods.

Activity: Transform data My previous work (section 4.5) has shown that such data transformations can be difficult (see Barrier B4 in section 4.9)). In this activity, once an analyst is unsatisfied with their exploration outcome, the analyst is provided with tables (Figure 6.6) containing a variety of possible data transformations to be applied—at a network-, node- and/or link-level. Then, the analyst, based on their case, chooses the appropriate data transformation. This is done by looking for the data transformation with a similar description of the analyst's case. Each data transformation is described by the application case, including when to apply it and an example of how to apply it. For example, if no interesting patterns were identified in the first exploration iteration due to many different link types, a data transformation of grouping existing link types is suggested. Other data transformations cover examples of data refinement, abstraction, and representing missing nodes, which are provided in the appendix (section C.7).

Possible Network **Transformations** List

Affects	Exploration Status/Outcome	Comments	Example	Proposed Transformation(s)
Network Level	Cannot identify coexistence of specific node types	Bipartite Network	Network of nodes of two types (e.g sites and bones)	Projection of Nodes
Node Level	Network has few nodes and/or links	Nodes are Textual pieces	Peace agreements, artifacts documents	Extract nodes and links from within the document (NLP Tokenization)
Link Level	High number of redundant links	What is your point of focus or looking for?	Social network dataset	(1) Remove redundant links (rows) if unnecessary. (2) Transform redundant links into a single link where the weight of the link=number of redundant links
Missing Data	Would like to represent missing or indeterminate values	Missing values are unknown at the moment, whereas indeterminate values are values that we expect but cannot identity it yet	In a care home dataset, to investigate the role of the person(s) who is mostly communicating with patients but the the role of such persons is not yet known	Dealing with absence of value by adding a node/link representing expected value
Data Abstraction	Most of the nodes have few links	Are we interested in the node itself or what it represents?	Are you interested in Dr. Smith himself or his specialization as an ENT Doctor	Represent the node by its abstraction

Figure 6.6: An Example List of Data Transformation and Wrangling Options on Different Scales

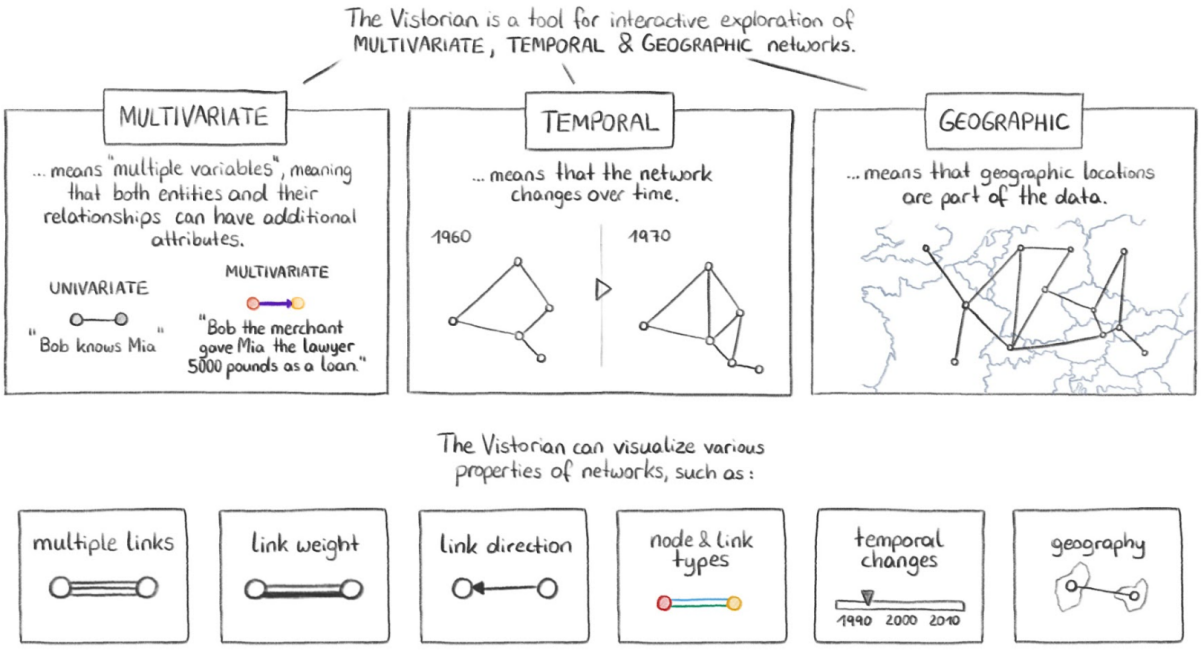


Figure 6.7: An excerpt from a comic explaining the features of the Vistorian. The panels can be read sequentially, but eye-catching headers facilitate quick navigation over the page when the comic is used as a reference while exploring the tool on a computer.

6.4.4 Design of Learning Resources

Activities are supported through a set of materials that analysts can consult at their own pace and revisit on demand, which are:

- ***Tutorial videos*** (5–8 min.) explain different steps in the roadmap and point to the corresponding activities, often walking analysts through detailed examples.
- ***Slides*** with explanations and activity examples or tool tutorials as a quick reference for analysts without the need to re-watch videos.
- ***Description cards*** of exploration strategies to provide step-by-step instructions on how to apply a strategy and what it measures. A detailed description of the strategies and their description can be found in the earlier (subsection 6.3.1).
- ***Data Comics for Networks*** to provide a quick reference to different exploration steps (see Figure 6.7), inspired by data comics [16]. Data comics are distinguished by their ability to convey information through combining storytelling and visualizations. Comics' plot unfolds through a sequence of panels that guide the reader through the narrative. The pace and flow of the story are controlled by the size, shape, and arrangement of the panels. For example, the gutters or the spaces between the panels play a crucial role in revealing how time and movement between scenes can be inferred. There are four essential components for a data comic: the visualizations, the flow, the narration, and

the words and pictures[16]. The visualization component ranges from realistic pictures to conceptual visualization. The flow component varies from a directed sequence to an undirected order of the panels. The narration component ranges from factual descriptions to fully narrated storytelling. The word and picture components range from verbal to nonverbal comics. The data comics in our context have been designed in collaboration with my colleague Magdalena Boucher. In the coaching program, I have used the data comics for networks as part of the supplementary material provided to the analysts to teach how to create and explore networks.

The distribution of help resources along the exploration process is as follows:

Steps	**Tutorial Videos**	**Data Comics**	**Interactive Slides**	**Description Cards**
Intro to Visual Exploration	1	1	1	-
Explaining the Roadmap	1	-	1	-
Understanding Data and Concept Mapping	9	1	2	2 (Sketching Symbols Cheatsheet)
Strategy Selection	1	-	1	16
Network Tables Creation	8	1	1	-
Observations Logging and Evaluation	1	-	1	1 (Logging Card)
Data Transformations	1	-	1	-
Additional Examples	3	-	3	-

Table 6.2: Learning Resources & Supporting Material Per Step.

Examples design plays a crucial step in teaching and transferring skills and promoting self-regulated learning. Examples were designed to cover creating a variety of network schemes that include: a node table, a link table where nodes have *single value* attributes, a link table where nodes have *composite-value* attributes, a link table where nodes are connected through multiple links types, and tables with derived or calculated attributes. In our educational material, we included the following examples:

1. Marie Boucher dataset: is about a set of letters and documents exchanged by a trader called Marie Boucher, and the goal is to explore her trading network.

2. Connections Network: the dataset of family members and their best friends. We aim to explore their relationship network and if they intersect.

3. Job Applications Network: the dataset is about a number of graduates who applied for open jobs. Each applicant can have 1-3 majors under their background umbrella, and each job application can be accepted/rejected.

4. Patients Referral Network: the dataset is about the number of patients who were referred to a service on a specific date. Each patient has an address that is composed of a building number, street name, and postcode. And each service has a type and a provider. Also, we add information about who referred the patient to the service requested. The name of the referring person and his/her workplace.

5. Additional example on how to extract nodes and links from textual documents (graduation book). The example discusses exploring the dataset to answer research questions. The example can be found in (subsection C.8.3)

6.5 Coaching Program Study

I have conducted 1:1 coaching sessions with real-world data analysts, some with their own data, in order to study how well the roadmap and its associated strategies, pathways, activities, and material could support analysts (R2) as part of guided self-paced exploration. In this coaching program, I introduced the roadmap as a tool to help analysts navigate the exploration process. I used the individual sessions to help overcome any obstacle the analysts may encounter. It allowed me to observe analysts' progress through the roadmap and their process of learning about exploration. Coaching sessions provided me with feedback on the design of the roadmap and learning materials, which informed design iterations on both.

6.5.1 Coaching Program & Coaching Sessions

Over the course of two months, I offered a free coaching program with individual coaching sessions open to everyone. Analysts, once signed up for the program, could book an arbitrary number of free 1:1 coaching sessions, each between 30 and 60 minutes. Each session was held online and led by myself to 'coach', assist analysts, and provide them with any help needed. After discussions with my supervisors, we designed three types of sessions. In the first 30-minute *kick-off session*, as the coach, I present and explain the nature of our coaching program, what it will offer, and how sessions can be booked. We also - myself and the analyst - discuss the analyst's data and analysis goals—if any—to inform their pathway (Section 6.3.2). Consequently, I introduce the analyst to their respective pathway, its exploration steps, as well as the corresponding material and activities. The material for each pathway is tailored based on the needs of each pathway chooser.

As part of subsequent *follow-up sessions*, we provide assistance through the different network exploration steps as needed. Based on the analyst's pathway, any outputs or particular observations shall be discussed up to that point. For example, in the first follow-up session with an analyst on a free-form exploration pathway, the first discussed output was a number of concept maps they had created in the create concept map step. We discussed further opportunities for exploration, such as sketching alternative concept maps or how to represent a specific concept, relationship, or data attribute. As part of these follow-up sessions, we also encouraged analysts to provide feedback on the roadmap design and the materials provided to support individual steps. Eventually, we invited each analyst to a *feedback session* to collect final feedback on the program, roadmap, pathways, material, and activities, as well as the coaching sessions themselves.

6.5.2 Coaching Program Participants

`Participants Recruitment`—We circulated a public online invitation to our coaching program through different online communities, including prior collaborators in network exploration, a social network analysis research group, and participants of our previous network exploration course [2]. The registration form asked for demographic data, professional background, and potential analysis goals and data. `Participants' demographics` are as follows: 10 analysts participated in our program. Participants' network exploration varied: 4 analysts had explored 3+ networks, 2 analysts had explored 1-2 networks, and 4 analysts had no previous experience in network exploration. For additional information on the participants' background, please refer to section 3.6. Analysts' distribution across pathways was as follows:

- **Learning about Concept & Tools (ConceptLearner) (no data, no goals; 2)**: *ConceptLearner1* was a researcher in education who explored 3+ networks. They were interested in learning about the Vistorian. *ConceptLearner2* was a freelancer with expertise in network visualization and was interested in learning about the Vistorian and the visualizations it offers.
- **Informing Data Collection (InformData) (goal(s), no data; 2)**: *InfomCollection1* was a researcher in social sciences and *InfomCollection2*'s background was in medical imaging and computer sciences. Both had no experience in network visualization but were interested in learning how the visual network exploration could potentially help their (future) research and the data to collect.
- **Free-form Exploration (Freeform) (data, no goals; 2)**: *Freeform2* had a background in computer science and social sciences, with experience in network visualizations. Their goal was to explore their dataset which focuses on peace agreements for interesting patterns. *Freeform1* had a background in computer science and design with no experience in network

visualization. They aimed to explore data collected from a qualitative study and wanted to discover what insights network visualizations can reveal.

- **Goal-Oriented Exploration (GoalOriented)(data and goal(s); 4)**: *GoalOriented1* was a researcher in social sciences, interested in visualizing survey votes through answering research questions. *GoalOriented2* was a researcher in the humanities and politics with limited experience in network visualization. They were interested in visualizing protest posts on social media by answering research questions. *GoalOriented3* had a background in history with no experience in network visualization. They were interested in exploring their data by generating a specific network visualization and answering a specific research question. *GoalOriented4* was a freelancer with limited experience in network visualizations. Their aim was to explore their novel dataset using circular network visualization.

6.5.3 Data collection & Analysis

In total, we spent 30 hours in coaching sessions across all analysts, running 30 sessions in total. Each analyst attended around 3 sessions on average (min: 1, max:6). Additionally, some participants preferred email communication for asynchronous communication. Across the coaching program, we collected the following data:

- Participant background in the form of **entry questionnaires**.
- **Audio and video data** from coaching sessions & final feedback interviews.
- **Analysts' sketches of concept maps and network schema** that were shared by analysts during coaching sessions.

Video and audio data from the coaching sessions and interviews was transcribed and then analyzed by myself using a thematic analysis approach [49, 113], focusing in particular on participants' feedback and reflections on the roadmap design and content, learning resources, and the visual exploration process. This included their goals, the nature of their data, the activities they engaged in, their experience with these, and their feedback on iterations over roadmap design.

6.6 Findings

Findings from our coaching program shed light on how participants used and experienced our roadmap and teaching materials and how our approach to facilitating visual network exploration had an impact on participants' exploration strategies.

6.6.1 Coaching Sessions and Personal Success Stories

The coaching program supported self-regulated learning alongside the individual coaching sessions and helped analysts grasp new knowledge about network exploration. For example, *Freeform1* said that "discussing my outcomes and questions through sessions works as cross-checking and reassures my understanding". After each step and output, the analyst would meet with us to discuss progress, questions, and next steps. We discussed possible next steps with them and discussed additional outputs, e.g., concept maps, where necessary, to make an informed decision about any next steps. These conversations opened up perspectives and ignited new ideas for exploration, and analysts were able to establish their first understanding of their data and processes.

7/10 analysts stated in the feedback session that they reached the goal they had set themselves for the coaching program. Analysts described their personal achievements as follows. *Freeform2*, shifted from vague goals at the beginning to more defined questions for a suitable network schema. They learned to distinguish important data attributes from non-important ones, to ask more questions, and how to overcome some of the tool's (The Vistorian) limitations by adjusting their network schema. *GoalOriented1* stated that they learned what was in the data and achieved the ability to reconsider their knowledge and assumptions about their data, their analysis approach, and any possible visualization options. *GoalOriented2* found that they gained the ability to explore new aspects of their data that they had not been aware of. Other analysts described similar success stories and achievements, such as a better understanding of the complexity and structure of their data through free-form exploration, ideas for possible data to collect and what analysis processes would be needed for data collection, and an understanding of new visualization and exploration techniques.

6.6.2 Roadmaps and their Contribution to Visual Network Exploration

The Layout and outline of the roadmap support the iterative process of data exploration. The breakdown of the visual exploration process into clearly defined and linked steps, as well as clearly marked decision-making points, helped analysts plan their process based on their personal analytical skills and goals. For example, *Freeform1* stated: *"the roadmap is very clear, especially, [the] different approaches to exploration. Starting [my] exploration [...] I was able to decide [...] where do I stand"*. This resulted in a *progressive* nature of exploration that encouraged analysts to explore different facets of their data through several iterations. *Freeform2* stated: *"the roadmap helped me to answer questions and ask more questions, e.g., by changing network schemes. I can understand a lot from the data, even if I am not a domain expert."* Analysts found the roadmap includes all the steps needed for their exploration.

We iterated the roadmap design alongside the coaching sessions based on our observations and direct feedback. Analysts agreed that the number of details shown on the map should be adaptable and expandable to the analyst's experience. The more the analysts are aware of the role of each step, the more information about the decisions and activities implemented at that step could be revealed and vice versa. In other words, a reduced version of the roadmap should be shown to analysts new to the exploration process. *GoalOriented1* describes each step as "a single digestible block" but that it can be overwhelming to see all steps in detail from the beginning. Analysts suggested breaking down the detailed roadmap into blocks or sections highlighting how analysts transition from one step to another. Following this and further feedback, we limited the details shown in some steps and added further symbols to the roadmap, for example, to represent different outcomes.

The step-by-step design of the roadmap supports analysts in an incremental learning process. For example, *Freeform1* stated: *"the network's schema creation step helped in having a better understanding of the previous step of concept mapping."* This shows how the visual exploration process is not a set of isolated steps. Rather, it is a process that relies on a set of linked steps where each step cumulatively builds on the outcomes of the previous ones and supports effective learning[245]. This enriched analysts' experience by deepening their understanding of both the exploration process and their data.

6.6.3 Strategies' Impact on Visual Exploration

Some exploration pathways and strategies support progressive exploration. As analysts start their exploration from different pathways, choosing a certain pathway and/or a strategy impacts the exploration and the learning process. As *Freeform1* states: *"I made the decision to use the roadmap with my own dataset. From a learner's perspective, I would say, I was wondering at the start on what pathway I would be able to pick up more knowledge."* Thus, in the following, we discuss individual analysts' *trajectories*. we define an exploration **trajectory** as the actual set of steps and iterations followed by the analyst in their exploration in comparison to the expected pathway. An example of such a personal trajectory is shown in Figure 6.1 for *Freeform2*.

- **Analysts starting on the Goal-Oriented Exploration pathway (4)**: 2/4 analysts who started with this pathway stuck to the strategy they chose at the beginning. The other two explored one or more strategies over the course of the coaching program. Some strategies have encouraged analysts to look into other representations of their data. For example, analysts who aimed for a specific visualization of a specific part of their data from the beginning did not try to explore other facets of their data (e.g., *GoalOriented3*). Whereas

analysts who used a task-driven strategy to answer a research question tried at least one additional strategy (e.g., *GoalOriented2* and *GoalOriented1*).

- **Analysts starting on the Free-Form Exploration pathway (2)**: explored their data with more than two strategies as they did not set a specific goal initially and wanted to obtain different perspectives. As new insights were obtained, they started trying different strategies and specifying questions.

- **Analysts starting on the Learning Concepts & Tools (2)**: were able to learn and work with the tool but did not engage in the same way as analysts who had their own data. This was not surprising to us.

- **Data Collectors (2)**: We were not able to collect sufficient feedback on this experience since both analysts did not pursue any actual exploration of the demo datasets.

Strategies helped analysts think about their data from different angles. *Freeform2* said *"Visual exploration using strategies helps in changing perspectives. That is, if we use statistical analysis only, we do not think about changing nodes/links or changing views."* Analysts were interested in understanding more in-depth how different strategies can inform their exploration process and how to decide which strategy best serves their goal and which strategy to single out. This opens new research directions in how different strategies can serve the analysts' goals the best. *Freeform1* said *"[I learned] that there were open-ended exploration strategies and task-driven exploration strategies. And I knew I could do either to find new insights from scratch [(open-ended exploration)] or go with my goal and see what new insights I can find from there [(task-driven exploration)]. So, different approaches were used to obtain the same data."*

6.6.4 Design of Activities & Learning Resources to support self-regulated Learning

Validating the concept map was a crucial step for most of the analysts. This was mostly done by re-asking what sort of questions the sketched concept map can answer and on what basis. *Reasoning* is also an important process that develops and contributes to findings interpretation through trial and error and through examining different possibilities at different steps. For example, reasoning decisions were taken when sketching a concept map and reflecting on how it will affect the following steps such as network schema creation, interactive visual exploration, and data transformations.

Biases in sketching concept maps is something that was mentioned by some analysts. Analysts were aware that biases and presumptions about the data might influence the sketching concepts maps, i.e., decisions on nodes and links. However, they still found themselves sketching

concept maps only about aspects they already knew rather than thinking of new perspectives on their data. When asked about whether they encountered any obstacles, *Freeform1* stated: *"When creating a concept map or a network, I felt I had a lot of bias when choosing and describing nodes and links from the qualitative data. It affected the sort of insights the visualization can reveal, instead of thinking of new ways to identify and connect nodes and links looking for new insights"*.

The data collection process has an impact on possible data transformation options. By discussing possible data transformation with analysts, some options were impossible to pursue as they would conflict with the specific data context. For example, when discussing some possible data transformation options with *GoalOriented1*, the analyst drew our attention to the fact that such operations are not applicable in their case. The proposed data transformation options contradict how the survey responses are collected and how response sample design works in social sciences. This can result in affecting the data's validity for analysis purposes. Thus, data transformation operations do not apply to all exploration trajectories.

Different types of learning material serve different purposes. In the feedback sessions, analysts agreed that different types of learning materials, as described in Section 6.4.4, were helpful as they served different purposes. 3/10 analysts stated that they preferred data comics to learn about a new topic because they provide understandable explanations. *Freeform1* stated *"data comics [are] good as introductory material because they grab your attention and they are so simple [to understand]"*. For detailed explanations of each step, videos seemed to work the best for analysts, as they can walk the analyst through different examples and how to apply each step. Description cards and slides served best to revisit specific information. Analysts found that all our resources complemented each other, and the combination of them enriched their learning process. *GoalOriented1* stated: "It's a combination of inputs (sic) that makes you learn effectively."

6.6.5 The Role of Logging Observations and Reflecting on the Exploration Process

Analysts could capture observations through their exploration and used their notes to interpret findings as we intended (see Section 6.4.3). To facilitate log observations, a checklist was provided to analysts that contained a list of questions asking about the expected visual patterns based on the network schema and the exploration goal. This reflection process helped analysts to understand more about their decisions at the network level as well as at the visualization level. However, analysts varied in how they reflected and logged their observations. *Freeform2* used Miro boards to collect screenshots of different network visualizations, added notes, and

linked findings, resulting in a large storyboard. *GoalOriented1* used paper sticky notes to log observations from different visualizations. They worked on merging and/or discarding observations as they explored more of their data to interpret and refine findings.

Interestingly, we found that some analyst who were not logging their observations felt more unsure about validating their progress on their own. *Freeform1* stated *"To be able to explore on my own, I need to have cross-checked points through the roadmap. To check after each step, if I can move on to the next step."* Analysts felt the need to ensure they were on the correct track through sessions by asking about their sketches, mappings, or visualization explorations. Thus, the logging and reflection process helped analysts validate their exploration process and turn observations into findings.

6.6.6 Validation of Strategies and Trust of Visual Exploration in Networks

Analysts found that developing trust in their observations is not about only interpreting observations from a visualization. Rather, it is about the ability to decide on the questions the visualizations can answer and how. Trust in the exploration process evolved by understanding more about the data's structure and values. *GoalOriented1* stated "The good way to look at it after several iterative exploration rounds is that I have like reached to define a set of few specific questions, which is actually really useful when you do exploratory data analysis. Due to the fact that good questions are based on some certainties and findings.".

6.7 Discussion

We set out with our two research questions: *R1) How can we characterize the visual network exploration process(es) that individual analysts perform?* and *R2) How can we facilitate the exploration in a flexible way tailored towards analysts' diverse individual goals and strategies?* In particular, we were seeking to check our roadmap design, steps, and strategies for completeness and possibly amend (R1) as well as assess the usefulness of these materials for analysts' exploration and personal learning (R2).

6.7.0.0.1 Main findings. Our roadmap, its steps, and activities, as well as our strategies and pathways, paint a rich picture of exploration processes in unprecedented detail. To the best of our knowledge, this is the **first time that we obtained a holistic picture of steps and decisions involved in network exploration**, including the range of preparatory steps (R1). Our findings and materials can be used to directly **inform other analysts planning and reflecting on their process** (R2). Compared to our observations from more formal top-down workshops in our previous work chapter 4, the guided, personalized approach we promote

in this chapter and through our materials enhances exploration and learning in a unique and complementary way. While alongside our previous formal workshops in which we likewise offered plenty of opportunities for 1:1 drop-in sessions, discussions in these sessions were lacking the **well-defined and shared terminology, concepts, and overview we provide** with this chapter. Roadmaps, steps, strategies, activities, and material **provide tangible means** to both *a)* guide and inform exploration in a self-paced and self-informed way and b) to pin down questions and bring them to the sessions.

We found analysts' logging of their observations in a structural way to be vital in particular because it encourages reflection and constant evaluation against set goals. We believe this prevents analysts from getting 'lost in exploration' and encourages more directed exploration. While the general notion of exploration since Tuckey is that of an open-ended Grand Tour [258], **especially novice analysts require some guidance and direction**. In fact, an open-ended Grand Tour implies a lot of work both in data preparation, creating and interacting with visualizations, interpreting findings, and iterating. **Starting an exploration with some high-level goal in mind can help analysts learn and navigate processes** new to them, make faster progress, and provide a sense of achievement. The emphasis on exploration strategies, in particular, the data-driven strategies and task-driven **strategies, has proven a highly effective part of our roadmap to inform action**. In that sense, we can think of these strategies as specific design patterns, an idea expressed by the fact that each strategy was described in the same format on a card, ready to be used in both virtual and physical learning settings as typical to design workshops.

6.7.0.0.2 Overcoming Existing Barriers to Exploration.

Connecting the roadmap research back to the barriers in network exploration we have identified in our previous work [2], we can provide some suggestions on how these could be addressed as (Table 6.3) shows. For example, the roadmap assisted analysts in overcoming several barriers such as *Missing goals and questions (B1)* as the analysts were put on a specific pathway to identify these goals and questions by being exposed to different strategies. Overcoming *preconceived ideas and mental images (B2)* that can bias exploration processes could partially be solved by sketching concept maps, asking questions, and offering different ways of thinking about data. Our findings also indicate that the roadmap seemed to promote some awareness of how biases and presumptions may influence exploration processes. *(B3) Deciding on the network structure* and *(B5) choosing appropriate data formats* was facilitated through sketching concept maps as well and transforming them through a set of clear instructions into network schemes. *(B6) Importing and Defining Schema* was further facilitated through supporting material such as data comics, which opens exciting avenues to further explore this medium for explanation. Eventually, our emphasis on logging observations can be seen as a start to *(B8) establishing trust in visual exploration* which encouraged reflection

Barrier	**Assistive Approach and Resources**
B1: Missing Goals & Questions	Roadmap, Strategies Description Cards, Explained Exploration Examples
B2: Pre-conceived Ideas & Mental Images	Sketching Concept Map Activity, Strategies Description Cards
B3: Deciding on a Network Structure	Sketching Concept Map Activity, Create a Network Schema Activity
B4: Choosing The Right Level of Abstraction	Data Transformation Activity, Explained Exploration Examples
B5: Choosing the Appropriate Data Formats	Sketching Concept Map Activity, Create a Network Schema Activity, Explained Exploration Examples
B6: Importing Data and Defining Schemata	Step-by-step Tutorials (e.g. Data Comics, short videos)
B7: Interpreting Visual Patterns in Visualization	Tutorials, Explained Exploration Examples
B8: Establish Trust in a Network Visualization	Roadmap, Logging and Validating Observations Activities

Table 6.3: Overcoming Barriers in Networks Visual Exploration

and iterative exploration.

6.7.0.0.3 Informing Future Guidance Systems. All our materials (including the roadmap, strategy cards, activity descriptions, slides, comics, and pathways) are provided in Appendix C. Our findings invite opportunities for future research to look more into how we can **extend the range of materials** and how we can provide **more personalized resources** for analysts. For example: *How can resources be tailored to more advanced analysts? What practices are common in communities and disciplines that are not represented by our participants? Can we build guidance mechanisms as presented here directly into exploration tools?* A starting point for such an exploration might be recent concepts in guidance in visual analytics systems [250] and network visualization [161]. In particular, in their automatic exploration tool for networks, Wang et al. [161] provide a range of low-level exploration strategies for visual exploration, focusing on detailed information for a given network such as 'exploring an ego-network' or 'exploring nodes along a path'. In this study, we focused on higher-level strategies that include data preparation, making observations, and iterations. However, these mentioned **lower-level strategies** can be seen as a natural extension if provided in a similar format to our activities and strategy cards. Likewise, our roadmap, strategies, and activities could be part of **guidance embedded into exploration tools** we have just mentioned.

Our results can inform the design of such systems in many ways. For example, based on our findings from this research, we can **make recommendations** for the design of self-regulated material and learning resources for visual exploration: *(a)* expose learners to various facets and options that can ignite ideas about how the data may be explored; *(b)* clarify how data transformations are applied to allow analysts to exclude those that contradict the data domain or scientific methodology; *(c)* encourage reflection on data and the exploration process through logging observations, linking outcomes, verifying findings, and exploring alternatives; *(d)* provide different types of materials such as introductory material, detailed explanations, interactive examples, and references; *(e)* provide different types of material (such as comics, slides, videos, potentially cheat sheets [268]) to support different tasks: learning new concepts, look-up information, etc.

6.7.0.0.4 Balance Openness and Guidance Likewise, exploration options can—and should—be constrained as an analyst learns new concepts. For example, we found that recommending lots of visualizations and steps at an early stage can overwhelm analysts. For example, a system should **reveal new visualizations and other options gradually**, provided with **sufficient explanation** (and additional guidance), and support **analysts make a decision on how to proceed**. The question is how to support interactions with analysts beyond the types of coaching sessions we have conducted. How self-sustained can material for guidance and exploration be? **The design of the visualization tools should rely more on the basis of dialogue design**. In other

words, it should be based on questions and answers to understand better analysts' intent and goals rather than a set of options to choose from. For example, scoring systems could be built and used to recommend visualizations to analysts based on a number of questions, such as those described in Section 6.4.3 (e.g., which visualization to choose?). Artificial intelligence might be an open field for exploration in this context. In the future, we envision automatic approaches to personalize the exploration process.

6.7.0.0.5 Limitations & Open Questions The roadmap design works on describing the ground base of the visual network exploration process rather than considering other types of visualization techniques and exploration processes. Furthermore, to limit the number of variables analysts are exposed to, I have chosen to adopt a single visualization tool (The Vistorian) throughout our research. This raises additional questions and opportunities for future research. **How does our approach in the context of visual network exploration expand to other areas of data visualization and visual analytics?** And: **How do our materials need to be adapted to integrate with other exploration tools that support different workflows, other types of data transformation, interactive exploration, and visualization types**. There is a lot of potential in exploring these types of questions in the future and also in sharing and discussing materials, including alternative approaches to roadmaps and the types of activities we propose here, to support visual exploration processes in general.

6.8 Conclusion

Our resources—roadmaps, pathways, materials, and strategy descriptions—proved to support analysts in their exploration process by offering step-by-step guidance and on-demand explanations. Using these resources in real-life analysis scenarios with 10 analysts, provided us with a better understanding of analysts' practices and a wealth of suggestions for future work. Such as designing resources that support self-regulated learning and work *with* the analyst along their *specific* data, questions, and skills. I hope this research will inspire future exploration tools that strike a balance between open-ended exploration and helping to learn routines for exploration by building a dialog with the analyst.

Chapter 7

Discussion, Future Work & Conclusion

In this chapter, we discuss our findings and how they support the visual exploration process of networks. We start in section 7.1, describing the user types and barriers we identified in network visual exploration. Then, in section 7.2, we discuss our understanding of the visual exploration process and the approaches we used to study it in the wild. Thus, to support network visual exploration based on our findings, we worked on identifying analytical metrics and KPIs to measure network visual exploration and users' advancement through it as section 7.3 discusses. We also discuss how we can characterize strategies in network visual exploration in section 7.4. Then, in section 7.5, we discuss the development of the network visual exploration roadmap and its impact on promoting the design of self-regulated educational material for such exploration purposes. Finally, I reflect on how our findings might relate to other visualization types in subsection 7.6.1, the current state of evaluation methods for visualization research in subsection 7.6.2, and how we can support guidance in network visual exploration within the context of knowledge generation model in subsection 7.6.3.

7.1 RQ(1): What do analysts struggle with when they use visualization tools to explore their network?

In chapter 4, we have identified eight **barriers** analysts struggle within the process of networks visual exploration, which are B1: missing goals and questions, B2: preconceived ideas and mental images, B3: deciding on network structure, B4: choosing the right level of abstraction, B5: choosing appropriate data formats, B6: importing and defining data schemes, B7: interpreting visual patterns, and B8: establish trust in visualization. Barriers' identification enhances tools' design, evaluation, and the design of educational efforts. Each barrier has its own impact on the analyst's data exploration process. In addition, in chapter 4, I have described the four **user types** I identified, which are demo users, data strugglers, single-session explorers, and multi-session

explorers. Identifying user types helps distinguish different trends and user needs. This also has enhanced the KPIs and monitoring dashboard design to evaluate users' exploration process (see chapter 5).

Further studies can investigate the existence of other barriers by experimenting with other tools, for example. Barriers can be classified based on the exploration pathway they exist in, the visualization, and even the domain in which they occur. Barriers can be used to design new supportive material customized for the nature of that barrier. In addition, different classifications of user types can be performed based on other factors (e.g., data domain). Educational efforts can be designed to accommodate user types and their needs rather than treating the different needs of all users as one.

7.2 RQ(2): How do we capture what analysts are doing through their exploration process in the wild?

We can capture how analysts explore networks in the wild by defining new metrics describing the process. That is, by studying the process of network visual exploration in the wild by applying several approaches described in chapter 3. The approaches included collecting and analyzing data from interactions logging, an intensive course, a mini-questionnaire and network visualization's state annotation, a coaching program, and analytical dashboard design. As a result, we can compile two sets of metrics: (1) a set of metrics to measure network visual exploration as a *process* (subsection 7.2.1) and (2) set of metrics to evaluate network visualization *tools* (subsection 7.2.2) in addition to standard evaluation metrics described in section 2.3.

7.2.1 Metrics to Evaluate Visual Exploration Process

As discussed in subsection 2.3.5, defining metrics to assess the network visual exploration process is tricky. As there are no tangible values to measure the progress in network visual exploration similar to the learning process in education (e.g., pass/fail mark). Based on this thesis's research, I compile the following **metrics that represent means of advancement and progress** in network visual exploration as follows:

- **The duration required to create a network schema**: can be used to calculate a threshold value (e.g., the average or median) for flagging analysts who exceed this time. These raised flags or indicators can help identify analysts struggling with creating a network or formatting their data. This process ensures timely assistance and enhances overall efficiency in network schema creation.
- **The first explored visualization**: by an analyst can serve as a key indicator of progress in

the exploration process. For each data model (e.g., network data), different visualizations can represent the data from various perspectives, such as connectivity or clusters. Typically, there is a base or primitive visualization that analysts initially think about and create to represent a specific data model. For example, in network data analysis, node-link diagrams are often the base visualizations used to explore network data, among other options like adjacency matrices, network-over-map views, and others. Therefore, the first visualization explored can help distinguish analysts who start with the base visualization (e.g., node-link diagram) from those who begin with more advanced, non-base visualizations. This distinction provides insight into the analyst's ability to interpret complex visualizations. Analysts who start with non-base visualizations (e.g., adjacency matrix) may have prior experience with the network or be more familiar with the exploration process.

- **Number of explored network schemes**: as a metric representing the number of exploration rounds the analyst takes. An exploration round or iteration starts by creating a network schema, exploring it visually, applying any optional steps, and starting again. Creating and exploring different networks from the same dataset can indicate to what extent the analyst is skilled.
- **The export** of wrangled datasets and explored visualizations: exporting datasets or visualizations show the analyst's interest in the resulting visualization and/or wrangled dataset that might contain insights.
- The creation of **bookmarks**: by capturing and annotating network states as the analyst explores. This represents a means of advancement in the exploration process, as the analyst can locate insights and visual patterns.
- **Applying data transformation operations**: show that the analyst has reached some understanding about his data or insights and would like to explore another form of the network (e.g., network projection). It reflects the analyst's understanding of the exploration process's repetitive nature, where knowledge results from several cycles of exploration and analytical tasks.
- Exploring **coordinated-views interactively**: for lengthy duration represents the analyst's ability to interpret and link information between different visualizations.

Such metrics allow us to capture signs of the analyst's advancement in their exploration process. The more signs are found, the more the exploration process can be described as being a *progressive exploration*. Further metrics can be discovered to provide a better understanding of the analyst's exploration advancements and needs. Furthermore, we will investigate how and why an analyst transitions from one user type to another (e.g., from a data struggler to an explorer), identifying the resources that assisted them and the obstacles they encountered.

7.2.2 Metrics to Evaluate Networks Visualization Tools

In section 2.3, we reviewed methods and metrics to evaluate the design of visualization tools (e.g., design of interfaces, workflows) [244] such as consistency and error-free performance of common tasks. Based on our research, we add the following metrics to evaluate the design of network visualization tools' features, help resources, and educational efforts:

- The **ratio between user types in tool's usage**: The ratio between user types represents the tool's performance. For example, if the data strugglers are higher than the explorers, then it shows the tool has a problem in network creation and modeling. If single-session explorers are higher than multi-session explorers, then it shows that analysts either did not find what they wanted or were interested in a single use of a specific visualization/feature.
- **Statistics of interactive features usage**: reflects which tool features analysts are more aware of their purposes and how to use them in the context of their exploration goal.
- Analyzing **trends in users visits along with educational efforts**: to investigate the impact of such efforts through the number of users' visits per type. For example, if the number of explorers increases by the end of the course, it indicates a positive impact of the course.
- **Evaluate the usage of help resources**: by analyzing which help resources are used, in which step of the visual exploration process, and how different user types approached it.
- **Means of Supporting Visual Exploration**: Does the tool provide analysts with features that facilitate the visual exploration process? such as tracking exploration history, offering exploration suggestions, and enabling collaborative exploration.

The state of current network visualization tools is not up to analysts' needs. Currently, most of the network visualization tools provide some means of network visual exploration support. Such as offering: various network visualizations (section 2.2), network modeling (subsection 2.2.1), data wrangling (subsection 2.2.2), or guidance (subsection 2.5.2). However, no tool exists that offers a complete, rich experience of supporting the visual exploration process from the beginning to the end. Such as having a tool that offers network modeling options, recommends visualization types, and suggests suitable data wrangling and guidance. Nonetheless, offering reasoning for the system's guidance by explaining how and why they were suggested.

The evaluation of network visualization tools should assess the tool's features and ability to support the advancement in the visual exploration process. Thus, we shall consider and benchmark the tools' exploration features to reflect the analysts' expertise. For example, we can classify the features based on their usage in the context of visual exploration as basic, intermediate, and advanced features. *Basic* features to include the minimal set of exploration features (e.g., data filtering), *intermediate* features to support specific exploration tasks (e.g., data wrangling), and *advanced* features to guide various exploration tasks(e.g., collaborative

exploration). Understanding the exploration process further and how to support the minimal literacy an analyst needs is important.

7.3 RQ(3): How can we enhance visualization tools design to support analysts in their visual exploration of networks?

The design of visualization tools can be enhanced to support analysts in their visual exploration of networks by: (1) ensuring to overcome barriers identified (chapter 4), (2) defining further metrics and KPIs that describe the network's visual exploration process, how users advance within it (subsection 7.2.1), and evaluate tools using them (subsection 7.2.2), (3) representing metrics and KPIs through analytical dashboards to reflect different users types needs and behavior (chapter 5), and (4) designing educational guides that promote self-regulated learning such roadmaps and exploration steps' activities (chapter 6). Through the roadmap's design, we identified four exploration pathways (chapter 6) to consider when creating tools and educational material, which are learning about concept & tools, informing data collection, free-form exploration, and goal-oriented exploration. However, further pathways can be identified or explored based on the classification criteria. Pathways not only help analysts in their visual exploration of networks but also help tool designers in addressing different needs.

The design of visualization tools, educational material, or evaluation studies focused on network visual exploration shall consider *users' types and their needs, learning pathways, exploration strategies, type of help resources provided, the level and type of guidance provided by the tool, and the tool's workflows*. Tool workflows and roadmaps should be considered as the two sides of the same coin. However, tools' evaluation studies should additionally consider the types of interaction logs and their categorization, the neutrality of the tool's design (not imposing a certain way of exploration), and which means will be used to capture the exploration goals. The design and evaluation process should focus on the coherence of the exploration process rather than separated steps.

Learning analytics and dashboards assist in describing the analysts' behavior and the whole exploration process. That is, by helping in extracting new metrics and KPIs needed in: (1) understanding users' decisions and (2) designing actionable dashboards that support decision-making(chapter 5). Thus, we can define the **visual exploration dashboard** as *a multi-page display where each page represents indicators and metrics through one or multiple visualizations about the user, the tool, and the visualizations within the context of visual exploration of data*. The visualizations designed in our dashboard were customized to reflect the visual exploration process, such as the exploration timeline. However, further visualizations can be designed and customized to reflect the visual exploration process and metaphors similar to Schwendimann et

al. [237] recommendation in designing visualizations for learning dashboards.

Further efforts can be taken to enhance the tool's design. For example, exploration maps or decision trees can be created to track analysts' steps and decisions while exploring data. This can inform the design of self-regulated material and recommendation systems. Also, additional dashboards can be designed to support analysts in their exploration process, similar to dashboards for students [29]. That is, by providing analysts with a better understanding of their own exploration process, they will gain credibility and trust in how they reached their findings. Dashboards represent and fulfill what Shneiderman and Plaisant described [244] as *"the need to have an actionable evidence of measurable benefits"*.

7.4 RQ(4): How can we characterize this exploration process in the matter of interactions and strategies followed by analysts?

In chapter 6, we have extracted and described 16 strategies in network visual exploration. Then, we classified the strategies into 3 main categories based on their goal as follows: 3 data-driven strategies, 10 task-based strategies, and 3 open-ended strategies. Each strategy is described using a description card that contains the strategy's goals, where, and how to apply it. The complete strategy description cards are in the appendix (section C.3). As an educational guide, the pseudolanguage used in describing the strategies enables analysts to understand exploration steps and how they are applied without using technical terms. That is, Harrison [118], and Sperrel et al. [250] recommend the use of natural or structured languages (e.g., pseudocode) in providing instructional guides. Strategy description cards educate analysts on navigating the exploration process, reflecting on their observations, and considering how to act if non-desired results are reached.

Strategies are characterized according to the general steps they share in network visual exploration. However, strategies differ in several aspects that should be considered as follows. Firstly, **the strategy's applicability** on specific data type only or generally for any data type. For example, temporal analysis can be applied only to data with time/dates, whereas gradual exploration can be applied to any dataset. Secondly, **the clarity of the strategy steps' description**. In other words, how straightforward patterns can be found using a specific strategy? For example, locating the node with the most links can be directly found. whereas comparing clusters based on their sizes or comparing two networks might not be straightforward, and it might require additional steps to be taken (e.g., investigating how two different network schemes were created or their data wrangled). The strategy's clarity can be measured by different means, such as if it

can be applied by different user types, expertise, or by the number of iterations needed to achieve the exploration goal. For example, the number of times the network schema was changed or data was wrangled toward reaching a goal. Thirdly, **the optimality of a strategy**, that is, if several strategies can be applied toward a specific goal, which strategy can lead toward better outcomes? Is it the strategy that requires shorter or longer exploration time, fewer iterations, or the one that reveals more insights?

The characterization of strategies can be enhanced further by analyzing exploration practices. For example, how can strategies' descriptions differ in the level of detail provided based on the user type's needs (demo users vs. multi-session explorers)? In addition, how and when additional actions can describe the strategies. For example, interaction logs analysis showed that analysts explore following various behaviors, such as exploring single vs. multiple visualizations, in sequential vs. parallel order, and whether they start from a primitive (e.g., from node-link diagram) vs. non-primitive visualizations (e.g., adjacency matrix). Thus, a better description of strategies can be provided by learning more about how analysts explore.

Data transformations are also important, as they impact the strategies and allow a better understanding of the data. Different studies suggested various data transformations such as: using hierarchies in clustering [259]. Thus, strategies can include further details about data wrangling and data transformation suggestions (e.g., based on the domain). Strategies description forms a ground base for describing exploration strategies and informing the design of automated guidance features (e.g., graph neural networks (GNNs) and notion detection). It can also be used to create data specification language (DSL) for exploration purposes. DSL languages can support describing data and logical relations between data attributes and exploration goals/questions. We shall examine the impact of different strategies on the exploration process and what additional strategies are to be identified and extracted from different mediums.

7.5 RQ(5): How can we teach or prompt self-regulated learning of visual exploration of networks through the tool?

To understand and teach networks visual exploration, we identified some of the barriers analysts face and their needs (section 7.1). We also extracted metrics to measure analysts progress (section 7.2), created educational resources (section 7.3), and provided description of some exploration strategies (section 7.4). However, most analysts explore independently, so how can we design the resources and guides to support them in exploring and learning at their own pace? Thus, to prompt **self-regulated** learning in networks, visual exploration, tools, and resources should support a more structured *personal learning* opportunity. As described by Zoss et al., [282] *personal learning* represents an unstructured learning opportunity within an informal

learning environment as in museums. In comparison, formal learning environments offer highly structured learning opportunities, such as online courses. However, visual exploration of networks cannot be taught through the highly structured environment as it requires the flexibility of the exploration process. Thus, by using the roadmap and coaching program (chapter 6) we aim to support *self-regulated personal learning* through an environment that blends *informal* and *formal* learning opportunities, which is described by Dabbagh and Kitsantas' [79] as *personal learning environment*.

Designing and delivering a self-regulated learning opportunity in visual exploration of networks requires metacognition as explained in subsection 2.5.5. As a starting point, the roadmap provides a *scaffolding* opportunity towards better knowledge transfer of network visualization skills and literacy [282]. This enables the roadmap to provide the metacognitive regulation needed to orchestrate the visual exploration process. In addition, the roadmaps raise self-knowledge as part of metacognitive knowledge, which enables analysts to know the strengths and weaknesses of their learning process of network visual exploration. Also, step activities offer task knowledge and exploration strategies, and guidance offers strategy knowledge on exploring data based on different goals. Thus, self-knowledge (roadmap), task-knowledge (steps' activities), and strategy-knowledge (strategies description cards) represent the metacognition required to form the basis of self-regulated learning as described by Schraw [234].

The description and evaluation of the self-regulated learning process in networks visual exploration process is distinguished from self-regulated learning in education in general by the following aspects as Table 7.1 shows. Firstly, the *self-assessment in education is more assertive* and the achievement of a task can be clearly assessed (e.g., choosing the correct answer, calculating specific value). Whereas in the visual exploration process, the *assessment of progress is more relative through the analyst's self-reflection* through asking questions (e.g., did the visual patterns answer a question?) or deciding on the need to apply other actions (e.g., data transformation). Secondly, the assessment of learning in education can be bound to time (e.g., the duration of solving a problem). Whereas the learning and assessment of visual exploration can be bound to other aspects, such as the number of complete exploration rounds or iterations (e.g., the number of different network schemes created). Thirdly, the content assessed in education (e.g., course) is usually fixed. Meanwhile, in the visual exploration process, analysts vary in the visualizations they explore (e.g., layouts and patterns) according to the data, schema, and data transformation they apply. Finally, learning analytics assesses the student's progress through different tangible metrics. Whereas in the visual exploration process, learning analytics used to measure the progress differ due to its nature, such as the duration of creating a schema or exporting various means of reaching insights (e.g., bookmarks, visualizations, and wrangled datasets). Thus, further formalization of the process and its metrics should be considered to

evaluate the visual exploration process.

Criterion	Self-regulated Learning in Education	Self-regulated Visual Exploration in Networks
Evaluation	Self-assessment (e.g., correct answer provision)	Self-reflection (e.g., findings reached by exploring and reflecting, no single correct answer)
Progress Bound	Time-bound	Iteration-bound as the higher iteration, the better to learn about the tool/concept/dataset
Content	Fixed-content (e.g. lessons) for all learners	Varying content based on explored dataset and/or goal
Learning Analytics and Metrics	Well-defined in education such as: assessment mark, learning sequence, average students marks	Analytics for visual exploration successfulness includes means of advancement as discussed in subsection 7.2.1

Table 7.1: Personal Learning in Education vs. Visual Exploration

To support self-regulated learning of network visual exploration, the *roadmap should*: mirror the exploration process step-by-step, inform the analyst about the data semantics and underlying structure, ignite ideas on how to think about data, provoke analysts to iterate the exploration process and experiment and to establish trust in the process. Also, ***the design*** *of a roadmap should* ***consider*** the following aspects: capturing exploration goal(s), customizing pathways, personalizing trajectories, and considering the tool's neutrality (not designed for a specific tool). Roadmaps can be studied further to customize them based on industry stakeholders and evaluate the suitability of network visualization tools in their own domain. Analysts can spend different durations on each step. Thus, a heat map can be built above the roadmap, showing areas of interest in the roadmap. For example, which set of steps the analysts return to and which steps and activities most traffic and transition happen between. Those areas reflect where most exploration attempts and/or uncertainty took place. The tool's support and coverage of the roadmap can be used by itself as an evaluation method.

The roadmap, steps, and activities collectively contribute to networks of visual exploration literacy, including representational, metaphoric, and topological literacy [282]. That is the roadmap's representation of how network visual exploration informs analysts with representational literacy. Steps and activities as sketching networks and transforming them into network

schema contribute to metaphoric literacy. Steps and activities related to choosing, creating, and interpreting network visualizations assist in learning topological literacy. Thus, as the analyst explores, the accumulated knowledge enriches the analyst to make better decisions in the future on how to explore data sets with similar structure or nature by acquiring various related network literacy (see subsection 2.5.1). Through our work with participants, we gathered numerous examples and stories illustrating how analysts' knowledge evolved and how they achieved a deeper understanding of their data, which subsequently informed their future decisions. For an in-depth look at one of these stories, please refer to section A.3.

The **limitations of the studies of this thesis** include the following. All of our studies were conducted using a single network visualization tool. Thus, including other network visualization tools can inform about additional barriers, learning analytics, and other aspects to consider in supporting self-regulated learning in network visual exploration. In addition, the size of the graphs investigated by participants was limited. Most of the participants worked with networks of small or medium size (number of links less than 5000); thus, making sense and looking into large network exploration requires additional efforts [67]. In addition, our studies covered only the four visualizations provided by the Vistorian. Other network visualizations might expose other needs or barriers.

7.6 Reflection on the Findings in Visualization Research

In this section, I briefly reflect on the broader implications of my work and my thesis's contribution to the data visualization community. In the first part, I reflect on the findings and discuss how they prompt us to consider the needs and potential in visualization research (subsection 7.6.1). In the second part, I reflect on the evaluation methods for visualization researchers and their current state(subsection 7.6.2). Finally, I discuss how guidance shall support network visual exploration in the context of knowledge generation model(subsection 7.6.3).

7.6.1 Generalization of Findings on Visualization Tools

The outcomes of this work are not confined to network visualizations and can likely be examined across a broader range of visualizations. Therefore, I will discuss the main findings of my thesis individually, reflecting on their implications and generalizability.

Through the first study (section 4.4), I identified four user types: demo users, data strugglers, single-session explorers, and multi-session explorers. Those user types resulted from log events analysis based on the log event types I captured anonymously (subsection 3.2.2). Those user types are not limited to network visualizations or the tool I experimented with -the Vistorian -.

User types are important to offer a better understanding of users' needs, visualizations, and tool usage, as discussed in chapter 5. Such user types can be identified across different visualization tools and visualization types with similar or different needs, trends, or characteristics. Further user types may be identified by capturing additional or different log event types, exploring different visualization types that may require different steps, or even with different data types. Some user types might even be broken down into sub-types through an extended analysis of each user type's logs. User types can also enhance log analysis by identifying which visualizations analysts start with for each user type and uncovering common patterns in their exploration history, also known as clusters. Visualization tools should offer customized experiences based on the user type and needs rather than treating all users similarly.

Also, we identified barriers in network visual exploration through the in-depth course study (section 4.5). The identified barriers are not restricted to networks' visual exploration, but many can be generalized across visualization types. *Some barriers apply to most visualization types*, such as B1:missing goals and questions and B2:pre-conceived ideas and mental images. *Some can be linked to specific visualizations*, such as B3: deciding on a network structure. Studying barriers across complex or non-simple visualizations can play an important role in enhancing how visualization tools can be built or how visual exploration processes can be taught. In addition, certain barriers, such as B8, open up intriguing research opportunities, focusing on establishing trust not only in the findings from network visualizations but also in the overall visual exploration process. The barriers can be examined across tools, visualizations, pathways, user types, and data domains. They can also be used to build tailored educational material. Barriers can also be used to define new metrics and KPIs that represent tool users' actual skills and state and their understanding.

A recent study by Nobre et al. [187] has investigated barriers to visualization literacy with 12 different chart types. They have characterized barriers as either conceptual or operational barriers. Also, they classified them into 3 categories which are: translation barriers (when an analyst cannot translate from the research question to a visual query), encoding barriers (when an analyst cannot comprehend visual encodings of a visualization), and decoding barriers (when an analyst cannot extract the information presented through visual channels correctly). Many of the barriers found in our work chapter 4 can be mapped to one of the categories. However, the barriers categories mentioned by Nobre et al. [187] are more limited to visualization literacy and how analysts interpret visualization with predesigned tasks. Such experience differs from visualizing and exploring data using network and other complex visualizations. Thus, we can see some of the barriers we identified in network visual exploration cannot be classified into any of their categories. Also, some of the barriers are faced before or after the visualization cycle's translation, decoding, and encoding parts. Thus, a wider categorization can take place to cover

visualization exploration, explanation, and literacy.

Key performance indicators (KPIs) discussed and developed in our study (see chapter 5) show the real need to have new metrics or ways to assess visualization tools (subsection 7.2.2) and understand the visual exploration process (subsection 7.2.1). This should include understanding the visual exploration process and finding new means of knowing when users successfully reach their goals and when to flag that they need help. In addition, they present additional and new ways to evaluate visualization tools, as I will discuss further in subsection 7.6.2. The proposed KPIs can be used across visualization types and tools as they are, or they may need customization to accommodate different needs and usage.

Pathways identified and discussed in subsection 6.3.2 emphasize the importance of addressing different analysts' needs. That is, by differentiating the four pathways and how analysts act toward their own goals. The four pathways based on this thesis's analysis are Learning about Concepts and Tools, Informing Data Collection, Free-form Exploration, and Goal-oriented Exploration. Those pathways were characterized based on having or not having data and/or exploration goals. Other pathways can be distinguished by selecting different factors. The factors can be related to the visualizations, tools, or the mental model and how analysts visualize and explore data.

Network visual exploration roadmap (see chapter 6) is one of the main contributions of this thesis. It shows the impact and important role of such guides in assisting analysts to understand how to move from one step to another, how to overcome some obstacles, what the analysts' options are (e.g., exploration strategies, data transformation operations), and others. It helps developers and researchers understand what different pathways share and what they differ in. Roadmaps actually emphasize the visualization tools' current state and that they are behind the analysts' actual needs. That is, most of the tools do not support all of the visualization steps. In other words, most visualization tools currently ask for the data files and the visualization types to create one. But how many of those tools ask the analysts what question they are addressing, what their exploration strategy is, and how they aim to achieve their goal? Thus, creating exploration roadmaps for other visualizations can help find roadmaps' commonalities and differences across visualization types.

Studies on creating exploration roadmaps for various visualizations should be extended to cover associated activities with each step. That is, each step of the visual exploration process (see section 6.4) should be examined across visualization types. For example, concept map sketching associated with network visual exploration roadmap may be suitable for other visualization types or substituted with other diagrams. In addition, to find the appropriate symbols to use in the diagrams' sketching, what do they convey to learners and explorers, and what might they mean? Sketching concept maps for network visual exploration is an important

activity that not only teaches the underlying concept but also offers a better understanding of the data. In the concept map sketching activity (full material provided here section C.2), analysts are introduced to two ways to start sketching, either from data or from the research question. That is, starting from data as a whole teaches the analysts more about what sort of data they have and what it describes. This opens the analyst's eyes to a broader perspective about the possible questions to investigate or answer. Meanwhile, sketching concept maps based on the research questions might limit the scope of analysts' aim to investigate, how it can be achieved, and what if all required data is present. Thus, it is important to have and decide on the various symbols needed to represent different cases using the concept map (Symbols table is provided in section C.2). For example, what is the best way to represent a relationship among three interconnected nodes, or how should one depict an expected but missing node? That is, networks are not only simple nodes represented as circles and links represented as lines. They are much more than that and tell us more about data. For example, what type of nodes is connected to which type, what is the type of link connecting, how dense is their connectivity, is the date, for example, describing the links or nodes, and much more. Also, it is important to look into the extent to which the symbols used in concept map sketching are clear and can be easily converted into node/link tables. Thus, it is important to formalize and regulate the process of concept map sketching and transformation into network schemes to establish better practices.

The design of help and support resources should examine and consider the analysts' preferences for each step of the visual exploration process. For example, different resources should be provided, such as data comics and videos. Exploration strategies should be mined from various resources, described in detail, and classified (subsection 6.3.1), including how and when to apply the strategy and how to evaluate its effectiveness. The full set of the 16 strategies described in my study is provided in section C.3. This makes us think of how we can create a strategies pool and allow it to inform analysts and provide tools with a resource that provides automated help for visual exploration. Strategies should be collected and described in a structured language (e.g., Pseudo language). Such strategies can be found by working with analysts and extracting from literature. Another crucial function of the roadmap is that it highlights the need for a systematic approach to the decisions that analysts routinely make. This structured framework ensures consistency, enhances the decision-making process, and improves the overall quality of the analysis. For example, how can we assist analysts in deciding which visualizations to explore and in what order? When explained and presented, such decision trees provided by the roadmap help lift the burden from analysts, clarifying which visualization will best inform their understanding of the data, in which order, and why. Another way of regulating and formalizing the exploration process is to inform analysts that what they visualize, explore, and interact with is what they describe in their network schema. Analysts, when using visualization tools, are not

always fully aware of how to use interactive features to explore and understand their data. For example, how to filter based on the node type or how to represent several links between two nodes in a way that serves the purpose best. That is, the schema used to create the visualization is the backbone of the exploration process, and this is what analysts need to focus on.

Finally, we cannot truncate the analysis and visualization of any dataset from its data collection process. Thus, any data transformations or manipulations should not be applied or suggested to the user without promoting them, as this might impact the data interpretation. The same applies to the tool's design and to what extent it imposes ways of exploring data. Thus, finding user types, barriers, KPIs, and sketching roadmaps for different visualizations can help in understanding more about the visual exploration process as a whole. It can inform visualization researchers about how analysts visualize data, reach insights, and reach their research goals. In addition, it shows that currently, no tool exists that offers a complete, rich experience of supporting the visual exploration process from the beginning to the end. Such as having a tool that offers network modeling options, recommends visualization types, and suggests suitable data wrangling and guidance. Nonetheless, offering reasoning for the system's guidance by explaining how and why they were suggested.

7.6.2 Views and Recommendations on the Evaluation Methods for Visualization Researchers

Researchers in the visualization field have experimented with several evaluation methods in various settings, as discussed in section 2.3, subsection 2.3.1, and subsection 2.3.2. Each evaluation method has its pros, cons, and ideal circumstances within which to conduct it. For example, some methods are more suitable to evaluate the visualization's design and literacy, whereas others are more fit to evaluate the visualization tool's workflow or usability. Thus, choosing the correct evaluation method(s) plays a major role in the study's success. Thus, in this part, I reflect on what I have found about the evaluation methods for visualization researchers.

As the cornerstone of any study, determining the **study's scope** helps select the proper evaluation method(s). That is, by choosing whether we are *evaluating aspects* related to *the visualization* itself (e.g., readability vs. construction of visualization, creating simple bar charts or complex network visualizations), *the visualization tool* (e.g., workflow and how users' interactions), or the *mental model and perceptions around the visualization process* (e.g., exploring data visually and how analysts choose their actions, explaining results through visualizations). Such decisions impact which methods are more suitable than others for a specific scope. For example, a task-based or insight-based evaluation method might be better if a study aims to evaluate how analysts interpret a newly designed visualization. If a study aims to

evaluate the usability of a visualization tool, logging and analyzing user interactions or think-aloud protocols along with video recording might be preferable. However, when working with a larger study scope, such as the mental model behind visualization construction, researchers examine how analysts create visualizations from their raw data, interact with the tool, overcome obstacles, and arrive at insights. Evaluating such a wide scope cannot be accomplished by only logging user interactions or pre-designed task-based evaluation. It requires using evaluation methods that complement each other and are aligned toward assessing how analysts reach their goals step-by-step. To do so, I discuss the main factors to consider when running a study and applying evaluation methods.

After determining the study scope, deciding on **the study setting** is one factor that impacts the outcomes' nature and how informative they are. For example, conducting a study in the wild versus a controlled setting exposes the study to different circumstances. The controlled settings and some evaluation methods limit the study outcomes or their quality. In task-based evaluation methods, analysts are given predesigned tasks aimed at testing specific skills. Consequently, researchers might overlook key insights if the tasks do not encompass them. In contrast, evaluation methods conducted in the wild are open-ended, allowing researchers to observe how analysts independently use their tools and to identify the challenges they might encounter. Such *openness to various types of outcomes* is important to understand further about real-world scenarios. Using network visual exploration as a case study, it becomes evident that this process is not a simple plug-and-play operation. It necessitates that the network schema created by the analyst produces visualizations that effectively serve the intended exploration goals. Therefore, every step in the visual exploration process is interconnected and cannot be isolated from the preceding steps. Consequently, capturing this process in non-controlled, real-world settings is essential.

Interpreting the outcomes of a study requires a crucial step, which is **capturing the context** of the user's visualization and analysis process. This factor is important in analyzing data collected in various study settings, especially in the wild. As there is no direct communication with users about what they want to do. The context is about knowing the *analyst's goal* of the visualization he is creating or exploring, which is crucial to understanding and interpreting the analyst's decisions. The context includes what the analysts want to do, how they are willing to do it (aka strategy), and what interaction events are associated with achieving such a goal. So, to evaluate the visual exploration process, we need to capture the analyst's goal using a wizard or a grammar language, for example (see section 7.7). In addition, we need to capture the analyst's actions, for example, by logging interaction events. Even logging user interactions should not be tracked as simple clicks or entries but should be tokenized in a way that offers a meaningful interaction. For example, clicking on a node in a network node-link diagram can be logged as a

simple node click as one way to do it. Whereas a more meaningful way is to log it as *selecting an ego-network in a cluster*[1], for example, where the interaction is tokenized with the context and its interpretation. Furthermore, although any user's identity is anonymized, all decisions and actions of an individual should be tracked and logged as a single profile. This offers a better understanding of the analyst's decisions and why they are taken.

Another factor is deciding on *how close we need to work with analysts*. For example, even though studies conducted in the wild aim to assess their objectives remotely, they should be accompanied by efforts to work closely with the analysts and collect qualitative data. Such as observations, discussions with analysts, surveys, and interviews to allow researchers to reflect on the evaluation process and remove obstacles that forbid understanding the whole picture. For example, by running in-depth courses of multiple sessions (see section 4.5) or coaching programs (see section 6.5) where sessions are conducted, and support is provided based on the analyst's needs. This offers a deeper understanding of users' needs and how they act in real-world scenarios rather than studies designed for success. Studies should involve analysts who differ in their background, such as their disciplines, expertise (e.g., practitioners vs. beginners), user types (see section 4.4), and pathways (see subsection 6.3.2). This allows for a better understanding of various analysts' needs. For example, analysts from the data collection pathway will offer opportunities to learn about their usage, which might differ from free-form exploration analysts' usage. We shall have both multidisciplinary and intradisciplinary studies. This allows us to differentiate domain requirements (e.g., how genes are explored) from requirements related to the visualization technique itself (e.g., not having two nodes with the same ID), what is applicable, and what is not. The more the study is **inclusive** and participants' profiles are **diverse**, the factor that allows observations and findings to be more informative and generalizable.

Newer evaluation metrics should be developed and extracted to evaluate the visualization process(subsection 7.2.1) and tools(subsection 7.2.2). Metrics and key performance indicators (KPIs) help evaluate aspects related to visualization research, depending on more tangible measures. For example, to evaluate the tool's usability from a different perspective, such as the number of multi-session explorers versus the number of data strugglers. Metrics and KPIs also help understand trends in different user types' usage along the visualization process (e.g., which support resources and educational material each user type relies on). A main contribution here is that we can characterize the visual exploration process and how analysts advance and provide help as needed (see section 7.2 and chapter 5).

Evaluation methods should consider new aspects. For example, how many challenges does an analyst face when using a certain visualization tool or when reading/creating a specific type

[1]The interpretation of a user action varies based on the tool features being used.

of visualization? Identifying potential barriers through the evaluation process (see section 4.9) is crucial for knowing if the analyst is facing difficulty with the tool itself or the exploration process. Another way is to use visual exploration roadmaps (see chapter 6) to benchmark visualization tools and evluate them. That is, the design of the exploration roadmaps does not only contribute to the literacy of visualizations but also to the evaluation of visualization tools. That is, by examining to what extent visualization tools support the mental model behind the exploration process. And how tools walk analysts step-by-step from the beginning to the end, which steps they support and which ones they do not. In addition, the communication style or medium should be evaluated. For example, does the tool use dialogue-based communication, or is it a pure yes/no question? That is, the input interfaces that do not allow system-user dialogue can cause difficulties for the analysts to explain what they want to do.

Biases, mental images, and pre-assumptions represent, as a whole, a factor that plays a major role in the findings' validity. For example, the tool's design might impact the research methodology by imposing a specific way of representing data (e.g., date and time to be visualized as changes through timeline only). Such unintended messages delivered by the tool's design are taught to the user silently and impact his understanding. For example, if time is explored using a time-slider, some analysts might mistakenly believe that this is the only method available for temporal analysis. This impacts how the analyst understands their data and the exploration process. Limiting biases is a hard task, and this is bound not only to the tool's design but also to the design of educational materials and support resources. Thus, it is important to consider possible biases and pre-assumptions when applying evaluation methods to have more accurate results. Furthermore, training examples and demos are crucial for participants' training purposes on how to understand and apply concepts. It is important that such resources do not build any biases and, on the contrary, train analysts on how to think freely about their data. Additionally, evaluation methods should consider to what extent the tool informs the analyst and educates them as they use it. That is by explaining the consequences of some decisions or actions or explaining the differences between multiple choices and how to select among them. For example, to inform the user about choosing which visualization to explore first among network visualizations. This is important to spread good practices, skills, and knowledge in visual exploration of networks.

Further efforts are needed to understand the mental model behind the exploration process (see section F.4), what is generalizable, and what is not. Also, how to use our understanding to support the sensemaking process (see section F.2) and design educational guides (see Appendix F). Pirolli and Card [204] have described that several *leverage* points impact the sensemaking process as either part of the foraging loop or the sensemaking loop (section 2.4). They suggested that such points should be considered as part of the design principles of new technologies and the development of evaluation studies and metrics. Another suggestion is

to look into Bloom's taxonomy and how each category is distributed among the knowledge acquisition cycle in network visualization (see section F.3). This shall help in understanding each user type's needs in the visual exploration process and how to support them in advancing in both visualization literacy and skill acquisition.

To conclude, the evaluation methods are a bit behind the actual visualization research needs. They still do not offer the ability to capture context around the analyst's goal. In addition, most of the studies in the literature lack a deep classification of how effective the evaluation methods are outside controlled settings. Controlled settings are important in several situations, such as assessing visualization literacy, but they fall behind in the wild when analysts are on their own. Thus, all of the factors I mentioned here play an important role in reconsidering how evaluation methods in visualization research are conducted and how they shall be conducted.

7.6.3 Toward Explainable Guidance in Networks Visual Exploration

In this part, I describe our findings in network visual exploration within the context of the knowledge generation model and how to provide guidance. The knowledge generation model (see section 2.4) is influenced by the selected exploration strategy during exploration. It defines how the strategy impacts the way the user reaches their findings. In network visual exploration, analysts start generating knowledge even before interacting with the visualization. That is when the analyst starts deciding which nodes and links will form the network schema and enable the analyst to understand data further. After settling on the schema, the user chooses which visualization to create. Then, the user can take any of the following actions: **explore** by interacting with the visualization interface, **wrangle** data by transforming data to reach the targeted network visualization, and **select a strategy** by selecting an initial strategy or changing the current one used. Each one of those actions generates further knowledge about the data and how to understand it.

There are many efforts to support such a process. However, guidance efforts in network visual exploration are scattered over educational content and automated guidance. Educational efforts aim to guide the analysts but cannot inform them about overcoming obstacles when using visualization tools. On the other hand, the visualization tool can recommend actions, visualizations, or data transformation operations, but it does not explain why such recommendations were provided. Guidance should be tailored to the analyst's goal (e.g., recommending strategy based on goals). Guidance can be improved by annotating data attributes to describe data semantics. Thus, such efforts are uncoordinated, and they suffer from the following:

- Analysts cannot decide which educational resource works the best in their case.
- Efforts are unsynchronized. This means that educational content might provoke a spe-

cific strategy or action, whereas the visualization tool or computational guidance might recommend something else. This results in the analyst's inability to decide on which recommendation to go with, distrusting the whole exploration process, and is against self-regulated learning.

- The guidance process is best when it mixes educational and computational system-based guidance. That is, if the analysts rely only on educational content, it costs them time and effort to explore data and achieve their goals. If analysts rely only on system-based guidance, they cannot understand how and why the visualization shows what they see.

Thus, combining both educational and computational system-based guidance in designing what we can call *explainable guidance* is preferred. Explainable guidance systems shall educate and inform analysts about possible exploration strategies, data wrangling operations, visualizations selection and interpretations, and verifying their findings in a non-technical language. Current system-based guidance [61] offers 3 degrees of guidance, which are: *orienting* through visual cues, *directing* by providing a ranked list of suggestions, and *prescribing* by the system selecting the optimal suggestion. Based on our discussion, explainable guidance should provide further guidance degrees such as:

- *Igniting*: ideas on how to think about data and possible ways of exploration,

- *Informing*: the analyst about how strategies can be applied step-by-step,

- *Reasoning*: why certain actions should be taken and why (e.g., why a specific visualization is explored or why a data transformation is applied).

System-based guidance only assists from the system side without considering the user's interactions as mixed-initiative guidance. Mixed-initiative guidance provides certain automated responses based on the user's interactions [80]. However, mixed-initiative guidance fails to capture the user's strategy and educate them on how to apply and navigate through the exploration process step-by-step. Also, suggestions based on previous users' interactions [80] are helpful, but in some cases, they might be misleading or incorrect. Due to that, the suggestion process should be relied on context and goal. Thus, we should aim toward explainable guidance, where help is provided and explained, allowing users to learn as they go. We should always keep in mind that the design of guides and supportive material can either ignite ideas on how to explore the network data or limit it. For further discussion about how to support the mental model through guidance, please refer to section F.4.

7.7 Future Work

Future work can aim to investigate the existence of other barriers using other tools, define and extract new metrics to describe the exploration process, enrich analysts with educational material that supports their goals, and use visualization tools that support the exploration process. To start with, the roadmap's design can be enriched with additional customization. For example, the roadmap can be tailored further based on analysts' type (e.g., freelancers, students, researchers). Also, roadmaps can be enhanced by looking into customized pathways based on commonalities in domains such as business (e.g., ecosystems networks [19]) and cybersecurity (e.g., network intrusion detection [164]).

The design of network visualization tools can be improved by providing additional features. For example, to recommend different network schemes based on data and goal (using Machine Learning), extract network schema from research question (using Natural Language Processing), and create wizards to walk analysts step-by-step through the exploration process. This supports future work in gamification, data journalism, and self-explanatory network visualization through strategies and exploration dashboards to offer insights about exploration. In addition, tool design can be supported by adding techniques such as graph summarizing [177] to enhance strategies implementation (e.g., visual question answering, chart accessibility). Emphasizing visual cues of an analyst's interest can be supported by incorporating subjective feedback [209].

To improve the logging process, it is essential to include mechanisms that capture users' strategies and intent, particularly in the wild. It enables us to understand further without impacting how data is collected as Matcha et al. [171] said *"the use of learning trace data allows for understanding the students' actual learning behavior without intervening in their learning or inadvertently"*. For example, capturing the network schema specification can be linked to the exploration goal. That is, each time the analyst changes the schema (e.g., source node, target node, or link types), we can examine how the analysts interact with the visualizations and to what extent they advance. Enhancing the logging process is very important to enable deciding whether network visual exploration was a successful process or not.

The network visual exploration needs to have a means of raising trust and validation of the visual exploration (proof-of-concept). This can be done, for example, by declaring grammar for visual exploration. This shall give analysts a logical ground on how to describe data, transform data, query data, and derive findings. Also, how to support trust in visual exploration findings based on providing history, such as a receipt to present evidence on how reasoning was reached. Further guidelines can be set to align the exploration process so that the analysts can validate their progress. In general, network visualization tools' design should capture beyond just the data and which visualization the analyst creates. Visualization tools should capture the analyst's strategy, how the data attributes are linked, and in which context or semantics. This will enable

visualization tools to provide better support by offering recommended visualizations and possible data transformations. Thus, visual exploration roadmaps support not only analysts' exploration process but also tools designers to offer better tools.

As a conclusion, this thesis's work and studies in investigating network visual exploration have informed our understanding and how we can support analysts by identifying barriers(section 4.9) and types of users(section 4.4) of network visualization tools, determining KPIs(chapter 5), and metrics to measure the network visual exploration process(section 7.2), and designing educational materials that support self-regulated learning in networks visual exploration (the roadmap, pathways, and steps' activities, and exploration strategies description(chapter 6). Our studies underscore the importance of user-centric design in data visualization tools. Developers can use these findings to create more intuitive and accessible tools for a broader range of analysts.

Appendix A

Supporting Material of Barriers Network Visual Exploration Study

A.1 One-to-One Interview Questions: Reflection on the Course and the Vistorian Tool

A.1.1 General Feedback on the Course

1. Can you describe your initial motivations and goals that led you to take the course?
    - Do you feel these motivations and goals changed as part of the course or after having taken the course?
    - What was your starting point with regards to your data:
        - Did you have a particular question, focus, or visualization of the data in mind? If so, what was it?
        - If so, did this question/focus/mental image change over the course?
2. Can you briefly describe what you took away from the course?
    - What do you feel you have achieved, also in the light of your expectations/goals/questions when you started the course?
    - What role did the Vistorian tool play as part of this?
    - Did the course and/or your interactions with the Vistorian change the perspective on your data, your focus of interest, and/or the scope of your questions?
3. Were you able to create visualizations using the Vistorian that helped you learn more about your data and the questions you have about your data?

- If so, can you give us examples? What made these visualizations useful? How do they facilitate your research and questions now and in the future?
- If not, what was the problem? What got in the way of creating visualizations that were helpful?
- Could you share some screenshots of the - from your point of view - successful or unsuccessful - visualizations that you have created as part of the course?

4. Do you remember the first visualization you created as part of the course?
   - If so, can you describe it (do you have a screenshot)?
   - Do you remember your experience when you first saw it?
   - Did it match your expectations?
   - What did you take away from it?
5. Can you describe some strategies that you tried as part of the course to explore and visualize your data?
   - What strategies worked for you and why?
   - What strategies did not work and why?
   - Do you feel you changed your approach/exploration to your data exploration throughout the course?
6. What did you find the most challenging about the course and the visualization processes we explored as part of this?
   - can you give examples?
   - Did you feel the exercises in the course (e.g., sketching) helped address these challenges? If so, in what way?

A.1.2 Individual Features of the Vistorian

1. What features of the Vistorian did you find particularly useful and why?
2. What features of the Vistorian did you find particularly difficult to use? Please give examples.
3. What additional features would you like to see added to the Vistorian?
4. What are the most useful visualizations in your case?

A.1.3 Use of the Vistorian in Comparison to Other Tools

1. When starting the course, have you already worked with other visualization/data analysis tools or programming languages? If so, can you please name them? (to work on the data you have focused on in the course or also for other data sets)

2. How would you describe your experience with the Vistorian compared to these other tools?

    - In terms of formatting and loading your data into the tool
    - In terms of specifying how to visualize your data
    - In terms of general interaction and exploration techniques

A.2 Participants Profiles

Here, we describe participants' profiles, including their background, experience in network visualization, and their dataset domain. The brackets next to each participant first describe the participant's goal, followed by the type of journey his exploration has taken through the course. Goals and journeys types are described in the paper.

- **P-SupplyChain (G1,J1):** comes from a background in social sciences, and her dataset is in supply chain management. This is their first time exploring a network, and they are in a phase between collecting data and initiating the exploration as well. They are interested in understanding how innovation is diffused across a certain supply network and in exploring what type of link types they might have in their dataset.

- **P-UrbanStudies(G1,J1):** is an engineer with a background in social sciences. They are interested in social network analysis and exploration. They don't have a dataset to explore but are willing to learn about the concept and tool.

- **P-Letters (G3,J2):** comes from the humanities and history background. They have built a few networks using Gephi and Excel. Their dataset is based on a collection of ancient documents and letters. They have a research question and are interested in understanding the structure of their network and how to visually explore it.

- **P-Carers (G3,J2):** is a doctor and health researcher. They have collected interviews and would like to answer the question of who talks to whom and how this affects the care network. They are looking to compare different layers of networks to have a better understanding of how different categories in the network affect the care plan.

- **P-Dance (G2,J3):** comes from history and computer science background. They are working on dancing programs dataset which they explored partially before the course. They aimed to create a network that shows which dancers joined the same program.
- **P-Documents (G2,J3):** is a graduate student in computer science, researching Natural Language Processing. They are interested in the association between labels and documents and the semantic relationship between labels.
- **P-Laureates (G2,J3):** is a social scientist with a dataset about the recipients of prestigious prizes and awards. They are interested in finding patterns in their network.
- **P-Migrants (G2,J3):** comes from a background in humanities and history. They had no experience in network visualization. Their data consisted of migrants in a certain era and they were interested in understanding the movement of migrants to a specific city.
- **P-RocksStudies (G2,J3):** comes from a background in natural sciences. They have experience in network visualization. Their dataset is about studies and methodologies that describe a specific phenomenon related to rocks.
- **P-ElderlyCare (G3,J3):** comes from humanities and computer science background. They are interested in exploring the care network of elderly people as they have their research question.
- **P-Artifacts (G3,J3):** comes from a history and archeology background. They are interested in Archaeozoology, which is the study of animal bones from archaeological sites. They are looking to explore certain types of bones and which specific geographic area they belong to through the flow of different chronological periods.
- **P-Festivals (G3,J3):** background is in social sciences and business. They have worked with ArcGIS, UCINET, and Krackplot to explore a few networks. They are interested in exploring how festivals impact the growth of certain areas locally and globally along with other goals. They were interested in visualizing their data, and playing with their networks using different attributes.
- **P-Criminology (G3,J3):** comes from a social sciences background. Their dataset is in criminology and they are interested in how to visually explore as they have their network ready.
- **P-Agriculture (G3,J3):** comes from background agriculture and veterinary science. They have their dataset and research question ready and would like to learn how to explore the data visually.

- **P-Archaeo (G3,J3):** comes from a history and archeology background. They are interested in Archaeozoology, which is the study of animal bones from archaeological sites. They are looking to explore certain types of bones and which specific geographic area they belong to through the flow of different chronological periods.

- **P-Business (G3,J3):** comes from a computer science, social sciences, and humanities background. Compared to the other participants, their skills are relatively advanced, and they have experience with various visualization tools. Their dataset is in commercial businesses. They are interested in exploring and understanding how the network has changed over time through different types of nodes.

A.3 The Story of a Journey

P-Business (J3) Journey: It's all about Data P-Business comes from a computer science, social sciences, and humanities background. Compared to the other participants, their skills are relatively advanced, and they have experience with various network visualization tools. Their main focus is to explore how funding rounds assist in the growth of industries and business networks, connect people to each other, where they are located on the map, and what kind of connections they form. P-Business found said: "I thought the data was ready, but it turned out that it is not ready at all." They played with a sample of their dataset throughout the course to understand more about what they needed for a better exploration.

They had two possible models in mind where different mutations of practitioners and funding rounds take place as nodes or links. As the Vistorian is easy to use, P-Business decided to try both and compare. P-Business: the first model connects people to the funding round, and the second connects people in the same round to each other. However, P-Business shows an advanced awareness of the data formation process as they go. That is, P-Business decided to ignore the altered model as connecting all people in the same round to each other might not be the ideal case, as being in the same round does not always mean they are connected or know each other. Thus, they've decided to connect practitioners to funding rounds. But at the same time, they emphasized that this might cause a problem as well in case they wish to use the map due to where they will place funding rounds on a map. They found that the adjacency matrix provided an unusual experience and was interesting as it helped to highlight some of the questions they had through the degree of connectivity. For example, by removing funding rounds, it presented some interesting insights as it decreased the dimensionality.

P-Business said, "Getting the data prepared to analyze visually helped me realize things about the data that I didn't know. It was less about the actual visualizations at the end. It was a lot more about looking at the types of structures we had and thinking about the process of

making the decision that I would have to actually come up with an interesting visualization. Those pieces gave me some realizations and also some takeaway messages for my team to say, hey, we need the data to be set up like this." P-Business is interested in exploring data through time and location to explore data from different perspectives, such as how people's connections lasted with the funding organization in the long term. The challenge besides data is to maintain persistent representations of colors across all different types of visualizations if they would like to have a specific node type with the same color to ensure reliability in representing consistent results encodings. In addition, how to share visualizations with others, maintain privacy considerations, and validate data. P-Business is also looking to be able to adjust values from both the data view and visualizations view. Also, they are looking for assistance on the data side, such as how to reduce data volume and what features to choose are the more valuable parts. "There is a tradeoff between the tool's simplicity and the level of control granularity in visualizing a network," they say, and the selling point of the Vistorian is its simplicity.

A.4 6-Weeks Course Outline

In this section, you can find the outline of the topics covered weekly in our in-depth network visual exploration course (see Table A.1). The first three weeks covered data preparation and the network formation process. This included creating node/link tables and associated data-wrangling operations. In the second part, we discussed how to read and interpret four types of network visualizations: node-link diagram, adjacency matrix, network over map, and timeline.

Week	Topic	Outline
1st	Network Data Preparation	- Data set selection for Hands-on Activities - What is a Network? - What is Network Exploration? - Role of Visualizations in Network Exploration - Sketching your network - Network Concept Map - Creating Node and/or Link Tables - How can you format your network data?
2nd	Extra Session on Data Preparation	- Discussion on Challenges Faced from Concept Map and Network Tables - Examples on Company Network - Re-practice Activities on Creating Network Concept Maps and Tables
3rd	Data Shaping Techniques and Challenges	- Discussion on Challenges Faced from Concept Map and Network Tables - Examples on Visual Network Exploration - Ensuring Consistency of your Data - Common Challenges in Network Visualizations - Assistive Techniques in Shaping your Data
4th	Exploring with Node-link Diagrams	- What is visual network exploration? - How to close-reading a network diagram (visual patterns) - Asking questions to your data (questions + hypotheses) - Performing interactive exploration (strategies)
5th	Exploring with Adjacency Matrix	- Sparse vs. Dense Networks - What is an Adjacency Matrix? - Visual Encodings - Ordering Algorithms - Visual Patterns
6th	Exploring with Timeline, Map, and Coordinated View	- Timeline Visualization - Map over network - Using Coordinated Views in Exploring Networks - Group Exercise

Table A.1: Network Visual Exploration Course Topics Per Week

Appendix B

Analytics Dashboard Pages

Data used in the dashboard was last updated on 20/Apr/2023. Click here for dashboard demo video

B.1 Overview Page

Overview page (Figure B.1) uses three stratified rows to report overall trends and usage in the Vistorian as following:

- **Top bar** navigates three pages of the dashboard.
- **Session calculations (a): The blue shadings represent users' level of advancement, i.e., demo users are light users, while multiple session explorers are expert users of the Vistorian tool.**
  - (a1) The number of four types of user sessions counted,
  - (a2) User session monthly distribution chart with an average line,
  - (a3) Box plot of average session time spent across the four types of users,
  - (a4) Pie chart representing the number of return and non-return users with a line chart of its yearly ratio.
- **Vistorian Logging Timeline (b):** A combination of a weekly bar chart of total users and a line chart of four types of users, with each educational effort annotated as a colored box below the timeline. *Interactively* clickable line charts and tooltips for all data points.
- Feature usage **(c)**:
  - (c1) The count of key features used by explorers (single-session and multi-session explorers)

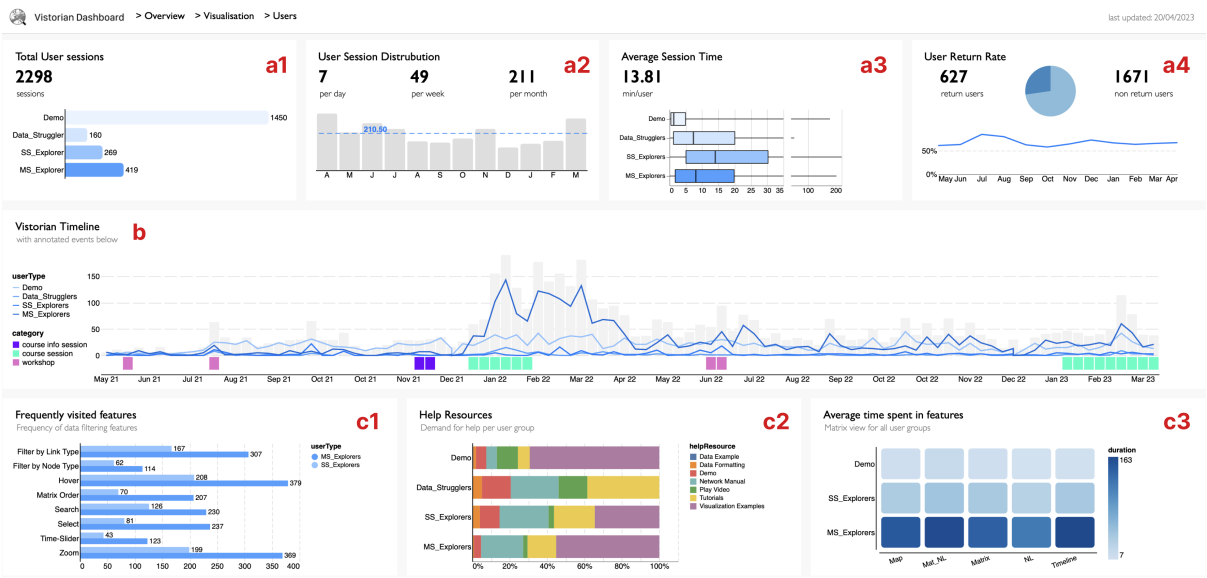


Figure B.1: Overview Page in Vistorian Monitoring Dashboard

- (c2) The percentage of different help resources used for each type of user. Help resources include data examples, data formatting, demos, network manual, play videos, tutorials, and visualization examples.
- (c3) Matrix showing duration for each visualization view explored by different user types.

B.2 Visualizations Page

Page Visualization (Figure B.2) uses a five-column table covering five views provided in the Vistorian, including node-link diagram, matrix diagram, timeline arches, map view and a coordinated view of node-link and matrix.

- (d1) The number of users for each view in the recent 12 months.
- (d2) Average time spent in each view among all users that explored this view in the recent 12 months.
- (d3, d4) The frequency of features used in each visualization view, categorized by ‘data filtering’ features and ‘visual representation’ features. For example, the time slider filters out data belonging to the data filtering group, while the labeling feature only changes the visual representation.

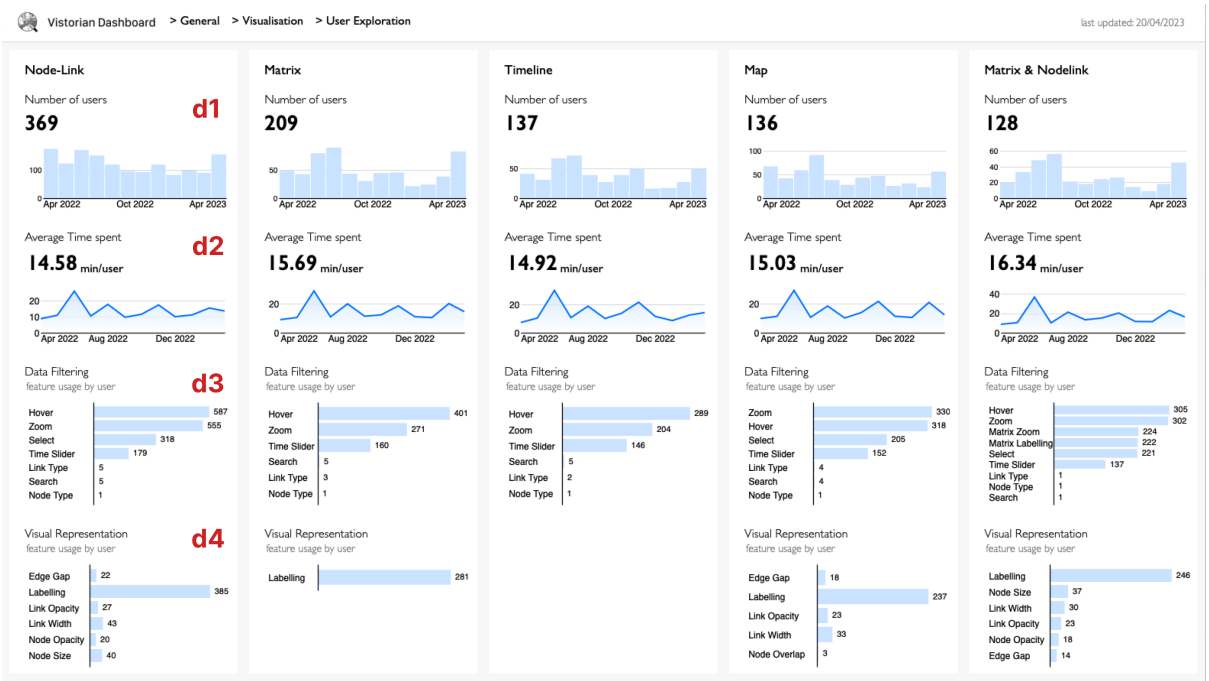


Figure B.2: Visualizations Page in the Vistorian Monitoring Dashboard

B.3 User Exploration Page

As an example, we discuss the multi-session user's timeline.

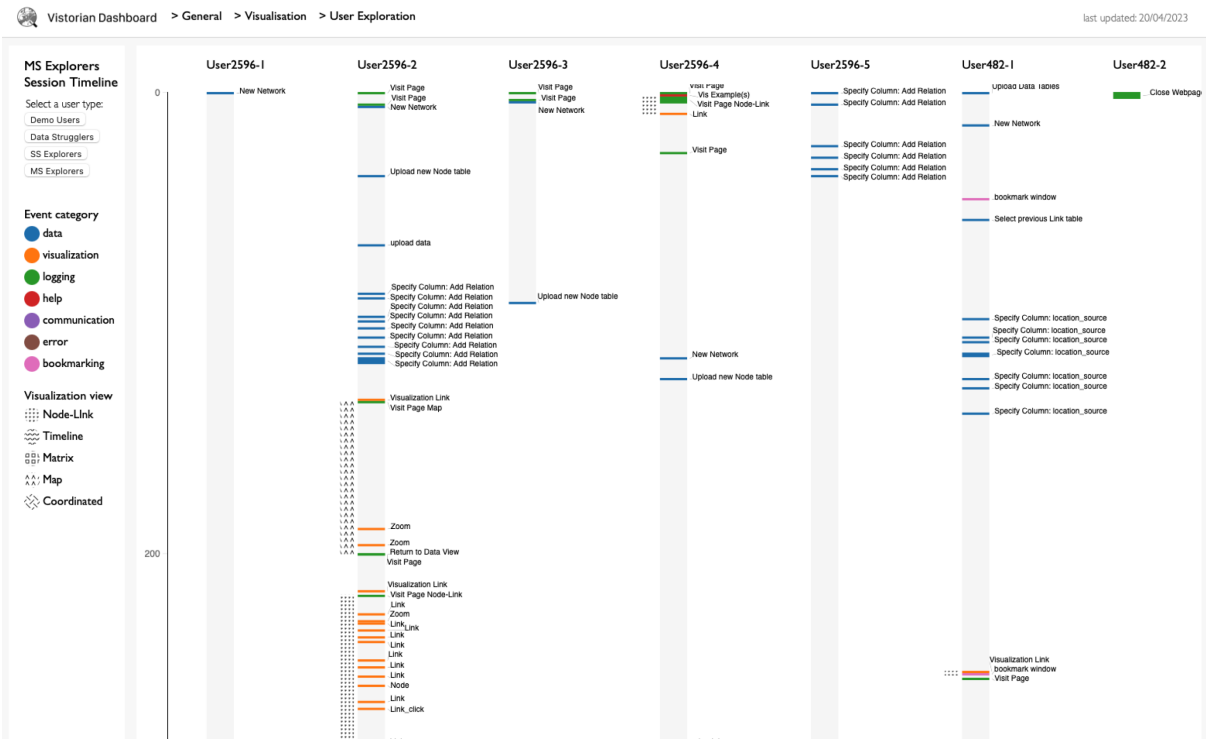


Figure B.3: User Exploration Page in the Vistorian Monitoring Dashboard

Page User visualizes interaction sequences for individual users grouped by their categories.

- Select the timelines of a particular user type by the top-left button group,
- Events are color-coded by their log categories,
- Textures represent the visualization view the user is in,

Interactivity Select event color legends to highlight particular categories of events as shown below.

B.3.0.0.1 Timeline for Multi-session Explorers Observed dense interaction histories in data management and visualization. As an example, we can see in Figure B.4 partial screenshots from the user 2596 sessions which we can read as follows:

- After creating the network in the (1) session, the user initiated the exploration process with a node-link diagram. In session (2), the user switched to a network over a map visualization. Then, the user explored a network visualization of a node-link diagram zooming and exploring different links. The sequence shows a behavior many users follow, starting with a node-link diagram and then branching to others.
- In the 3rd session, the user started creating a network by uploading a node table but didn't finalize it. In the following sessions, the user explored an example, created a network and explored it using a node-link, a matrix, and ending with a timeline.
- Although we didn't see any interaction with the time-slider, the timeline shows an interest in temporal exploration.

B.3.0.0.2 Timeline for Demo Users : Comparatively short session length, No sign of success in data management or their own data visualizations, Frequently observed help resource usage.

B.3.0.0.3 Timeline for Data-Strugglers : Multiple trials with data management events, with help resources used frequently, No sign of success in creating networks, Spend most of the time in data management.

B.3.0.0.4 Timeline for Single-session Explorer : Each session contains first data events and then visualization events, Some users show a long period of visual exploration over several views while some users end the exploration quickly.

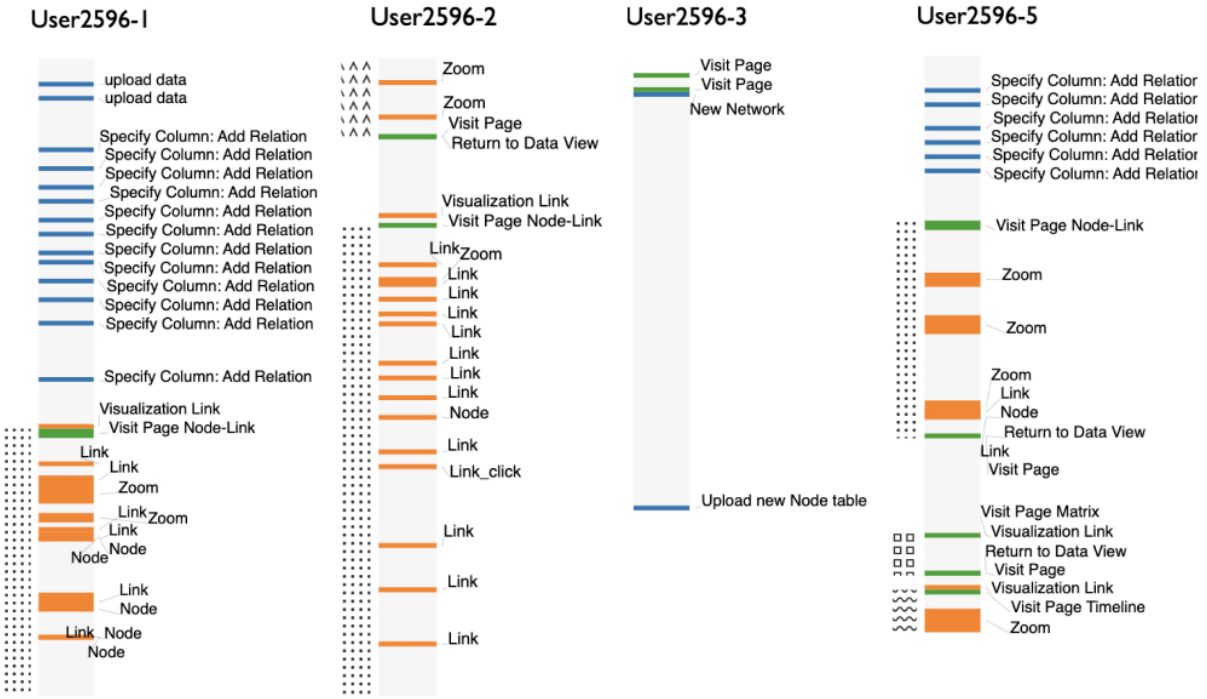


Figure B.4: An Example of Multi-Session User Exploration Sessions: User Scenario (User 2596)

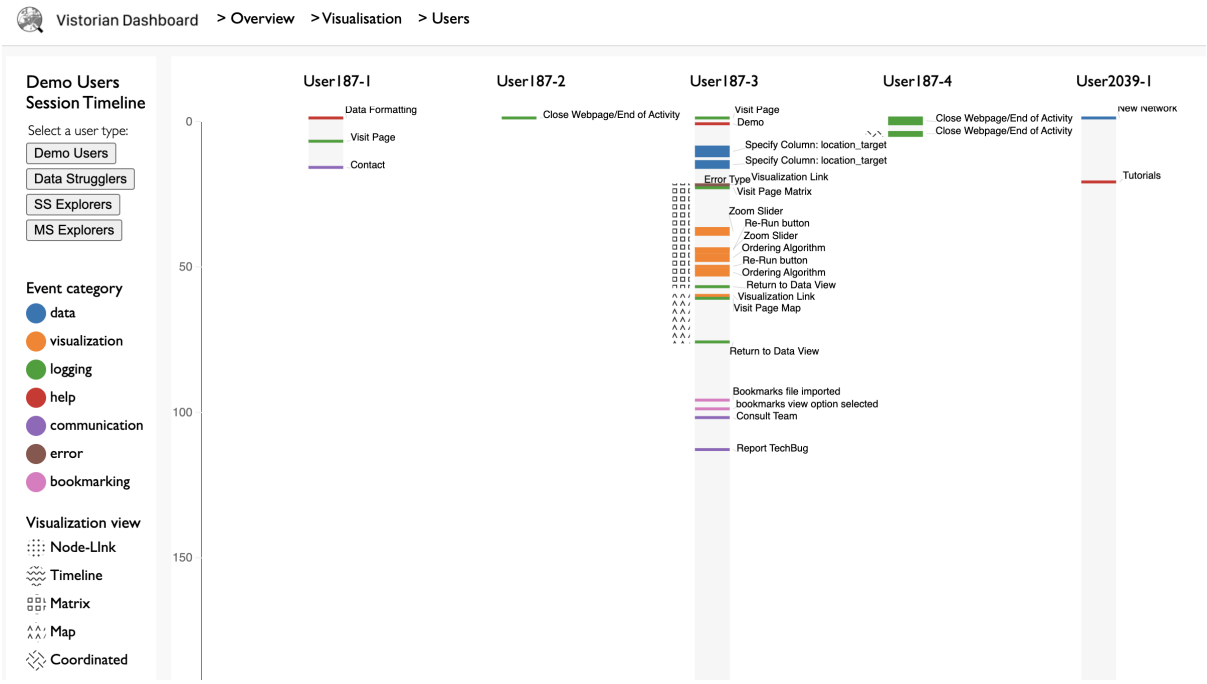


Figure B.5: An Example of Demo User Exploration Sessions

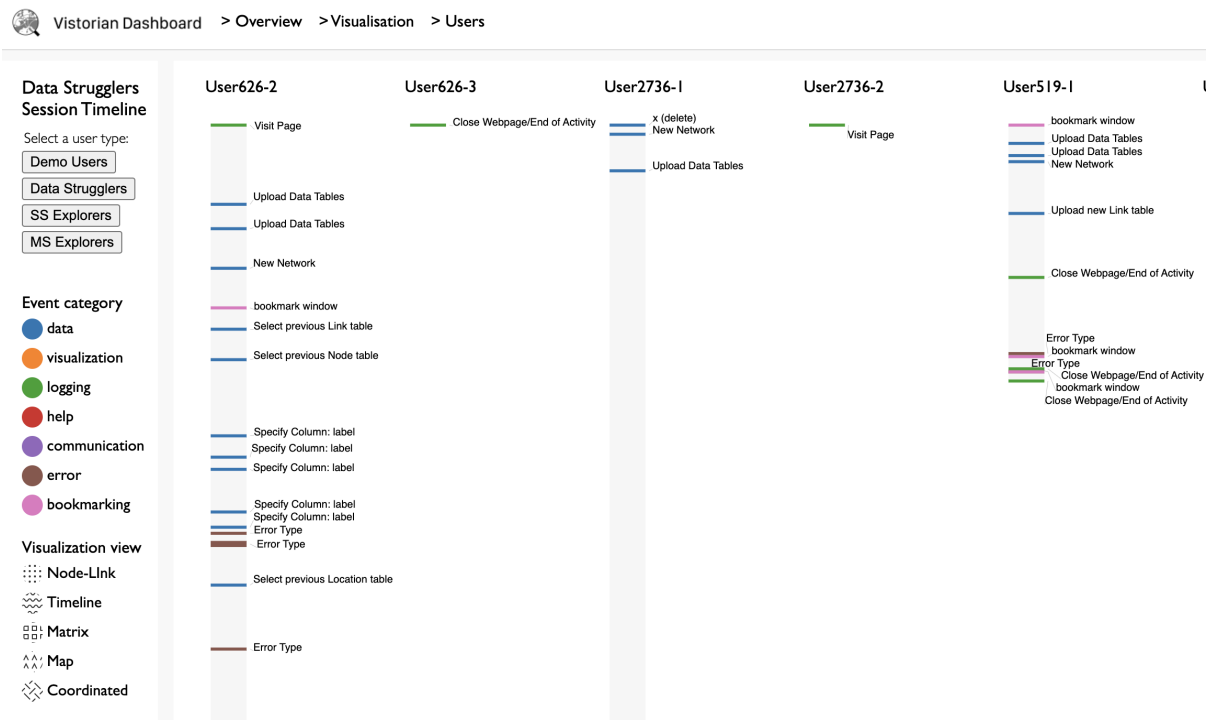


Figure B.6: An Example of Data-Strugglers Exploration Sessions

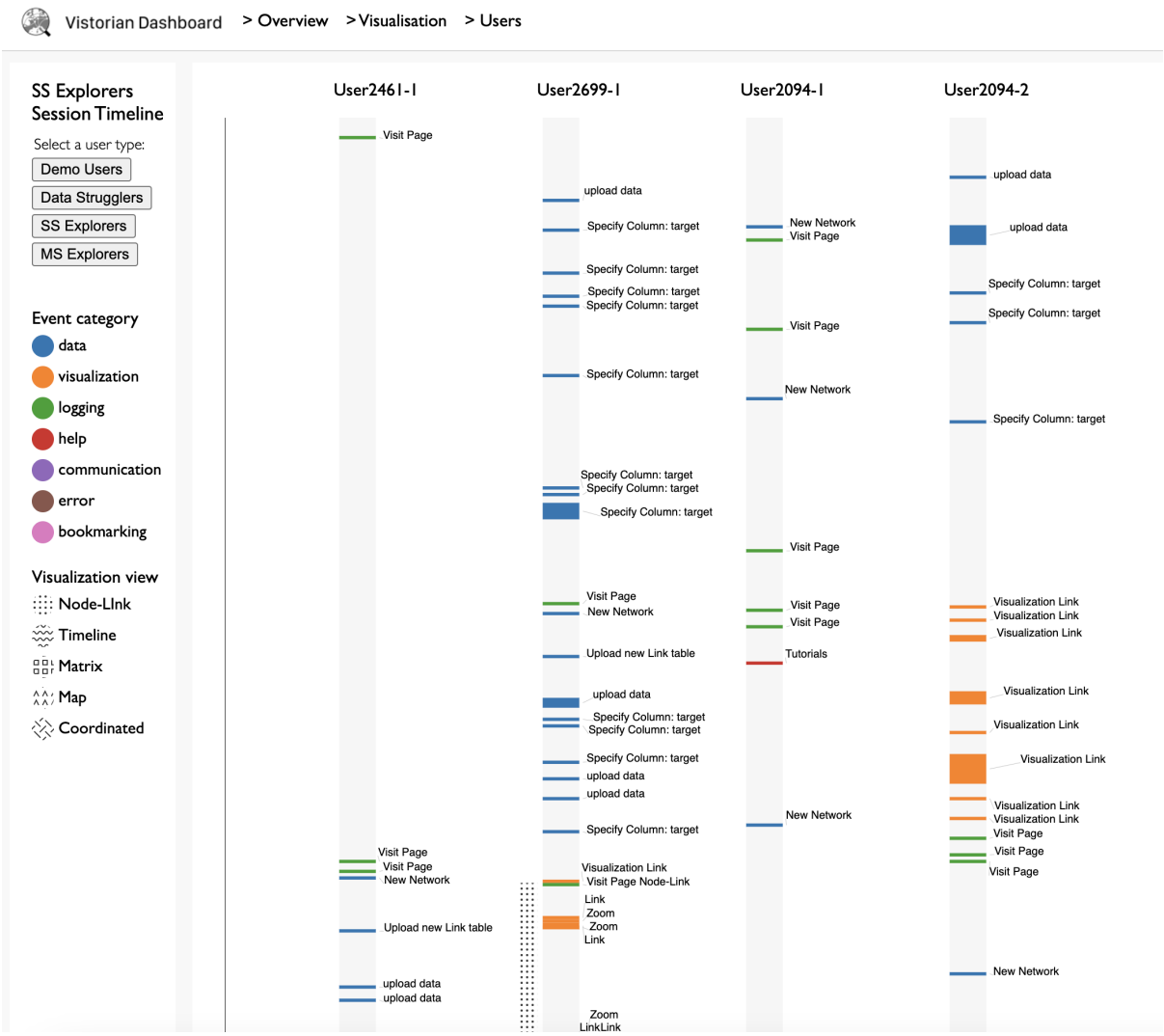


Figure B.7: An Example of Single-Session Exploration Sessions

Appendix C

Resources of Networks Visual Exploration Roadmap

Roadmap Steps & Activities The coaching program material contains the following:

- (section C.1) What is Networks Visual Exploration?
- (section C.2) Understanding your Data and Sketching Concept Maps
- (section C.3) The Selection and Description of Networks Visual Exploration Strategies
- (section C.4) From Concept Maps to Network Schema
- (section C.5) Which visualization to read and How?
- (section C.6) Logging & Validating your Observations
- (section C.7) Data Transformation & Wrangling

C.1 What is Networks Visual Exploration?

10/12/2023 R

Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

Interactive ***Network Visual Exploration***



Visual Network Exploration

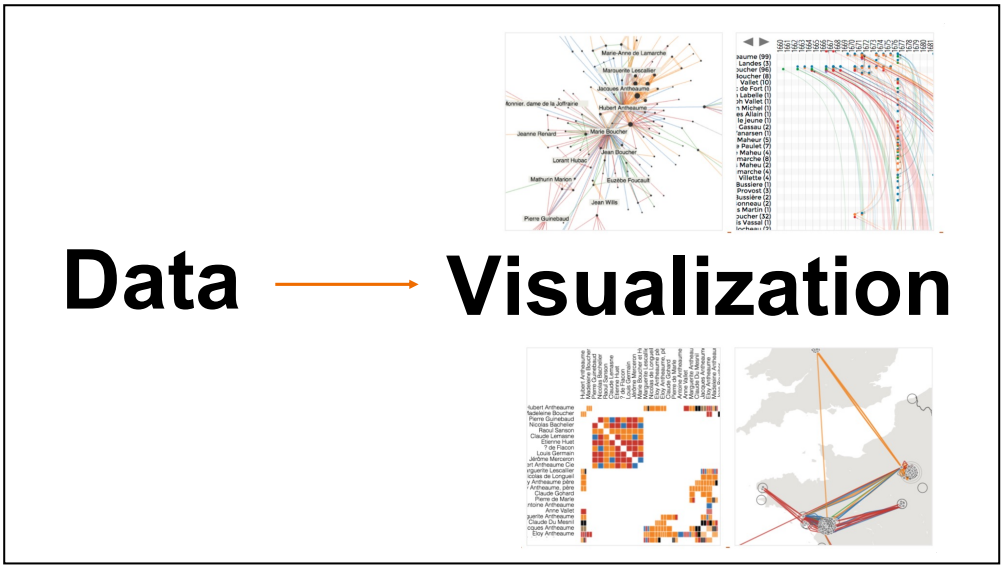
Coaching Program

1

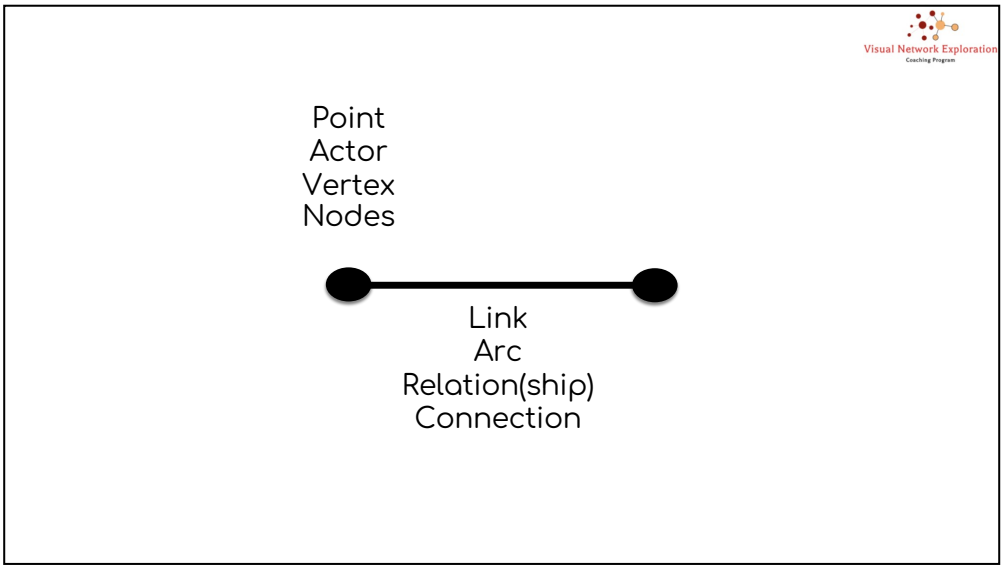
What is a **Network Visualization**?

2

10/12/2023 R

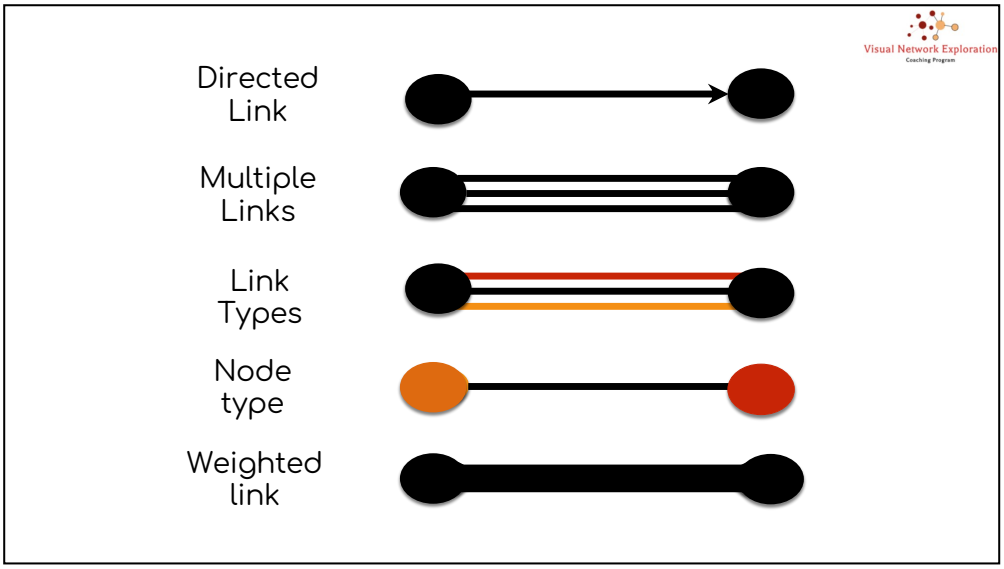


3

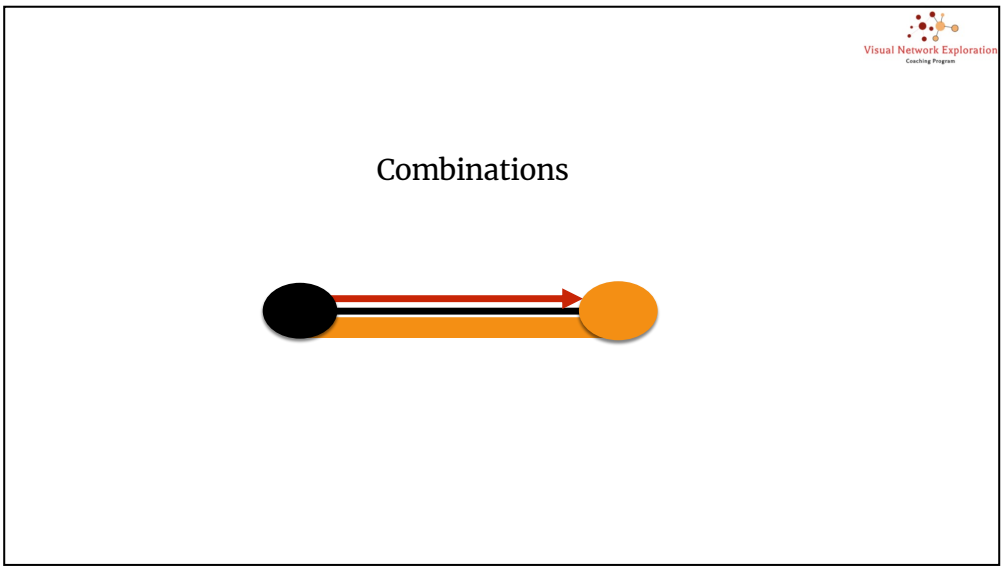


4

10/12/2023 R



5



6

10/12/2023 R

Network Visualizations: Node-Link Diagram



7

What is an Interactive **Network Visual Exploration**?

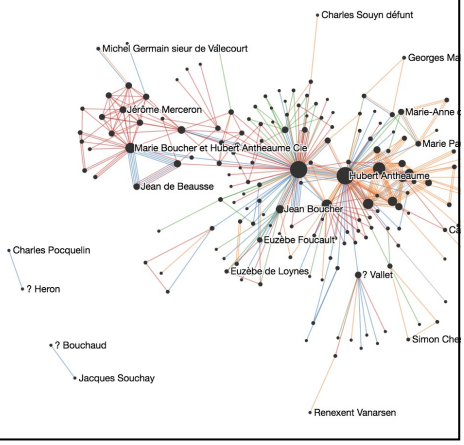
8

10/12/2023 R

What is Networks Visualizations Exploration?

Visual Network Exploration Coaching Program

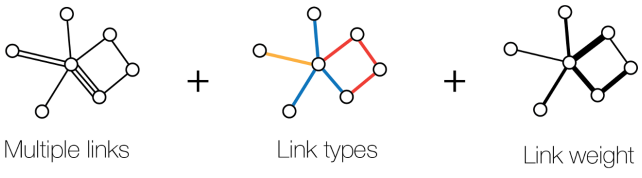
1. Describing **visual patterns**
2. Asking **questions**
3. Generating **hypotheses**



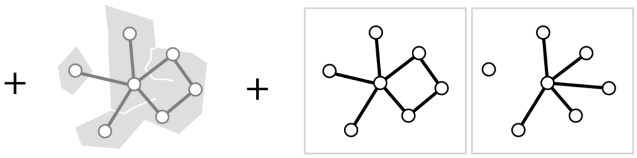
9

Topology + Attributes

Visual Network Exploration Coaching Program



Multiple links + Link types + Link weight

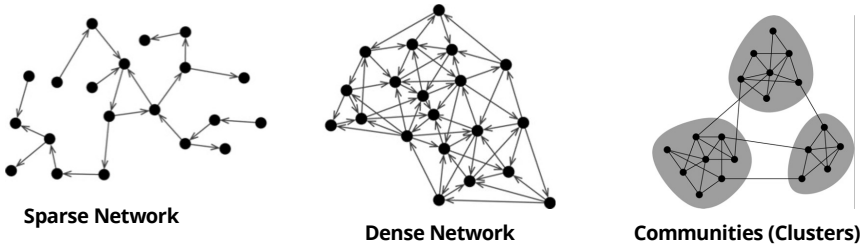


Geography + Changes over time

10

10/12/2023 R

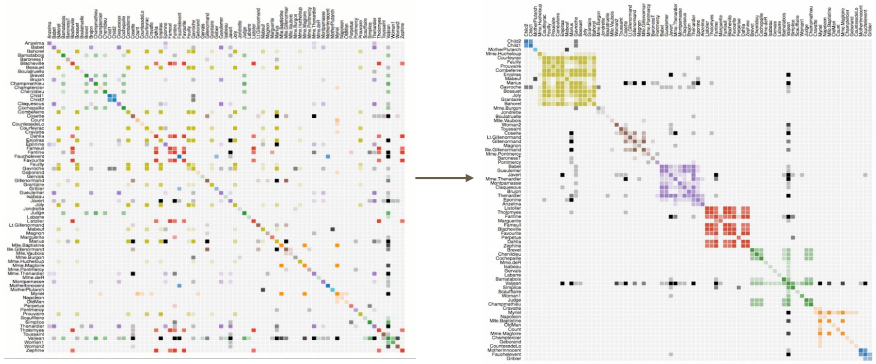
Example of Visual Patterns: Network Structure



11

Different Visualizations and Views:

Adjacency Matrix



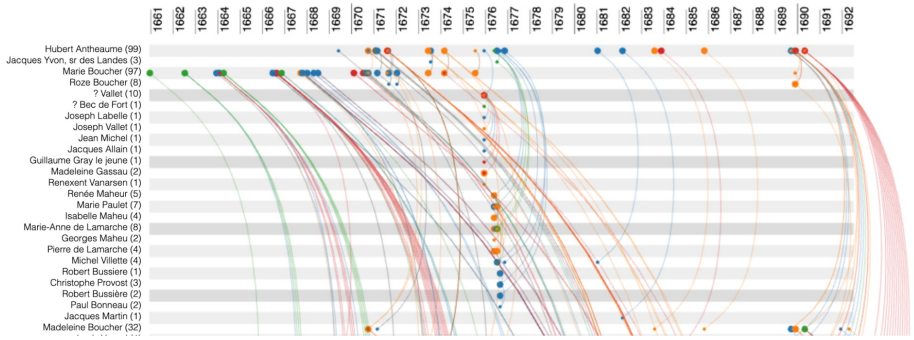
12

10/12/2023 R

Different Visualizations and Views

Visual Network Exploration Coaching Program

Timeline

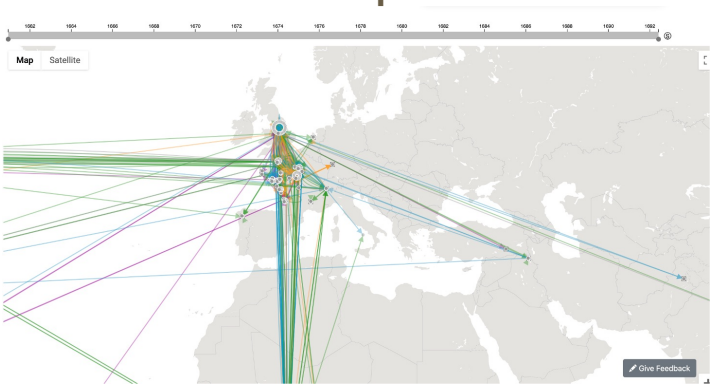


13

Different Visualizations and Views:

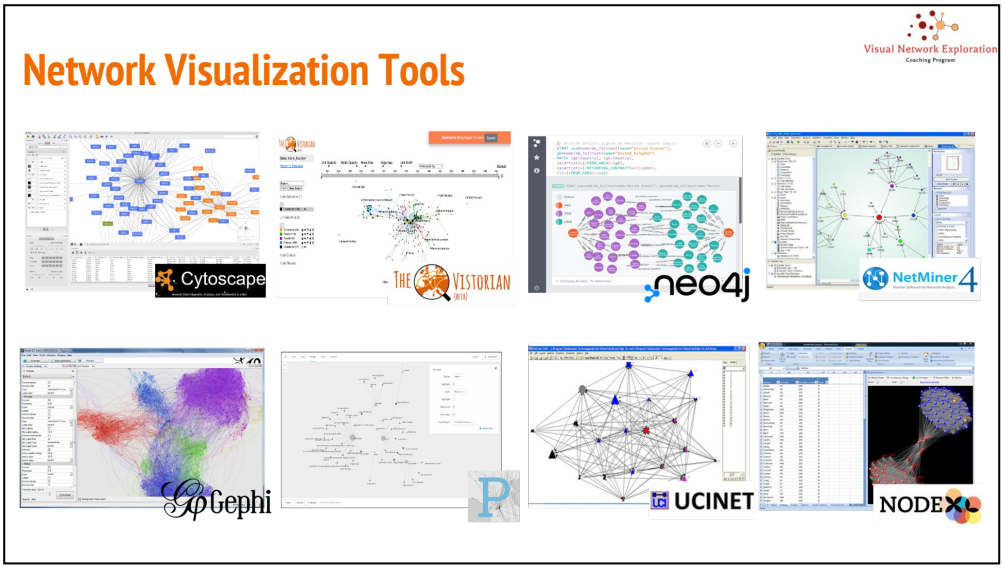
Visual Network Exploration Coaching Program

Map

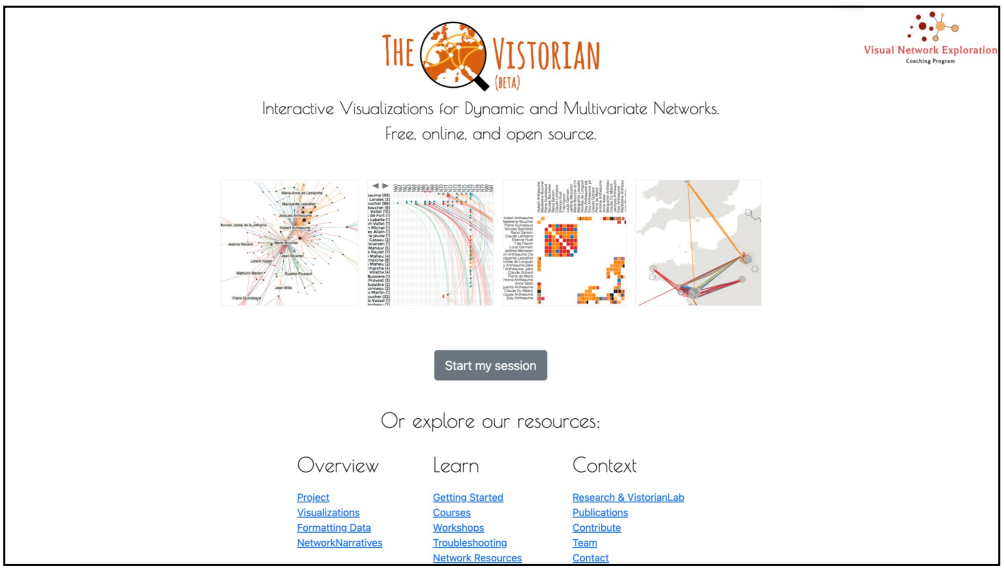


14

10/12/2023 R



15



16

C.2 Concept Map Sketching

C.2.1 Understanding your Data and Sketching Concept Maps (Starting from Data)

12/10/23

Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

Understanding your Data and *Sketching* **Concept Maps** (Starting from **Data**)



Visual Network Exploration

Coaching Program

Brief Intro to Network Sketching
(*Concept Maps)*

12/10/23

What do Networks consist of?



Nodes describes concepts (nouns) such as entities, events, people, subjects, ..

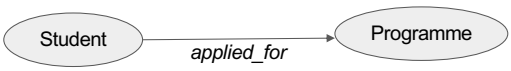


Links describe relationships between nodes (actions or verbs).
Links can be:

- directed → : where direction of the relationship is important
  (ex. James sent a letter to Mia)
  OR
- undirected — : where direction is not important in a relationship
  (ex. Sam is friend of James)

What do Networks consist of?

As an example: a network shows **Students** *applied for* a university **Programmes**. Then we can describe it as:
node *verb* *node*



What about if I have ***additional data attributes*** the describes ***our entities(nodes) and relationships(links)***?

For example:

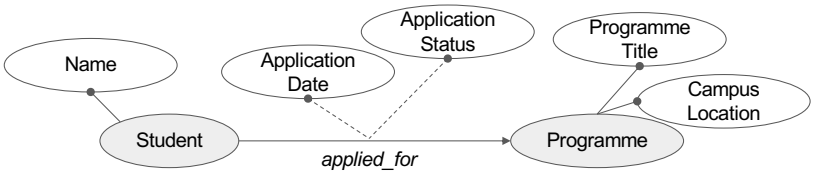
- *The Programme* has data attributes such as (Programme Title, Campus Location),
- *Student* has the data attribute(Name),
- and the *application*'s data attributes are(application date, status) showing whether it was accepted or not.

12/10/23

What do Networks consist of?

Data attributes can be added to describe nodes or links further. So, based on the example given will add the following:

- (Programme Title, Campus Location) to the **Programme** node,
- (Name) to the **Student** node,
- and (application date, status) will be added to the link **applied_for** as it describes the relationship.



Describe your Data

Concept Mapping

Describing your data through Concept Mapping

Describing data through **'Concept Maps'** enables us to:

- Identify role-playing **entities (nodes)** in your data
  ⇒ + what additional data *describes your nodes*
  (ex. Location of each node)
- Understand how those nodes are linked through what we call **(links)**
  ⇒ + what additional data describes such links (ex. Link type)

Concept Maps reveals: the **semantics** of your network + underlying **structure** of your **data**

12/10/23

How can I start *Concept Mapping?*

Visual Network Exploration Coaching Program

How can I start creating my concept map?

You can start creating your concept map based on having and/or :

Data
(ex. Formatted dataset(s), qualitative data to be formatted,..)

Describe all data using a Concept Map → generate possible question your concept map can answer

A research goal
that you aim to achieve
(ex. Answering research question, comparing 2+ datasets, ..)

Gradually build your concept map, starting from a single research goal → add more elements to your concept map as you.

12/10/23

Starting with the Data you have

- As a start, you can:
  - Exclude attributes with high number of null values or values that are unknown.
  - Start with identifying main *concepts as nodes* and *relationships as links* from data attributes.
  - Identify the data attributes that describes:
    - a **node** in a *persistent nature* (ex. Name of a person)
    - from those describing a **link's** changing nature (ex. Date of requesting a service)

Starting with the Data you have

Based on what you start from, your concept map will be created as a whole from the beginning or gradually as you go. So, for example:

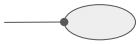
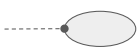
- If starting from data ⇒ Include all data attributes to create the whole concept map of your dataset, then generate possible research question.
- If starting from research goal(s) ⇒ Start creating a concept map for a single research goal, and add more to it gradually as you process a single research goal at a time.

12/10/23

Sketching *Concept Maps*
Symbols & Notions

Visual Network Exploration Coaching Program

Symbols of Concept Map Sketching

Symbol	Name	Description	Can be used as (network table)
	Concept	Represents any concept (ex.entities, events, persons)	Nodes
	Directed Relationship	Relationship that is an associations between or among concepts where direction is important	Directed Link
	Undirected Relationship	Relationship that is an associations between or among concepts where direction isn't important.	Undirected Link
	Concept Attribute	An attribute that describes a concept (ex. Name of a person, specialization of an employee)	Node / Node Type
	Relationship Attribute	An attribute that describes a relationship between concepts (ex. a date of when a letter was sent).	Link Type / Link Weight / Link ID

12/10/23

How to Describe and Sketch your Concept Map?

Each data attribute in your dataset can participate in describing your network's semantics and structure.

Which role an attribute can play, describe or represent in a network?

- A candidate node/link value/type.
- ***Location of*** node/link: location of where a node/link was created.
- ***Date of*** node/link: date where a node/link was created.
- ***Type of*** a of node/link: a set of values that can be used to distinguish different node/link types.
- ***Direction*** of a link: show how a link was created from which source node toward which target node.
- ***Weight*** of a link: numeric value that is reflected on the width of a link.

Examples on Sketching *Concept Maps* ***Starting from your Data***

12/10/23

Visual Network Exploration Coaching Program

Starting with the Data (Example 1: *Letters Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

What is this dataset about?

In our first sketching example, our dataset is about a set of letters and documents exchanged by a trader called Marie Boucher and the goal is to explore her trading network.

	Sender Name	Sender Place	Qualification	Receiver Name	Receiver Place	Date Letter Sent
1	Hubert Antheaun	St Nicolas, Nante	Commerce	Jacques Yvon, sr de	Logane, Ile St	03/08/1673
2	Marie Boucher	St Nicolas, Nante	Commerce	Hubert Antheaume	St Nicolas, N	06/08/1675
3	Marie Boucher	St Nicolas, Nante	Famille	Hubert Antheaume	St Nicolas, N	06/08/1675
4	Marie Boucher	St Nicolas, Nante	Famille	Roze Boucher	St Nicolas, N	15/09/1671
5	Marie Boucher	Paris	Commerce	Roze Boucher	St Nicolas, N	15/09/1671
6	Marie Boucher	Paris	Commerce	Hubert Antheaume	Paris	15/09/1671
7	? Vallet	Ile St Domingue	Travail	? Bec de Fort		01/01/1676
8	? Vallet	Ile St Domingue	Commerce	Joseph Labelle	Logane, Ile St	01/01/1676
9	? Vallet	Ile St Domingue	Famille	? Vallet	Ile Guadelou	01/01/1676
10	? Vallet	Ile St Domingue	Famille	Joseph Vallet	Brest	01/01/1676

Visual Network Exploration Coaching Program

Starting with the Data (Example 1: *Letters Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

Data's Nature:

- Names
- Places (*locations)*
- Communication *dates*
- Type of qualification

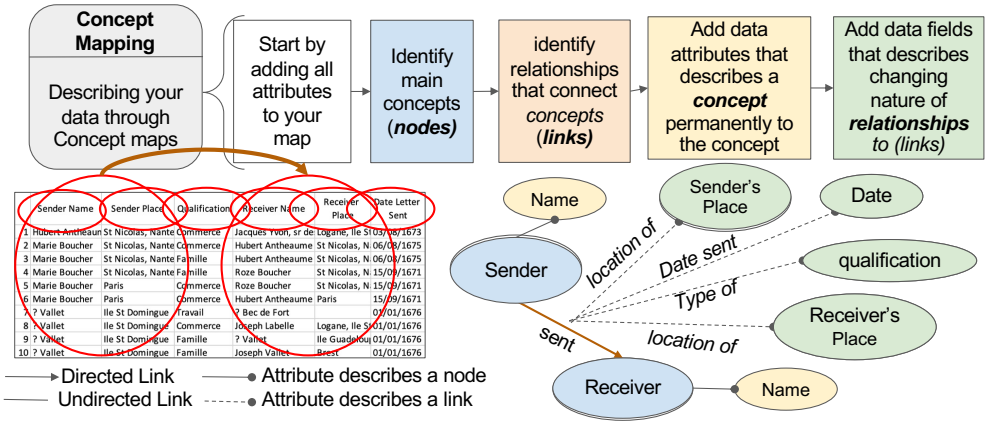
What questions the data can answer?

- Who contacted whom? And in which context (e.g. commerce)?
- When did most communication take place?
- Where did most of the messaging happened?
- Where most of the traders were?
- How does the network of 'family' connection differ from network of 'commerce' in matter of structure? Or density?
- And much more ..

	Sender Name	Sender Place	Qualification	Receiver Name	Receiver Place	Date Letter Sent
1	Hubert Antheaun	St Nicolas, Nante	Commerce	Jacques Yvon, sr de	Logane, Ile St	03/08/1673
2	Marie Boucher	St Nicolas, Nante	Commerce	Hubert Antheaume	St Nicolas, N	06/08/1675
3	Marie Boucher	St Nicolas, Nante	Famille	Hubert Antheaume	St Nicolas, N	06/08/1675
4	Marie Boucher	St Nicolas, Nante	Famille	Roze Boucher	St Nicolas, N	15/09/1671
5	Marie Boucher	Paris	Commerce	Roze Boucher	St Nicolas, N	15/09/1671
6	Marie Boucher	Paris	Commerce	Hubert Antheaume	Paris	15/09/1671
7	? Vallet	Ile St Domingue	Travail	? Bec de Fort		01/01/1676
8	? Vallet	Ile St Domingue	Commerce	Joseph Labelle	Logane, Ile St	01/01/1676
9	? Vallet	Ile St Domingue	Famille	? Vallet	Ile Guadelou	01/01/1676
10	? Vallet	Ile St Domingue	Famille	Joseph Vallet	Brest	01/01/1676

12/10/23

Concept Map Example 1: *Letters Network*



Starting with the Data (Example 2: *Connections Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

What is this dataset about?

In our second sketching example, our dataset is about a family members and best friend of number of people. And we would like to explore their relationships network and if they intersect.

Name	Father	Mother	Grandfather	Grandmother	Friend
Bob	Martin	Ana	Sam	Sara	James
Ali	Jamal	Suzen	Kamal	Sara	Mia
James	Sami	Poppy	Kai	Sangeeta	Layla
Mia	Mo	Ann	Roy	Summer	Pri

12/10/23

Starting with the Data (Example 2: *Connections Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

Data's Nature:

- Names
- Family connections (*type of connection*)
- Friends names

What questions the data can answer?

- Who share same grandfather/grandmother/..?
- Who are friends of ex. James or Mia?
- Who has the highest number of connections?
- Who is a role-playing node in this network? (ex. Connect high number of people together)
- Who has no connections to the network? (ex. isolated)
- And much more ..

Name	Father	Mother	Grandfather	Grandmother	Friend
Bob	Martin	Ana	Sam	Sara	James
Ali	Jamal	Suzen	Kamal	Sara	Mia
James	Sami	Poppy	Kai	Sangeeta	Layla
Mia	Mo	Ann	Roy	Summer	Pri

Concept Map Example 2: *Connections Network*

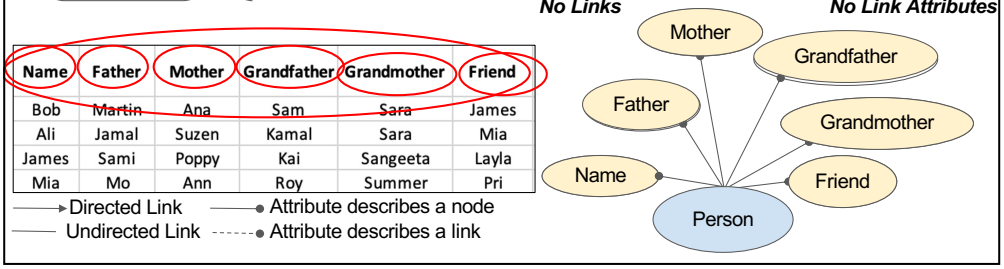
Concept Mapping

Describing your data through Concept maps

Start by adding all attributes to your map → Identify main concepts (***nodes***) → identify relationships that connect *concepts* (***links***) → Add data attributes that describes a ***concept*** permanently to the concept → Add data fields that describes changing nature of ***relationships*** *to (links)*

No Links ***No Link Attributes***

Name	Father	Mother	Grandfather	Grandmother	Friend
Bob	Martin	Ana	Sam	Sara	James
Ali	Jamal	Suzen	Kamal	Sara	Mia
James	Sami	Poppy	Kai	Sangeeta	Layla
Mia	Mo	Ann	Roy	Summer	Pri



Directed Link — Attribute describes a node
Undirected Link ------ Attribute describes a link

12/10/23

Starting with the Data (Example 3: *Job Applications Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

What is this dataset about?

In our third sketching example, our dataset is about a number of graduates who applied for opening jobs . Each applicant can have 1-3 majors under the their background umbrella. Each job application can be accepted/rejected.

Applicant Name	Background	Major 1	Major 2	Major 3	Latest Degree	Graduation Date	Job Title	Salary	Application Status
Alice	Science	Chemistry	Physics	NA	MSc	02/03/2021	Researcher	3000	Accepted
John	Science	Management	Business	Finance	BSc	07/04/2019	Audit	2500	Accepted
Lili	Science	Architecture	Design	NA	BSc	05/02/2022	Project Manager	3291	Rejected
Jerry	Science	Computer Science	Computer Engineering	Business	MBA	05/01/2022	Product Manager	3429	Accepted
Mia	Art	Art	History	NA	BA	01/09/2020	Researcher	2931	Accepted
James	Art	Humanities	Social Sciences	NA	MA	02/03/2022	Researcher	3292	Rejected
Samy	Art	Law	NA	NA	PhD	01/02/2023	Lawyer	3434	Accepted

Visual Network Exploration Coaching Program

Starting with the Data (Example 3: *Job Applications Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

Data's Nature:

- Names
- Background & majors (*types*)
- Graduation (*date*)
- Salary (*numeric*)
- Degree (*types*)
- Application status (*types*)

What questions the data can answer?

- Which jobs were highly applied for?
- Which degree most accepted applicants had?
- Which background/majors most applicants held?
- What were the qualifications of applicants who applied for 'Researcher' job?
- Did rejected applicants share any commonalities?
- And much more ..

Applicant Name	Background	Major 1	Major 2	Major 3	Latest Degree	Graduation Date	Job Title	Salary	Application Status
Alice	Science	Chemistry	Physics	NA	MSc	02/03/2021	Researcher	3000	Accepted
John	Science	Management	Business	Finance	BSc	07/04/2019	Audit	2500	Accepted
Lili	Science	Architecture	Design	NA	BSc	05/02/2022	Project Manager	3291	Rejected
Jerry	Science	Computer Science	Computer Engineering	Business	MBA	05/01/2022	Product Manager	3429	Accepted
Mia	Art	Art	History	NA	BA	01/09/2020	Researcher	2931	Accepted
James	Art	Humanities	Social Sciences	NA	MA	02/03/2022	Researcher	3292	Rejected
Samy	Art	Law	NA	NA	PhD	01/02/2023	Lawyer	3434	Accepted

12/10/23

Visual Network Exploration Coaching Program

Concept Map Example 3: *Job Applications Network*

Concept Mapping

Describing your data through Concept maps

Start by adding all attributes to your map

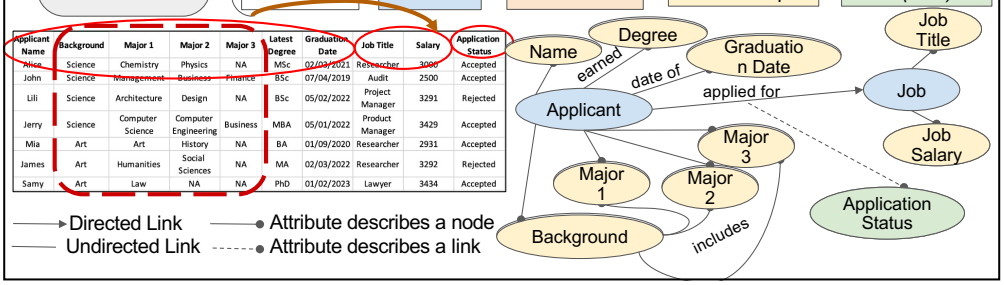
Identify main concepts (***nodes***)

identify relationships that connect *concepts* (***links***)

Add data attributes that describes a ***concept*** permanently to the concept

Add data fields that describes changing nature of ***relationships*** *to (links)*

Applicant Name	Background	Major 1	Major 2	Major 3	Latest Degree	Graduation Date	Job Title	Salary	Application Status
Alice	Science	Chemistry	Physics	NA	MSc	02/03/2021	Researcher	3090	Accepted
John	Science	Management	Business	Finance	BSc	07/04/2019	Audit	2500	Accepted
Lili	Science	Architecture	Design	NA	BSc	05/02/2022	Project Manager	3291	Rejected
Jerry	Science	Computer Science	Computer Engineering	Business	MBA	05/01/2022	Product Manager	3429	Accepted
Mia	Art	Art	History	NA	BA	01/09/2020	Researcher	2931	Accepted
James	Art	Humanities	Social Sciences	NA	MA	02/03/2022	Researcher	3292	Rejected
Samy	Art	Law	NA	NA	PhD	01/02/2023	Lawyer	3434	Accepted



Directed Link — Attribute describes a node

Undirected Link - - - Attribute describes a link

Visual Network Exploration Coaching Program

Starting with the Data (Example 4: *Patients Referral Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

What is this dataset about?

In our fourth sketching example, our dataset is about a number of patients who were referred to a service on a specific date. Each patient has an address that is composed of : building number, street name, postcode. And each service has a type and a provider.

Patient Name	Address	Building Number	Street	Postcode	Rerferall Date	Service Type	Service Provider
Mia	12 James Street, EH839L	12	James Street	EH839L	01/02/2022	Consultation	ENT Doctor
Lara	22 St. Andrews Square, LA325H	22	St. Andrews Square	LA325H	02/01/2023	Physiotherapy	Physiotherapy Department
Ali	18 Nicolas Stret, DW29LB	18	Nicolas Stret	DW29LB	02/02/2023	Physical Assessment	GP
Noah	6 Terrace, CE424L	6	Terrace	CE424L	19/02/2023	Regular Checkup	Dentist

12/10/23

Visual Network Exploration Coaching Program

Starting with the Data (Example 4: *Patients Referral Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

Data's Nature:

- Names
- Address (*locations*)
- Referral (*Dates)*
- Services (*types*)
- Service Provider (*types*)

What questions the data can answer?

- Which service(s) were highly referred to?
- Is there a dense area where most patients come from?
- When did most patient got their referrals?
- Which service provider got most referrals?
- And much more ..

Patient Name	Address	Building Number	Street	Postcode	Rerferall Date	Service Type	Service Provider
Mia	12 James Street, EH839L	12	James Street	EH839L	01/02/2022	Consultation	ENT Doctor
Lara	22 St. Andrews Square, LA325H	22	St. Andrews Square	LA325H	02/01/2023	Physiotherapy	Physiotherapy Department
Ali	18 Nicolas Stret, DW29LB	18	Nicolas Stret	DW29LB	02/02/2023	Physical Assessment	GP
Noah	6 Terrace, CE424L	6	Terrace	CE424L	19/02/2023	Regular Checkup	Dentist

Visual Network Exploration Coaching Program

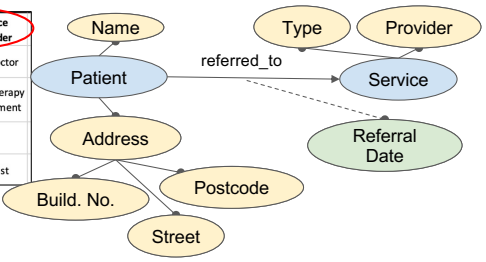
Concept Map Example 4: *Patients Referral Network*

Concept Mapping

Describing your data through Concept maps

Start by adding all attributes to your map → Identify main concepts (***nodes)*** → identify relationships that connect *concepts* (***links)*** → Add data attributes that describes a ***concept*** permanently to the concept → Add data fields that describes changing nature of ***relationships*** *to (links)*

Patient Name	Address	Building Number	Street	Postcode	Rerferall Date	Service Type	Service Provider
Mia	12 James Street, EH839L	12	James Street	EH839L	01/02/2022	Consultation	ENT Doctor
Lara	22 St. Andrews Square, LA325H	22	St. Andrews Square	LA325H	02/01/2023	Physiotherapy	Physiotherapy Department
Ali	18 Nicolas Stret, DW29LB	18	Nicolas Stret	DW29LB	02/02/2023	Physical Assessment	GP
Noah	6 Terrace, CE424L	6	Terrace	CE424L	19/02/2023	Regular Checkup	Dentist



Directed Link — Undirected Link — Attribute describes a node — Attribute describes a link

12/10/23

Visual Network Exploration Coaching Program

Starting with the Data (Example 5: *Patients Referral Network (extended)*)

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

What is this dataset about?

In our fifth sketching example, our dataset is similar to our dataset in example #4 about patients referral. The difference is we added information about who referred the patient to the service requested. The name of referring person and his/her workplace.

Patient Name	Address	Rerferall Date	Service Type	Service Provider	Referred by	Referring Workplace
Mia	12 James Street, EH839L	01/02/2022	Consultation	ENT Doctor	Dr Alice	Glenagles GP
Lara	22 St. Andrews Square, LA325H	02/01/2023	Physiotherapy	Physiotherapy Department	Dr Alice	Glenagles GP
Ali	18 Nicolas Stret, DW29LB	02/02/2023	Physical Assessment	GP	Social Worker John	Community Center
Noah	6 Terrace, CE424L	19/02/2023	Regular Checkup	Dentist	Dr James	Medows GP

Visual Network Exploration Coaching Program

Starting with the Data (Example 5: *Patients Referral Network (extended)*)

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

Data's Nature:

- Names
- Address (*locations*)
- Referral (*Dates)*
- Services (*types*)
- Service Provider (*types*)
- Referring workplaces (*types*)

What questions the data can answer?

- In addition to previous questions, which services did 'Dr Alice' mostly referred patients to?
- What do 'social workers' refer patients to?
- What do type of workplaces do most of the referrals are directed to?
- And much more ..

Patient Name	Address	Rerferall Date	Service Type	Service Provider	Referred by	Referring Workplace
Mia	12 James Street, EH839L	01/02/2022	Consultation	ENT Doctor	Dr Alice	Glenagles GP
Lara	22 St. Andrews Square, LA325H	02/01/2023	Physiotherapy	Physiotherapy Department	Dr Alice	Glenagles GP
Ali	18 Nicolas Stret, DW29LB	02/02/2023	Physical Assessment	GP	Social Worker John	Community Center
Noah	6 Terrace, CE424L	19/02/2023	Regular Checkup	Dentist	Dr James	Medows GP

12/10/23

Concept Map Example 5: *Patients Referral Network (extended)*

Concept Mapping
Describing your data through Concept Mapping

Start by adding all attributes to your map

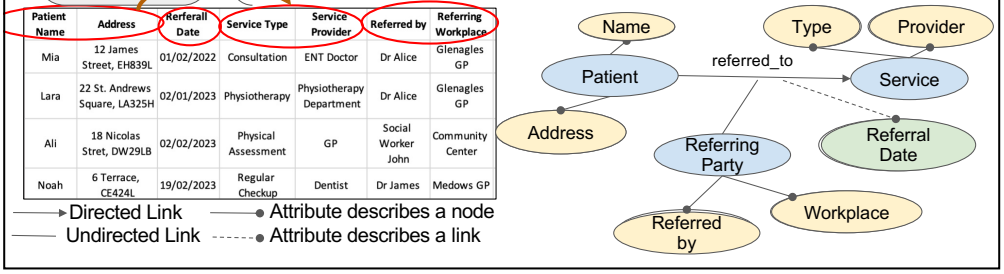
Identify main concepts (***nodes***)

identify relationships that connect *concepts* (***links***)

Add data attributes that describes a ***concept*** permanently to the concept

Add data fields that describes changing nature of ***relationships*** *to (links)*

Patient Name	Address	Rerferall Date	Service Type	Service Provider	Referred by	Referring Workplace
Mia	12 James Street, EH839L	01/02/2022	Consultation	ENT Doctor	Dr Alice	Glenagles GP
Lara	22 St. Andrews Square, LA325H	02/01/2023	Physiotherapy	Physiotherapy Department	Dr Alice	Glenagles GP
Ali	18 Nicolas Stret, DW29LB	02/02/2023	Physical Assessment	GP	Social Worker John	Community Center
Noah	6 Terrace, CE424L	19/02/2023	Regular Checkup	Dentist	Dr James	Medows GP



Directed Link
Undirected Link
Attribute describes a node
Attribute describes a link

Sketching *Concept Mapping*
Recap

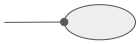
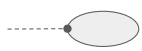
12/10/23

Recap on Concept Map Sketching

A network of links and nodes can describe through:

- ***Location of*** node/link
- ***Date of*** node/link
- ***Type of*** a of node/link (categorical)
- ***Direction*** of a relation
- **Link *Weight*** as a numerical value that your data might have
- **Link *ID***: a unique ID number that distinguish a link from other links. Usually used to describe how a specific link weight has changed through time.
- ***Derived Attribute***: An attribute that its values are derived from another attributes (ex. Calculated, composed, ..). This can be generated and used for both nodes/links (ex. node/link type where types could be grouped to reflect hierarchy in node/link types , link weight values). This could be used to inform decisions taken in the data wrangling and transformation step.

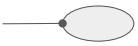
Symbols of Concept Map Sketching

Symbol	Name	Description	Can be used as (network table)
	Concept	Represents any concept (ex.entities, events, persons)	Nodes
	Directed Relationship	Relationship that is an associations between or among concepts where direction is important	Directed Link
	Undirected Relationship	Relationship that is an associations between or among concepts where direction isn't important.	Undirected Link
	Concept Attribute	An attribute that describes a concept (ex. Name of a person, specialization of an employee)	Node / Node Type
	Relationship Attribute	An attribute that describes a relationship between concepts (ex. date of when a letter was sent).	Link Type / Link Weight / Link ID

12/10/23

Visual Network Exploration Coaching Program

Symbols of Concept Map Sketching

Symbol	Name	Description	Can be used as (network table)
	Multi-Relation	A relationship that connects 3 or more node entities (ex. Ternary relation)	Link types derived from any of the participating nodes
	A missing Concept	A node that represents a concept that you expect to have or need but do not no about it for now at least.	Absent Node
	A sub-attribute	An attribute that can be partially considered part from another attribute such as composite attribute.	Node Type / Node
	Derived Attribute	An additional attribute that you can create by deriving its values from 1+ attributes (ex.computed	Node/ Link Type, Link Weight

Visual Network Exploration Coaching Program

Considerations for Concept Map Sketching

Here are some questions to consider and think about while sketching your concept map and transforming it into a network schema:

- Is my data consistent?
  - In concept maps: does each attribute hold the same data type?
  - In network schema: Does each column and row have the same format?
- Do I have unknown/absent values? How am I going to represent them in my map? Or should I ignore them for now?
- Am I interested in exploring the whole data network? Or networks of specific nodes (aka ego-networks)?
- Can my networks nodes/link types be:
  - broken-down into smaller categories?
  - Or grouped to higher level?

C.2.2 Understanding your Data and Sketching Concept Maps

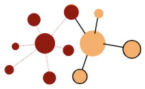
(Starting from a Research Goal/Technique)

Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

Understanding your Data and *Sketching* **Concept Maps**

*(Starting from **a Research Goal/Technique**)*



Visual Network Exploration

Coaching Program

How can I start *Concept Mapping?*

Visual Network Exploration
Coaching Program

How can I start creating my concept map?

You can start creating your concept map based on having and/or :

Data
(ex. Formatted dataset(s), qualitative data to be formatted,..)

Describe all data using a Concept Map → generate possible question your concept map can answer

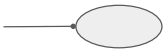
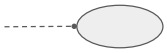
A research goal
that you aim to achieve
(ex. Answering research question, comparing 2+ datasets, ..)

Gradually build your concept map, starting from a single research goal → add more elements to your concept map as you.

Sketching *Concept Maps*
Symbols & Notions

Visual Network Exploration
Coaching Program

Symbols of Concept Map Sketching

Symbol	Name	Description	Can be used as (network table)
	Concept	Represents any concept (ex.entities, events, persons)	Nodes
	Directed Relationship	Relationship that is an associations between or among concepts where direction is important	Directed Link
	Undirected Relationship	Relationship that is an associations between or among concepts where direction isn't important.	Undirected Link
	Concept Attribute	An attribute that describes a concept (ex. Name of a person, specialization of an employee)	Node / Node Type
	Relationship Attribute	An attribute that describes a relationship between concepts (ex. a date of when a letter was sent).	Link Type / Link Weight / Link ID

Visual Network Exploration
Coaching Program

How to Describe and Sketch your Concept Map?

Each data attribute in your dataset can participate in describing your network's semantics and structure.

Which role an attribute can play, describe or represent in a network?

- A candidate node/link value/type.
- ***Location of*** node/link: location of where a node/link was created.
- ***Date of*** node/link: date where a node/link was created.
- ***Type of*** a of node/link: a set of values that can be used to distinguish different node/link types.
- ***Direction*** of a link: show how a link was created from which source node toward which target node.
- ***Weight*** of a link: numeric value that is reflected on the width of a link.

Examples on Sketching *Concept Maps* ***Starting from your Research Question***

Visual Network Exploration
Coaching Program

Starting with the Data (Example 1: *Letters Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

What is this dataset about?

In our first sketching example, our dataset is about a set of letters and documents exchanged by a trader called Marie Boucher and the goal is to explore her trading network.

	Sender Name	Sender Place	Qualification	Receiver Name	Receiver Place	Date Letter Sent
1	Hubert Antheaun	St Nicolas, Nante	Commerce	Jacques Yvon, sr de	Logane, Ile St	03/08/1673
2	Marie Boucher	St Nicolas, Nante	Commerce	Hubert Antheaume	St Nicolas, N	06/08/1675
3	Marie Boucher	St Nicolas, Nante	Famille	Hubert Antheaume	St Nicolas, N	06/08/1675
4	Marie Boucher	St Nicolas, Nante	Famille	Roze Boucher	St Nicolas, N	15/09/1671
5	Marie Boucher	Paris	Commerce	Roze Boucher	St Nicolas, N	15/09/1671
6	Marie Boucher	Paris	Commerce	Hubert Antheaume	Paris	15/09/1671
7	? Vallet	Ile St Domingue	Travail	? Bec de Fort		01/01/1676
8	? Vallet	Ile St Domingue	Commerce	Joseph Labelle	Logane, Ile St	01/01/1676
9	? Vallet	Ile St Domingue	Famille	? Vallet	Ile Guadeloup	01/01/1676
10	? Vallet	Ile St Domingue	Famille	Joseph Vallet	Brest	01/01/1676

Visual Network Exploration Coaching Program

Starting with the Data (Example 1: *Letters Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

Data's Nature:

- Names
- Places (*locations)*
- Communication *dates*
- Type of qualification

What questions the data can answer?

- Who contacted whom? And in which context (e.g. commerce)?
- When did most communication take place?
- Where did most of the messaging happened?
- Where most of the traders were?
- How does the network of 'family' connection differ from network of 'commerce' in matter of structure? Or density?
- And much more ..

	Sender Name	Sender Place	Qualification	Receiver Name	Receiver Place	Date Letter Sent
1	Hubert Antheaun	St Nicolas, Nante	Commerce	Jacques Yvon, sr de:	Logane, Ile St	03/08/1673
2	Marie Boucher	St Nicolas, Nante	Commerce	Hubert Antheaume	St Nicolas, N	06/08/1675
3	Marie Boucher	St Nicolas, Nante	Famille	Hubert Antheaume	St Nicolas, N	06/08/1675
4	Marie Boucher	St Nicolas, Nante	Famille	Roze Boucher	St Nicolas, N	15/09/1671
5	Marie Boucher	Paris	Commerce	Roze Boucher	St Nicolas, N	15/09/1671
6	Marie Boucher	Paris	Commerce	Hubert Antheaume	Paris	15/09/1671
7	? Vallet	Ile St Domingue	Travail	? Bec de Fort		01/01/1676
8	? Vallet	Ile St Domingue	Commerce	Joseph Labelle	Logane, Ile St	01/01/1676
9	? Vallet	Ile St Domingue	Famille	? Vallet	Ile Guadeloup	01/01/1676
10	? Vallet	Ile St Domingue	Famille	Joseph Vallet	Brest	01/01/1676

Visual Network Exploration Coaching Program

Concept Map Example 1: *Letters Network*

Concept Mapping

Describing your data through Concept maps

Start by adding **attributes** addressing your **strategy**

Identify main concepts (***nodes)***

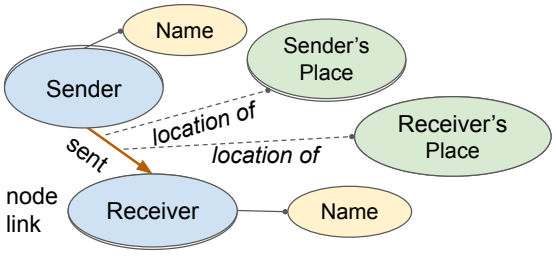
identify relationships that connect *concepts* (***links)***

Add data attributes that describes a ***concept*** permanently to the concept

Add data fields that describes changing nature of ***relationships*** *to (links)*

	Sender Name	Sender Place	Qualification	Receiver Name	Receiver Place	Date Letter Sent
1	Hubert Antheaun	St Nicolas, Nante	Commerce	Jacques Yvon, sr de:	Logane, Ile St	03/08/1673
2	Marie Boucher	St Nicolas, Nante	Commerce	Hubert Antheaume	St Nicolas, N	06/08/1675
3	Marie Boucher	St Nicolas, Nante	Famille	Hubert Antheaume	St Nicolas, N	06/08/1675
4	Marie Boucher	St Nicolas, Nante	Famille	Roze Boucher	St Nicolas, N	15/09/1671
5	Marie Boucher	Paris	Commerce	Roze Boucher	St Nicolas, N	15/09/1671
6	Marie Boucher	Paris	Commerce	Hubert Antheaume	Paris	15/09/1671
7	? Vallet	Ile St Domingue	Travail	? Bec de Fort		01/01/1676
8	? Vallet	Ile St Domingue	Commerce	Joseph Labelle	Logane, Ile St	01/01/1676
9	? Vallet	Ile St Domingue	Famille	? Vallet	Ile Guadeloup	01/01/1676
10	? Vallet	Ile St Domingue	Famille	Joseph Vallet	Brest	01/01/1676

Data-Driven: spatial distribution of senders and receivers



Directed Link — Attribute describes a node

Undirected Link - - - - Attribute describes a link

Visual Network Exploration Coaching Program

Starting with the Data (Example 2: *Connections Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

What is this dataset about?

In our second sketching example, our dataset is about a family members and best friend of number of people. And we would like to explore their relationships network and if they intersect.

Name	Father	Mother	Grandfather	Grandmother	Friend
Bob	Martin	Ana	Sam	Sara	James
Ali	Jamal	Suzen	Kamal	Sara	Mia
James	Sami	Poppy	Kai	Sangeeta	Layla
Mia	Mo	Ann	Roy	Summer	Pri

Visual Network Exploration Coaching Program

Starting with the Data (Example 2: *Connections Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

Data's Nature:

- Names
- Family connections (*type of connection*)
- Friends names

What questions the data can answer?

- Who share same grandfather/grandmother/..?
- Who are friends of ex. James or Mia?
- Who has the highest number of connections?
- Who is a role-playing node in this network? (ex. Connect high number of people together)
- Who has no connections to the network? (ex. isolated)
- And much more ..

Name	Father	Mother	Grandfather	Grandmother	Friend
Bob	Martin	Ana	Sam	Sara	James
Ali	Jamal	Suzen	Kamal	Sara	Mia
James	Sami	Poppy	Kai	Sangeeta	Layla
Mia	Mo	Ann	Roy	Summer	Pri

Visual Network Exploration
Coaching Program

Concept Map Example 2: *Connections Network*

Concept Mapping

Describing your data through Concept maps

Start by adding **attributes** addressing your **strategy**

Identify main concepts (***nodes)***

identify relationships that connect *concepts* (***links)***

No Links

Add data attributes that describes a ***concept*** permanently to the concept

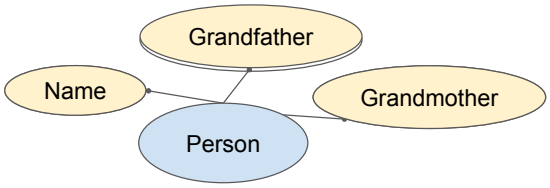
Add data fields that describes changing nature of ***relationships*** *to (links)*

No Link Attributes

Name	Father	Mother	Grandfather	Grandmother	Friend
Bob	Martin	Ana	Sam	Sara	James
Ali	Jamal	Suzen	Kamal	Sara	Mia
James	Sami	Poppy	Kai	Sangeeta	Layla
Mia	Mo	Ann	Roy	Summer	Pri

Directed Link
Undirected Link
Attribute describes a node
Attribute describes a link

Task-Driven(Clusters): who share same grandparents?



Starting with the Data (Example 3: *Job Applications Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

What is this dataset about?

In our third sketching example, our dataset is about a number of graduates who applied for opening jobs . Each applicant can have 1-3 majors under the their background umbrella. Each job application can be accepted/rejected.

Applicant Name	Background	Major 1	Major 2	Major 3	Latest Degree	Graduation Date	Job Title	Salary	Application Status
Alice	Science	Chemistry	Physics	NA	MSc	02/03/2021	Researcher	3000	Accepted
John	Science	Management	Business	Finance	BSc	07/04/2019	Audit	2500	Accepted
Lili	Science	Architecture	Design	NA	BSc	05/02/2022	Project Manager	3291	Rejected
Jerry	Science	Computer Science	Computer Engineering	Business	MBA	05/01/2022	Product Manager	3429	Accepted
Mia	Art	Art	History	NA	BA	01/09/2020	Researcher	2931	Accepted
James	Art	Humanities	Social Sciences	NA	MA	02/03/2022	Researcher	3292	Rejected
Samy	Art	Law	NA	NA	PhD	01/02/2023	Lawyer	3434	Accepted

Visual Network Exploration Coaching Program

Starting with the Data (Example 3: *Job Applications Network)*

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

Data's Nature:

- Names
- Background & majors (*types*)
- Graduation (*date)*
- Salary (*numeric*)
- Degree (*types*)
- Application status (*types*)

What questions the data can answer?

- Which jobs were highly applied for?
- Which degree most accepted applicants had?
- Which background/majors most applicants held?
- What were the qualifications of applicants who applied for 'Researcher' job?
- Did rejected applicants share any commonalities?
- And much more ..

Applicant Name	Background	Major 1	Major 2	Major 3	Latest Degree	Graduation Date	Job Title	Salary	Application Status
Alice	Science	Chemistry	Physics	NA	MSc	02/03/2021	Researcher	3000	Accepted
John	Science	Management	Business	Finance	BSc	07/04/2019	Audit	2500	Accepted
Lili	Science	Architecture	Design	NA	BSc	05/02/2022	Project Manager	3291	Rejected
Jerry	Science	Computer Science	Computer Engineering	Business	MBA	05/01/2022	Product Manager	3429	Accepted
Mia	Art	Art	History	NA	BA	01/09/2020	Researcher	2931	Accepted
James	Art	Humanities	Social Sciences	NA	MA	02/03/2022	Researcher	3292	Rejected
Samy	Art	Law	NA	NA	PhD	01/02/2023	Lawyer	3434	Accepted

Visual Network Exploration Coaching Program

Concept Map Example 3: *Job Applications Network*

Concept Mapping

Describing your data through Concept maps

Start by adding **attributes** addressing your **strategy**

Identify main concepts (***nodes)***

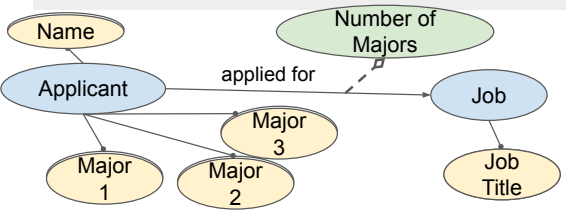
identify relationships that connect *concepts* (***links)***

Add data attributes that describes a ***concept*** permanently to the concept

Add data fields that describes changing nature of ***relationships*** *to (links)*

Applicant Name	Background	Major 1	Major 2	Major 3	Latest Degree	Graduation Date	Job Title	Salary	Application Status
Alice	Science	Chemistry	Physics	NA	MSc	02/03/2021	Researcher	3000	Accepted
John	Science	Management	Business	Finance	BSc	07/04/2019	Audit	2500	Accepted
Lili	Science	Architecture	Design	NA	BSc	05/02/2022	Project Manager	3291	Rejected
Jerry	Science	Computer Science	Computer Engineering	Business	MBA	05/01/2022	Product Manager	3429	Accepted
Mia	Art	Art	History	NA	BA	01/09/2020	Researcher	2931	Accepted
James	Art	Humanities	Social Sciences	NA	MA	02/03/2022	Researcher	3292	Rejected
Samy	Art	Law	NA	NA	PhD	01/02/2023	Lawyer	3434	Accepted

Task-Driven (hypothesis) : Applicants with higher number of held majors are the ones who applied for a researcher job.



Directed Link
Undirected Link
Attribute describes a node
Attribute describes a link

Visual Network Exploration Coaching Program

Starting with the Data (Example 5: *Patients Referral Network (extended)*)

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

What is this dataset about?

In our fifth sketching example, our dataset is similar to our dataset in example #4 about patients referral. The difference is we added information about who referred the patient to the service requested. The name of referring person and his/her workplace.

Patient Name	Address	Rerferall Date	Service Type	Service Provider	Referred by	Referring Workplace
Mia	12 James Street, EH839L	01/02/2022	Consultation	ENT Doctor	Dr Alice	Glenagles GP
Lara	22 St. Andrews Square, LA325H	02/01/2023	Physiotherapy	Physiotherapy Department	Dr Alice	Glenagles GP
Ali	18 Nicolas Stret, DW29LB	02/02/2023	Physical Assessment	GP	Social Worker John	Community Center
Noah	6 Terrace, CE424L	19/02/2023	Regular Checkup	Dentist	Dr James	Medows GP

Visual Network Exploration Coaching Program

Starting with the Data (Example 5: *Patients Referral Network (extended)*)

Data Understanding

1) Select a dataset.
2) Take an overview about data's nature.
3) Think about the questions it can answer.

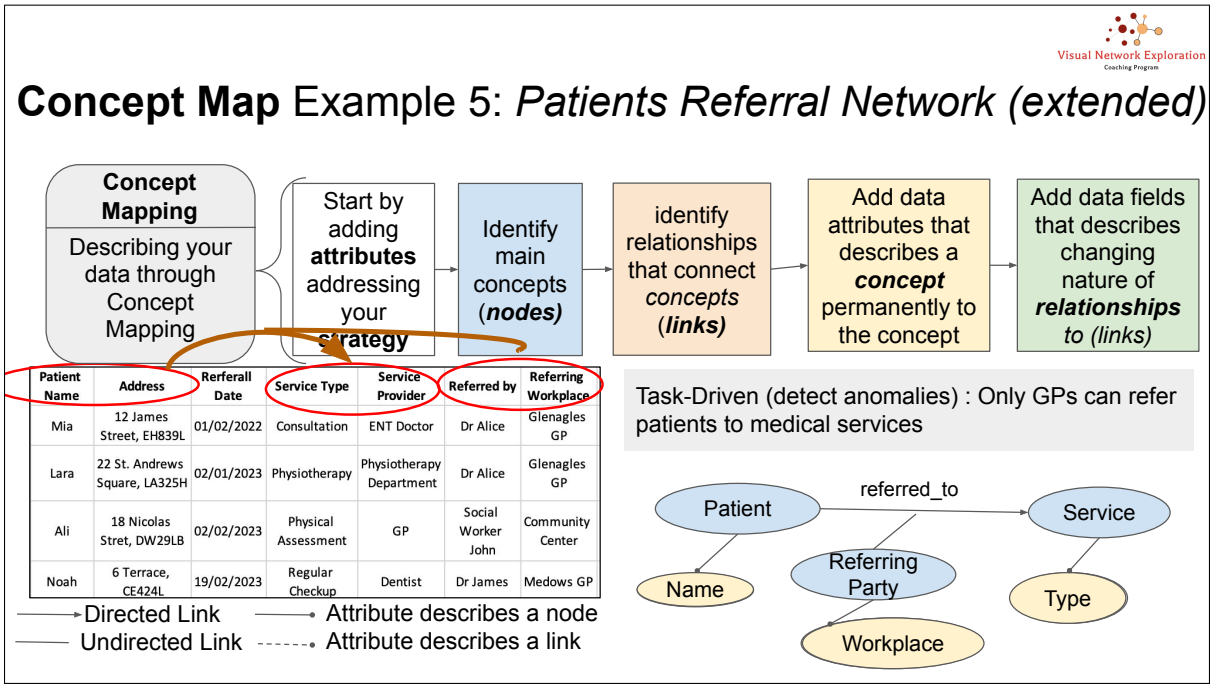
Data's Nature:

- Names
- Address (*locations*)
- Referral (*Dates)*
- Services (*types*)
- Service Provider (*types*)
- Referring workplaces (*types*)

What questions the data can answer?

- In addition to previous questions, which services did 'Dr Alice' mostly referred patients to?
- What do 'social workers' refer patients to?
- What do type of workplaces do most of the referrals are directed to?
- And much more ..

Patient Name	Address	Rerferall Date	Service Type	Service Provider	Referred by	Referring Workplace
Mia	12 James Street, EH839L	01/02/2022	Consultation	ENT Doctor	Dr Alice	Glenagles GP
Lara	22 St. Andrews Square, LA325H	02/01/2023	Physiotherapy	Physiotherapy Department	Dr Alice	Glenagles GP
Ali	18 Nicolas Stret, DW29LB	02/02/2023	Physical Assessment	GP	Social Worker John	Community Center
Noah	6 Terrace, CE424L	19/02/2023	Regular Checkup	Dentist	Dr James	Medows GP



Sketching *Concept Mapping* ***Recap***

Visual Network Exploration
Coaching Program

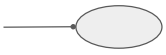
Recap on Concept Map Sketching

A network of links and nodes can describe through:

- ***Location of*** node/link
- ***Date of*** node/link
- ***Type of*** a of node/link (categorical)
- ***Direction*** of a relation
- **Link *Weight*** as a numerical value that your data might have
- **Link *ID***: a unique ID number that distinguish a link from other links. Usually used to describe how a specific link weight has changed through time.
- ***Derived Attribute***: An attribute that its values are derived from another attributes (ex. Calculated, composed, ..). This can be generated and used for both nodes/links (ex. node/link type where types could be grouped to reflect hierarchy in node/link types , link weight values). This could be used to inform decisions taken in the data wrangling and transformation step.

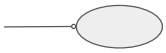
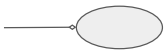
Visual Network Exploration
Coaching Program

Symbols of Concept Map Sketching

Symbol	Name	Description	Can be used as (network table)
	Concept	Represents any concept (ex.entities, events, persons)	Nodes
	Directed Relationship	Relationship that is an associations between or among concepts where direction is important	Directed Link
	Undirected Relationship	Relationship that is an associations between or among concepts where direction isn't important.	Undirected Link
	Concept Attribute	An attribute that describes a concept (ex. Name of a person, specialization of an employee)	Node / Node Type
	Relationship Attribute	An attribute that describes a relationship between concepts (ex. date of when a letter was sent).	Link Type / Link Weight / Link ID

Visual Network Exploration
Coaching Program

Symbols of Concept Map Sketching

Symbol	Name	Description	Can be used as (network table)
	Multi-Relation	A relationship that connects 3 or more node entities (ex. Ternary relation)	Link types derived from any of the participating nodes
	A missing Concept	A node that represents a concept that you expect to have or need but do not no about it for now at least.	Absent Node
	A sub-attribute	An attribute that can be partially considered part from another attribute such as composite attribute.	Node Type / Node
	Derived Attribute	An additional attribute that you can create by deriving its values from 1+ attributes (ex.computed	Node/ Link Type, Link Weight

Visual Network Exploration
Coaching Program

Considerations for Concept Map Sketching

Here are some questions to consider and think about while sketching your concept map and transforming it into a network schema:

- Is my data consistent?
  - In concept maps: does each attribute hold the same data type?
  - In network schema: Does each column and row have the same format?
- Do I have unknown/absent values? How am I going to represent them in my map? Or should I ignore them for now?
- Am I interested in exploring the whole data network? Or networks of specific nodes (aka ego-networks)?
- Can my networks nodes/link types be:
  - broken-down into smaller categories?
  - Or grouped to higher level?

C.3 The Selection and Description of Networks Visual Exploration Strategies

The strategies in the visual exploration vary based on their goal. Through our analysis, we were able to identify 16. The strategies per category are:

C.3.1 Data-Driven Exploration Strategies

Data-driven exploration strategies share the following characteristics: (1) they explore data partially or fully based on a specific analysis technique (ex. Temporal changes) and (2) each strategy will include a specific set of linked steps based on the chosen analysis technique toward creating the network schema that will shape the network's visualization.

Data-Driven
Exploration Strategies

Characteristics:

- To explore your data partially or fully based on a specific analysis technique.
- The exploration process will follow a specific set of linked steps based on the chosen analysis technique toward creating the network schema that will shape the network's visualization.
- Such strategy depends on mining for insights based on specific perspective (ex. Temporal changes).

Data-Driven Exploration Strategies

DS1 : Exploring by the Spatial Analysis Technique

Description: To explore data depending on the location of nodes. For example, distribution on a map or clusters among locations in an adjacency matrix.

Network Visualizations : Map when nodes' geolocations are used, Other (ex.node-link diagram) when locations are used as nodes

What questions to ask?

- Were you able to understand how locations were associated with specific nodes, node types or link types?
- Can you add other aspects such as exploring locations over time?

Application : *any dataset that contains locations (e.g. geolocation coordinates).*

Strategy Steps:

- ***Step 1:*** Sketch the concepts map of your dataset .
- ***Step 2:*** *Initially* decide on how you like to visually represent the location of your nodes (ex. *distribution of geolocation of nodes across a map* or *use locations names as nodes*).
- ***Step 3:*** Select the node attributes, their locations, and any additional attribute you wish to visualize (ex. Link type or date) from the concept map.
- ***Step 4:*** Transform selected attributes from concept map into a network schema.
- ***Step 5:*** Create/update network visualization(s) that are compatible with the selected data attributes' nature and explore them:
  - ex. Map to look into the geological distribution of nodes.
  - ex. Adjacency Matrix looking for clusters (ex. Clusters of cities where an event has taken place in).
  - Refer to the roadmap and supporting material on networks visual exploration.
- ***Step 6****: Log your observations based on visualized network (refer to DC2).*
- ***Step 7:*** *Evaluate and reflect on your observations (refer to DC2).*
- ***Step 8****: if any data transformation is required (ex. projection of nodes), apply data transformation then return to step (5), if not proceed to following step.*
- ***Step 9:*** *End your exploration and journal your findings (refer to DC4)*
- *P.S. In some cases, maps not always to reveal insights. Some other data transformations (DC3) help in thinking about your data from another angle.*

Data-Driven Exploration Strategies

DS2 : Exploring by the Temporal Analysis Technique

Description: To explore network data changes overtime. For example, changes in network layout, clusters, or location overtime.

Network Visualizations : *Depends on the schema. Overall,* any visualization supporting temporal exploration through interactive features (ex. time-slider)

What questions to ask?

- How did the network layout/ clusters change overtime?
- Did new links/nodes disappear/ appear overtime?

Application : *any dataset that contains datetime (e.g. date of immigration).*

Strategy Steps:

- ***Step 1:*** Sketch the concepts map of your dataset .
- ***Step 2:*** *Initially* decide on how you like to visually represent the datetime of your nodes (ex. *Datetime describing the immigration link between two cities nodes* or *the year as a node and city as a node*).
- ***Step 3:*** Select the nodes/links, datetime representation, and any additional attribute you wish to visualize (ex. Link type or date) from the concept map.
- ***Step 4:*** Transform selected attributes from concept map into a network schema.
- ***Step 5:*** Create/update network visualization(s) that are compatible with the selected data attributes' nature and explore them:
  - ex. Any visualization with time-slider if datetime describes links such as timeline.
  - ex. Adjacency Matrix (ex. Clusters of cities and year of immigration)
  - Refer to the roadmap and supporting material on networks visual exploration.
- ***Step 6****: Log your observations based on visualized network (refer to DC2).*
- ***Step 7:*** *Evaluate and reflect on your observations (refer to DC2).*
- ***Step 8****: if any data transformation is required (ex. projection of nodes), apply data transformation then return to step (5), if not proceed to following step.*
- ***Step 9:*** *End your exploration and journal your findings (refer to DC4)*
- *P.S. Some data transformations (DC3) help in thinking about your data from various perspectives.*

Data-Driven Exploration Strategies

DS3 : Exploring by the Structural Analysis Technique

Description: To explore data by the network's structure (ex. layout, connectivity, density)

Network Visualizations : node-link diagram (layout,density), adjacency matrix (clusters)

What questions to ask?

- What is the network layout? Is it a sparse or dense network?
- Does it change overtime?
- What are the highly connected nodes?

Application : *any dataset visualized as a network*

Strategy Steps:

- ***Step 1:*** Sketch the concepts map of your dataset .
- ***Step 2:*** *Initially* extract possible nodes/links from your dataset.
- ***Step 3:*** Select the nodes/links and any additional attribute you wish to visualize (ex. Link type or date) from the concept map.
- ***Step 4:*** Transform selected attributes from concept map into a network schema.
- ***Step 5:*** Create/update network visualization(s) that are compatible with the selected data attributes' nature and explore them:
  - ex. Node-link diagram (layout)
  - ex. Adjacency Matrix (Clusters)
  - Refer to the roadmap and supporting material on networks visual exploration.
- ***Step 6****: Log your observations based on visualized network (refer to DC2).*
- ***Step 7:*** *Evaluate and reflect on your observations (refer to DC2).*
- ***Step 8****: if any data transformation is required (ex. projection of nodes), apply data transformation then return to step (5), if not proceed to following step.*
- ***Step 9:*** *End your exploration and journal your findings (refer to DC4)*
- *P.S. Some data transformations (DC3) help in thinking about your data from various perspectives.*

C.3.2 Open-Ended Exploration Strategies

Open-ended exploration strategies share the following characteristics: (1) they explore data partially or fully in an open-ended way, looking for visual pattern(s) or insights of the analyst's interest, and they depend on mining for insights by playing freely with the network schema.

Open-Ended Exploration Strategies

Characteristics:

- To explore your data partially or fully in an open-ended way looking for interesting visual pattern(s) or insights.
- Such strategy depends on mining for insights by playing freely with the network schema.

Open-Ended Exploration Strategies

OS1: Starting with Minimal Data Attributes & Expand Schema Gradually

Description: To start exploring data with minimal number of data attributes and gradually expand the network schema by adding more attributes.

Network Visualizations: *Any network visualization(s)*

What questions to ask?

- What changes in the network's visual patterns by changing the schema? (*refer to DC2*)
- Which data attribute(s) tells the story of my data?(*refer to DC2*)

Application : *on any dataset.*

Strategy Steps:

- ***Step 1:*** Start by sketching the concepts map of your dataset.
- ***Step 2:*** Selecting the minimal number of data attributes from your concept map to represent 2 linked nodes (node1 and node2).
- ***Step 3:***Transform selected attributes from concept map into a network schema.
- ***Step 4:*** Create/update network visualization(s) that are compatible with the selected data attributes' nature and explore them:
  - ex. Node-link diagram looking into structure and role-playing nodes
  - ex. Adjacency Matrix looking for clusters
  - Refer to the supporting material on networks visual exploration.
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 6:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 7:*** Iterate the exploration process and change network schema by:
  - ***adding*** one or two data attributes from your concept map to the network schema. For example, add a link type to filter with or dates for temporal exploration.
  - or by changing the currently selected nodes.
- ***Step 8:*** Return to step(3) and implement it along with the following steps using the updated schema and observe the changes in your network.
- ***Step 9:*** End your exploration and journal your findings (*refer to DC4*)

Open-Ended Exploration Strategies

OS2: Starting with All Data Attributes & Exclude As you Go

Description: To start exploring your data with a schema of all data attributes. Then to exclude data attributes that are not informative or makes visualization hard to explore.

Network Visualizations: *Any network visualization(s)*

What questions to ask?

- Which data attribute(s) removal can make visualization more readable?
- Which data attribute(s) can I use a nodes/links?(*refer to DC2*)
- Can I break the dataset into chunks?

Application : *suits smaller datasets such as number of data attributes*

Strategy Steps:

- ***Step 1:*** Start by sketching the concepts map of whole dataset.
- ***Step 2:*** Select what concepts the represents linked nodes (node1 & node2).
- ***Step 3:***Transform attributes from concept map into a network schema.
- ***Step 4:*** Create/update network visualization(s) that are compatible with the data attributes' nature and explore them:
  - ex. Node-link diagram looking into structure and role-playing nodes
  - ex. Adjacency Matrix looking for clusters
  - Refer to the supporting material on networks visual exploration.
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 6:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 7:*** Iterate the exploration process and change network schema by:
  - ***removing*** one or two data attributes from your network schema. For example, remove attributes with similar values for most of the nodes. Or attributes that causes cluttered visualization.
  - or by changing the currently selected nodes.
- ***Step 8:*** Return to step(3) and implement it along with the following steps using the updated schema and observe the changes in your network.
- ***Step 9:*** End your exploration and journal your findings (*refer to DC4*)

Open-Ended Exploration Strategies

OS3: Starting with Sample of the Data Attributes

Description: To start exploring data with a sample of it. Either a sample of the selected attributes or a sample of the selected rows.

Network Visualizations: *Any network visualization(s)*

What questions to ask?

- Is the number of data attributes large? (Sample attributes) Is the number of rows large?(Sample records)
- What is the criteria to sample?

Application : *to have an initial understanding of the data explored*

Strategy Steps:

- ***Step 1:*** Start by sketching the concepts map of whole dataset.
- ***Step 2:***Transform attributes from concept map into a network schema.
- ***Step 3:*** Select a sample of data attributes/rows. A sampling criteria can be set (ex.attributes with no null values or rows with a specific value)
- ***Step 4:*** Create/update network visualization(s) that are compatible with the data attributes' nature and explore them:
  - ex. Node-link diagram looking into structure and role-playing nodes
  - ex. Adjacency Matrix looking for clusters
  - Refer to the supporting material on networks visual exploration.
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 6:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 7:*** Iterate the exploration process and change network schema by:
  - ***Selecting another sample*** whether it was based on rows/attributes. Try to select different samples to offer insight from different perspectives.
- ***Step 8:*** Return to step(3) and implement it along with the following steps using the updated schema and observe the changes in your network.
- ***Step 9:*** End your exploration and journal your findings (*refer to DC4*)

C.3.3 Task-Driven Exploration Strategies

Task-driven exploration strategies share the following characteristics: (1) they explore data partially or fully toward achieving a pre-selected exploration goal, (2) influenced by the designated goal, the network schema is created to support revealing associated visual patterns in the visualizations, (3) and each strategy focuses on evaluating the goal's outcome based on the captured visual insights.

Task-Driven
Exploration Strategies

Characteristics:

- To explore your data partially or fully toward achieving a pre-selected exploration goal.
- Influenced by the designated goal, the exploration process will follow a specific set of linked steps that will shape the network schema and support revealing associated visual patterns in the visualizations.
- Such strategy focuses on evaluating the goal's outcome based on the captured visual insights.

Task-Driven Exploration Strategies

TS1: Exploring to Answer Research Questions

Description: To explore data aiming to answer a specific research question(s) through network visualizations.

Network Visualizations : *compatible with data attributes types of the schema.*

What questions to ask?

- Were the research questions answered? (*refer to DC2*)
- Which visual patterns informed the addressing the question(s)? (*refer to DC2*)

Application : *on any dataset.*

Strategy Steps:

- ***Step 1:*** Select a single research question from your questions pool.
- ***Step 2:*** Sketch/update your concepts map by selecting the data attributes that addresses the currently selected question.
- ***Step 3:*** Create network schema(s) based on the concept map - changes- .
- ***Step 4:*** Create the network visualization(s) that are compatible with the nature of the selected data attributes and explore them. Refer to the supporting material on how to explore networks visually.
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 6:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 7:*** *if any data transformation is required*, *apply* data transformation then return to step (3), if not proceed to following step.
- ***Step 8:*** *if other research question(s) exist*, return to step(1) and implement it along with the following steps.
- ***Step 9:*** End your exploration and journal your findings (*refer to DC4*)
- P.S. you can create a single concept map for each question, or create/update a comprehensive concept map showing all questions' data. The later option opens more aspects to explore your network data through.

Task-Driven Exploration Strategies

TS2: Exploring by Evaluating Hypothesis

Description: To explore data by visualizing a hypothesis and evaluating it visually.

Network Visualizations : *compatible with data attributes types of the schema.*

What questions to ask?

- Does the visualization selected proves the hypothesis ?
- Does creating alternative network schemas offer better understanding? (*refer to DC2*)

Application : *on any dataset.*

Strategy Steps:

- ***Step 1:*** Start by sketching the concepts map of your dataset
- ***Step 2:*** Create network schema(s) by selecting the data attributes participating in forming in the hypothesis (*ex. Hypothesis: Most loan requests are for educational scholarships. Network Schema: loan Request → Spent on*).
- ***Step 3:*** Create the network visualization(s) that are compatible with the nature of the selected data attributes and explore them. Refer to the supporting material on how to explore networks visually. (ex. Node-link diagram to examine layout)
- ***Step 4:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 5:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 6:*** *if any data transformation is required*, *apply* data transformation then return to step (3), if not proceed to following step. (ex. Combining similar requests and add link weight= count of requests)
- ***Step 7:*** *if other hypothesies(s) exist*, return to step(1) and implement it along with the following steps.
- ***Step 8:*** End your exploration and journal your findings (*refer to DC4*)
- P.S. you can create a comprehensive concept map showing all data, or create a concept map that participate in the hypothesis only. The first option opens more aspects to explore your network data through.

Task-Driven Exploration Strategies

TS3: Exploring by Validating an Argument

Description: To explore data by evaluating and validating arguments visually. That is, to examine possible findings through arguments derivations and deductive rules .

Network Visualizations : *compatible with data attributes types of the schema.*

What questions to ask?

- Is the deductive arguments result in:
  - further understanding of the network ?
  - Incorrect/inconsistent results?
-

Application : *any dataset, mostly where looking into deduction and inference*

Strategy Steps:

- ***Step 1:*** Start by sketching the concepts map of your dataset for each argument. (*ex. A→B and B→C, then looking into findings from A→C*)
- ***Step 2:*** Create network schema by selecting the data attributes participating in forming in each argument .
- ***Step 3:*** Create the network visualization(s) that are compatible with the nature of the selected data attributes and explore them. Refer to the supporting material on how to explore networks visually. (ex. Node-link diagram to examine layout comparing findings from each argument visualization)
- ***Step 4:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 5:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 6:*** *if any data transformation is required*, *apply* data transformation then return to step (2), if not proceed to following step. (ex. Combining both A→B and B→C within the same visualization looking for central nodes and other patterns)
- ***Step 7:*** *if other arguments(s) exist*, return to step(1) and implement it along with the following steps.
- ***Step 8:*** End your exploration and journal your findings (*refer to DC4*)
- P.S. you can create a comprehensive concept map showing all arguments, or create a concept map for each argument. The first option opens more aspects to explore your network data through - if applicable-.

Task-Driven Exploration Strategies

TS4: Exploring by Measuring a Flow

Description: To explore data by measuring a flow in the network is to: explore and monitor the movement of data or interactions of data, entities, or relationships within a network visualization (ex. change of node's location or clusters).

Network Visualizations : *compatible with data attributes types of the schema.*

What questions to ask?
- How did the flow change overtime?
- What are the nodes or links affected the flow?
- What are the flow metrics measured? (ex. Directionality, volume, frequency, duration)

Application : *to investigate how data navigate or propagate through a network (ex. Trades, supply chains)*

Strategy Steps:
- ***Step 1:*** Start by sketching the concepts map of your dataset concepts.
- ***Step 2:*** Create network schema by selecting the data attributes participating in forming in the flow - if known-, or sketch the whole concept map
- ***Step 3:*** Set the flows metrics you are interested in measuring.
- ***Step 4:*** Create the network visualization(s) that are compatible with the nature of the selected data attributes, flow metrics and explore them. Refer to the supporting material on how to explore networks visually. (ex. Node-link diagram to examine flow within the layout)
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 6:*** Evaluate and reflect on your observations such as the metrics chosen.
- ***Step 7:*** *if any data transformation is required*, *apply* data transformation (ex. change of link type) then return to step (2), if not proceed to following step.
- ***Step 9:*** End your exploration and journal your findings (*refer to DC4*)
- P.S. you can create a comprehensive concept map showing all concepts, or create a concept map for each examined flow. The first option opens more aspects to explore your network data through - if applicable-.

Task-Driven Exploration Strategies

TS5: Exploring by Finding a Pattern

Description: To explore data by aiming to find specific patterns and trends within a network. For example, where two clusters do not overlap.

Network Visualizations : *compatible with the targeted pattern (ex. Map for nodes within same country).*

What questions to ask?
- What does the target pattern represents in your data?
- What alternative patterns can help in revealing insights?
- Can data wrangling be applied to reveal insights?

Application : *any dataset*

Strategy Steps:
- ***Step 1:*** Start by sketching the concepts map of your data that involved the target pattern of nodes and links
- ***Step 2:*** Create network schema by selecting the data attributes participating in forming the expected pattern .
- ***Step 3:*** Create the network visualization(s) that are compatible with the nature of the selected data attributes and explore them. Refer to the supporting material on how to explore networks visually.(ex. Adjacency matrix for non-overlapping clusters)
- ***Step 4:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 5:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 6:*** *if any data transformation is required*, *apply* data transformation (ex. Change selected nodes) then return to step (2), if not proceed to following step.
- ***Step 7:*** *if other pattern(s) are targeted*, return to step(1) and implement it along with the following steps.
- ***Step 8:*** End your exploration and journal your findings (*refer to DC4*)
- P.S. you can create a comprehensive concept map showing all data. This option opens more aspects to explore your network data through - if applicable-.

Task-Driven Exploration Strategies

TS6: Exploring through a Visual Image

Description: To explore data by aiming for a specific visual image of the network the analyst aims for.

Network Visualizations : *compatible with targeted visual image.*

What questions to ask?

- What does the visual image represent in your data analysis and exploration?
- Can nodes/link be changed to offer better understanding of the data?
- Can alternative visualization serve the goal better?

Application : *any dataset*

Strategy Steps:

- ***Step 1:*** Start by sketching the concepts map of the visual image you had.
- ***Step 2:*** Ensure that each concept in your concepts map is represented by the data attributes you have in your data. Otherwise, look for alternative attribute(s) or representation of the visual image.
- ***Step 3:*** Create network schema by selecting the data attributes participating in your concept map .
- ***Step 4:*** Create the network visualization(s) that are compatible with the nature of target visual image and explore it. Refer to the supporting material on how to explore networks visually.
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 6:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 7:*** *if any data transformation is required*, *apply* data transformation then return to step (2/3), if not proceed to following step. (ex. change of nodes/links)
- ***Step 8:*** *if other visual image(s) exist*, return to step(1) and implement it along with the following steps.
- ***Step 9:*** End your exploration and journal your findings (*refer to DC4*)
- P.S. you can create a comprehensive concept map showing all target visual images, or create a concept map for each visual image. The first option opens more aspects to explore your network data through - if applicable-.

Task-Driven Exploration Strategies

TS7: Exploring by Comparing and Contrasting

Description: To explore by comparing differences and contrasting similarities between networks of different data or different filters applied.Such in comparing ego-networks.

Network Visualizations : *compatible with data attributes types of the schema.*

What questions to ask?

- What are the differences/similarities between the *networks* visualizations of same type?
- What are main metrics to compare between (ex. Central nodes in a node-link diagram)?

Application : *two or more networks (e.g. of same data attributes but collected from different stakeholder types)*

Strategy Steps:

- ***Step 1:*** Start by sketching the concepts map of your dataset(s).
- ***Step 2:*** Create network schema by selecting the data attributes participating in the target network. P.S. Some data will require one network where data filters can help in examining different sub-networks (e.g. ego-networks), and some will require creating a network for each dataset sharing the same schema with different data.
- ***Step 4:*** Create the network visualization(s) that are compatible with the nature of your networks and explore them. Refer to the supporting material on how to explore networks visually. (e.g. Node-link diagrams for ego-networks central nodes)
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 6:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 7:*** *if any data transformation is required*, *apply* data transformation then return to step (2), if not proceed to following step. (e.g. changing the linked nodes)
- ***Step 8:*** *if other dataset(s) exist*, return to step(1) and implement it along with the following steps.
- ***Step 9:*** End your exploration and journal your findings (*refer to DC4*)

Task-Driven Exploration Strategies

TS8: Exploring through Association and Co-existence

Description: To explore data by looking into association between specific nodes. This could be in general or under specific conditions (e.g. in specific era -time- or area - location -)

Network Visualizations : *depending on the criteria investigated (ex. Coexistence in geographical area → maps)*

What questions to ask?

- What is the suitable visualization to explore? (ex. matrix)
- What is the best selection of nodes /links to investigate the association ?

Application : *any dataset that association is looked into (e.g. bones and sites in archaeology)*

Strategy Steps:

- ***Step 1:*** Start by sketching the concepts map of your data concepts.
- ***Step 2:*** Decide on what nodes and links will create your network.
- ***Step 3:*** Create network schema by the selected data attributes. .
- ***Step 4:*** Create the network visualization(s) that are compatible with the nature of the selected data attributes and explore them. Refer to the supporting material on how to explore networks visually. (ex. Adjacency matrix looking for clusters of bones at the same site)
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 6:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 7:*** *if any data transformation is required*, *apply* data transformation then return to step (2), if not proceed to following step. (ex. Projection of nodes)
- ***Step 8:*** *if there are other association(s) to investigate*, return to step(1) and implement it along with the following steps.
- ***Step 9:*** End your exploration and journal your findings (*refer to DC4*)
- P.S. you can create a comprehensive concept map showing all data, or create a concept map for each investigated association. The first option opens more aspects to explore your network data through - if applicable-.

Task-Driven Exploration Strategies

TS9: Exploring by Detecting Anomalies

Description: To explore data aiming to find anomalies and unexpected patterns, appearance, links, or nodes in the network.

Network Visualizations : *compatible with data attributes types of the schema.*

What questions to ask?

- What is the expected appearance/ representation of data?
- When does the unexpected representation might appear? Why? (e.g. through temporal exploration to find disconnected nodes)

Application : *any dataset*

Strategy Steps:

- ***Step 1:*** Start by sketching the expected concepts map of your dataset .
- ***Step 2:*** Describe the conditions your network should comply to (ex. One outbound link for certain type of nodes)
- ***Step 3:*** Create network schema based on your selected nodes and links.
- ***Step 4:*** Create the network visualization(s) that are compatible with the nature of the selected data attributes and explore them. Refer to the supporting material on how to explore networks visually. (ex. Node-link diagram to examine layout)
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 6:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 7:*** *if any data transformation is required*, *apply* data transformation then return to step (2), if not proceed to following step.
- ***Step 8:*** *if other possible network schema(s) exist*, return to step(2) and implement it along with the following steps.
- ***Step 9:*** End your exploration and journal your findings (*refer to DC4*)

Task-Driven Exploration Strategies

TS10: Exploring by Extracting Concepts from Various Content

Description: To explore data by extracting data that describes media, documents, and objects in general. For example, describing conditions in x-rays.

Network Visualizations : *compatible with data attributes types of the schema.*

What questions to ask?

- What are the characteristics that describe the objects?
- What are the main characteristics to be selected as nodes, links, or additional data to describe my data? Based on my goal

Application : *any objects that can be described*

Strategy Steps:

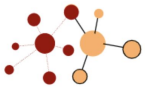
- ***Step 1:*** *Start by extracting features and characteristics that describe your objects.*
- ***Step 2:*** Start by sketching the concepts map of your data.
- ***Step 3:*** Create network schema by selecting the including: the proposed nodes, link and their attributes.
- ***Step 4:*** Create the network visualization(s) that are compatible with the nature of the selected data attributes and explore them. Refer to the supporting material on how to explore networks visually. (ex. Node-link diagram to show the association between diseases and symptoms)
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 6:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 7:*** *if any data transformation is required*, *apply* data transformation then return to step (3), if not proceed to following step. (ex. Change the selected data attribute to represent the node)
- ***Step 8:*** End your exploration and journal your findings (*refer to DC4*)

C.4 Network Schema Creation

Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

From Concept Maps to **Network Schema**



Visual Network Exploration

Coaching Program

What do we mean by
Network Schema?

Visual Network Exploration
Coaching Program

What is a Network Schema?

A network schema defines how data is organized to form a network. This includes:

- The **type of table(s)** : link table or/and node table (*Required*).
- **Values** selected to **represent nodes** (*Required*).
- **Directionality of a link**: directed or undirected.
- Association of time/location with links.
- Categorical values that can be used to specify node/link types.

A network schema should at least contain the description of a link between two nodes.

From a Concept Map to *a **Network Schema***

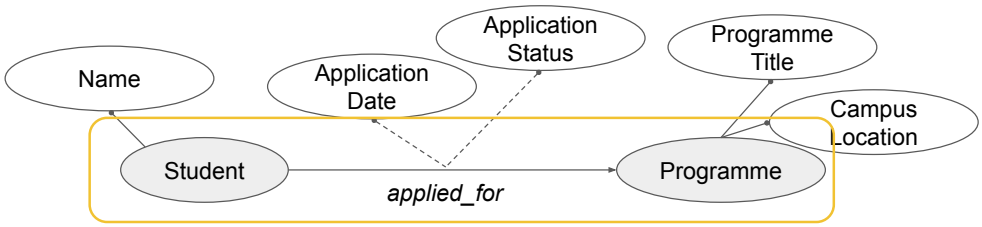
Visual Network Exploration
Coaching Program

Mapping Guidelines from Concept Maps to Schema

If your concept map contains:

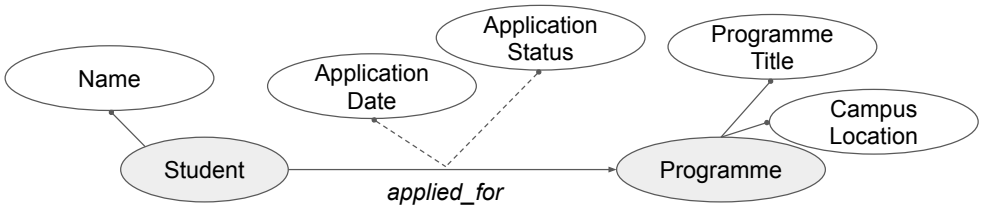
- **a link between two or more nodes** (concepts)⇒ create a ***Link Table***

  (Each row in a link table should describe a link including nodes, link type,..etc)



Visual Network Exploration
Coaching Program

Mapping Guidelines from Concept Maps to Schema



Node 1	Node 2			
Student Name	Programme Title			→ *Which student applied to which programme?*

Node 1	Node 2			
Student Name	Programme Location			→ *Which campus students applied to?*

Node 1	Node 2	Link Date	Link Type	
Student Name	Programme Title	Application Date	Application Status	→ *Filter students applications by status and date*

Node 1	Node 2	Link Type	
Student Name	Programme Location	Programme Title	→ *Filter students across campus location by programme*

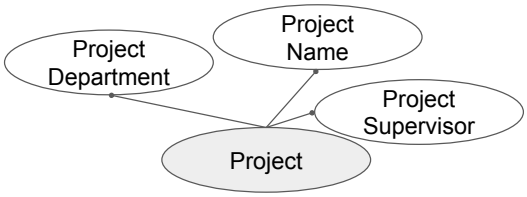
Visual Network Exploration Coaching Program

Mapping Guidelines from Concept Maps to Schema

If your concept map contains:

- **a node** (a concept with attributes) **but with no links⇒** create a ***Node Table***

  (Each row should describe a node's connections to other values)



Node Label	Relation 1	Relation 2
Project Name	Project Department	Project Supervisor

Visual Network Exploration Coaching Program

Mapping Guidelines from Concept Maps to Schema

If your concept map contains:

- both of cases mentioned previously⇒ ***create both Link Table & Node Table***
- *P.S. additional supporting tables can be provided such as:*
  - ***Node types table:*** *specifies the type of each node which allows filtering data based on type*
  - ***Geolocations table:*** *specifies the longitude and latitude of each node.*

Visual Network Exploration
Coaching Program

Shall I start Creating Table(s) for all of my Concept Map?

You can start creating a network schema:

- For the ***whole concept map*** :
  - and ***explore*** *your network* using any of the strategies of your choice such as *in free-form strategies or technique-based strategies.*
  - Mapping the whole concept map works more for concept maps with few data attributes.
- For ***part of the concept map*** :
  - And ***exploring*** part of the concept map based on a specific research question/ goal, or an exploration strategy.
  - You can expand the network schema as you go by adding more attributes.
  - Works better for large concept maps toward a more guided exploration.

Examples on *Mapping Concept Maps into* ***Network Schema***

Reminder

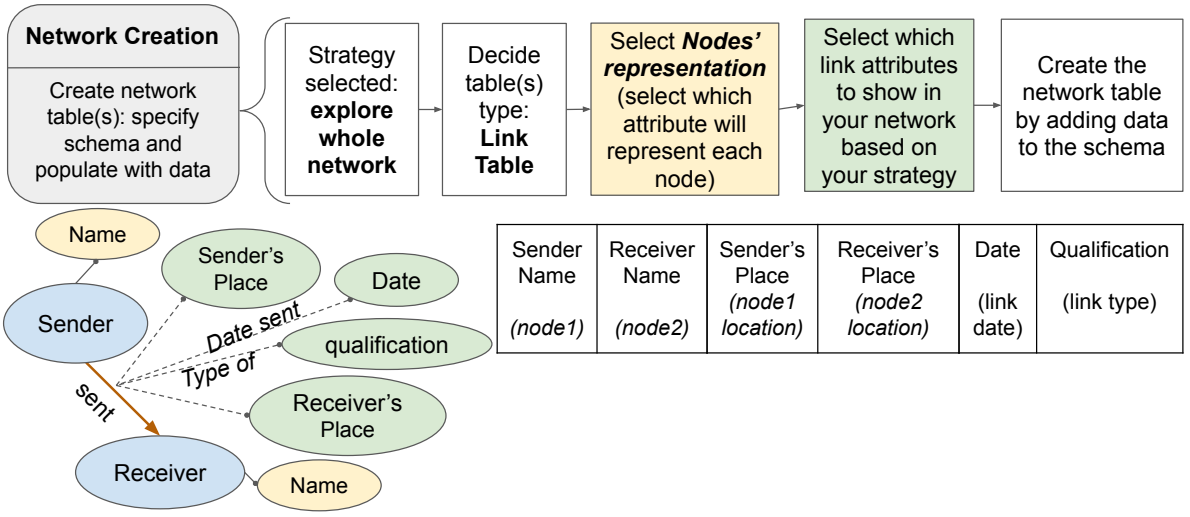
A **network** can be **explored** using **multiple** and different types of **strategies**.

Visual Network Exploration
Coaching Program

Example 1 : **Letters** Concept Map

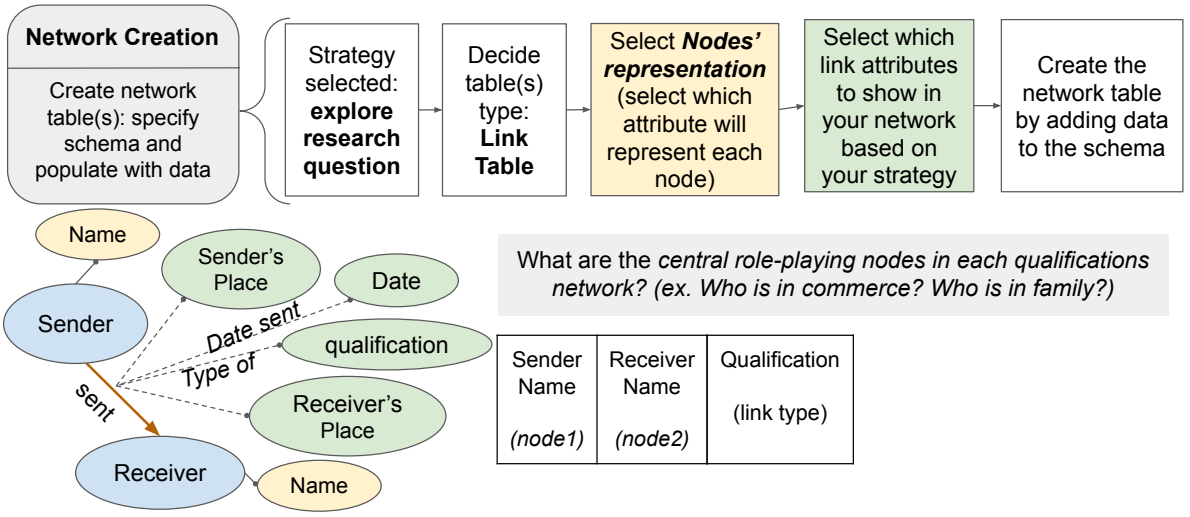
Visual Network Exploration Coaching Program

Network Schema Example 1: *Letters **whole** Concept Map*



Visual Network Exploration Coaching Program

Network Schema Example 1: *Letters **Part of** Concept Map*



Visual Network Exploration Coaching Program

Network Schema Example 1: *Letters* ***Part of Concept Map***

Network Creation

Create network table(s): specify schema and populate with data

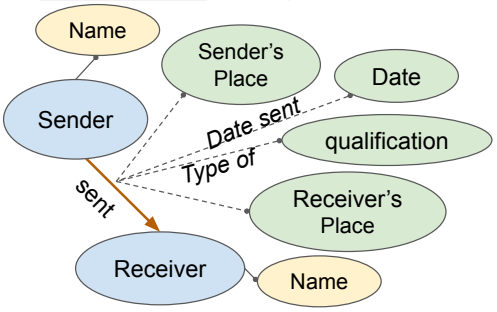
Strategy selected: **technique-based (**spatial distribution**)**

Decide table(s) type: **Link Table**

Select ***Nodes' representation*** (select which attribute will represent each node)

Select which link attributes to show in your network based on your strategy

Create the network table by adding data to the schema



What is the spatial distribution of senders and receivers ?

Sender Name (*node1*)	Receiver Name (*node2*)	Sender's Place (*node1 location*)	Receiver's Place (*node2 location*)

Visual Network Exploration Coaching Program

Example 2 : **Connections** Concept Map

Visual Network Exploration Coaching Program

Network Schema Example 2: *Connections* ***Concept Map***

Network Creation

Create network table(s): specify schema and populate with data

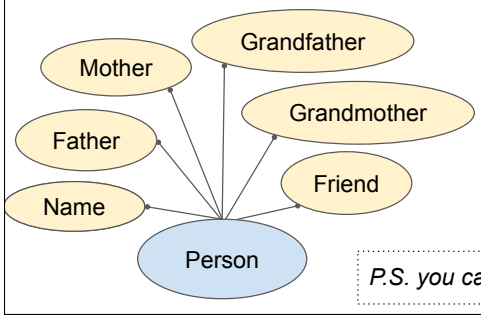
Strategy selected: **goal-based (**Association & Coexistence**)**

Decide table(s) type: **Node Table**

Select ***Nodes' representation*** (select which attribute will represent each node)

Select which link attributes to show in your network based on your strategy

Create the network table by adding data to the schema



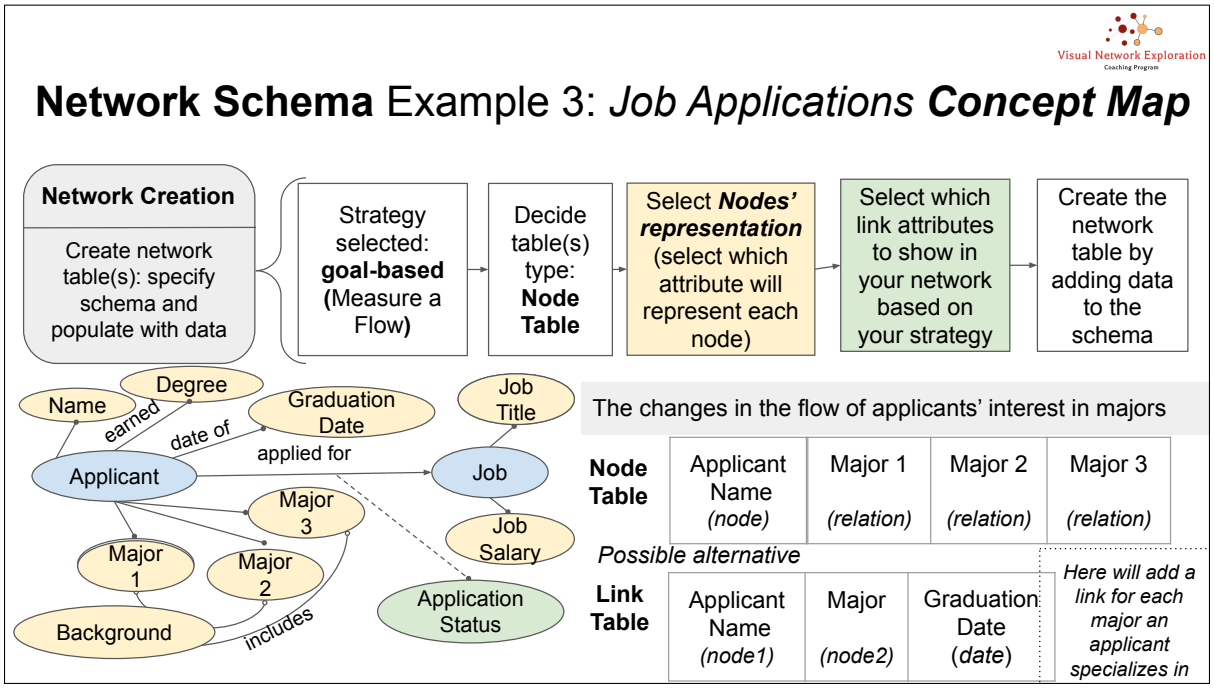
Who are the persons who share same friend or grandfather?

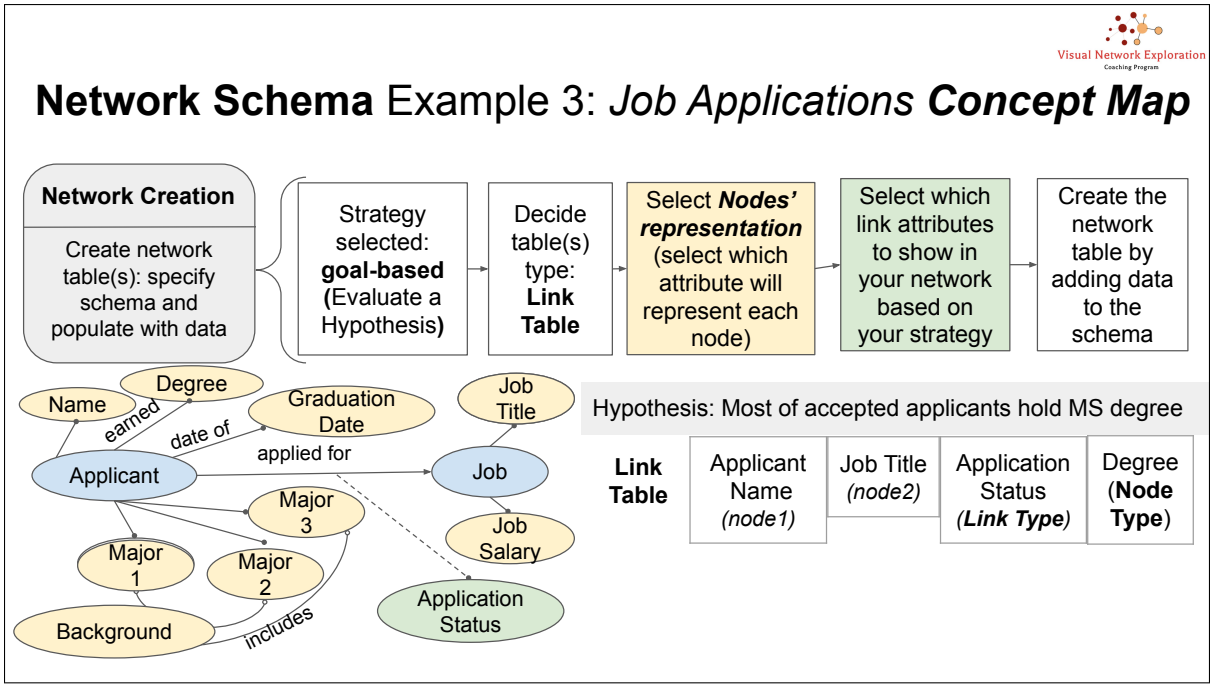
Person Name (node)	Friend (relation)	Grandfather (relation)

P.S. you can create different networks as a whole or part of the concept map

Visual Network Exploration Coaching Program

Example 3 : **Job Applications** Concept Map


Visual Network Exploration
Coaching Program
Network Schema Example 3: Job Applications Concept Map
Network Creation
Create network table(s): specify schema and populate with data
Strategy selected: goal-based (Measure a Flow)
Decide table(s) type: Node Table
Select Nodes' representation (select which attribute will represent each node)
Select which link attributes to show in your network based on your strategy
Create the network table by adding data to the schema
Name
Degree
earned
Graduation Date
date of
Job Title
applied for
Applicant
Job
Major 3
Job Salary
Major 1
Major 2
Application Status
Background
includes
The changes in the flow of applicants' interest in majors
Node Table
Applicant Name (node)
Major 1 (relation)
Major 2 (relation)
Major 3 (relation)
Possible alternative
Link Table
Applicant Name (node1)
Major (node2)
Graduation Date (date)
Here will add a link for each major an applicant specializes in


Visual Network Exploration
Coaching Program
Network Schema Example 3: Job Applications Concept Map
Network Creation
Create network table(s): specify schema and populate with data
Strategy selected: goal-based (Evaluate a Hypothesis)
Decide table(s) type: Link Table
Select Nodes' representation (select which attribute will represent each node)
Select which link attributes to show in your network based on your strategy
Create the network table by adding data to the schema
Name
Degree
earned
Graduation Date
date of
Job Title
applied for
Applicant
Job
Major 3
Job Salary
Major 1
Major 2
Application Status
Background
includes
Hypothesis: Most of accepted applicants hold MS degree
Link Table
Applicant Name (node1)
Job Title (node2)
Application Status (Link Type)
Degree (Node Type)

Visual Network Exploration Coaching Program

Network Schema Example 3: *Job Applications* ***Concept Map***

Network Creation

Create network table(s): specify schema and populate with data

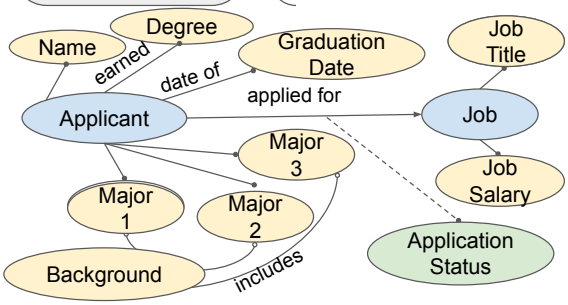
Strategy selected: **technique -based (*structure*)**

Decide table(s) type: **Link Table**

Select ***Nodes' representation*** (select which attribute will represent each node)

Select which link attributes to show in your network based on your strategy

Create the network table by adding data to the schema



Do we have clusters of applicants of a specific degree applying to a specific job? Does applicant's background or job's salary play a role?

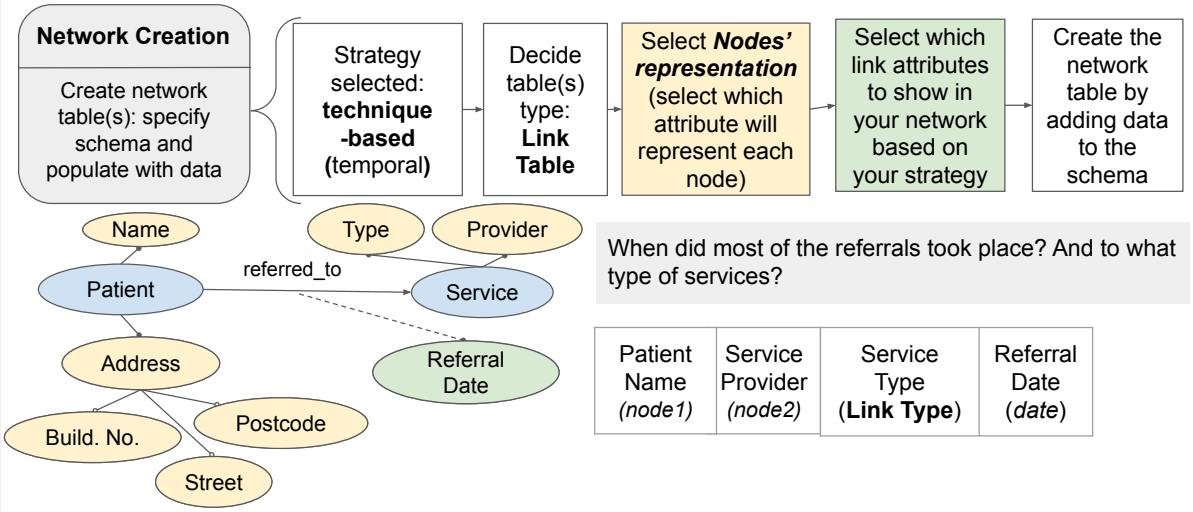
Link Table	Applicant **Degree** *(node1)*	Job Title *(node2)*	Background (**Node Type**)	Job Salary ***(Link Weight)***

Visual Network Exploration Coaching Program

Example 4 : **Patients Referral** Concept Map

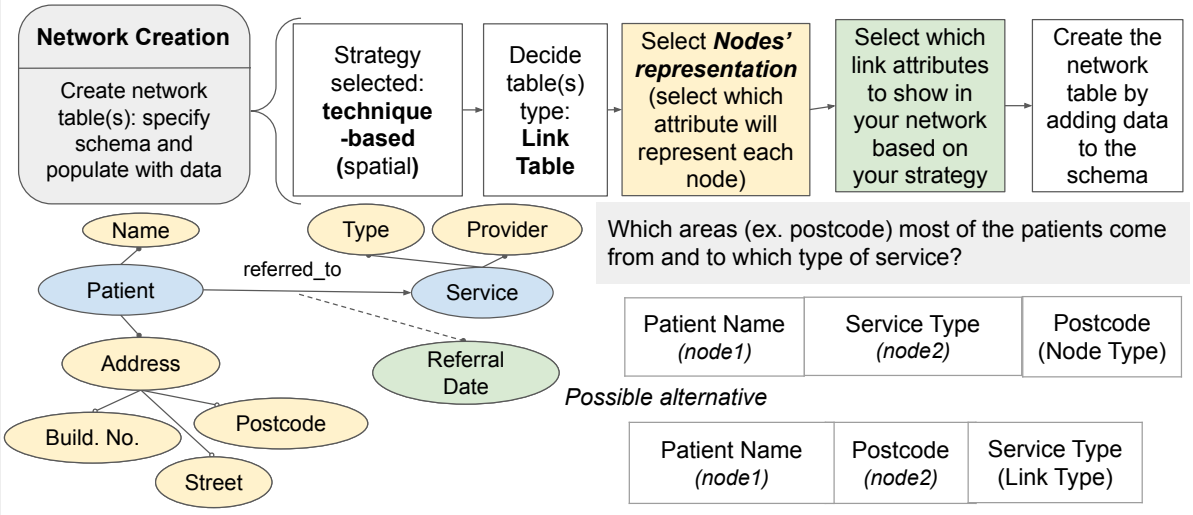
Visual Network Exploration Coaching Program

Network Schema Example 4: *Patients Referral* ***Concept Map***



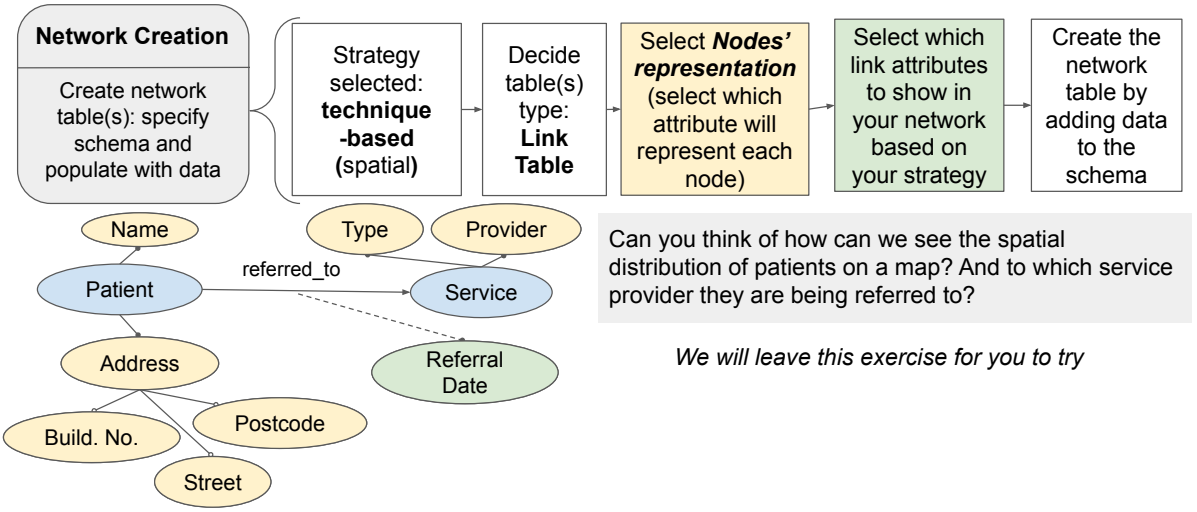
Visual Network Exploration Coaching Program

Network Schema Example 4: *Patients Referral* ***Concept Map***



Visual Network Exploration Coaching Program

Network Schema Example 4: *Patients Referral* ***Concept Map***



Can you think of how can we see the spatial distribution of patients on a map? And to which service provider they are being referred to?

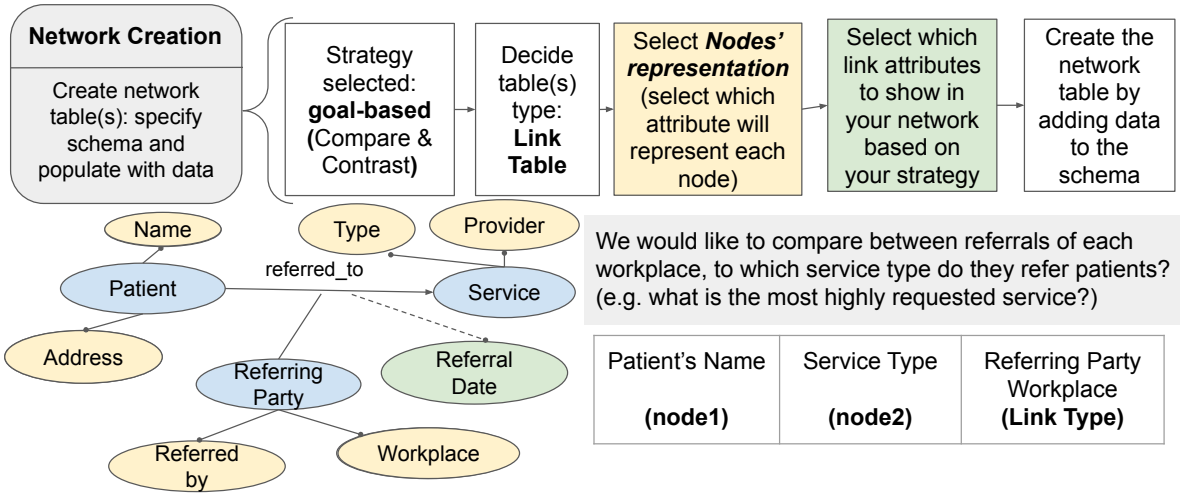
We will leave this exercise for you to try

Visual Network Exploration Coaching Program

Example 5 : Extended Patients Referral Concept Map

Visual Network Exploration Coaching Program

Network Schema Example 5: Extended *Patients Referral*



We would like to compare between referrals of each workplace, to which service type do they refer patients? (e.g. what is the most highly requested service?)

Patient's Name **(node1)**	Service Type **(node2)**	Referring Party Workplace **(Link Type)**

Reminder

A **network** can be **explored** using **multiple** and different types of **strategies**.

Visual Network Exploration Coaching Program

From Concept Map to Network Schema (Recap)

- Based on strategy are you mapping can cover the:
  - whole concept map
  - part of the concept map
- Based on the your strategy and how you see the data through your concept map, you can decide the type of the network table to create
  - Link table
  - and/or Node table
  - Optional supporting tables:
    - Node types table
    - Geolocations tables
- Remember that each features such as link/node types, time, location, link weight, and even ID help you to filter your data through your exploration journey. It depends on how you represent your data.

Visual Network Exploration Coaching Program

Supporting your Exploration through Network Schema

Network Schema Attribute *as*	**Description**	**Interactive Exploration Feature**
Node type	An attribute describes the type of a node	Filtering by node type (ex. Teacher vs student)
Link Type	An attribute describes a link (e.g. link importance)	Link Type (ex. Commerce vs family connections)
Link ID	An attributes defines the ID number of a link	Used to show the changes in the weight of link overtime
Link Time	States the time of when a link was created or took place	Filter network using temporal changes of link
Location	Geolocation of a node	Distribution of nodes over geographical areas

Visual Network Exploration
Coaching Program

Resources on Network Schema and Data Formation

- Data Formats Webpage: https://vistorian.github.io/formattingdata.html
- Data Comic (C2) : Difference between Node Tables and Link Tables

C.5 Which visualization to read and How?

C.5.0.0.1 Which visualization to read?

Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

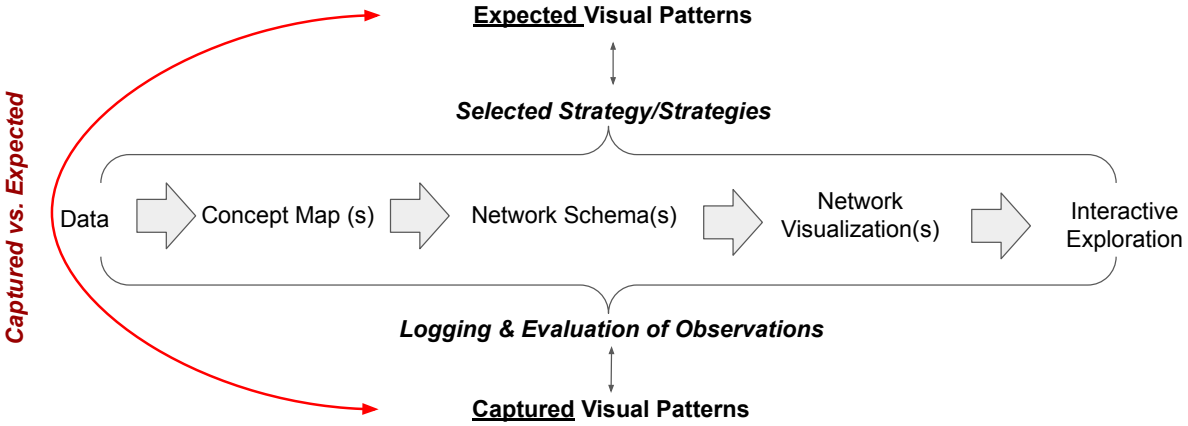
Which ***Network Visualization to Explore****? How to* ***Read*** *it?*



Visual Network Exploration

Coaching Program

What to consider when choosing **Visualizations to Explore**?



Through the visual exploration process we develop an expected visual image about our network, and how we will measure or analyze what we see through visual patterns. By comparing expected visual patterns to those captured we understand more about our decisions, the data we are exploring and the strategy.

How to decide what **Visualizations to Explore**?

	Questions to Ask	Explanation & Implementation
1	**What is your strategy(ies)?**	Some strategies can directly point you to a specific visualization to explore, whereas others might involve multiple visualizations to explore and get insights through visual patterns.
2	**What is your chosen network schema?**	Data types of attributes in your schema plays a role in the decision of which visualization to choose (ex. Including time then consider temporal exploration).
3	**Is it your first acquaintance with the dataset?**	Whether you are a beginner or not, if this your first time to explore your network start with node-link diagram to have an overview about your data. Otherwise, If you are looking for other facets to explore your data, choose visualizations supporting your strategy.
4	**Is it your first time to explore networks?**	For beginners, it's recommended to explore using node-link diagram to understand more about networks exploration and analysis concept.
5	**Are you looking for more analytical operations (ex. comparing and contrasting?)**	• If you were comparing/contrasting between different sub-/datasets (ex. ego-networks), compare between same visualization type of each network. • If you were complementing/interpreting different visualizations of the same network, make use of coordinated views or even create your own side-to-side view. • If you were comparing/contrasting between different networks (different schemas created from the same dataset) log your observations from each visualizations and build upon them.

Choosing a Visualization based on a Strategy

Open-Ended Exploration Strategies

- Any visualization that is compatible with network schema data types.
- Skim all visualizations and explore each in-depth and interactively.

Task-Driven Exploration Strategies

- Based on your selected goal, specific visualizations can help to reveal interesting patterns more (ex. Show undesired links in anaomiles detection strategy)
- The concept map and network schema you've created allows you to think about what visual patterns to expect and what visualization to explore.

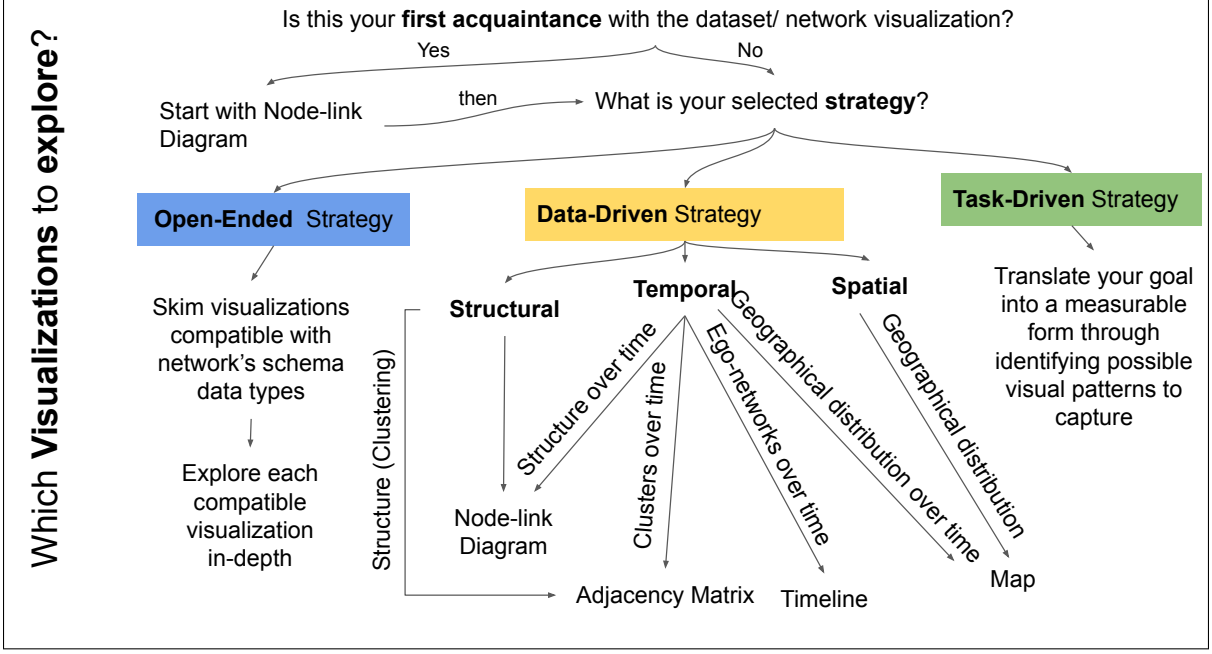
Data-Driven Exploration Strategies

- Based on the technique, a direct decision can be taken:
  - ex. for Temporal exploration of structure use node-link diagrams
  - ex. for clusters patterns adjacency matrix/node-link diagram.
- Techniques may overlap to explore a visualizations, thus multiple visualizations might help in offering different insights.

Which Visualization to choose based on ***Visual Patterns*** you're looking for?

Visual Pattern	**Node-link Diagram**	**Adjacency Matrix**	**Timeline**	**Map**
Structure	✓			Structure overmap
Clustering	✓	✓		Clusters overmap
Highly connected nodes	✓	✓	✓	✓
Bridges	✓			✓
Isolated nodes	✓	✓		Isolated nodes overmap
Neighbors	✓	✓		✓
Paths	✓			
Motifs	✓			
Ego-networks	✓		✓	
Spatial	✓ (location as a node)	✓ (location as a node)		✓ (geo-location)
Temporal	✓	✓	✓	✓
Task-driven Strategy	visualization to be chosen based on the goal and how it can be explored visually			

Which **Visualizations** to **explore**?

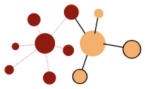


C.5.0.0.2 Exploring Node-link Diagram

Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

Exploring and Reading a ***Node-link Diagram***



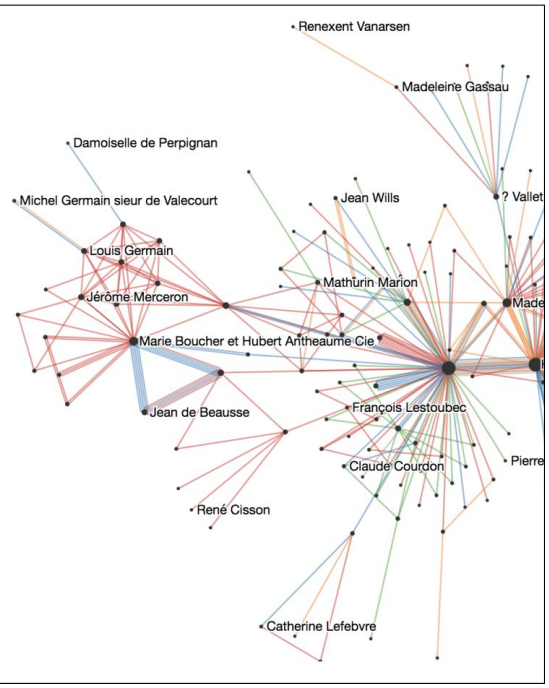
Visual Network Exploration

Coaching Program

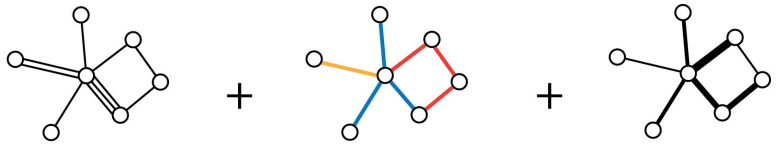
Outline

- What is visual network **exploration**
- How to **close-reading** a network diagram (visual patterns)
- Asking **questions** to your data (questions + hypotheses)
- Performing **interactive** exploration (strategies)

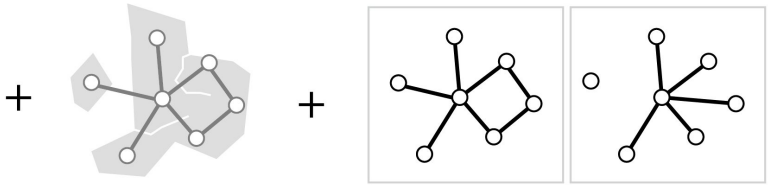
What is visual exploration?



Topology + Attributes



Multiple links + Link types + Link weight



Geography + Changes over time

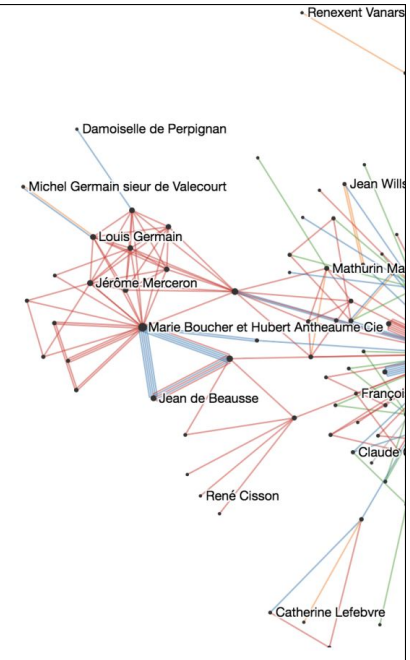
Exploratory Data Analysis (EDA)

	Univariate *Looks at one variable (data column) at a time*	**Multivariate** *Looks at two or more variables at a time to explore relationships*
Non- Graphical (quantitative and objective)	• Sample vs population distribution • Outlier detection	show the relationship between two or more variables in the form of statistics
Graphical (more qualitative and subjective)	• Sample's distributional characteristics (e.g. histogram)	Summarizes the data in a diagrammatic or pictorial way (e.g. Network topology, temporal changes)

It is almost always a good idea to perform univariate EDA on each of the components of a multivariate EDA before performing the multivariate EDA.

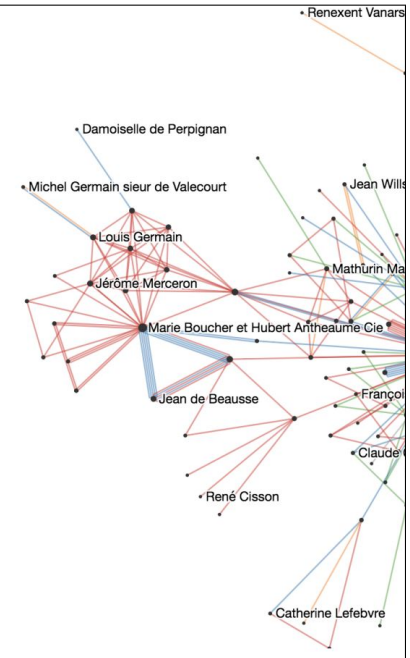
Visual Exploration:

1. Describing **visual patterns**



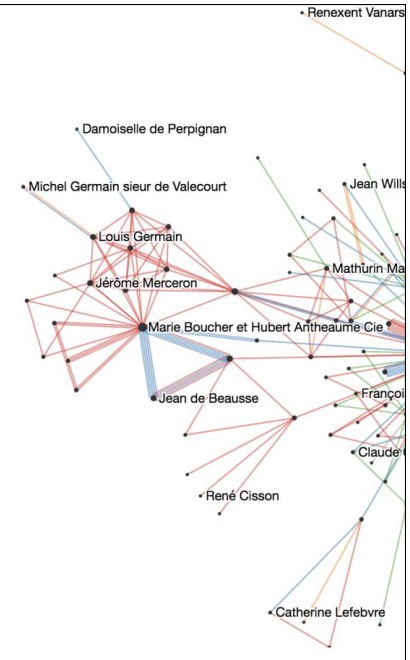
Visual Exploration:

1. Describing **visual patterns**
2. Asking **questions**



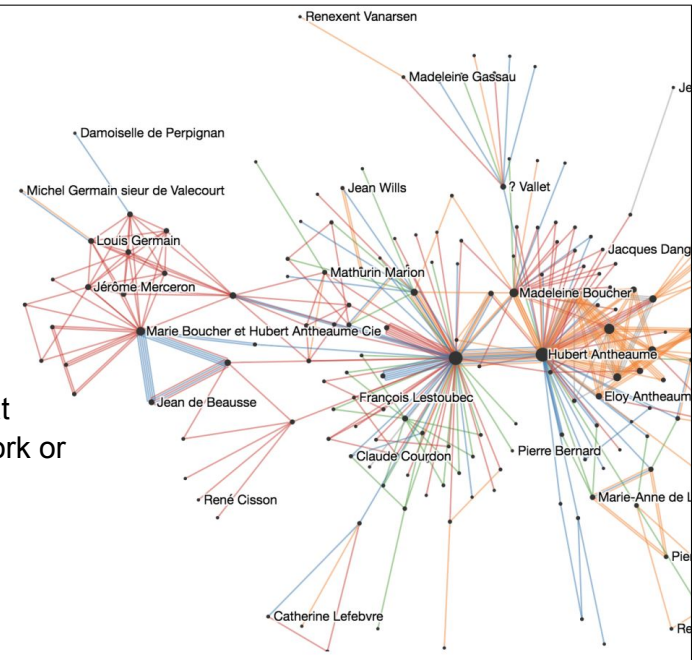
Visual Exploration:

1. Describing **visual patterns**
2. Asking **questions**
3. Generating **hypotheses**



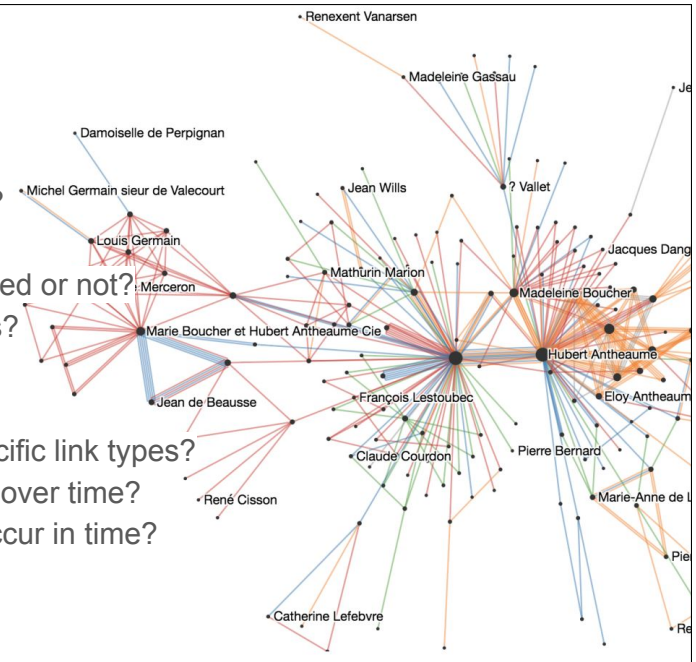
1. Visual Patterns

- Highly connected nodes
- Clusters
- Bridges (connect clusters)
- Neighbors
- Paths
- Motifs: sub-graphs that repeat themselves in a specific network or among various networks.
- Multiple links
- Attributes
- > *Questions?*



2. Questions

- What are *other* central nodes?
- What are the isolated nodes?
- Are the central nodes connected or not?
- Does the network has clusters?
- How are these clusters?
- Are clusters connected?
- What is the distribution of specific link types?
- How does the network evolve over time?
- When do specific link types occur in time?
- How do communities form?
- ....

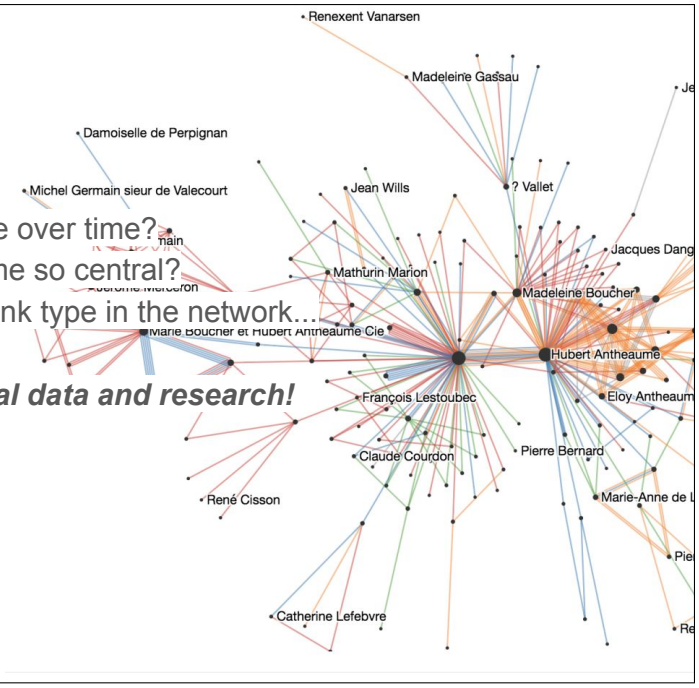


Additional Questions

- **Do we have a flows across your network?**
  - E.g. innovation can be interpreted as flow of patents or startups through relations and identifying large nodes for example with high number of relations
  - E.g. Care can be interpreted as flow of services or referrals through network relations
  - E.g. money laundering involves looking at financial flows among people and companies and identifying areas that don't make sense, either because they have more inflows than outflows or because they appear central in a network when they shouldn't be. (source: Visualizing Graph Data)

3. Hypotheses

- What makes up clusters?
- Why do specific things change over time?
- Why did specific nodes become so central?
- What is the role of a specific link type in the network...
- ...
- ***These may require additional data and research!***



Strategies

- **Overview**: Gain an overview of the entire network.
- **Zoom** : Zoom in on items of interest
- **Filter**: filter out uninteresting items (e.g., links)
- **Details-on-demand**: Select an item or group and get details when needed (e.g., neighbors)
- **History**: explore temporal evolution
- **Compare:** e.g., clusters, nodes, ...
- ...

Exercise : What did you find?

- **What are you expecting?**
  - Dense or sparse?
  - Nodes with high degree?
  - Certain topology
- **What are you NOT expecting?**
  - E.g. certain clusters
  - E.g. certain association

Assistive Data Transformations in Exploring your Network

- **Helping Techniques :**
  - Filter nodes by using centrality score called Betweenness to hide the less influential bides and focus only on the more central nodes. This includes:
    - Degree Centrality
      - Inbound Centrality
      - Outbond Centrality
    - Betweenness Centrality
    - Path Finder
  - Filter by node type
  - Filter by link type
  - Aggregate nodes

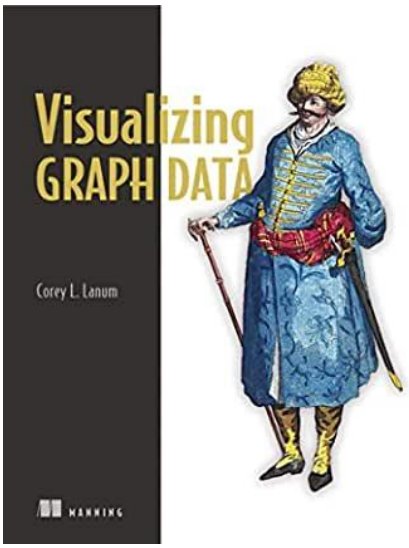
References

- Social network analysis - Introduction to structural thinking: Dr Bernie Hogan https://www.youtube.com/watch?v=2ZHuj8uBinM&t=5188s
- A Tour through the Visualization Zoo

Useful Guides

Examples on :

- Intelligence and law enforcement graphs
- Financial and online review fraud graphs
- Cyber security
- Sales and marketing graphs



Readings

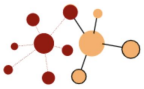
- Beck, Fabian, et al. "A taxonomy and survey of dynamic graph visualization." *Computer Graphics Forum*. Vol. 36. No. 1. 2017: http://dynamicgraphs.fbeck.com/
- Herman, Ivan, Guy Melançon, and M. Scott Marshall. "Graph visualization and navigation in information visualization: A survey." *IEEE Transactions on visualization and computer graphics* 6.1 (2000): 24-43.
- Von Landesberger, Tatiana, et al. "Visual analysis of large graphs: state-of-the-art and future research challenges." *Computer graphics forum*. Vol. 30. No. 6. Oxford, UK: Blackwell Publishing Ltd, 2011.
- Hadlak, Steffen, Heidrun Schumann, and Hans-Jörg Schulz. "A Survey of Multi-faceted Graph Visualization." *EuroVis (STARs)*. 2015.
- Schöttler, Sarah, Tobias Kauer, and Benjamin Bach. "Geographic Network Visualization Techniques: A Work-In-Progress Taxonomy." *Computer Graphics* 20 (2014): 2043-2052.: https://geographic-networks.github.io

C.5.0.0.3 Exploring Adjacency Matrix

VIS HUB Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

Exploring and Reading an ***Adjacency Matrix***



Visual Network Exploration

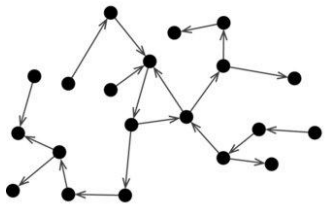
Coaching Program

Outline

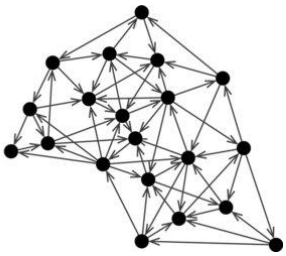
- Sparse vs. Dense Networks
- What is an Adjacency Matrix?
- Visual Encodings
- Visual Patterns

Networks Structure

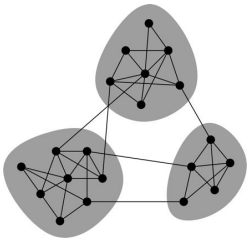
- Networks vary from very sparse (e.g., genealogy trees) to very dense (e.g., migration networks) including a locally dense category.



Sparse Network



Dense Network



Communities (Clusters)

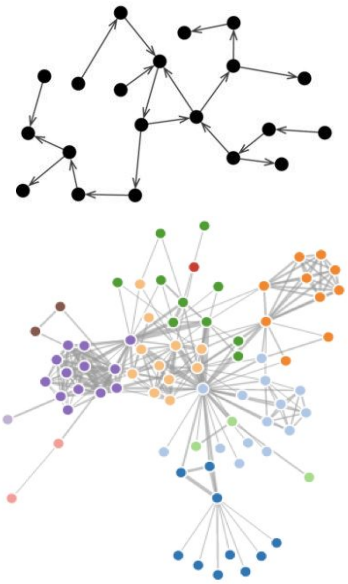
Sparse Networks

Sparse network

- Not-highly dense
- Topology can be clearly seen

Better view using Node-link diagram:

+ Path finding / following
+ Outliers
+ Disconnected components (if sparse)
+ Inbound/outbound relations
- No clear topology in Highly-dense network



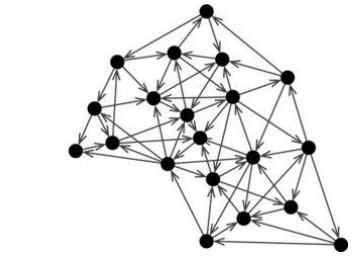
Dense Networks

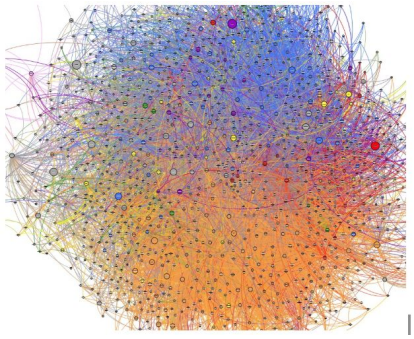
Dense Networks

- Number of edges is close to max number of edges
- Shows 1-to-1 relations

Use **Adjacency Matrix** for more readable view:

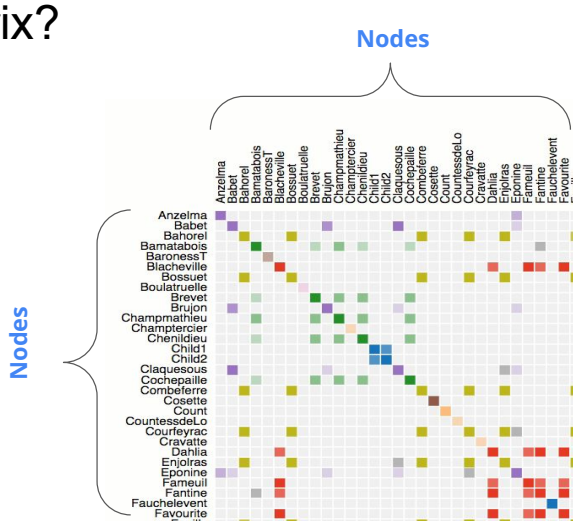
+ Clusters: identity communities
+ Missing links in clusters
+ Highly connected nodes
+ Ordering: Change order of rows and columns
- Does not show the topology





What is an Adjacency Matrix?

- An adjacency matrix is a table with rows and columns, you can represent nodes and edges.
- All rows and columns will contain the nodes names.
- Edges will be represented as the intersection of any row and column, where each cell ij represents an edge from node i to node j.



Co-Authorship Network Example (Author-to- Paper)

Undirected

Author	Paper
John	Paper A
Sara	Paper A
Kyle	Paper A
Jamal	Paper B
Sara	Paper B



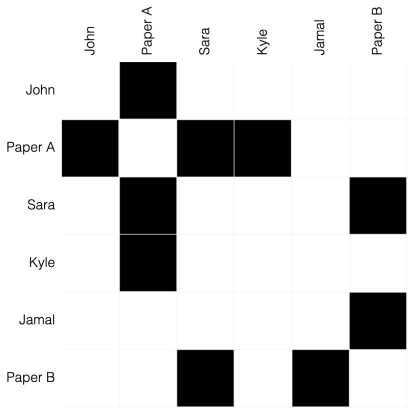
	Paper A	Paper B	John	Sara	Kyle	Jamal
Paper A	0	0	**1**	**1**	**1**	0
Paper B	0	0	0	**1**	0	**1**
John	**1**	0	0	0	0	0
Sara	**1**	**1**	0	0	0	0
Kyle	**1**	0	0	0	0	0
Jamal	0	**1**	0	0	0	0

Co-Authorship Network Example (Author-to- Paper)

Undirected

Author	Paper
John	Paper A
Sara	Paper A
Kyle	Paper A
Jamal	Paper B
Sara	Paper B





Directed vs Undirected Adjacency Matrix

Co-Authorship Network Example (Author-to- Paper)

Directed

Author	Paper
John	Paper A
Sara	Paper A
Kyle	Paper A
Jamal	Paper B
Sara	Paper B

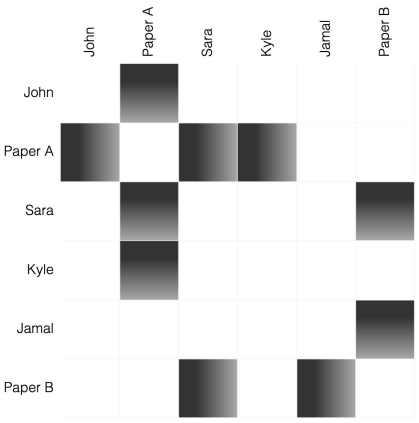


	Paper A	Paper B	John	Sara	Kyle	Jamal
Paper A	0	0	**0**	**0**	**0**	0
Paper B	0	0	0	**0**	0	**0**
John	**1**	0	0	0	0	0
Sara	**1**	**1**	0	0	0	0
Kyle	**1**	0	0	0	0	0
Jamal	0	**1**	0	0	0	0

Co-Authorship Network Example (Author-to- Paper)

Directed

Author	Paper
John	Paper A
Sara	Paper A
Kyle	Paper A
Jamal	Paper B
Sara	Paper B



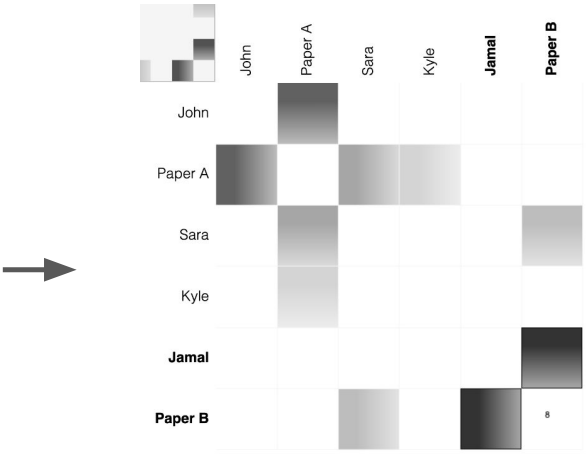
Adjacency Matrices & Weighted Graphs

- You simply replace the 1s with the weight of the edge, and 0s with Null because 0 can be a valid weight.

Co-Authorship Network Example (Author-to- Paper)

Weighted Graphs

Author	Paper	Contr/
John	Paper A	6
Sara	Paper A	3
Kyle	Paper A	1
Jamal	Paper B	8
Sara	Paper B	2



Co-Authorship Network Example (Author-to- Paper)

Weighted Graphs

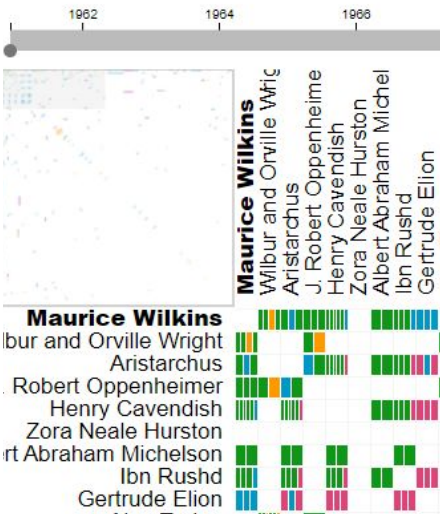
Author	Paper	Weight
John	Paper A	6
Sara	Paper A	3
Kyle	Paper A	1
Jamal	Paper B	8
Sara	Paper B	2

	Paper A	Paper B	John	Sara	Kyle	Jamal
Paper A	Null	Null	**Null**	**Null**	**Null**	Null
Paper B	Null	Null	Null	**Null**	Null	**Null**
John	**6**	Null	Null	Null	Null	Null
Sara	**3**	**2**	Null	Null	Null	Null
Kyle	**1**	Null	Null	Null	Null	Null
Jamal	Null	**8**	Null	Null	Null	Null

Visual Encodings

Visual Encodings

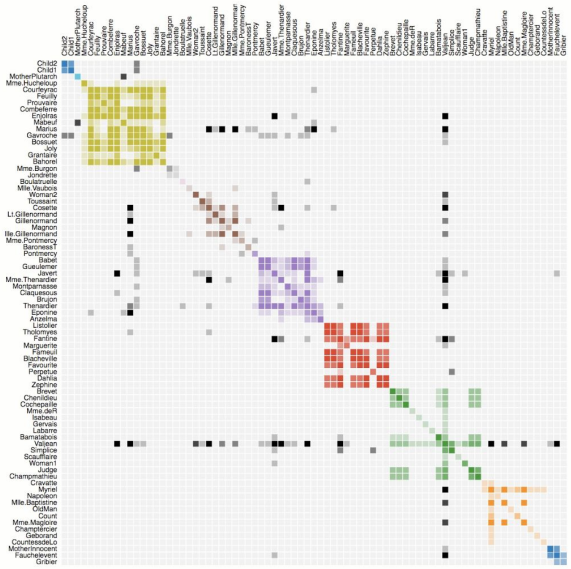
- **Squared cells** = single connections
- **Split cells** = multiple connections between nodes
- **Cell color** = connection type



Ordering

- **Label Ordering**: Defines the ordering of rows and columns
  - **Alphanumerical**: Order by name
  - **Reverse Alphanumerical**: by name, reversed
  - **Node degree**: according to number of connections, descending
  - **Similarity**: groups nodes with similar connectivity to reveal patterns in the matrix.

Les Misérables Co-occurrence
By Mike Bostock
https://bost.ocks.org/mike/miserables/



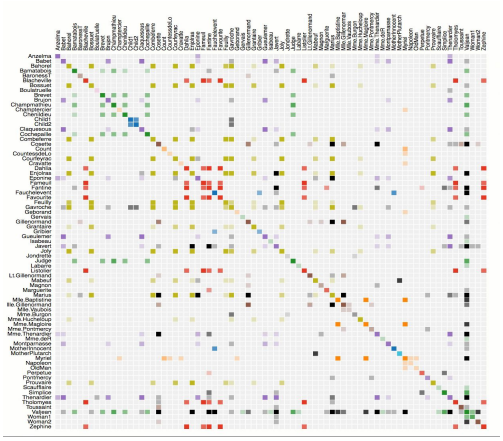
Visual Patterns

- Visual patterns inside the matrix are defined by a reordering algorithm.

- *Patterns* include:
  - Block Pattern
  - Off-diagonal Block Pattern
  - Line/Star Pattern
  - Bands Pattern
  - Noise Anti-Pattern
  - Bandwidth Anti-Pattern

Visual Patterns (Random Pattern)

- Also called *salt-and-pepper*
- The classic pattern for a matrix plot.
- It can be found whenever the row/column ordering is not able to reveal the underlying graph topology or if simply no structure exists.

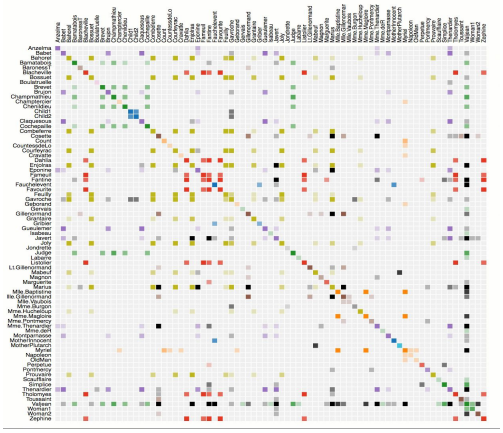


Visual Patterns (Random Pattern)

- Also called *salt-and-pepper*
- The classic anti-pattern for a matrix plot.
- It can be found whenever the row/column ordering is not able to reveal the underlying graph topology or if simply no structure exists.

What can I know about my data with this pattern?

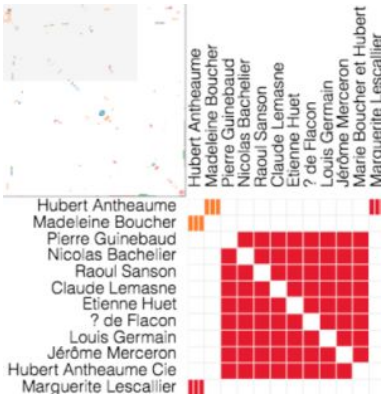
- A matrix can be *noisy* or show *structure on different levels*:
  - *locally*, i.e. for subgraphs (submatrices),
  - and **globally**, i.e. the entire graph (matrix).



Visual Patterns (Block Pattern)

Block along the diagonal of the matrix indicate:

- Groups of densely connected nodes in the network
- Each node indicated in the rows is connected to each node in the column.



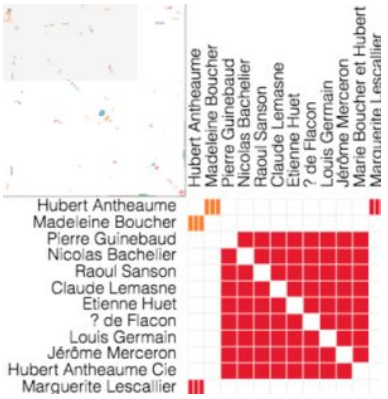
Visual Patterns (Block Pattern)

Block along the diagonal of the matrix indicate:

- Groups of densely connected nodes in the network
- Each node indicated in the rows is connected to each node in the column.

What can I know about my data with this pattern?

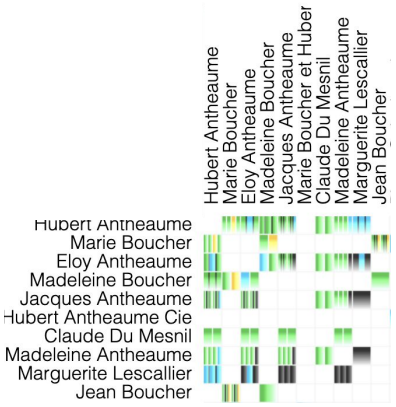
- These blocks would be referred to as *cohesive groups or clusters*.
- Clear block patterns help:
  - C**ounting clusters**,
  - **estimate cluster overlap** and
  - identify larger and smaller clusters.



Visual Patterns (Block Pattern)

Incomplete squares along the diagonal are

- almost squares but are missing some cells,
- i.e. not all nodes in the rows are connected to all nodes in the columns.
- Such incomplete squares are *clusters of densely connected nodes*



Visual Patterns (Block Pattern)

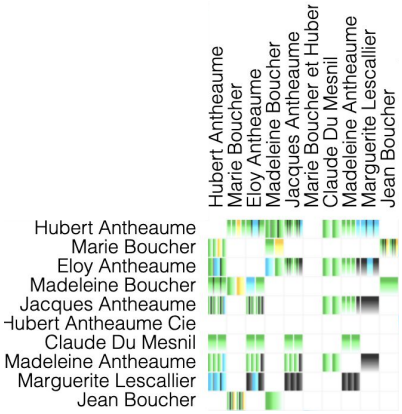
Incomplete squares along the diagonal are

- almost squares but are missing some cells,
- i.e. not all nodes in the rows are connected to all nodes in the columns.
- Such incomplete squares are *clusters of densely connected nodes*

What can I know about my data with this pattern?

Show block patterns with missing cells, meaning:

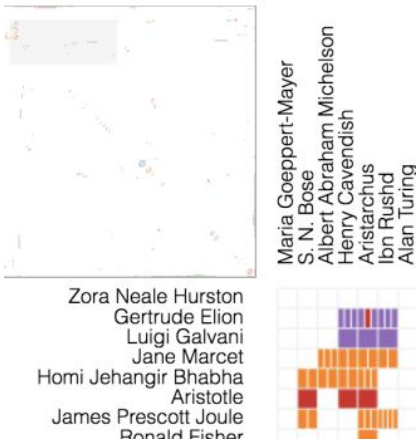
- Clusters have missing connections
- or being connected to other clusters (i.e., off-diagonal dots)



Visual Patterns (Block Pattern)

Complete or incomplete squares **off the diagonal indicate bigraphs**.

- In bigraphs, nodes from one set are only connected to nodes from the other set.
- In the figure from the left (almost all) nodes in the rows are connected to (almost all) nodes in the columns.

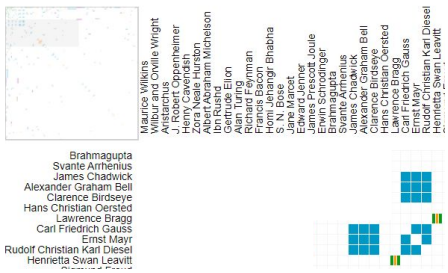


Visual Patterns (Off-Diagonal Block Pattern)

- Off-diagonal coherent areas correspond to either sub-patterns of a block pattern or relations in a bi-graph.

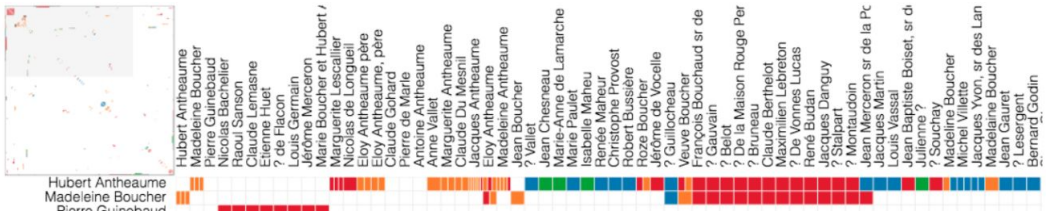
What can I know about my data with this pattern?

- Helps in understanding how groups/entities are connected.



Visual Patterns (Line/Star Pattern)

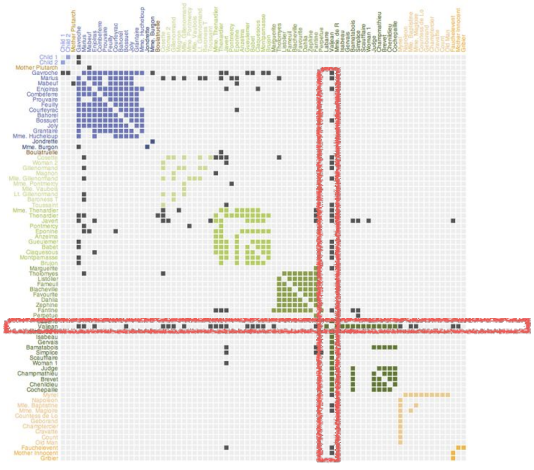
- A row or column with lots of cells indicates a highly connected node (Dense Rows or Columns) .
- In the figure below, the node Hubert Antheaume is connected almost all the nodes in the visible columns. The node can be called a **hub**.



Visual Patterns (Line/Star Pattern)

What can I know about my data with this pattern?

- This pattern helps you to analyze the reason on the general connectivity aspects within the network.
- The length of a line thereby indicates the number of connections (node degree)



Visual Patterns (Band Pattern)

- Off-diagonal continuous lines refer to paths and cycles in a network.
- They represent a set of vertices with a few connections to other vertices

What can I know about my data with this pattern?

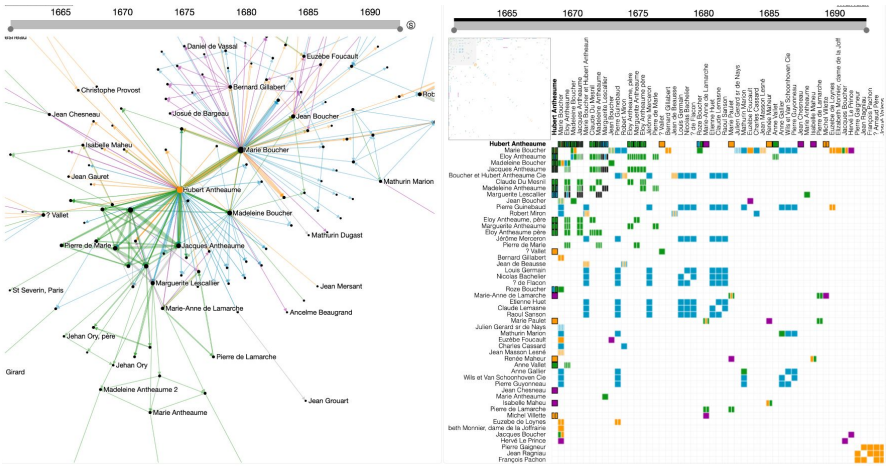
- Bands would refer connection paths and transition chains, where the width of the band visually depicts how many distinct paths could be taken through the network.



Notes

- Real world graphs exhibit a mixture of overlapping patterns appearing at different scales.
- The visual patterns describe are not always clearly distinguishable and may appear merged together.
- Reordering algorithms take into consideration different aspects of the topology, inducing different patterns.

Multiple-Coordinates View in The Vistorian



Demo

- **Zoom**
- **Pan**
- Show **connection weight**
- Changing **labelling order (Ordering algorithm)**

General Questions?

- **Homework:**
  Run an exploration session with your insights/findings from your exploration with the Node-link diagram and compare them to those you may find through the adjacency matrix.

References

- https://vistorian.github.io/visualizations.html
- Matrix Cheat Sheets
- Matrix Reordering Methods for Table and Network Visualization (inria.fr)
- Task taxonomy for graph visualization

C.5.0.0.4 Exploring Timeline and Map

Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

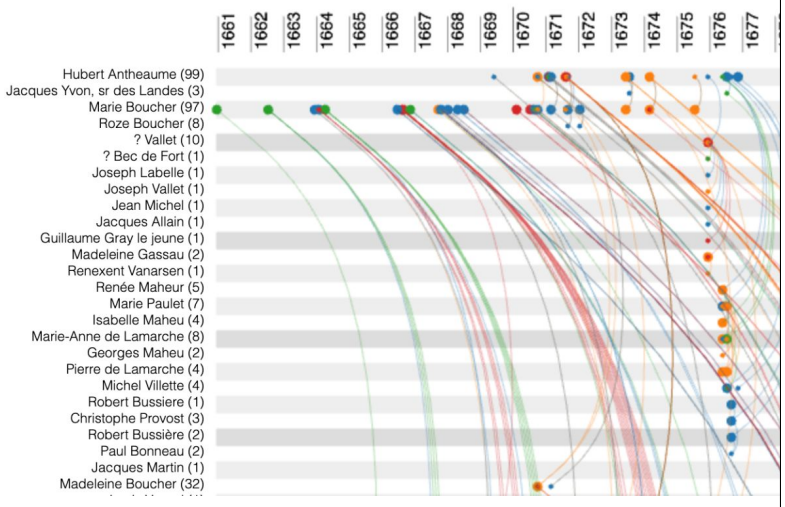
Exploring and Reading a ***Network Timeline***

Visual Network Exploration
Coaching Program

Timeline Visualization

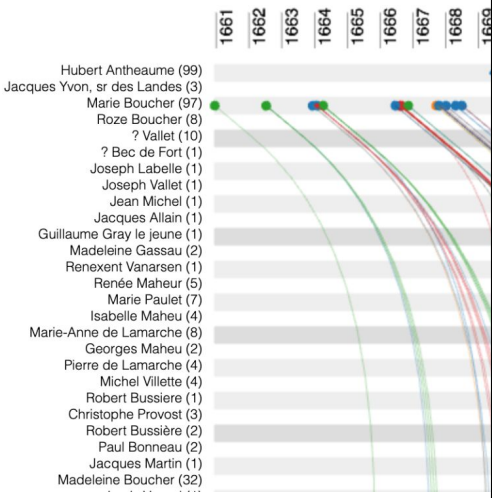
Visual Encoding:

- **Arc**
- **Arc colors**: Different colors mean different types of relation.
- **Numbers in brackets at node labels:** Node degree
- **Link start (Larger) circle and end circle (smaller) size**



Visual Patterns & Questions to Ask in a Timeline

- How does the highly connected nodes links change overtime?
- What are the links (inbound,outbound) of a specific node (ego-network)?
- How does those links change through time? (ex. Periods with no links)
- Which node initiated links?
- Are there specific nodes where most links took place?



Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

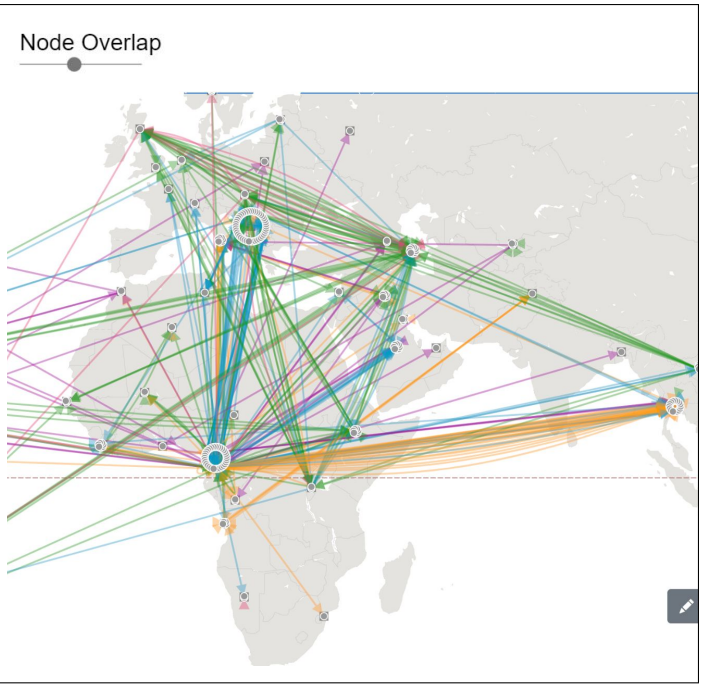
Exploring and Reading a ***Network Map***

Visual Network Exploration
Coaching Program

Map Visualization

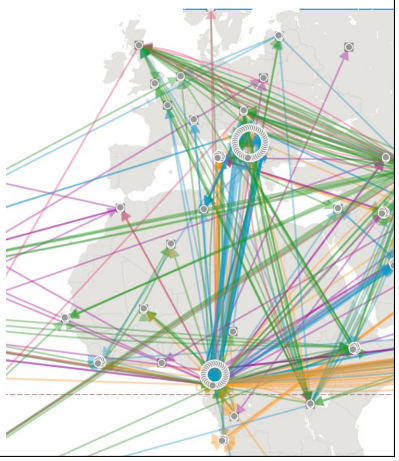
Visual Encoding:

- **Node Transparency**: Transparent nodes are free, i.e. they do not have geographic positions associated with them.
- **Node Overlap**: defines the spatial distance between nodes with the same position.

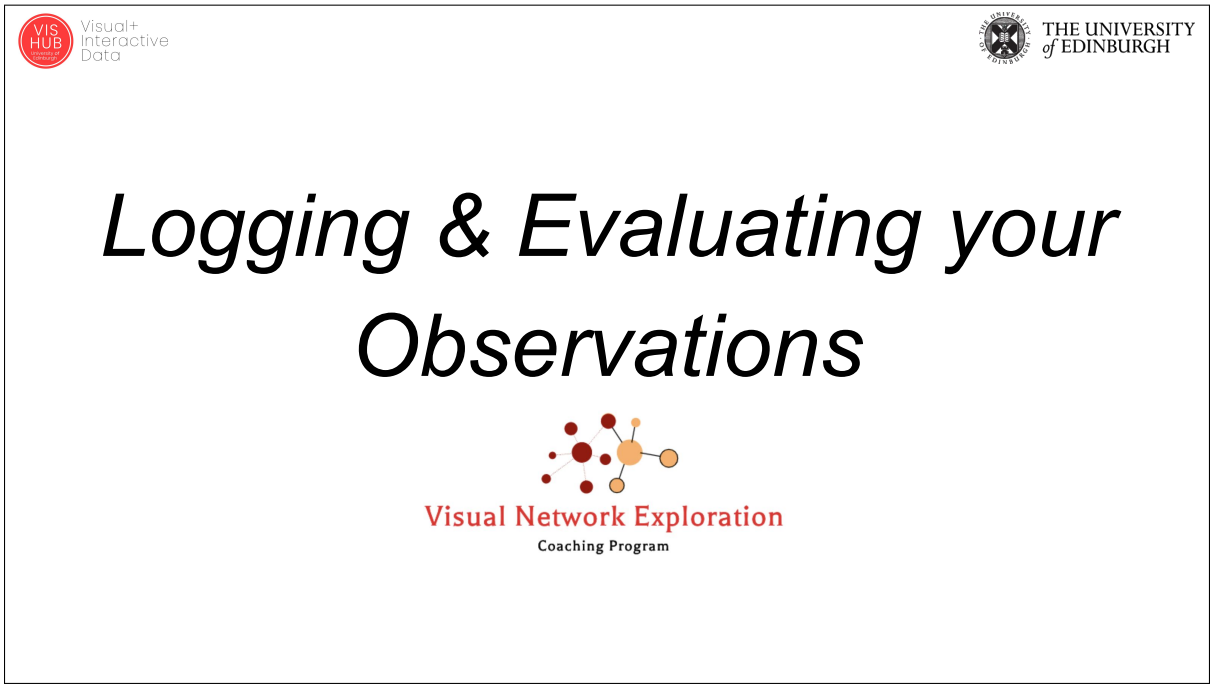


Visual Patterns & Questions to Ask in a Map

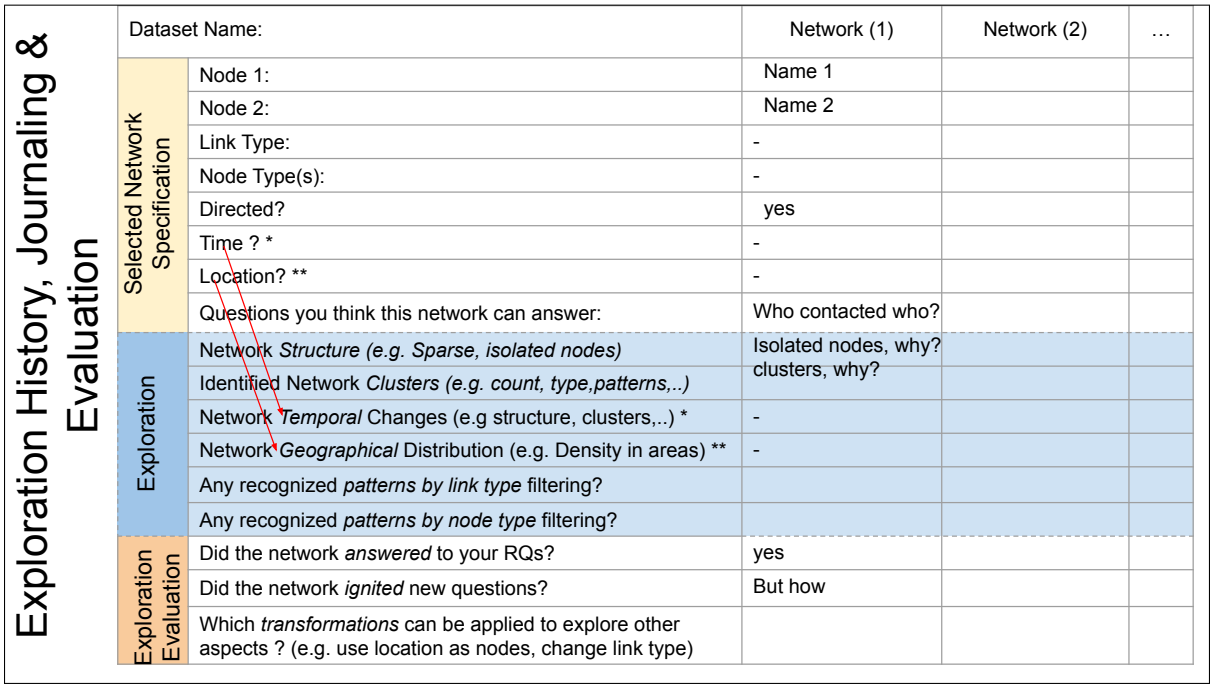
- Where are the highly connected nodes located?
- How does the distribution of nodes change over location - and maybe over overtime - ?
- What are the links (inbound,outbound) of a specific node (ego-network)?
- Do we have clusters based on location
- What is the location of first nodes/links?
- What are the links that relate clusters?



C.6 Logging & Validating your Observations

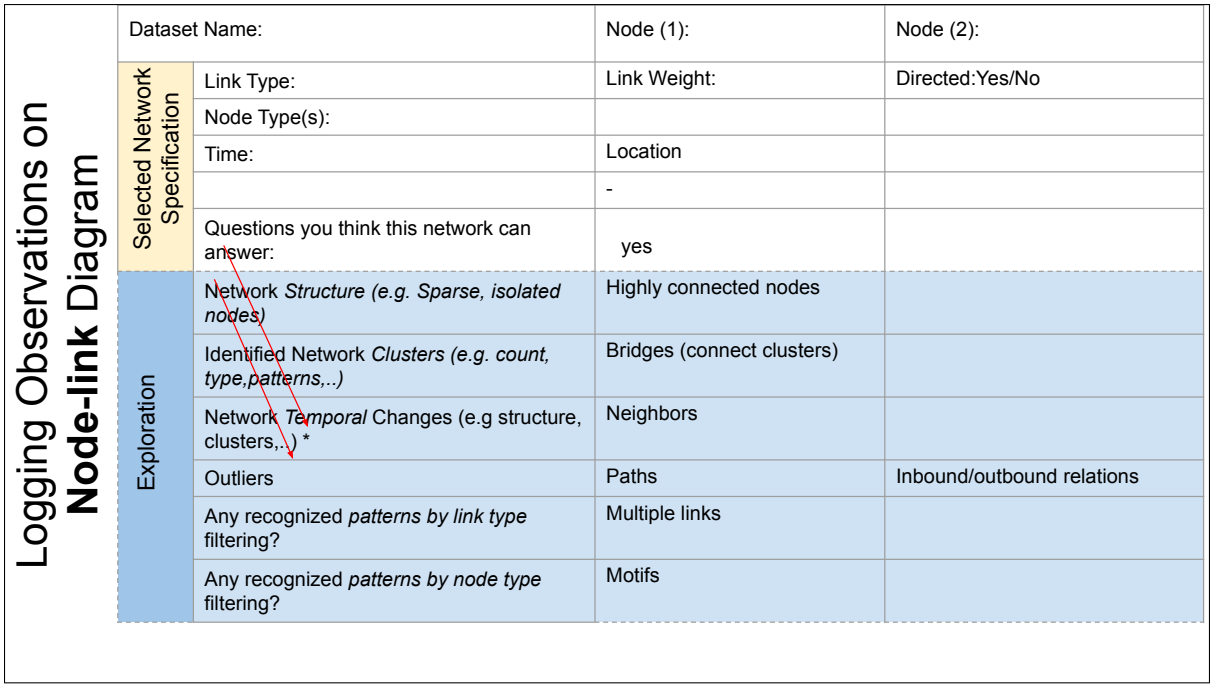


Exploration History, Journaling & Evaluation		Dataset Name:	Network (1)	Network (2)	…
	Selected Network Specification	Node 1:			
		Node 2:			
		Link Type:			
		Node Type(s):			
		Directed?			
		Time ? *			
		Location? **			
		Questions you think this network can answer:			
	Exploration	Network *Structure (e.g. Sparse, isolated nodes)*			
		Identified Network *Clusters (e.g. count, type,patterns,..)*			
		Network *Temporal* Changes (e.g structure, clusters,..) *			
		Network *Geographical* Distribution (e.g. Density in areas) **			
		Any recognized *patterns by link type* filtering?			
		Any recognized *patterns by node type* filtering?			
	Exploration Evaluation	Did the network *answered* to your RQs?			
		Did the network *ignited* new questions?			
		Which *transformations* can be applied to explore other aspects ? (e.g. use location as nodes, change link type)			



Exploration History, Journaling & Evaluation

	Dataset Name:	Network (1)	Network (2)	…
Selected Network Specification	Node 1:	Name 1		
	Node 2:	Name 2		
	Link Type:	-		
	Node Type(s):	-		
	Directed?	yes		
	Time ? *	-		
	Location? **	-		
	Questions you think this network can answer:	Who contacted who?		
Exploration	Network *Structure (e.g. Sparse, isolated nodes)*	Isolated nodes, why? clusters, why?		
	Identified Network *Clusters (e.g. count, type,patterns,..)*			
	Network *Temporal* Changes (e.g structure, clusters,..) *	-		
	Network *Geographical* Distribution (e.g. Density in areas) **	-		
	Any recognized *patterns by link type* filtering?			
	Any recognized *patterns by node type* filtering?			
Exploration Evaluation	Did the network *answered* to your RQs?	yes		
	Did the network *ignited* new questions?	But how		
	Which *transformations* can be applied to explore other aspects ? (e.g. use location as nodes, change link type)			



Logging Observations on **Node-link** Diagram

	Dataset Name:	Node (1):	Node (2):
Selected Network Specification	Link Type:	Link Weight:	Directed:Yes/No
	Node Type(s):		
	Time:	Location	
		-	
	Questions you think this network can answer:	yes	
Exploration	Network *Structure (e.g. Sparse, isolated nodes)*	Highly connected nodes	
	Identified Network *Clusters (e.g. count, type,patterns,..)*	Bridges (connect clusters)	
	Network *Temporal* Changes (e.g structure, clusters,..) *	Neighbors	
	Outliers	Paths	Inbound/outbound relations
	Any recognized *patterns by link type* filtering?	Multiple links	
	Any recognized *patterns by node type* filtering?	Motifs	

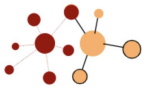
Logging Observations on Adjacency Matrix

Dataset Name:		Node (1):	Node (2):
Selected Network Specification	Link Type:	Link Weight:	Directed:Yes/No
	Node Type(s):		
	Location:		
	Time:		
	Questions you think this network can answer:		
Exploration	Number of Clusters: Are those clusters overlapping or isolated?		
		Bridges (connect clusters)	
	Network *Temporal* Changes (e.g structure, clusters,..) *	Neighbors	
		Paths	
	Any recognized *patterns by link type* filtering?	Multiple links	
	Any recognized *patterns by node type* filtering?	Motifs	

Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

Journaling *your* ***Findings***



Visual Network Exploration

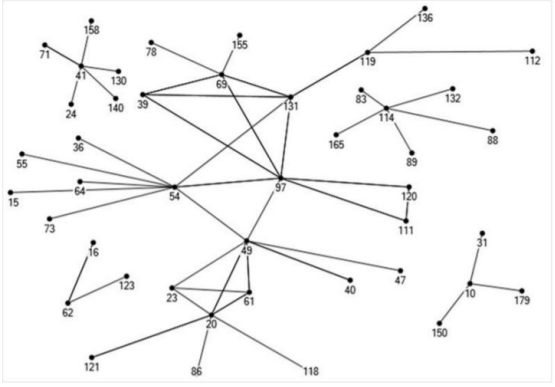
Coaching Program

Exploration History, Journaling & Evaluation

	Dataset Name:	Network (1)	Network (2)	…
Selected Network Specification	Node 1:			
	Node 2:			
	Link Type:			
	Node Type(s):			
	Directed?			
	Time ? *			
	Location? **			
	Questions you think this network can answer:			
Exploration	Network *Structure (e.g. Sparse, isolated nodes)*			
	Identified Network *Clusters (e.g. count, type,patterns,..)*			
	Network *Temporal* Changes (e.g structure, clusters,..) *			
	Network *Geographical* Distribution (e.g. Density in areas) **			
	Any recognized *patterns by link type* filtering?			
	Any recognized *patterns by node type* filtering?			
Exploration Evaluation	Did the network *answered* to your RQs?			
	Did the network *ignited* new questions?			
	Which *transformations* can be applied to explore other aspects ? (e.g. use location as nodes, change link type)			

Role of Visualizations in Network Exploration

Role	Betweeness (Unique links to others in the network)	Degree (Connected to many individuals)	Characteristics
Gatekeepers	↑ Higher	↓ Lower	• May play an important role in activity, but not much information is held on them • Removal may fragment networks
Highly visible figures	↓ Lower	↑ Higher	• May have information about many others in the network • May be involved in lots of activity in the network, but do not play a unique role
Central figures	↑ Higher	↑ Higher	• Very visible and central role • Key figures that may be focused on to fragment networks and to gather information



Social Network Analysis of local gang issues by the Home Office

Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

Visual Patterns in Visualizations

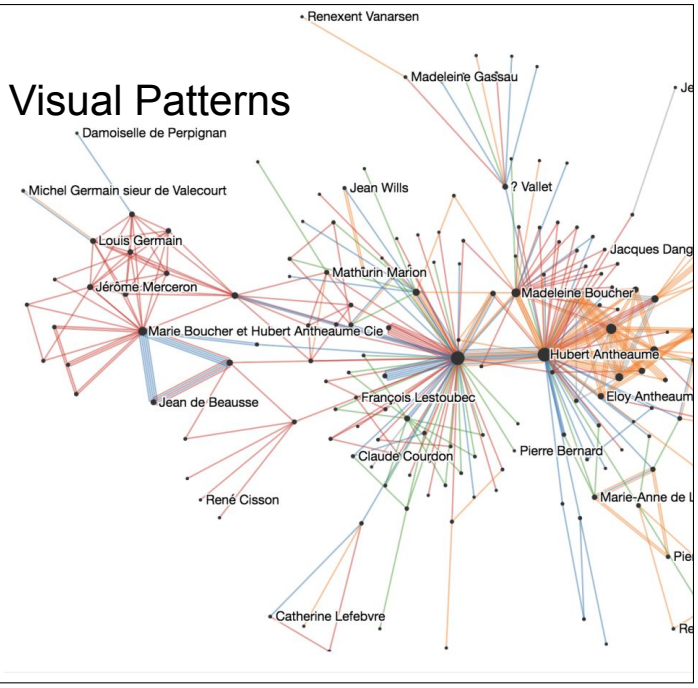
(Quick Reference)

Visual Network Exploration

Coaching Program

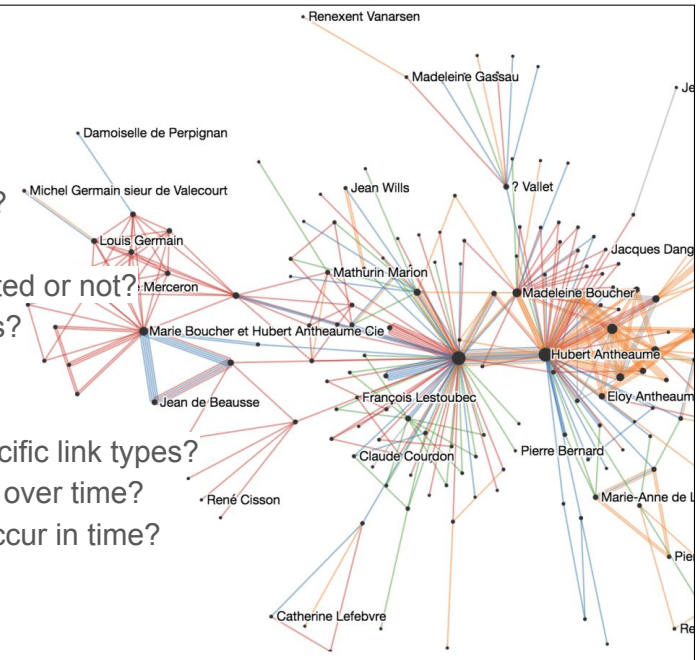
1. Node-Link Diagrams: Visual Patterns

- Highly connected nodes
- Clusters
- Bridges (connect clusters)
- Neighbors
- Paths
- Motifs
- Multiple links
- Attributes
- > *Questions?*



2. Questions

- What are *other* central nodes?
- What are the isolated nodes?
- Are the central nodes connected or not?
- Does the network has clusters?
- How are these clusters?
- Are clusters connected?
- What is the distribution of specific link types?
- How does the network evolve over time?
- When do specific link types occur in time?
- How do communities form?
- ....

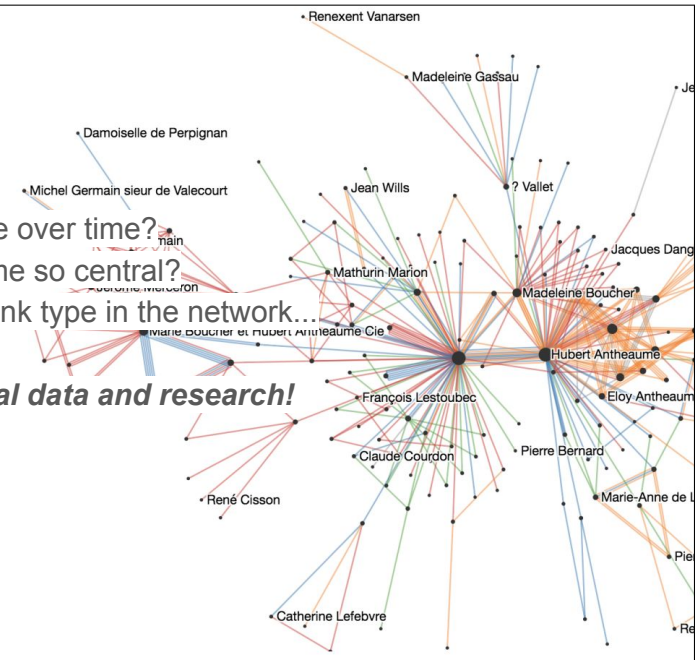


Additional Questions

- **Do we have a flows across your network?**
  - E.g. innovation can be interpreted as flow of patents or startups through relations and identifying large nodes for example with high number of relations
  - E.g. Care can be interpreted as flow of services or referrals through network relations
  - E.g. money laundering involves looking at financial flows among people and companies and identifying areas that don't make sense, either because they have more inflows than outflows or because they appear central in a network when they shouldn't be. (source: Visualizing Graph Data)

3. Hypotheses

- What makes up clusters?
- Why do specific things change over time?
- Why did specific nodes become so central?
- What is the role of a specific link type in the network...
- ...
- ***These may require additional data and research!***

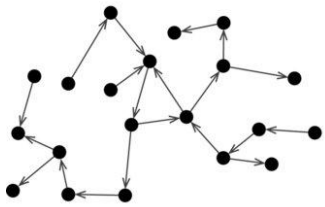


Strategies

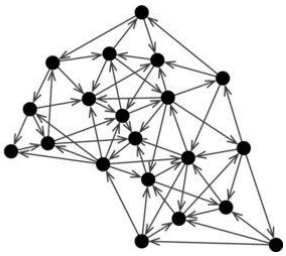
- **Overview**: Gain an overview of the entire network.
- **Zoom** : Zoom in on items of interest
- **Filter**: filter out uninteresting items (e.g., links)
- **Details-on-demand**: Select an item or group and get details when needed (e.g., neighbors)
- **History**: explore temporal evolution
- **Compare:** e.g., clusters, nodes, ...
- ...

Networks Structure

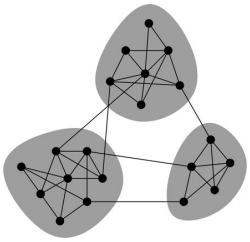
- Networks vary from very sparse (e.g., genealogy trees) to very dense (e.g., migration networks) including a locally dense category.



Sparse Network



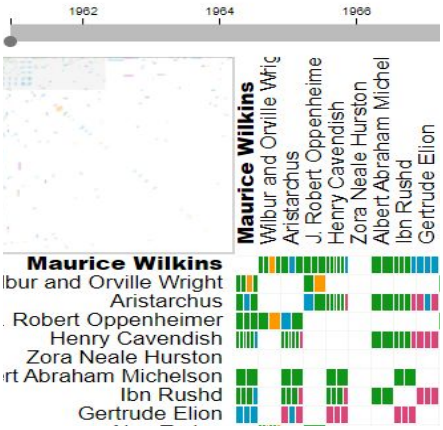
Dense Network



Communities (Clusters)

Adjacency Matrix: Visual Encodings

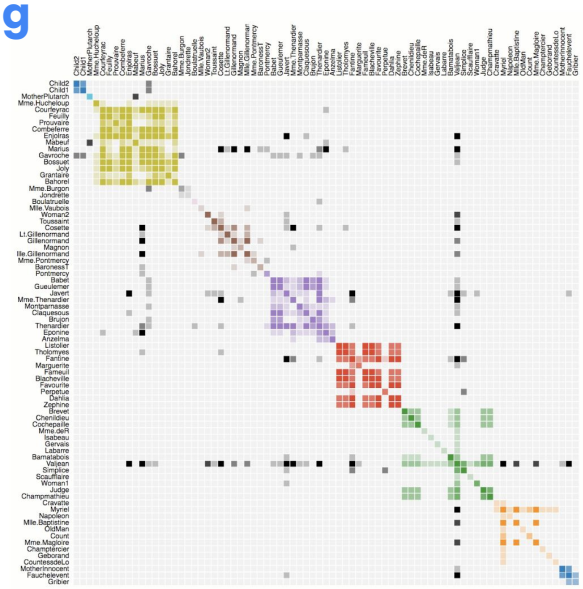
- **Squared cells** = single connections
- **Split cells** = multiple connections between nodes
- **Cell color** = connection type



Adjacency Matrix: Ordering

- **Label Ordering**: Defines the ordering of rows and columns
  - **Alphanumerical**: Order by name
  - **Reverse Alphanumerical**: by name, reversed
  - **Node degree**: according to number of connections, descending
  - **Similarity**: groups nodes with similar connectivity to reveal patterns in the matrix.

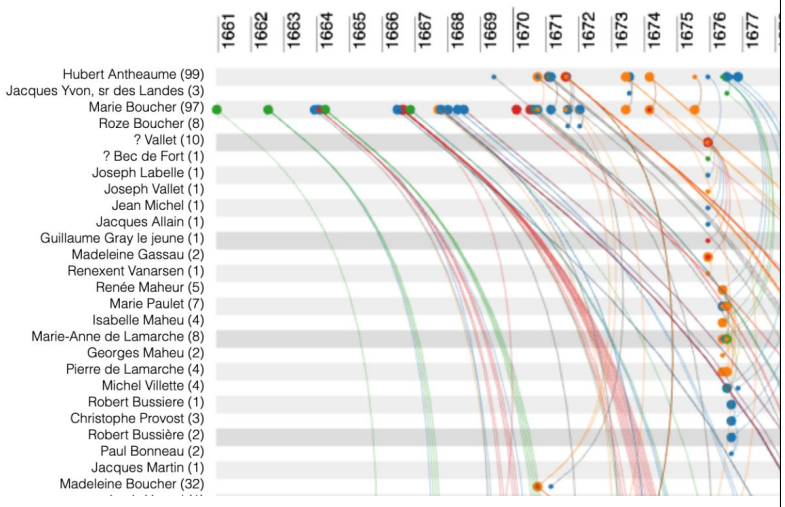
Les Misérables Co-occurrence
By Mike Bostock
https://bost.ocks.org/mike/miserables/



Timeline Visualization

Visual Encoding:

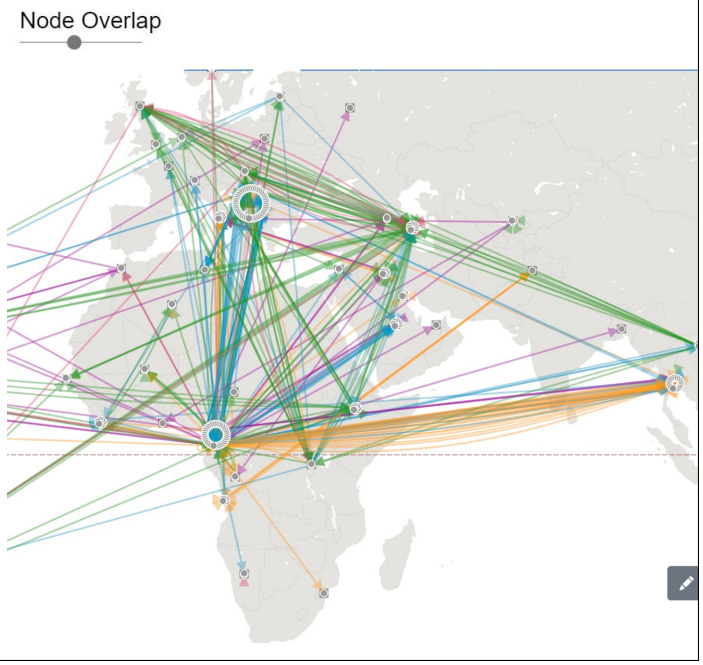
- **Arc**
- **Arc colors**: Different colors mean different types of relation.
- **Numbers in brackets at node labels:** Node degree
- **Link start (Larger) circle and end circle (smaller) size**



Map Visualization

Visual Encoding:

- **Node Transparency**: Transparent nodes are free, i.e. they do not have geographic positions associated with them.
- **Node Overlap**: defines the spatial distance between nodes with the same position.



C.7 Data Transformation and Wrangling

Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

Data Transformation & Wrangling



Visual Network Exploration

Coaching Program

Possible Network **Transformations** List

Affects	**Exploration Status & Outcome**	**Nature**	**Example**	**Proposed Transformation(s)**
Network Level	Cannot identify coexistence of specific node types	Bipartite Network	Network of nodes of two types (e.g sites and bones)	Projection of Nodes
	Nodes geographical distribution doesn't lead to clusters identification		Events across the globe	* Use locations as nodes instead of geolocation * Group locations into higher level
	No patterns were identified due the large size of the data	Interested in high-level interpretation of the network's structure .	Replace cities of a specific country, with the country name if you are looking for disease transmission in a global scale.	Reduce the size of the network through: ● Aggregating links ● Aggregating nodes ● Sampling: ○ Explore top level of data hierarchy ● Breaking-down the network into smaller groups

Possible Network **Transformations** List

Affects	**Exploration Status & Outcome**	**Nature**	**Example**	**Proposed Transformation(s)**
Community Level	Too many communities that are overlapped and no clear patterns can be identified in between.	Aiming to find how communities or sub-clusters are connected	Patients per disease group connectivity to specific symptoms vs all patients connectivity	• Create nodes out of clusters

Affects	**Exploration Status & Outcome**	**Nature**	**Example**	**Proposed Transformation(s)**
Node Level	Network has few nodes and/or links	Nodes are Textual pieces	e.g. peace agreements, artifacts documents	Extract nodes and links from within the document (NLP Tokenization)
	No patterns were identified	Examine different exploration strategies	Connection between tablet users' age and cognitive problems (kids vs. adults)	• Filter by node type • Aggregation of nodes to higher level

Possible Network **Transformations** List

Affects	**Exploration Status & Outcome**	**Nature**	**Example**	**Proposed Transformation(s)**
Node Level	Network has few nodes and/or links	Nodes are Textual pieces	e.g. peace agreements, artifacts documents	Extract nodes and links from within the document (NLP Tokenization)
	No patterns were identified	Hairball	E.g. Studying phenomenon among students from several grades and ethnicities	• Filter by node type • Aggregation of nodes to higher level • Extract different nodes

Possible Network **Transformations** List

Affects	**Exploration Status & Outcome**	**Comments**	**Example**	**Proposed Transformation(s)**
Link Level	No patterns were identified	Check your network schema and which data attribute can assist in revealing insights.	Replace the link type from grade to ethnicity or gender in a dataset of mutual friends and level of activity in a school.	● Change chosen link type ● Filter by link type ● Aggregation of link types to higher level ● Breaking-down links into links with subtypes. ● Network extrapolation: identify new links through logical rules (e.g. transitive)
	High number of redundant links	What is your point of focus or looking for?	E.g. Social network dataset	● Remove redundant links (rows) if unnecessary. ● Transform redundant links into a single link where the weight of the link=number of redundant links

Possible Network **Transformations** List

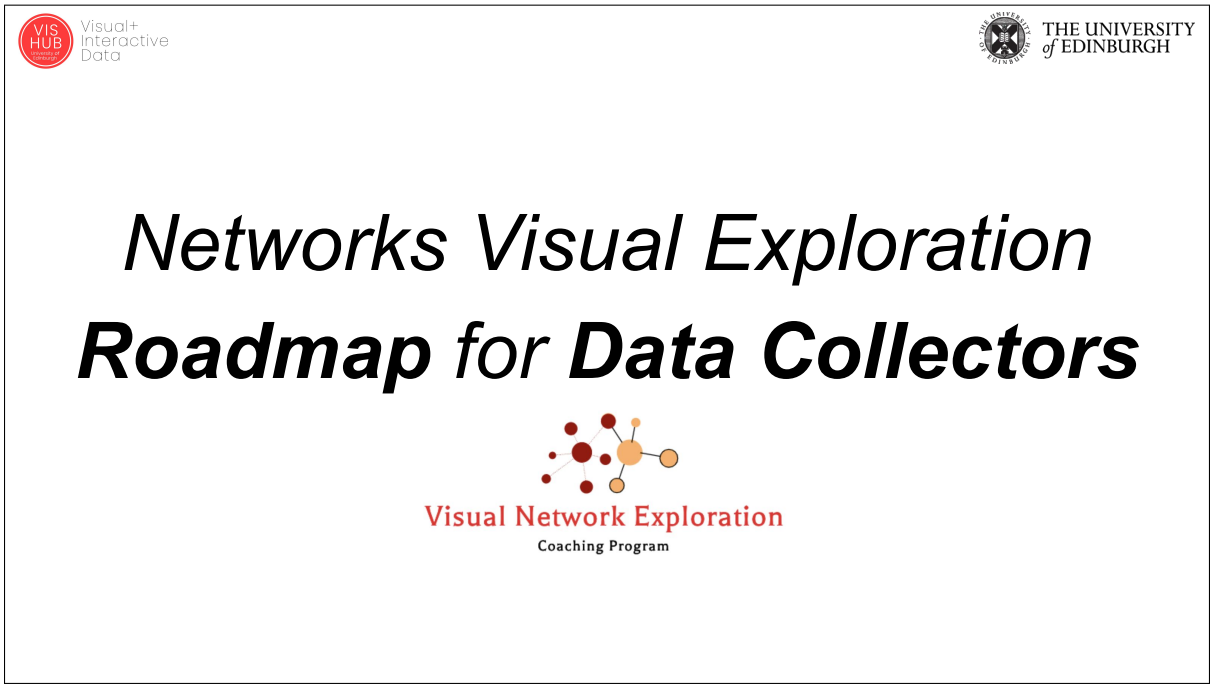
Affects	**Exploration Status & Outcome**	**Comments**	**Example**	**Proposed Transformation(s)**
Missing Data	Would like to represent missing or indeterminate values	Missing values are unknown at the moment, whereas indeterminate values are values that we expect but cannot identity it yet	E.g. we expect a person who is mostly communicating with patients but we don't the role of this person yet	Dealing with absence of value by adding a node/link representing expected value
Data Abstraction	Representing single nodes with few links	Are we interested in the node itself or what describes it?	for example, are you interested in Dr. Smith himself or his specialization ENT Doctor	Represent the node by its abstraction
	Nodes geographical distribution doesn't show clustering or interesting pattern	Are we interested in the distribution of the nodes among map? or we are more interested on finding clusters of those nodes?		Use locations as nodes

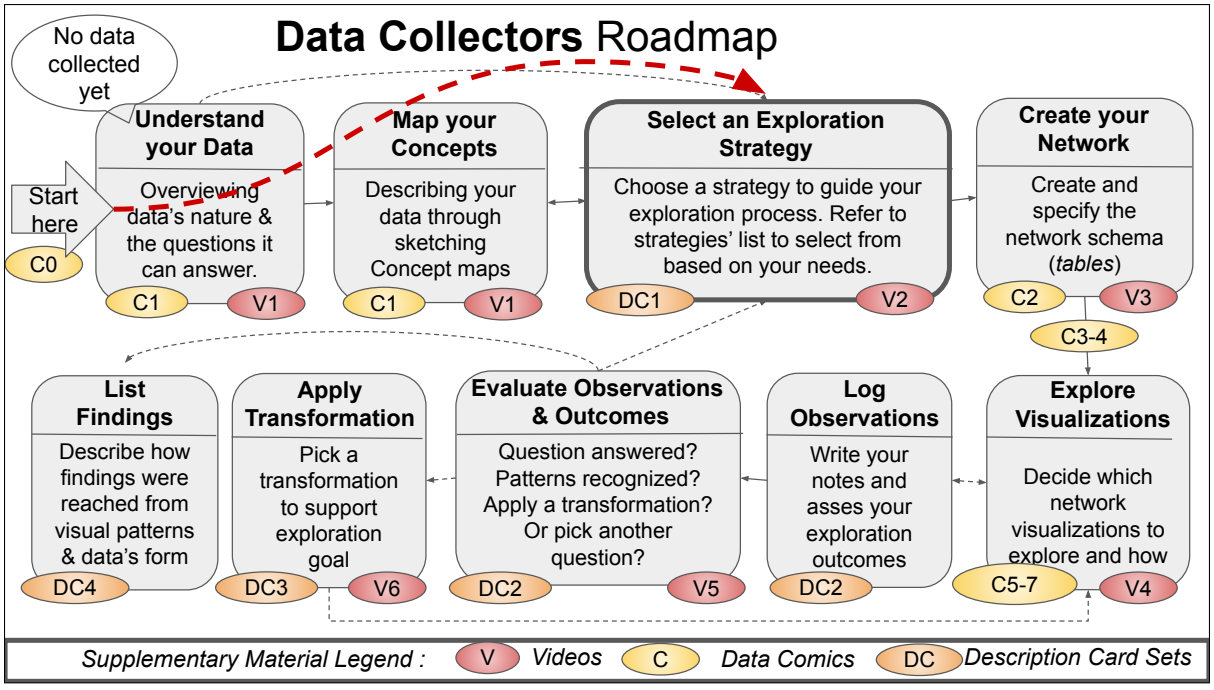
Possible Network **Transformations** List

Affects	Exploration Status/Outcome	Comments	Example	Proposed Transformation(s)
Network Level	Cannot identify coexistence of specific node types	Bipartite Network	Network of nodes of two types (e.g sites and bones)	Projection of Nodes
Node Level	Network has few nodes and/or links	Nodes are Textual pieces	Peace agreements, artifacts documents	Extract nodes and links from within the document (NLP Tokenization)
Link Level	High number of redundant links	What is your point of focus or looking for?	Social network dataset	(1) Remove redundant links (rows) if unnecessary. (2) Transform redundant links into a single link where the weight of the link=number of redundant links
Missing Data	Would like to represent missing or indeterminate values	Missing values are unknown at the moment, whereas indeterminate values are values that we expect but cannot identity it yet	In a care home dataset, to investigate the role of the person(s) who is mostly communicating with patients but the the role of such persons is not yet known	Dealing with absence of value by adding a node/link representing expected value
Data Abstraction	Most of the nodes have few links	Are we interested in the node itself or what it represents?	Are you interested in Dr. Smith himself or his specialization as an ENT Doctor	Represent the node by its abstraction

C.8 Visual Exploration Walkthroughs

C.8.1 Networks Visual Exploration Roadmap for Data Collectors


Visual+ Interactive Data
THE UNIVERSITY of EDINBURGH
Networks Visual Exploration
Roadmap for Data Collectors
Visual Network Exploration
Coaching Program


Data Collectors Roadmap
No data collected yet
Start here
C0
Understand your Data
Overviewing data's nature & the questions it can answer.
C1
V1
Map your Concepts
Describing your data through sketching Concept maps
C1
V1
Select an Exploration Strategy
Choose a strategy to guide your exploration process. Refer to strategies' list to select from based on your needs.
DC1
V2
Create your Network
Create and specify the network schema (tables)
C2
V3
C3-4
Explore Visualizations
Decide which network visualizations to explore and how
C5-7
V4
Log Observations
Write your notes and asses your exploration outcomes
DC2
Evaluate Observations & Outcomes
Question answered? Patterns recognized? Apply a transformation? Or pick another question?
DC2
V5
Apply Transformation
Pick a transformation to support exploration goal
DC3
V6
List Findings
Describe how findings were reached from visual patterns & data's form
DC4
Supplementary Material Legend :
V Videos
C Data Comics
DC Description Card Sets

Note

The Data Collectors *roadmap* follows the same track of the general exploration roadmap, but with the following choices being taken:

1. No ***dataset*** : No data has been collected yet. Thus, the goal is to create a dummy dataset to test through sketching a concept map and exploring a proposed network based on the selected strategy.
2. The ***selected strategy*** is : The selected strategy helps in deciding what data to be collected (ex. Answering a Research Question). A description card for each strategy is described in detail in its own section of this course.
3. Feel free to try and play with any other dataset or exploration strategy.

Our Example for Data Collection

Let us assume that a research center wants to collect data about their researchers' interest and what are the topics they are mostly interested in. The would like to know more about their level of expertise, how do they present their findings (ex. papers, workshops,..), and what kind of collaboration happens internally and externally with other research centers.

So, in order to do that they've decide to start with a writing down a seet of research questions to explore.

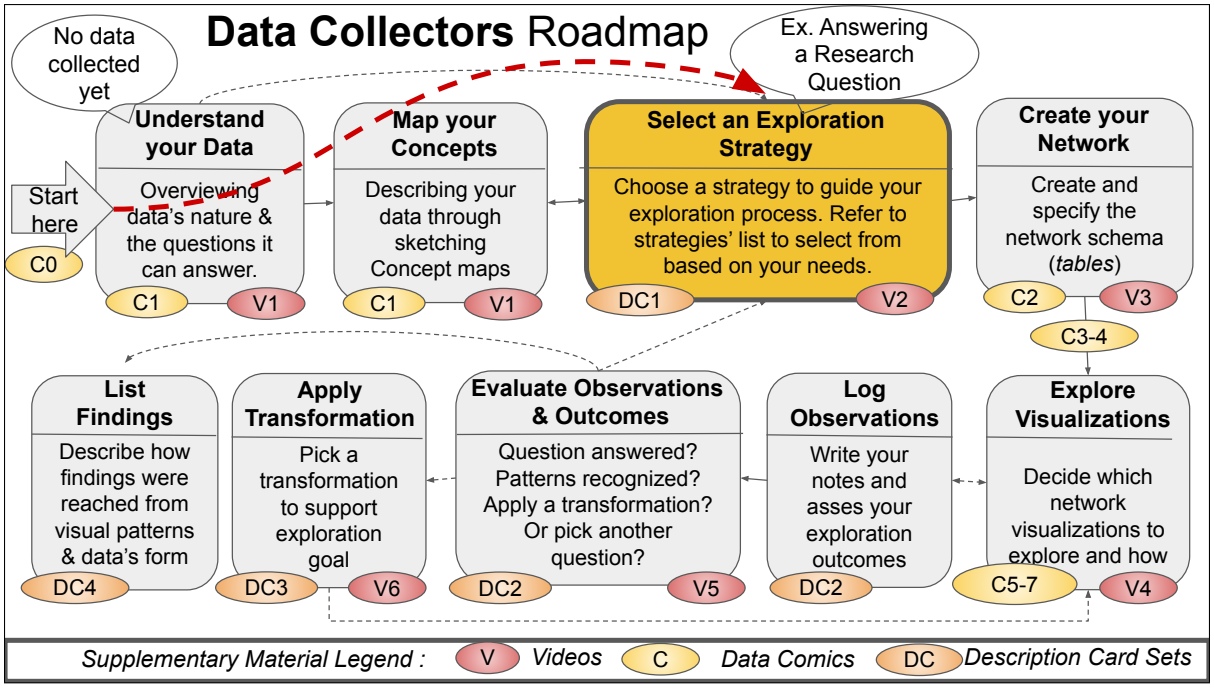
Visual Network Exploration
Coaching Program

Pick a Question (Activity)

Research Goal Definition

Transform the research goal into a measurable form through a strategy (e.g. Research Question)

- Which topic most of the researchers are interested in?
- What level of expertise each researchers hold?
- What are the contributions of each researcher? And of what type?
- Which researchers collaborate with others internally ?
- Which researchers collaborate with others externally?



Goal-based Exploration Strategies

GS1: Exploring to Answer Research Questions

Description: To start exploring your data by selecting the data attributes and schema that the research question addresses.

Network Visualizations : *compatible with the type of schema data attributes.*

What questions to ask?

- Were the research questions answered? (*refer to DC2*)
- Which visual patterns informed the formulation of the question(s)? (*refer to DC2*)

Application : *on any dataset.*

Strategy Steps:

- ***Step 1:*** Select a single research question from your questions pool.
- ***Step 2:*** Sketch/update your concepts map by selecting the data attributes that addresses the currently selected question.
- ***Step 3:*** Create network table(s) based on the concept map - changes- .
- ***Step 4:*** Create the network visualization(s) that are compatible with the nature of the selected data attributes and explore them. Refer to the supporting material on how to explore networks visually.
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2)*.
- ***Step 6:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 7:*** *if any data transformation is required*, *apply* data transformation then return to step (3), if not proceed to following step.
- ***Step 8:*** *if other research question(s) exist*, return to step(1) and implement it along with the following steps by selecting another research question.
- ***Step 9:*** End your exploration and journal your findings (*refer to DC4*)
- P.S. you can create a single concept map for each question, or create/update a comprehensive concept map showing all questions' data. The later option opens more aspects to explore your network data through.

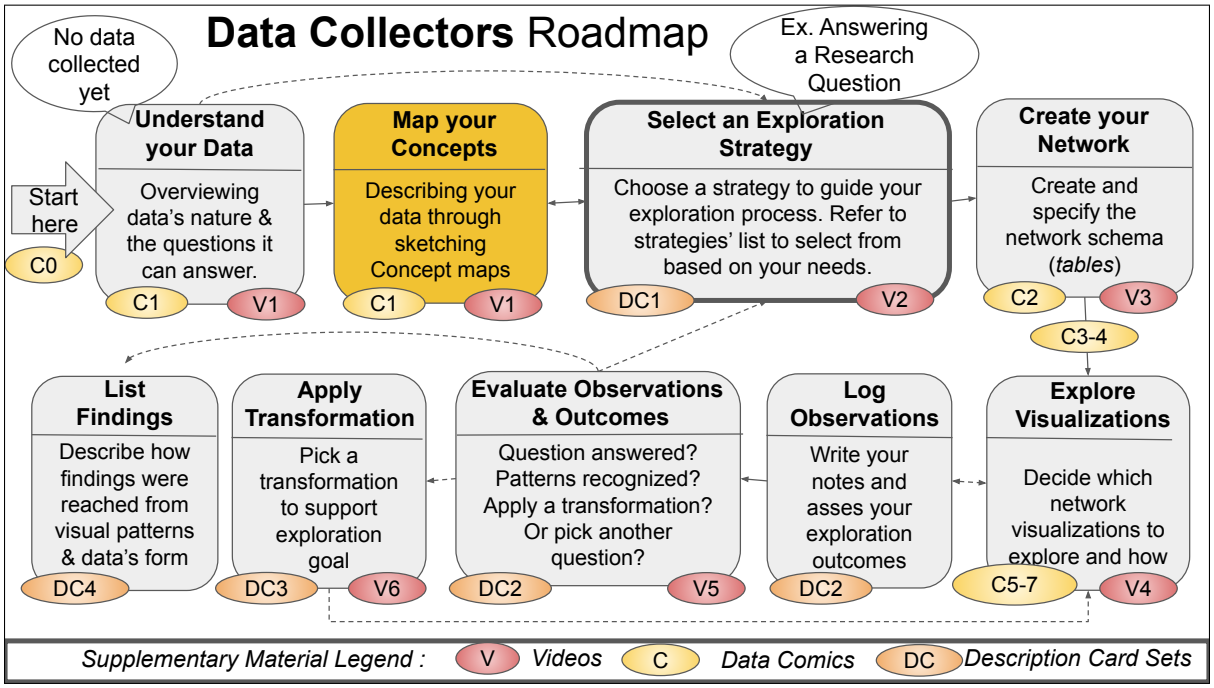
Visual Network Exploration
Coaching Program

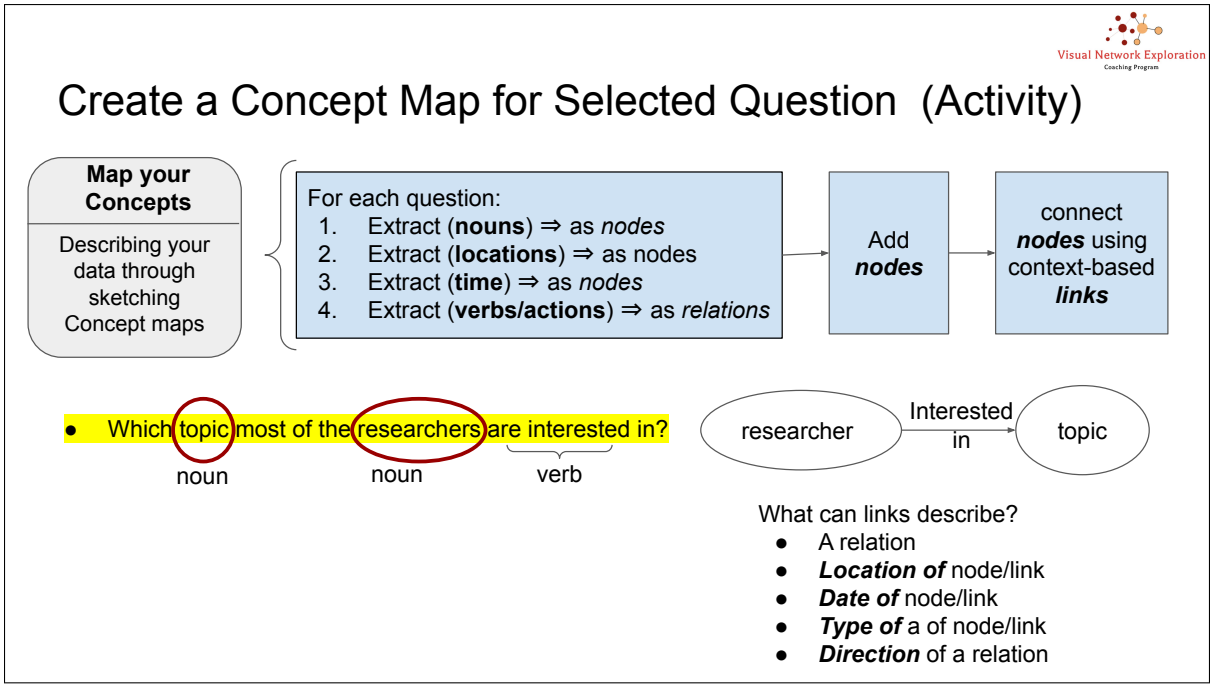
Pick a Question (Activity)

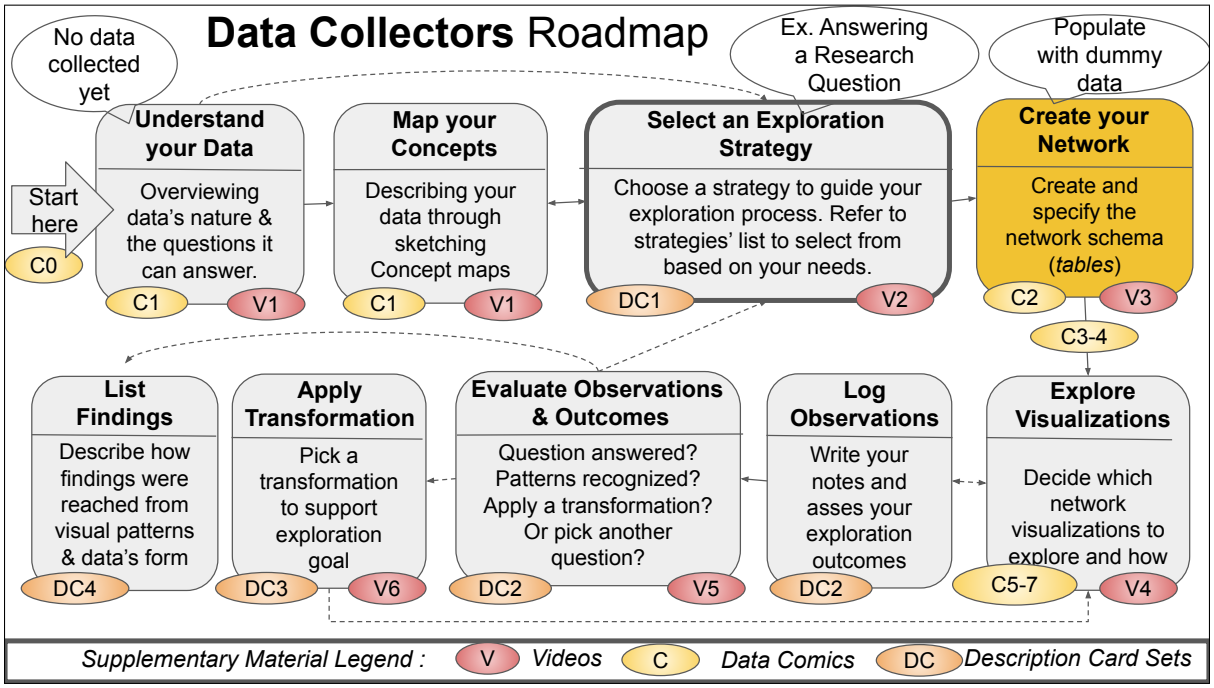
Research Goal Definition

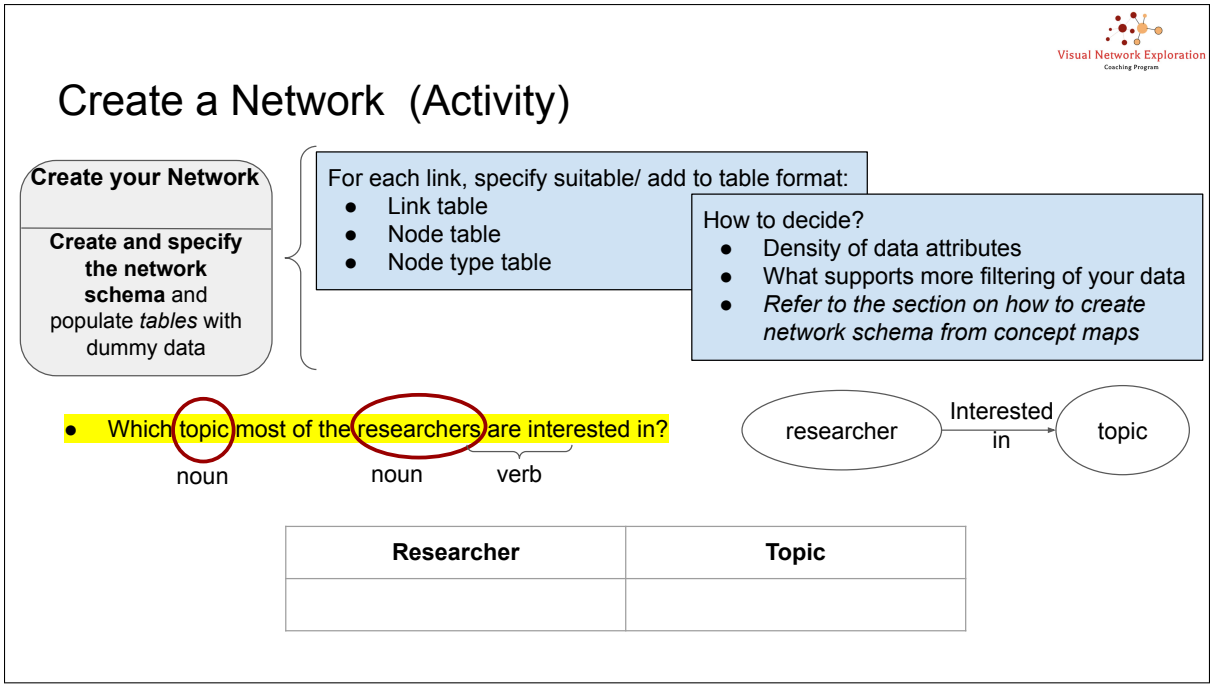
Transform the research goal into a measurable form through a strategy (e.g. Research Question)

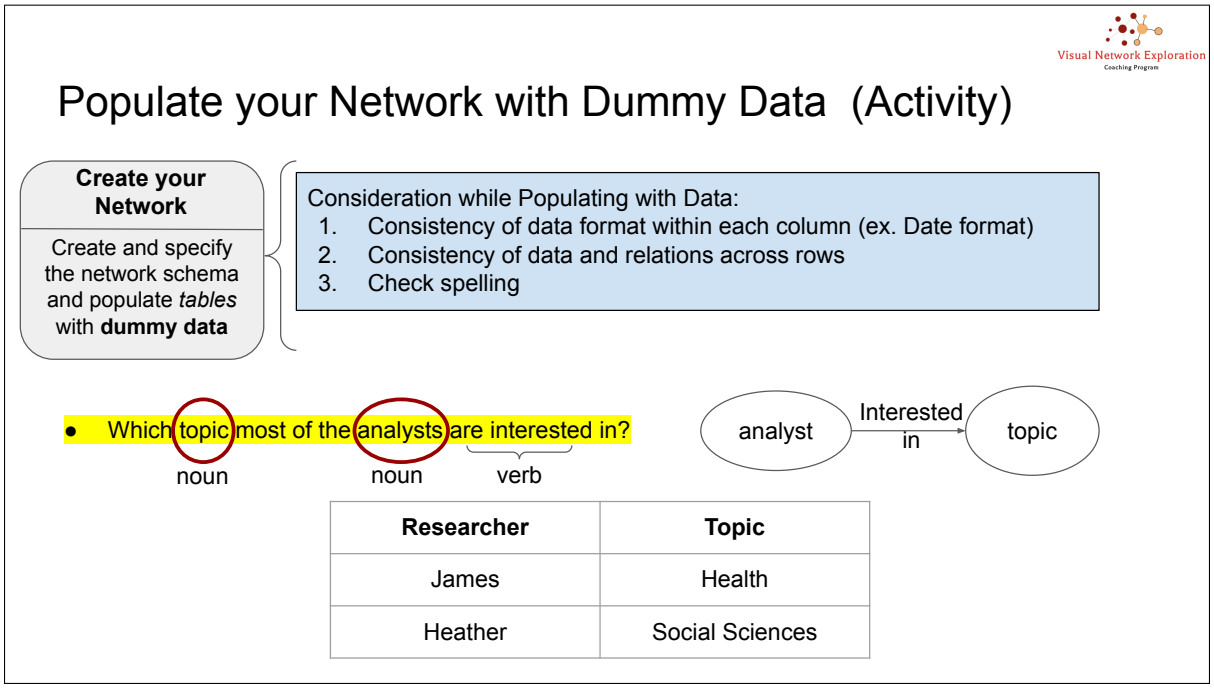
- Which topic most of the researchers are interested in?

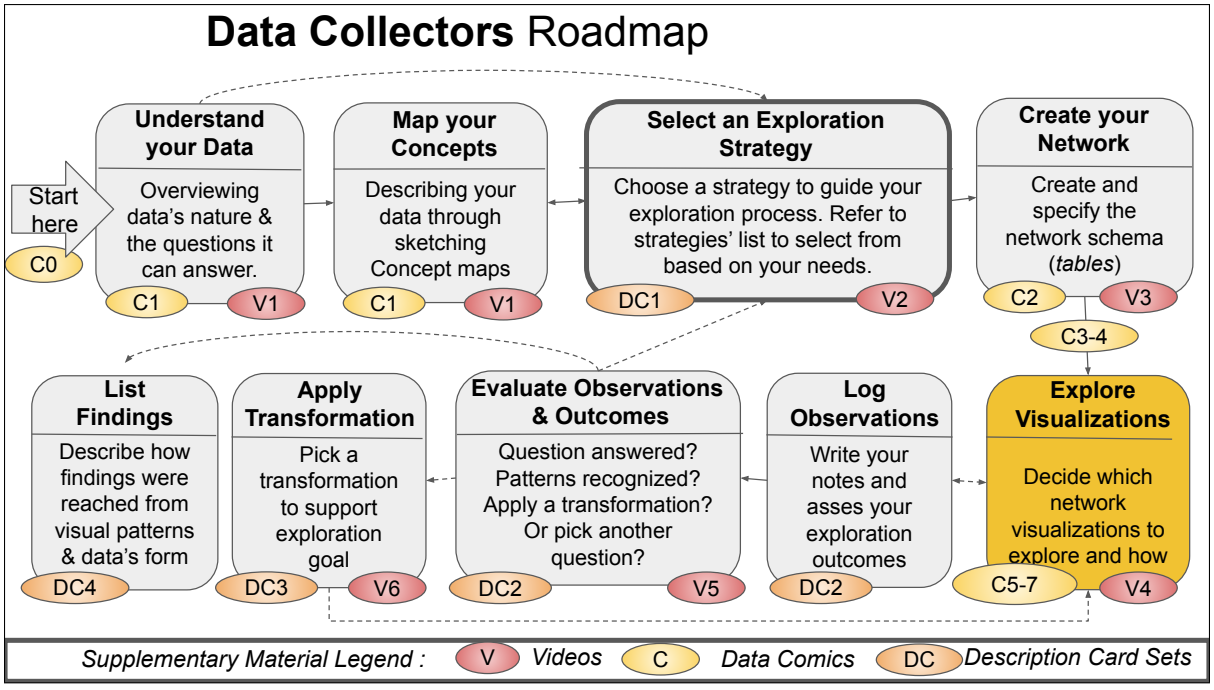

Data Collectors Roadmap
No data collected yet
Ex. Answering a Research Question
Start here
C0
Understand your Data
Overviewing data's nature & the questions it can answer.
C1
V1
Map your Concepts
Describing your data through sketching Concept maps
C1
V1
Select an Exploration Strategy
Choose a strategy to guide your exploration process. Refer to strategies' list to select from based on your needs.
DC1
V2
Create your Network
Create and specify the network schema (tables)
C2
V3
C3-4
List Findings
Describe how findings were reached from visual patterns & data's form
DC4
Apply Transformation
Pick a transformation to support exploration goal
DC3
V6
Evaluate Observations & Outcomes
Question answered? Patterns recognized? Apply a transformation? Or pick another question?
DC2
V5
Log Observations
Write your notes and asses your exploration outcomes
DC2
Explore Visualizations
Decide which network visualizations to explore and how
C5-7
V4
Supplementary Material Legend :
V Videos
C Data Comics
DC Description Card Sets


Visual Network Exploration
Coaching Program
Create a Concept Map for Selected Question (Activity)
Map your Concepts
Describing your data through sketching Concept maps
For each question:
1. Extract (nouns) ⇒ as nodes
2. Extract (locations) ⇒ as nodes
3. Extract (time) ⇒ as nodes
4. Extract (verbs/actions) ⇒ as relations
Add nodes
connect nodes using context-based links
Which topic most of the researchers are interested in?
noun
noun
verb
researcher
Interested in
topic
What can links describe?
A relation
Location of node/link
Date of node/link
Type of a of node/link
Direction of a relation


Data Collectors Roadmap
No data collected yet
Ex. Answering a Research Question
Populate with dummy data
Start here
C0
Understand your Data
Overviewing data's nature & the questions it can answer.
C1
V1
Map your Concepts
Describing your data through sketching Concept maps
C1
V1
Select an Exploration Strategy
Choose a strategy to guide your exploration process. Refer to strategies' list to select from based on your needs.
DC1
V2
Create your Network
Create and specify the network schema (tables)
C2
V3
C3-4
List Findings
Describe how findings were reached from visual patterns & data's form
DC4
Apply Transformation
Pick a transformation to support exploration goal
DC3
V6
Evaluate Observations & Outcomes
Question answered? Patterns recognized? Apply a transformation? Or pick another question?
DC2
V5
Log Observations
Write your notes and asses your exploration outcomes
DC2
Explore Visualizations
Decide which network visualizations to explore and how
C5-7
V4
Supplementary Material Legend :
V Videos
C Data Comics
DC Description Card Sets


Visual Network Exploration
Coaching Program
Create a Network (Activity)
Create your Network
Create and specify the network schema and populate tables with dummy data
For each link, specify suitable/ add to table format:
Link table
Node table
Node type table
How to decide?
Density of data attributes
What supports more filtering of your data
Refer to the section on how to create network schema from concept maps
Which topic most of the researchers are interested in?
noun
noun
verb
researcher
Interested in
topic
Researcher
Topic


Visual Network Exploration
Coaching Program
Populate your Network with Dummy Data (Activity)
Create your Network
Create and specify the network schema and populate tables with dummy data
Consideration while Populating with Data:
1. Consistency of data format within each column (ex. Date format)
2. Consistency of data and relations across rows
3. Check spelling
Which topic most of the analysts are interested in?
noun
noun
verb
analyst
Interested in
topic
Researcher
Topic
James
Health
Heather
Social Sciences


Data Collectors Roadmap
Start here
C0
Understand your Data
Overviewing data's nature & the questions it can answer.
C1
V1
Map your Concepts
Describing your data through sketching Concept maps
C1
V1
Select an Exploration Strategy
Choose a strategy to guide your exploration process. Refer to strategies' list to select from based on your needs.
DC1
V2
Create your Network
Create and specify the network schema (tables)
C2
V3
C3-4
Explore Visualizations
Decide which network visualizations to explore and how
C5-7
V4
Log Observations
Write your notes and asses your exploration outcomes
DC2
Evaluate Observations & Outcomes
Question answered?
Patterns recognized?
Apply a transformation?
Or pick another question?
DC2
V5
Apply Transformation
Pick a transformation to support exploration goal
DC3
V6
List Findings
Describe how findings were reached from visual patterns & data's form
DC4
Supplementary Material Legend :
V Videos
C Data Comics
DC Description Card Sets

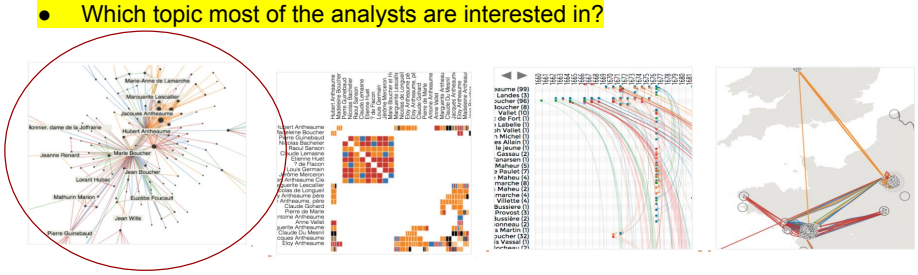
Visual Network Exploration Coaching Program

How to decide on which visualization to explore? (Activity)

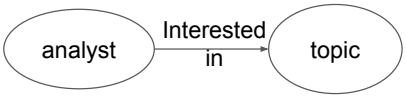
Explore Visualizations

Decide which network visualizations to explore and how

- Which topic most of the analysts are interested in?



Researcher	Topic
James	Health
Heather	Social Sciences

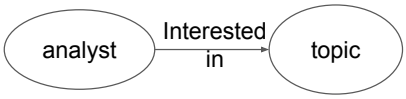


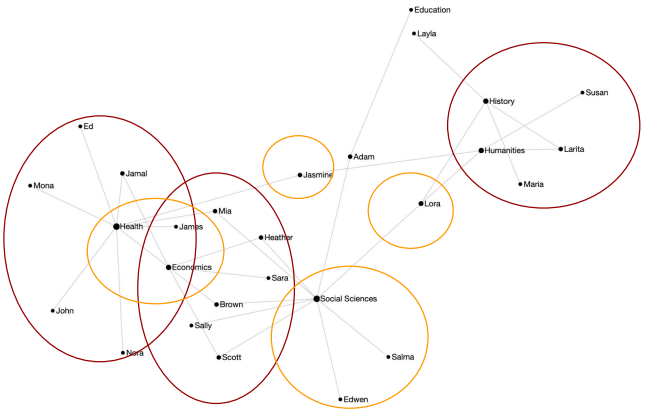
Visual Network Exploration Coaching Program

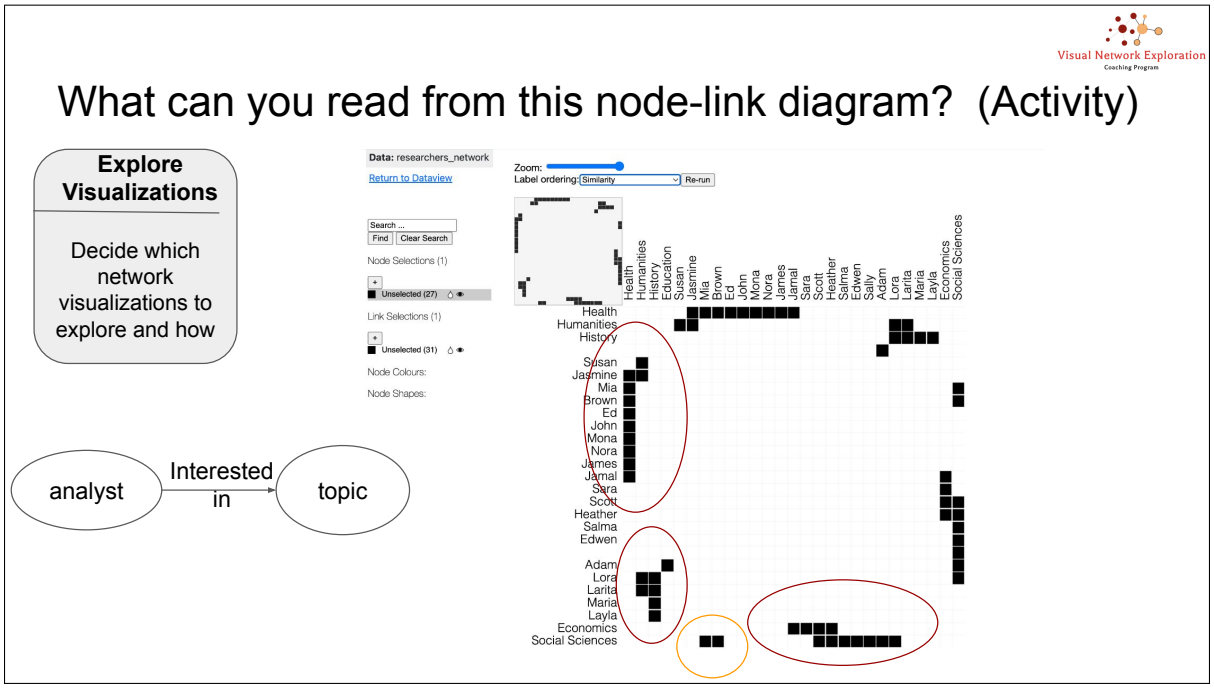
What can you read from this node-link diagram? (Activity)

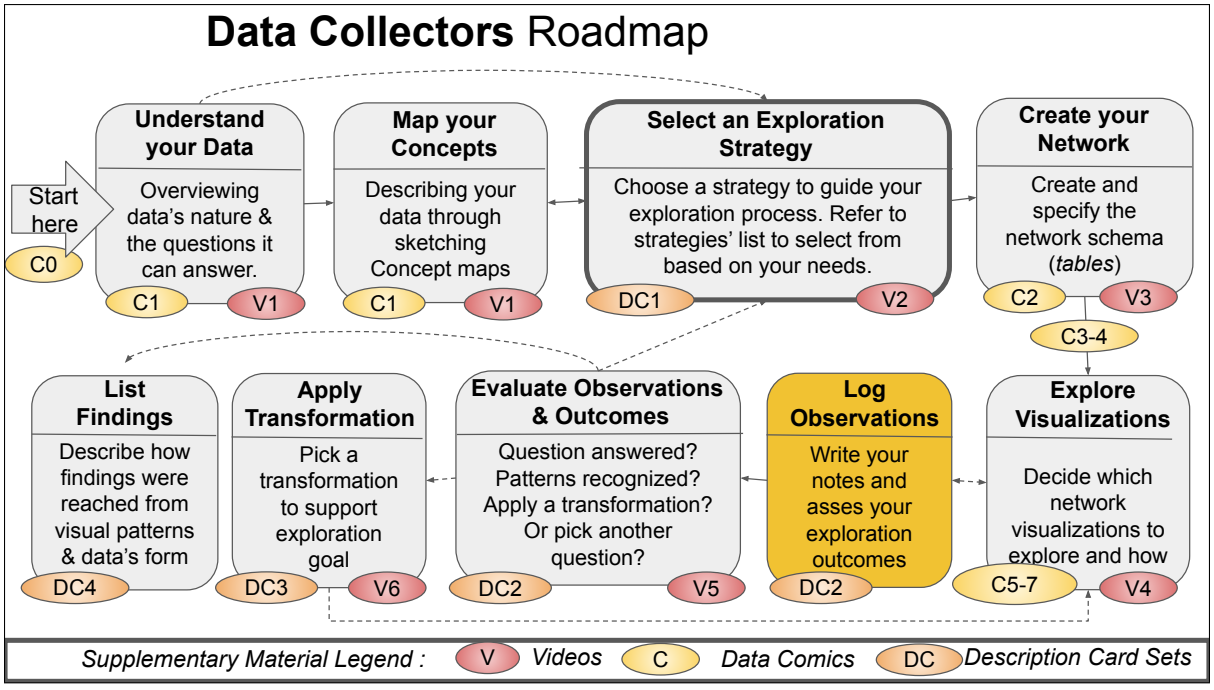
Explore Visualizations

Decide which network visualizations to explore and how



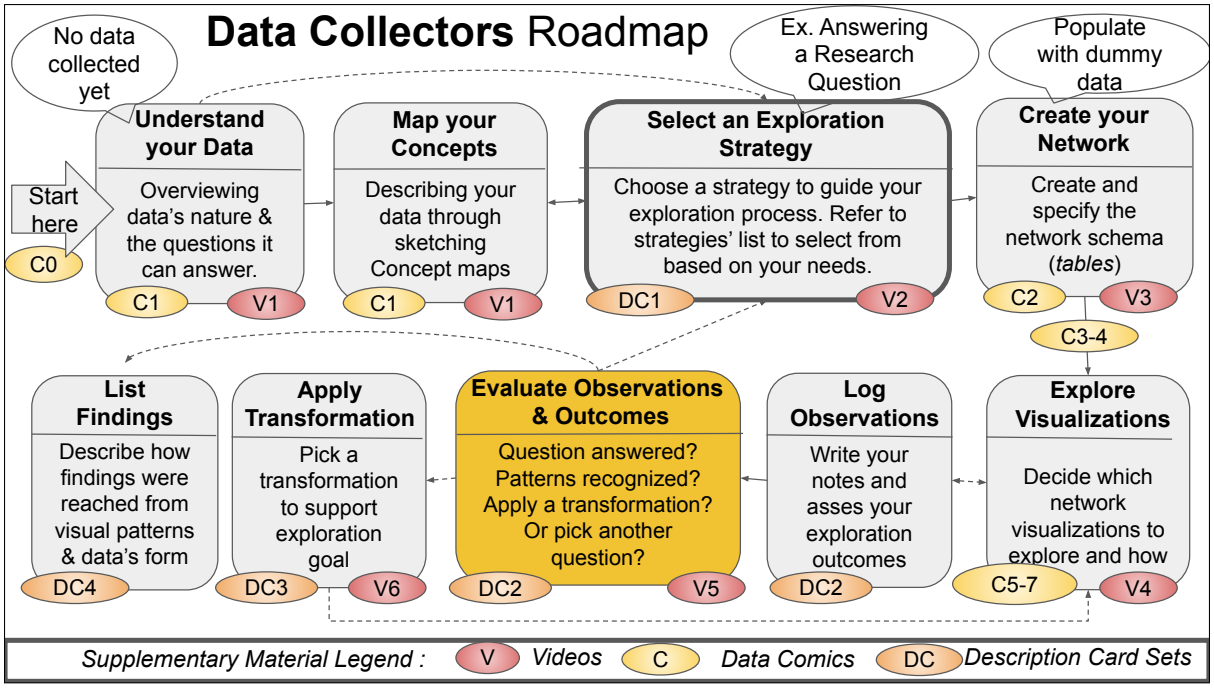



Visual Network Exploration
Coaching Program
What can you read from this node-link diagram? (Activity)
Explore Visualizations
Decide which network visualizations to explore and how
analyst
Interested in
topic
Data: researchers_network
Return to Dataview
Search ...
Find
Clear Search
Node Selections (1)
Unselected (27)
Link Selections (1)
Unselected (31)
Node Colours:
Node Shapes:
Zoom:
Label ordering:
Similarity
Re-run
Health
Humanities
History
Education
Susan
Jasmine
Mia
Brown
Ed
John
Mona
Nora
James
Jamal
Sara
Scott
Heather
Salma
Edwen
Sally
Adam
Lora
Larita
Maria
Layla
Economics
Social Sciences


Data Collectors Roadmap
Start here
C0
Understand your Data
Overviewing data's nature & the questions it can answer.
C1
V1
Map your Concepts
Describing your data through sketching Concept maps
C1
V1
Select an Exploration Strategy
Choose a strategy to guide your exploration process. Refer to strategies' list to select from based on your needs.
DC1
V2
Create your Network
Create and specify the network schema (tables)
C2
V3
C3-4
Explore Visualizations
Decide which network visualizations to explore and how
C5-7
V4
Log Observations
Write your notes and asses your exploration outcomes
DC2
Evaluate Observations & Outcomes
Question answered? Patterns recognized? Apply a transformation? Or pick another question?
DC2
V5
Apply Transformation
Pick a transformation to support exploration goal
DC3
V6
List Findings
Describe how findings were reached from visual patterns & data's form
DC4
Supplementary Material Legend :
V Videos
C Data Comics
DC Description Card Sets

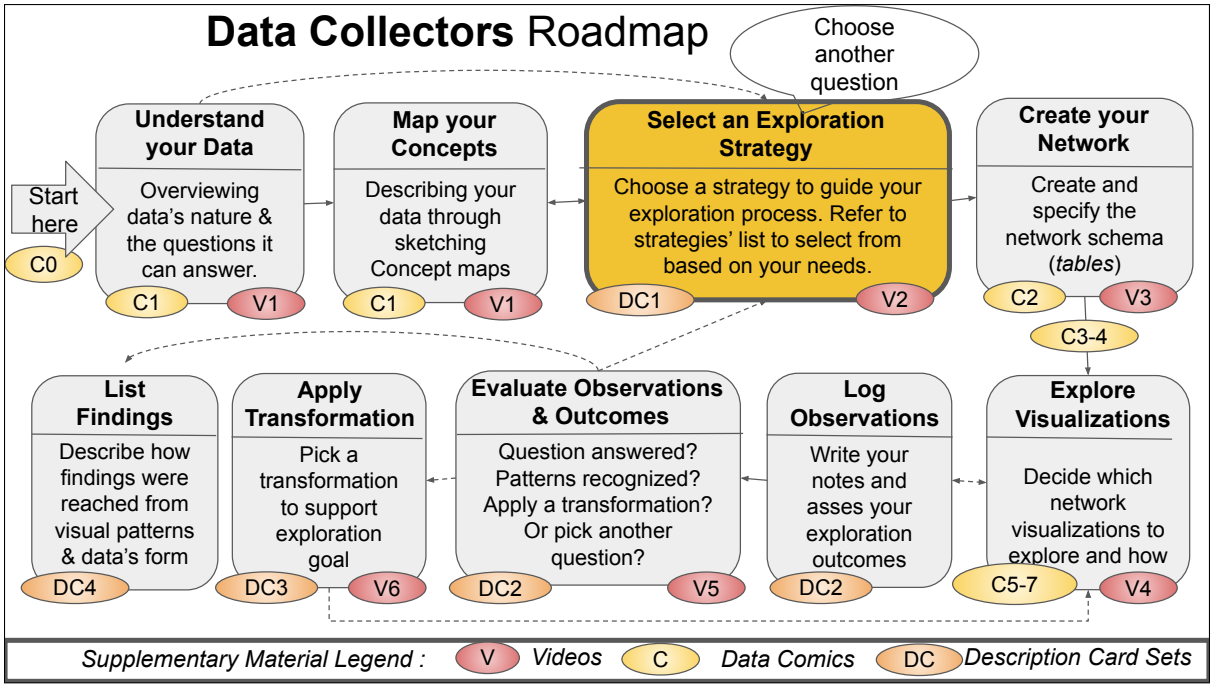
Exploration History, Journaling & Evaluation

Dataset Name: Researchers interest dummy dataset		Network (1)	Network
Selected Network Specification	Node 1:	Researcher	
	Node 2:	Topic	
	Link Type:	-	
	Node Type(s):	-	
	Directed?	No	
	Time ? *	-	
	Location? **	-	
	Questions you think this network can answer:	Which topics researchers are interested in?	
Exploration	Network *Structure (e.g. Sparse, isolated nodes)*	Shared nodes, how many?	
	Identified Network *Clusters (e.g. count, type,patterns,..)*	Which nodes hold the highest degree? clusters, why?	
	Network *Temporal* Changes (e.g structure, clusters,..) *	-	
	Network *Geographical* Distribution (e.g. Density in areas) **	-	
	Any recognized *patterns by link type* filtering?		
	Any recognized *patterns by node type* filtering?		
Exploration Evaluation	Did the network *answered* to your RQs?	partially	
	Did the network *ignited* new questions?	But what about expertise?	
	Which *transformations* can be applied to explore other aspects ? (e.g. use location as nodes, change link type)		



Exploration History, Journaling & Evaluation

Dataset Name: Researchers interest dummy dataset		Network (1)	Network
Selected Network Specification	Node 1:	Researcher	Researcher
	Node 2:	Topic	Topic
	Link Type:	-	Level of Expertise
	Node Type(s):	-	Researcher vs Topic
	Directed?	No	
	Time ? *	-	
	Location? **	-	
	Questions you think this network can answer:	Which topics researchers are interested in?	
Exploration	Network *Structure (e.g. Sparse, isolated nodes)*	Shared nodes, how many?	
	Identified Network *Clusters (e.g. count, type,patterns,..)*	Which nodes hold the highest degree? clusters, why?	
	Network *Temporal* Changes (e.g structure, clusters,..) *	-	
	Network *Geographical* Distribution (e.g. Density in areas) **	-	
	Any recognized *patterns by link type* filtering?		
	Any recognized *patterns by node type* filtering?		
Exploration Evaluation	Did the network *answered* to your RQs?	partially	
	Did the network *ignited* new questions?	But what about expertise?	
	Which *transformations* can be applied to explore other aspects ? (e.g. use location as nodes, change link type)		



Visual Network Exploration Coaching Program

Pick a Question (Activity)

Research Goal Definition

Transform the research goal into a measurable form through a strategy (e.g. Research Question)

- Which topic most of the researchers are interested in?
- What level of expertise each researchers hold?

Visual Network Exploration Coaching Program

Pick a Question (Activity)

Map your Concepts

Describing your data through sketching Concept maps

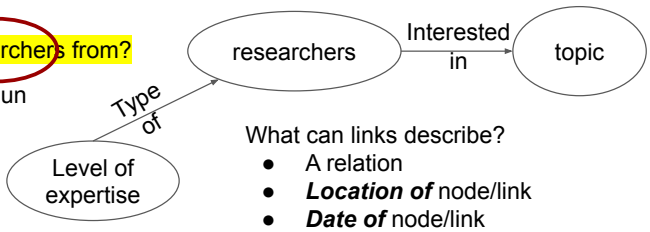
For each question:
1. Extract (**nouns**) ⇒ as *nodes*
2. Extract (**locations**) ⇒ as nodes
3. Extract (**time**) ⇒ as *nodes*
4. Extract (**verbs/actions**) ⇒ as *relations*

Add ***nodes***

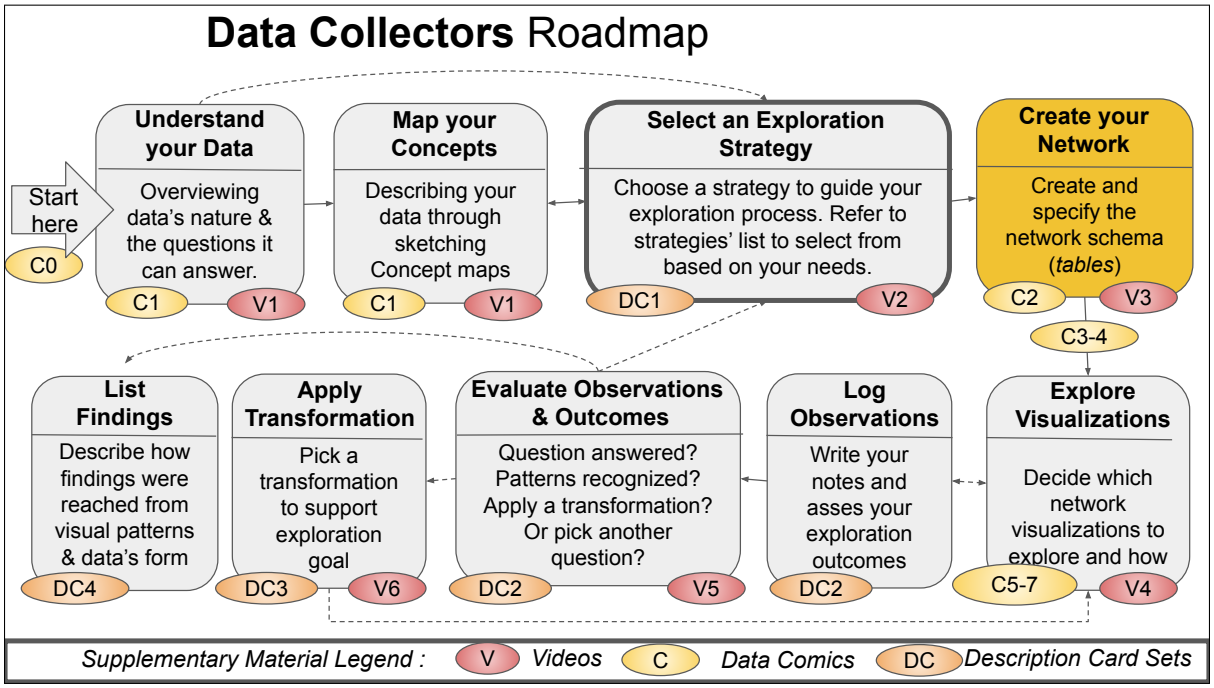
connect ***nodes*** using context-based ***links***

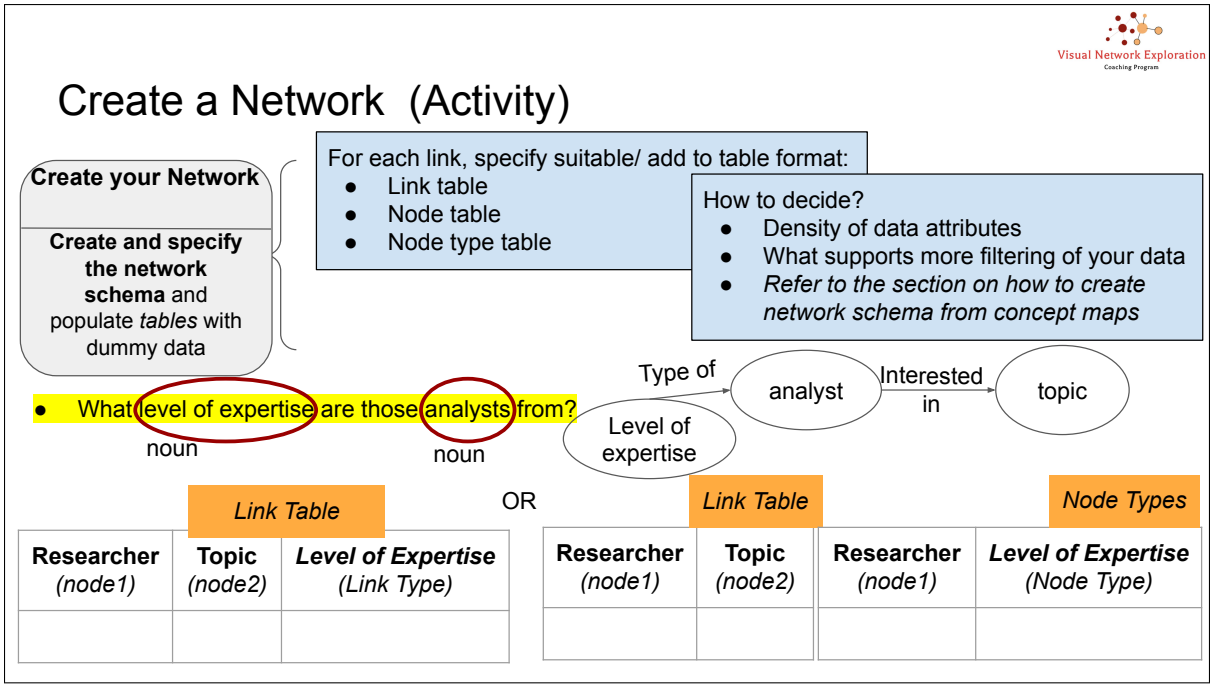
- What level of expertise are those researchers from?

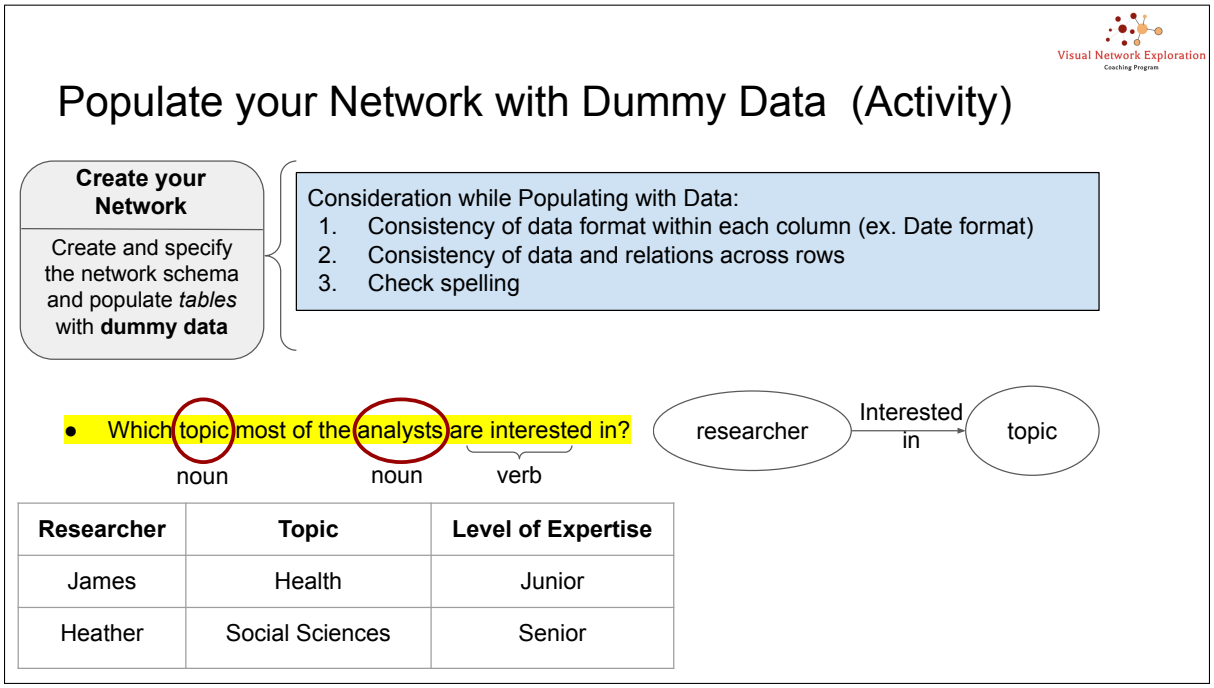
noun noun

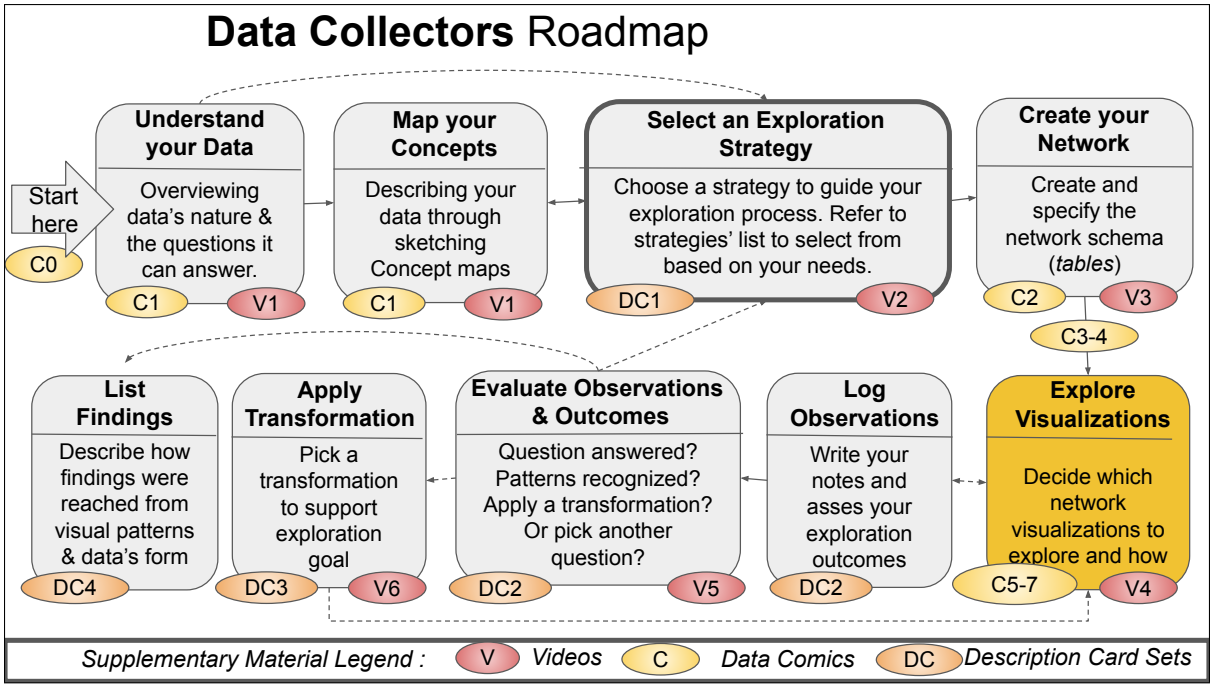


What can links describe?
- A relation
- ***Location of*** node/link
- ***Date of*** node/link
- ***Type of*** a of node/link
- ***Direction*** of a relation


Data Collectors Roadmap
Start here
C0
Understand your Data
Overviewing data's nature & the questions it can answer.
C1
V1
Map your Concepts
Describing your data through sketching Concept maps
C1
V1
Select an Exploration Strategy
Choose a strategy to guide your exploration process. Refer to strategies' list to select from based on your needs.
DC1
V2
Create your Network
Create and specify the network schema (tables)
C2
V3
C3-4
Explore Visualizations
Decide which network visualizations to explore and how
C5-7
V4
Log Observations
Write your notes and asses your exploration outcomes
DC2
Evaluate Observations & Outcomes
Question answered? Patterns recognized? Apply a transformation? Or pick another question?
DC2
V5
Apply Transformation
Pick a transformation to support exploration goal
DC3
V6
List Findings
Describe how findings were reached from visual patterns & data's form
DC4
Supplementary Material Legend :
V Videos
C Data Comics
DC Description Card Sets


Visual Network Exploration
Coaching Program
Create a Network (Activity)
Create your Network
Create and specify the network schema and populate tables with dummy data
For each link, specify suitable/ add to table format:
Link table
Node table
Node type table
How to decide?
Density of data attributes
What supports more filtering of your data
Refer to the section on how to create network schema from concept maps
What level of expertise are those analysts from?
noun
noun
Type of
Level of expertise
analyst
Interested in
topic
Link Table
Researcher (node1)
Topic (node2)
Level of Expertise (Link Type)
OR
Link Table
Researcher (node1)
Topic (node2)
Node Types
Researcher (node1)
Level of Expertise (Node Type)


Visual Network Exploration
Coaching Program
Populate your Network with Dummy Data (Activity)
Create your Network
Create and specify the network schema and populate tables with dummy data
Consideration while Populating with Data:
1. Consistency of data format within each column (ex. Date format)
2. Consistency of data and relations across rows
3. Check spelling
Which topic most of the analysts are interested in?
noun
noun
verb
researcher
Interested in
topic
Researcher | Topic | Level of Expertise
James | Health | Junior
Heather | Social Sciences | Senior


Data Collectors Roadmap
Start here
C0
Understand your Data
Overviewing data's nature & the questions it can answer.
C1
V1
Map your Concepts
Describing your data through sketching Concept maps
C1
V1
Select an Exploration Strategy
Choose a strategy to guide your exploration process. Refer to strategies' list to select from based on your needs.
DC1
V2
Create your Network
Create and specify the network schema (tables)
C2
V3
C3-4
Explore Visualizations
Decide which network visualizations to explore and how
C5-7
V4
Log Observations
Write your notes and asses your exploration outcomes
DC2
Evaluate Observations & Outcomes
Question answered? Patterns recognized? Apply a transformation? Or pick another question?
DC2
V5
Apply Transformation
Pick a transformation to support exploration goal
DC3
V6
List Findings
Describe how findings were reached from visual patterns & data's form
DC4
Supplementary Material Legend :
V Videos
C Data Comics
DC Description Card Sets

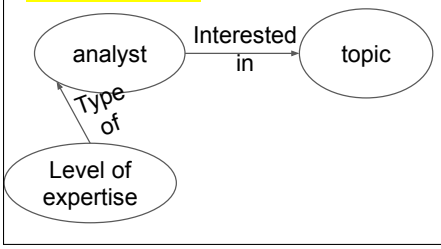
Visual Network Exploration
Coaching Program

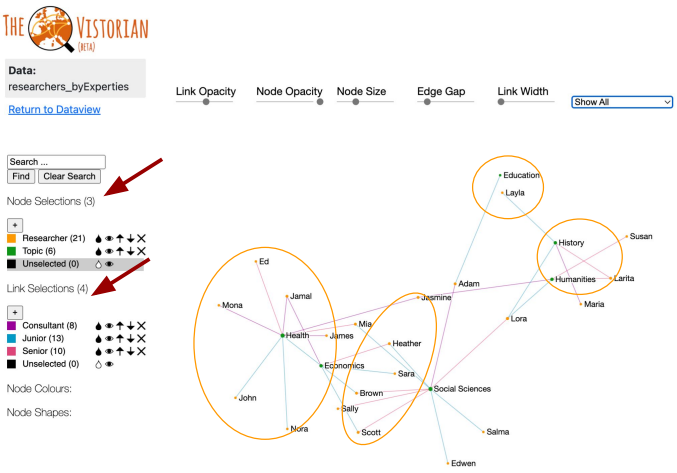
What can you read from this node-link diagram? (Activity)

Explore Visualizations

Decide which network visualizations to explore and how

What level of expertise each researchers hold





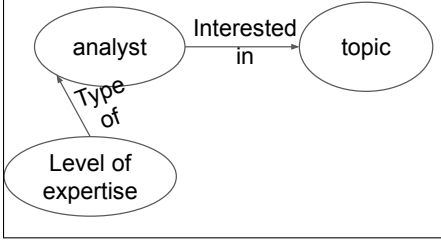
Visual Network Exploration
Coaching Program

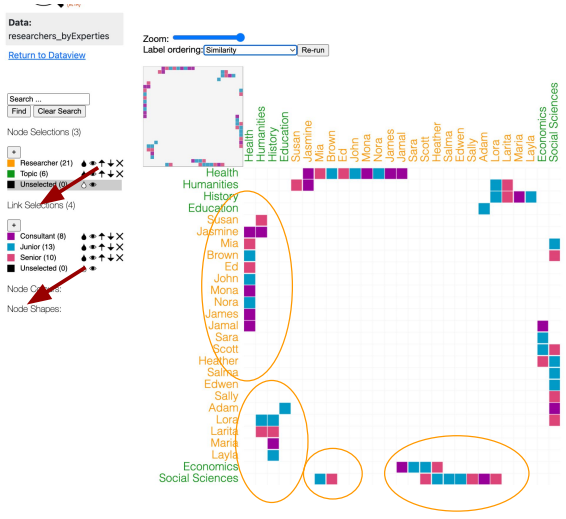
What can you read from this node-link diagram? (Activity)

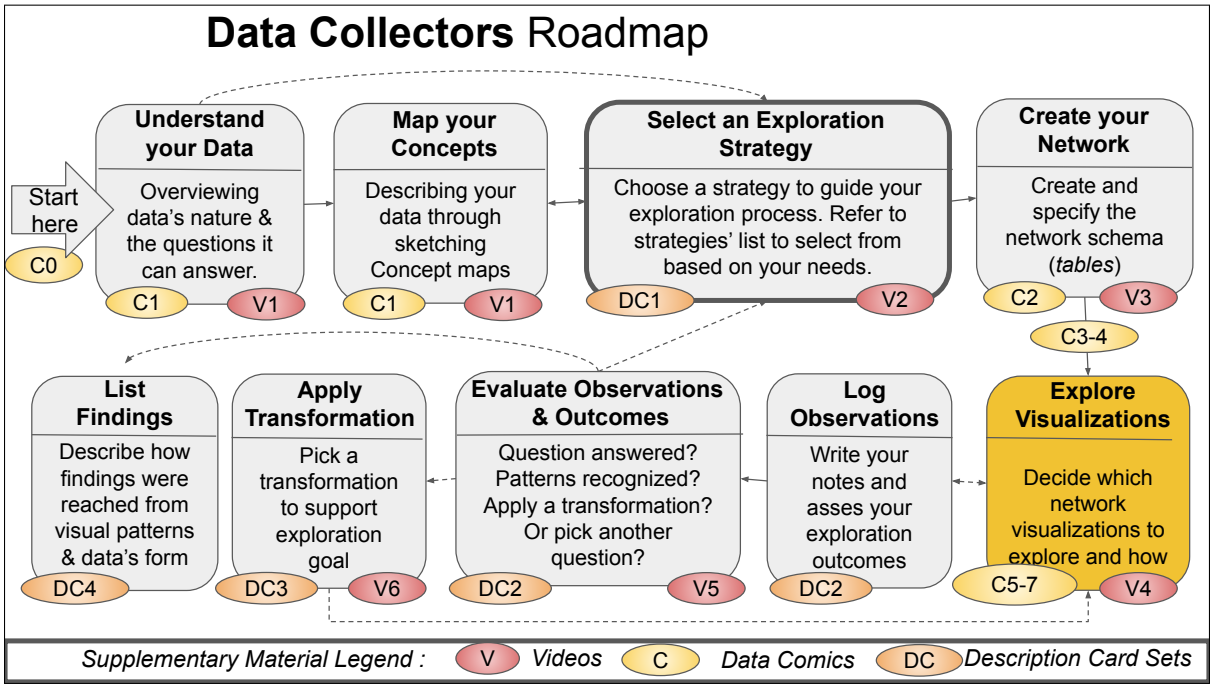
Explore Visualizations

Decide which network visualizations to explore and how

What level of expertise each researchers hold







Exploration History, Journaling & Evaluation

	Dataset Name: Researchers interest dummy dataset	Network (1)	Network
Selected Network Specification	Node 1:	Researcher	Researcher
	Node 2:	Topic	Topic
	Link Type:	-	Level of Expertise
	Node Type(s):	-	Researcher vs Topic
	Directed?	No	
	Time ? *	-	
	Location? **	-	
	Questions you think this network can answer:	Which topics researchers are interested in?	
Exploration	Network *Structure (e.g. Sparse, isolated nodes)*	Shared nodes, how many?	
	Identified Network *Clusters (e.g. count, type,patterns,..)*	Which nodes hold the highest degree? clusters, why?	
	Network *Temporal* Changes (e.g structure, clusters,..) *	-	
	Network *Geographical* Distribution (e.g. Density in areas) **	-	
	Any recognized *patterns by link type* filtering?		
	Any recognized *patterns by node type* filtering?		
Exploration Evaluation	Did the network *answered* to your RQs?	partially	
	Did the network *ignited* new questions?	But what about expertise?	
	Which *transformations* can be applied to explore other aspects ? (e.g. use location as nodes, change link type)		

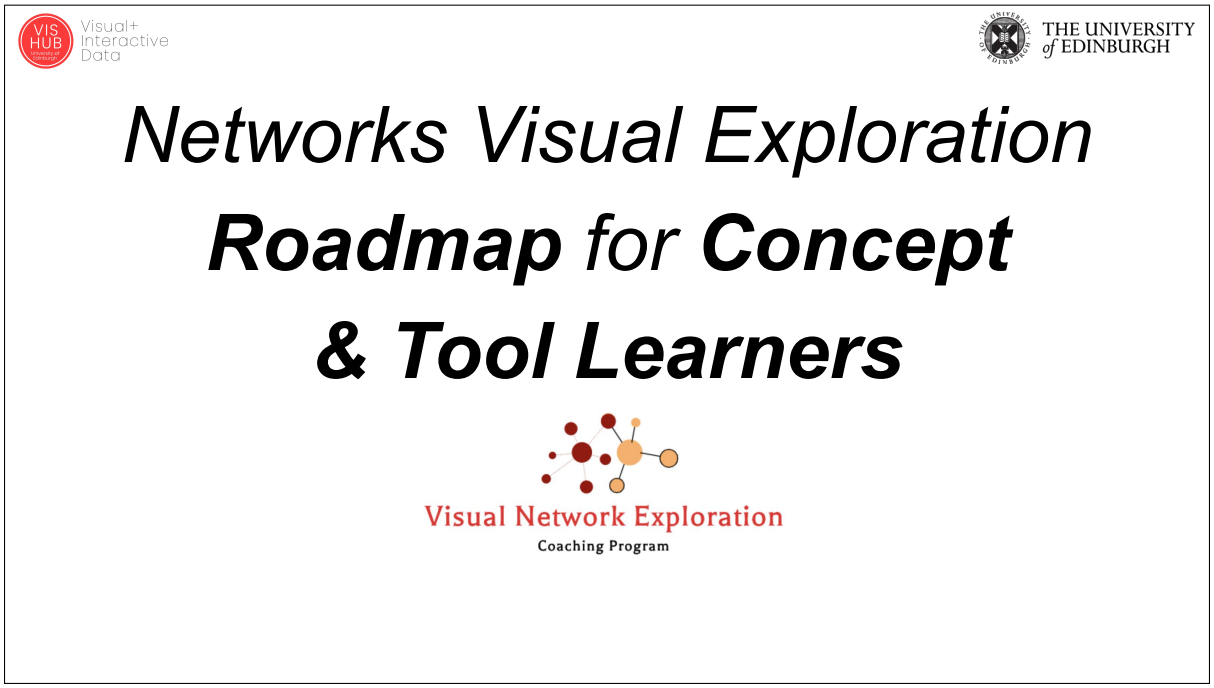
Visual Network Exploration Coaching Program

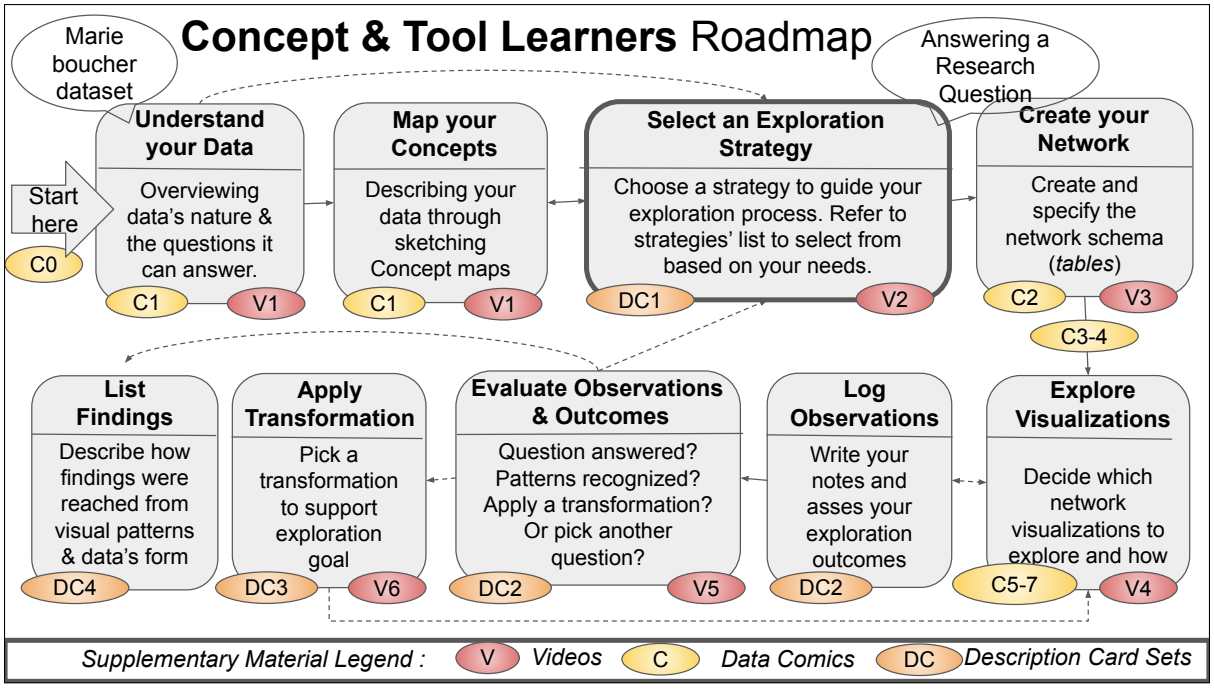
Visual Network Exploration
Coaching Program

Summary: Steps to Form your Network Exploration

1. Start by selecting a strategy
2. Identify *possible* nodes based on the selected strategy
3. Identity *possible* links based on the selected strategy
   a. Connect nodes through links
   b. Describe your data: What kind of links do you have?
4. Abstraction Decisions to be considered
5. Explore through different visualizations
6. Repeat by examining different possible network aiming for ones with special patterns or insights.
7. Decide on what you might be interested in exploring (e.g. Ego network vs whole network, choosing a specific link type,.. etc)

C.8.2 Networks Visual Exploration Roadmap for Concept & Tool Learners

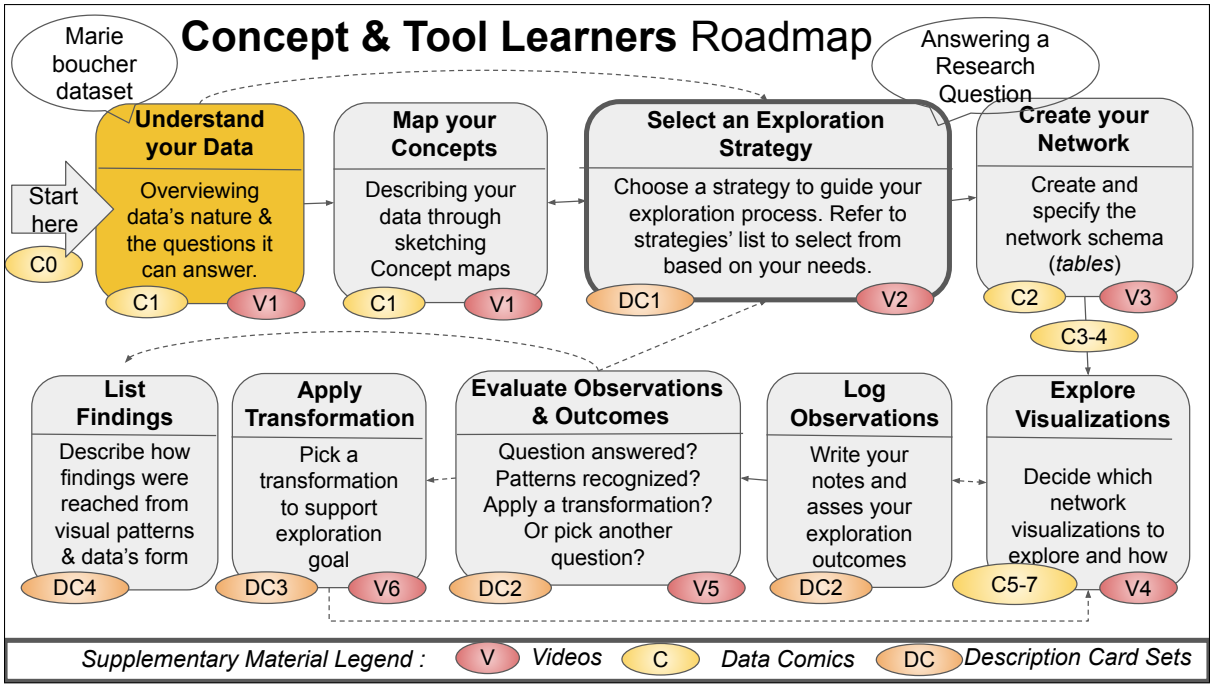

Visual+ Interactive Data
THE UNIVERSITY of EDINBURGH
Networks Visual Exploration
Roadmap for Concept
& Tool Learners
Visual Network Exploration
Coaching Program


Concept & Tool Learners Roadmap
Marie boucher dataset
Start here
C0
Understand your Data
Overviewing data's nature & the questions it can answer.
C1
V1
Map your Concepts
Describing your data through sketching Concept maps
C1
V1
Select an Exploration Strategy
Choose a strategy to guide your exploration process. Refer to strategies' list to select from based on your needs.
DC1
V2
Answering a Research Question
Create your Network
Create and specify the network schema (tables)
C2
V3
C3-4
Explore Visualizations
Decide which network visualizations to explore and how
C5-7
V4
Log Observations
Write your notes and asses your exploration outcomes
DC2
Evaluate Observations & Outcomes
Question answered? Patterns recognized? Apply a transformation? Or pick another question?
DC2
V5
Apply Transformation
Pick a transformation to support exploration goal
DC3
V6
List Findings
Describe how findings were reached from visual patterns & data's form
DC4
Supplementary Material Legend :
V Videos
C Data Comics
DC Description Card Sets

Note

The Conc*ept & Tool Learners roadmap* follows the same track of the general exploration roadmap, but with the following choices being taken:

1. The ***demonstrated dataset*** *is* : Marie Boucher dataset (csv file provided in same section at the course page)
2. The ***selected strategy*** is : Answering a Research Question (GS1 description card - strategies are described in detail in own section of this course)
3. ***Additional Practice Dataset***: Alumni dataset (file provided in the Extra Practice Dataset section at the course page)
4. Feel free to try and play with any other dataset or exploration strategy.



Visual Network Exploration Coaching Program

Starting with your Data

Understand your Data

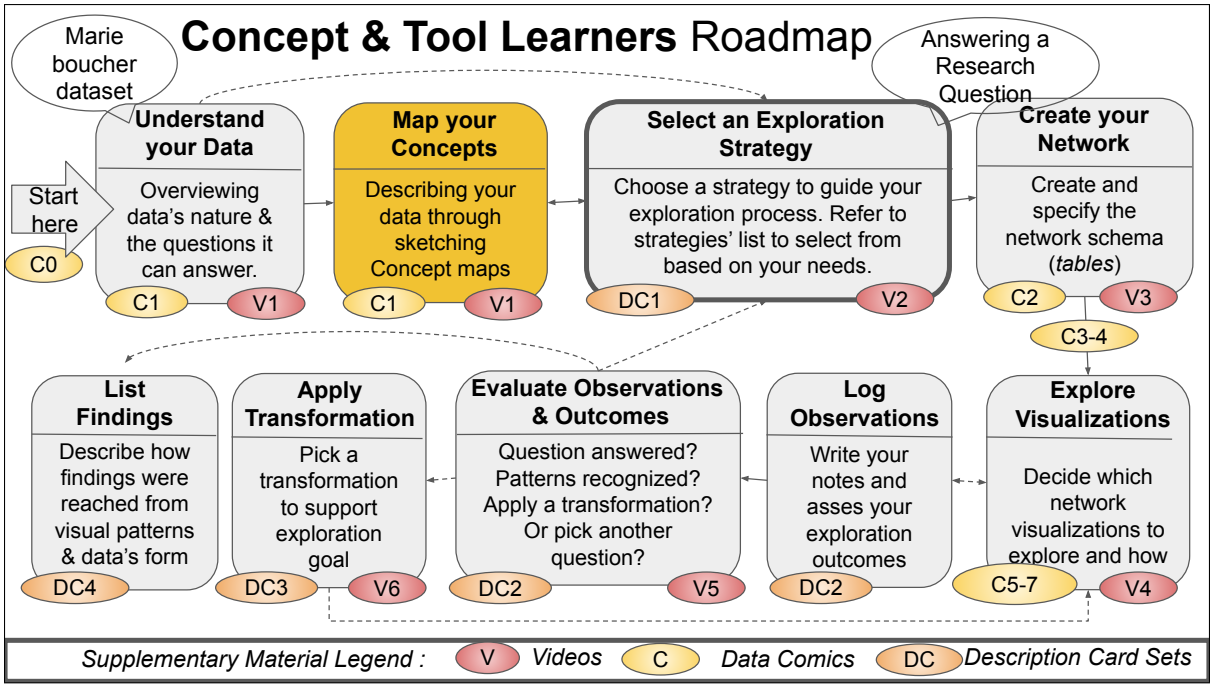
Overviewing data's nature & the questions it can answer.

What is your data about?

Our 1st demo dataset is about the letters exchanged by a trader called Marie Boucher and the goal is to explore her network.

To download the dataset (.csv format) click here.

A	B	C	D	E	F	G
Index	**Name1**	**Place1**	**Qualification**	**Name2**	**Place2**	**Date**
0	Hubert Antheaume	Nantes	Commerce	Jacques Yvon, sr des l	Santo Domingo	03/08/1673
1	Marie Boucher	Nantes	Commerce	Hubert Antheaume	Nantes	06/08/1675
2	Marie Boucher	Nantes	Famille	Hubert Antheaume	Nantes	06/08/1675
3	Marie Boucher	Nantes	Famille	Roze Boucher	Nantes	15/09/1671
4	Marie Boucher	Paris	Commerce	Roze Boucher	Nantes	15/09/1671
5	Marie Boucher	Paris	Commerce	Hubert Antheaume	Paris	15/09/1671
6	? Vallet	Santo Domingo	Travail	? Bec de Fort		01/01/1676



Visual Network Exploration Coaching Program

Describe your Data

Concept Mapping

Describing your data through Concept Mapping

Describing data through **'Concept Maps'** enables us to:

- Identify role-playing **entities (nodes)** in your data
  ⇒ + what additional data *describes your nodes*
  (ex. Location of each node)
- Understand how those nodes are linked through what we call **(links)**
  ⇒ + what additional data describes such links (ex. Link type)

Concept Maps reveals: the **semantics** of your network + underlying **structure** of your **data**

Visual Network Exploration Coaching Program

Concept Map of Marie Boucher Letters *Network*

Concept Mapping

Describing your data through Concept maps

Start by adding all attributes to your map

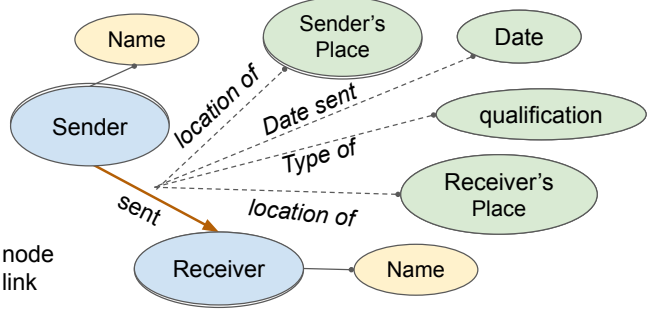
Identify main concepts (***nodes)***

identify relationships that connect *concepts* (***links)***

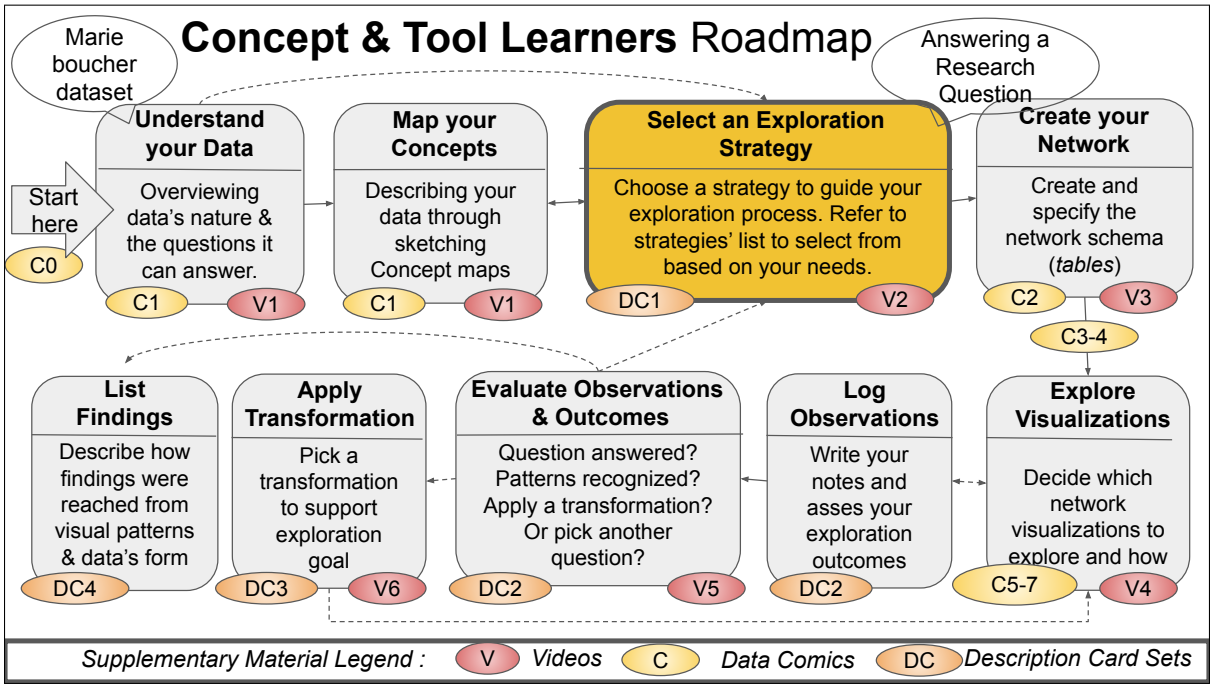
Add data attributes that describes a ***concept*** permanently to the concept

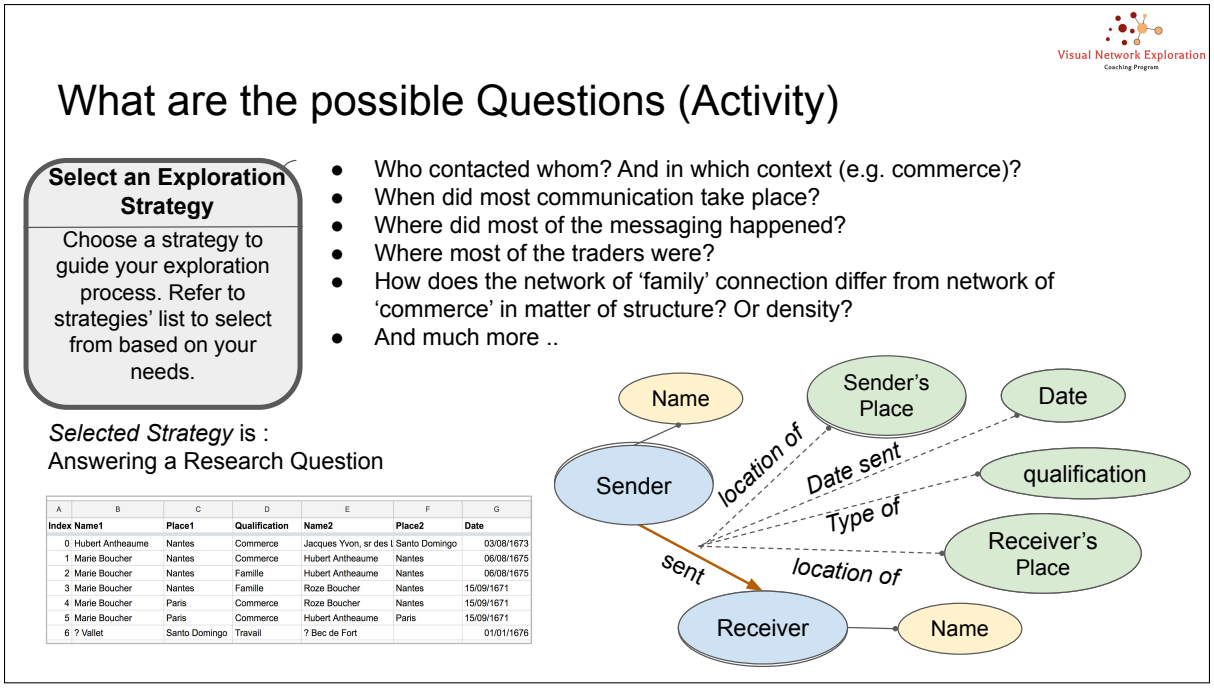
Add data fields that describes changing nature of ***relationships*** *to (links)*

	Sender Name	Sender Place	Qualification	Receiver Name	Receiver Place	Date Letter Sent
1	Hubert Antheaun	St Nicolas, Nante	Commerce	Jacques Yvon, sr de:	Logane, Ile St	03/08/1673
2	Marie Boucher	St Nicolas, Nante	Commerce	Hubert Antheaume	St Nicolas, N	06/08/1675
3	Marie Boucher	St Nicolas, Nante	Famille	Hubert Antheaume	St Nicolas, N	06/08/1675
4	Marie Boucher	St Nicolas, Nante	Famille	Roze Boucher	St Nicolas, N	15/09/1671
5	Marie Boucher	Paris	Commerce	Roze Boucher	St Nicolas, N	15/09/1671
6	Marie Boucher	Paris	Commerce	Hubert Antheaume	Paris	15/09/1671
7	? Vallet	Ile St Domingue	Travail	? Bec de Fort		01/01/1676
8	? Vallet	Ile St Domingue	Commerce	Joseph Labelle	Logane, Ile St	01/01/1676
9	? Vallet	Ile St Domingue	Famille	? Vallet	Ile Guadeloup	01/01/1676
10	? Vallet	Ile St Domingue	Famille	Joseph Vallet	Brest	01/01/1676



Directed Link
Undirected Link
Attribute describes a node
Attribute describes a link





Index	Name1	Place1	Qualification	Name2	Place2	Date
0	Hubert Antheaume	Nantes	Commerce	Jacques Yvon, sr des L	Santo Domingo	03/08/1673
1	Marie Boucher	Nantes	Commerce	Hubert Antheaume	Nantes	06/08/1675
2	Marie Boucher	Nantes	Famille	Hubert Antheaume	Nantes	06/08/1675
3	Marie Boucher	Nantes	Famille	Roze Boucher	Nantes	15/09/1671
4	Marie Boucher	Paris	Commerce	Roze Boucher	Nantes	15/09/1671
5	Marie Boucher	Paris	Commerce	Hubert Antheaume	Paris	15/09/1671
6	? Vallet	Santo Domingo	Travail	? Bec de Fort		01/01/1676

Task-Driven Exploration Strategies

GS1: Exploring to Answer Research Questions

Description: To start exploring your data by selecting the data attributes and schema that the research question addresses.

Network Visualizations : *compatible with the type of schema data attributes.*

What questions to ask?

- Were the research questions answered? (*refer to DC2*)
- Which visual patterns informed the formulation of the question(s)? (*refer to DC2*)

Application : *on any dataset.*

Strategy Steps:

- ***Step 1:*** Select a single research question from your questions pool.
- ***Step 2:*** Sketch/update your concepts map by selecting the data attributes that addresses the currently selected question.
- ***Step 3:*** Create network table(s) based on the concept map - changes- .
- ***Step 4:*** Create the network visualization(s) that are compatible with the nature of the selected data attributes and explore them. Refer to the supporting material on how to explore networks visually.
- ***Step 5:*** Log your observations based on visualized network(*refer to DC2*).
- ***Step 6:*** Evaluate and reflect on your observations (*refer to DC2*).
- ***Step 7:*** *if any data transformation is required*, *apply* data transformation then return to step (3), if not proceed to following step.
- ***Step 8:*** *if other research question(s) exist*, return to step(1) and implement it along with the following steps by selecting another research question.
- ***Step 9:*** End your exploration and journal your findings (*refer to DC4*)
- P.S. you can create a single concept map for each question, or create/update a comprehensive concept map showing all questions' data. The later option opens more aspects to explore your network data through.

Visual Network Exploration Coaching Program

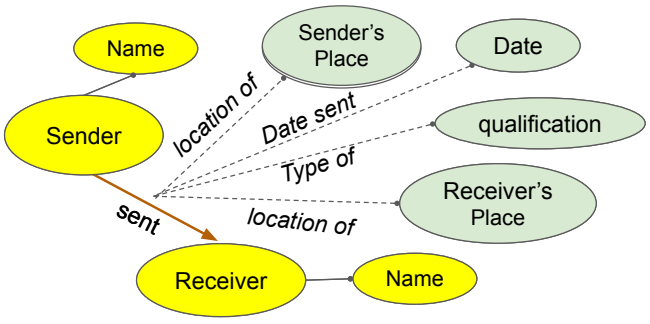
Pick a Question (Activity)

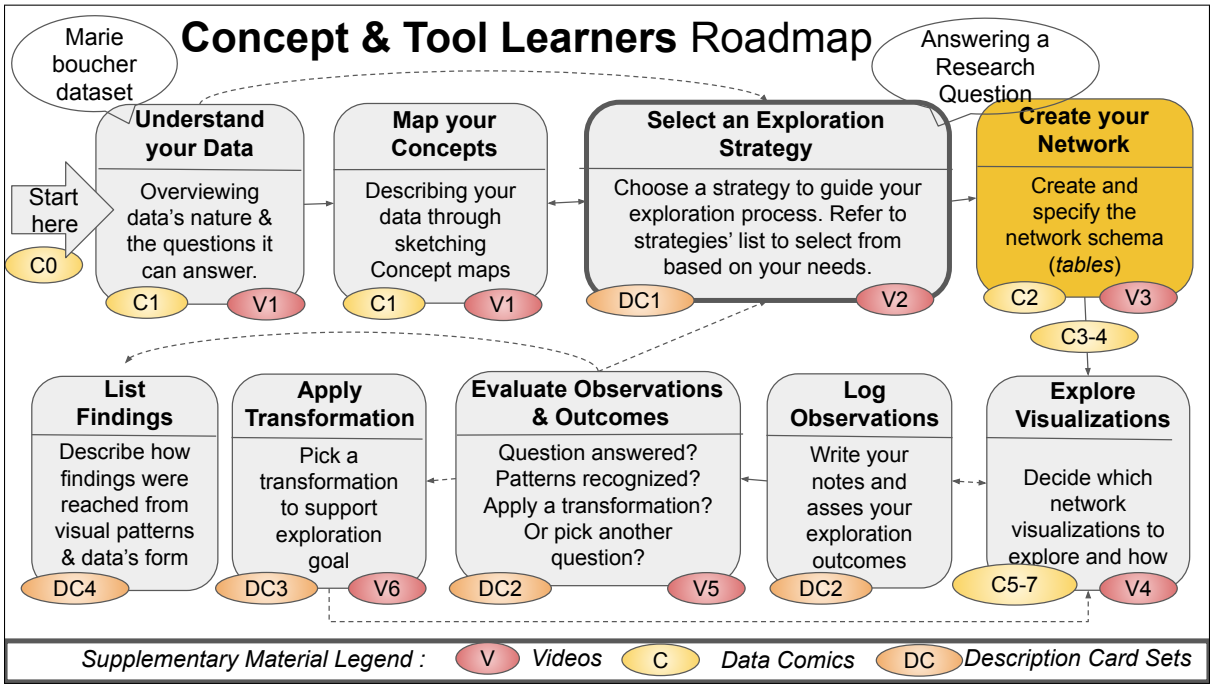
Explored Question

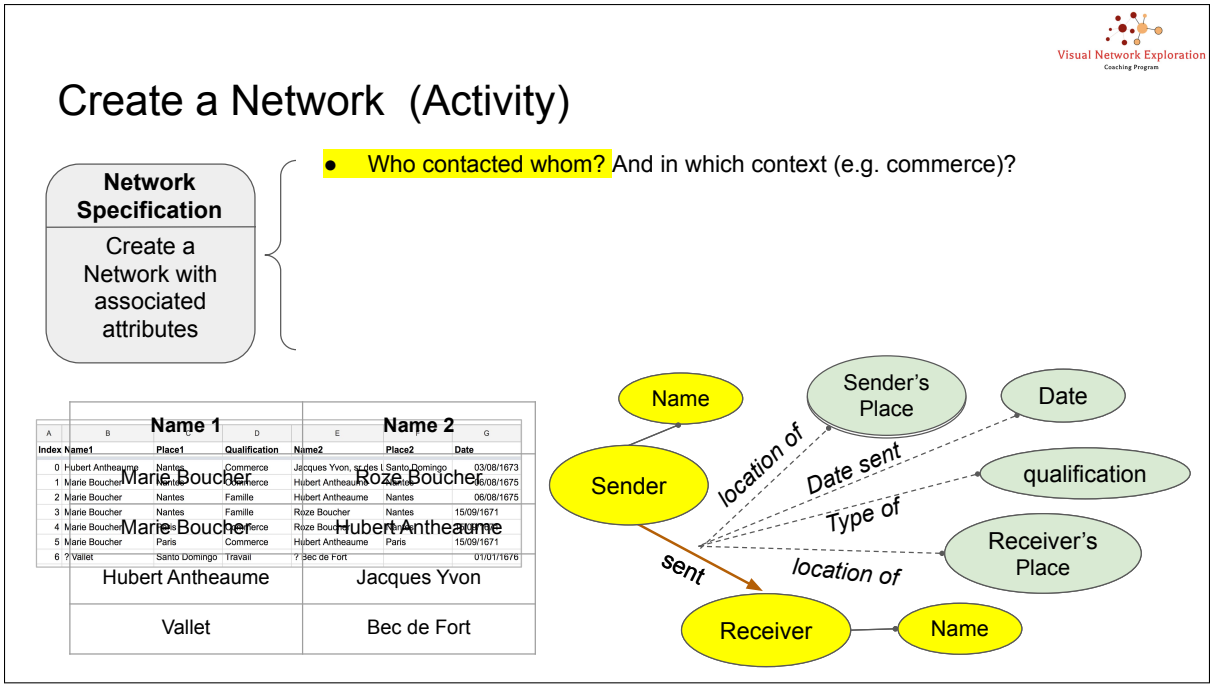
Pick a question to explore & associated attributes

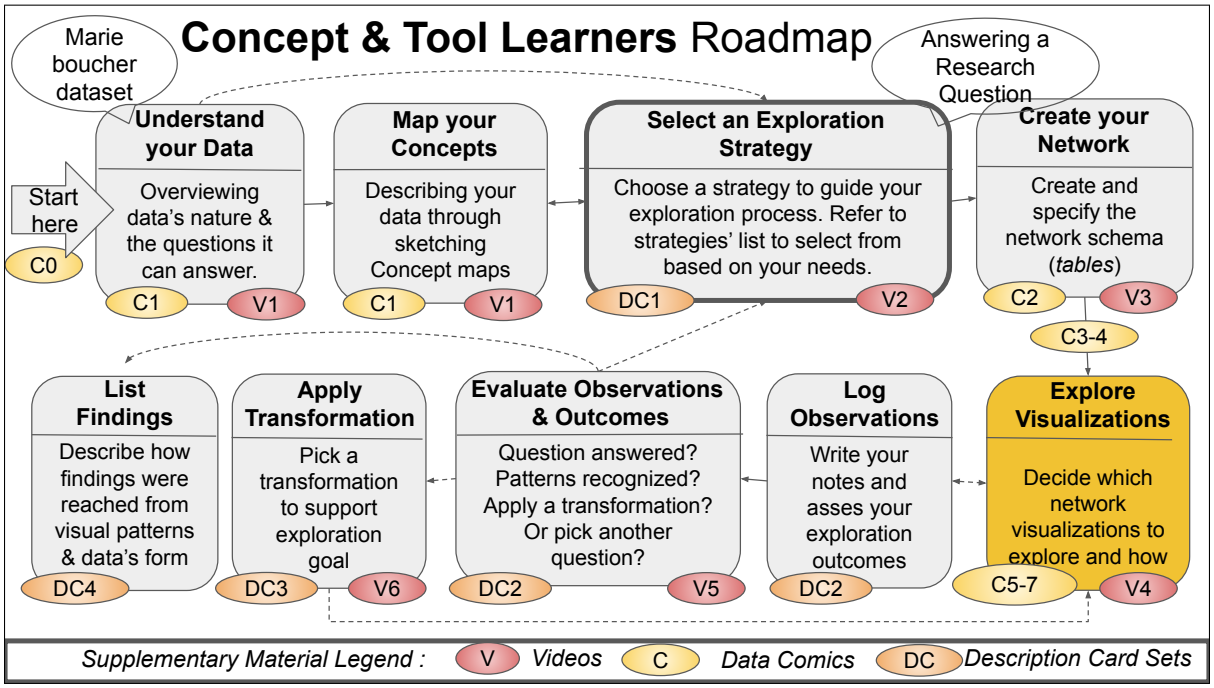
- Who contacted whom? And in which context (e.g. commerce)?

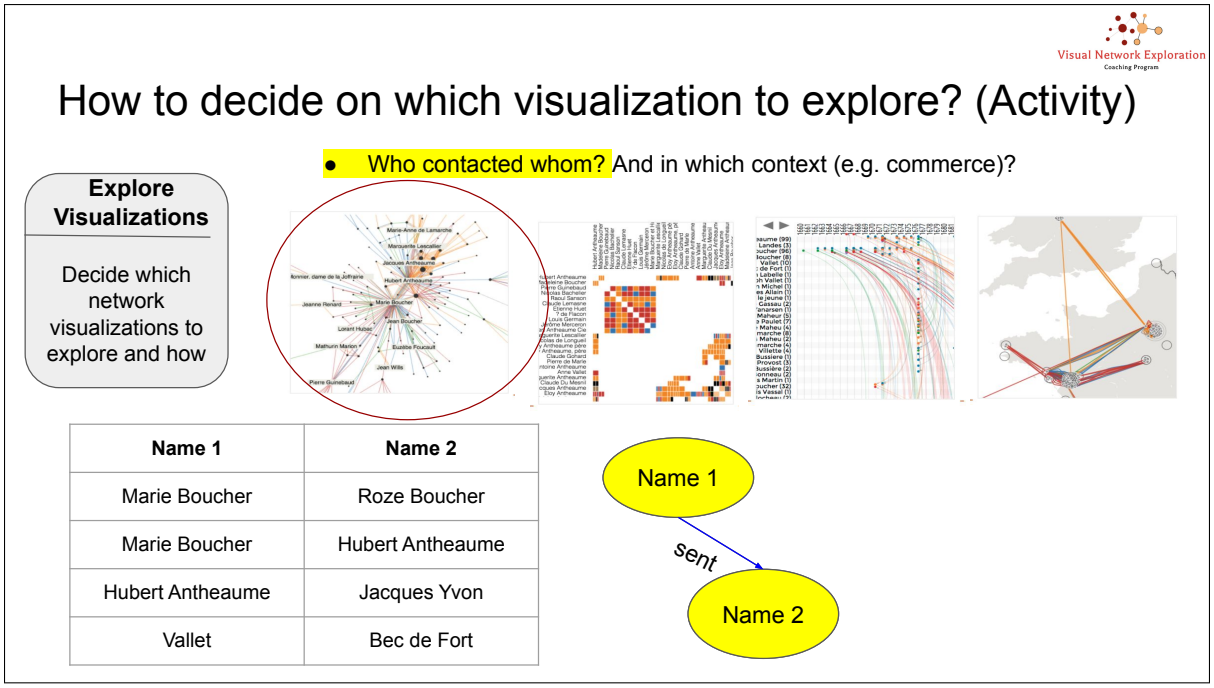
	Index	Name1	Place1	Qualification	Name2	Place2	Date
0	0	Hubert Antheaume	Nantes	Commerce	Jacques Yvon, sr des l	Santo Domingo	03/08/1673
1	1	Marie Boucher	Nantes	Commerce	Hubert Antheaume	Nantes	06/08/1675
2	2	Marie Boucher	Nantes	Famille	Hubert Antheaume	Nantes	06/08/1675
3	3	Marie Boucher	Nantes	Famille	Roze Boucher	Nantes	15/09/1671
4	4	Marie Boucher	Paris	Commerce	Roze Boucher	Nantes	15/09/1671
5	5	Marie Boucher	Paris	Commerce	Hubert Antheaume	Paris	15/09/1671
6	6	? Vallet	Santo Domingo	Travail	? Bec de Fort		01/01/1676



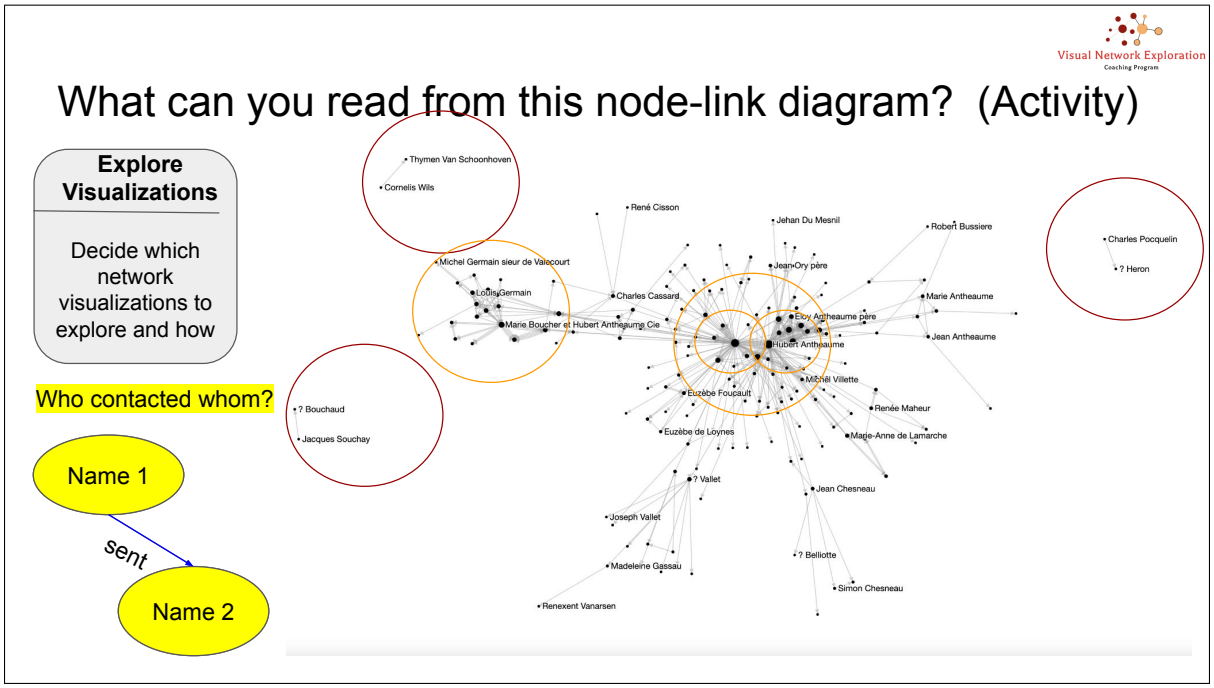

Concept & Tool Learners Roadmap
Marie boucher dataset
Answering a Research Question
Start here
C0
Understand your Data
Overviewing data's nature & the questions it can answer.
C1
V1
Map your Concepts
Describing your data through sketching Concept maps
C1
V1
Select an Exploration Strategy
Choose a strategy to guide your exploration process. Refer to strategies' list to select from based on your needs.
DC1
V2
Create your Network
Create and specify the network schema (tables)
C2
V3
C3-4
List Findings
Describe how findings were reached from visual patterns & data's form
DC4
Apply Transformation
Pick a transformation to support exploration goal
DC3
V6
Evaluate Observations & Outcomes
Question answered? Patterns recognized? Apply a transformation? Or pick another question?
DC2
V5
Log Observations
Write your notes and asses your exploration outcomes
DC2
Explore Visualizations
Decide which network visualizations to explore and how
C5-7
V4
Supplementary Material Legend :
V Videos
C Data Comics
DC Description Card Sets

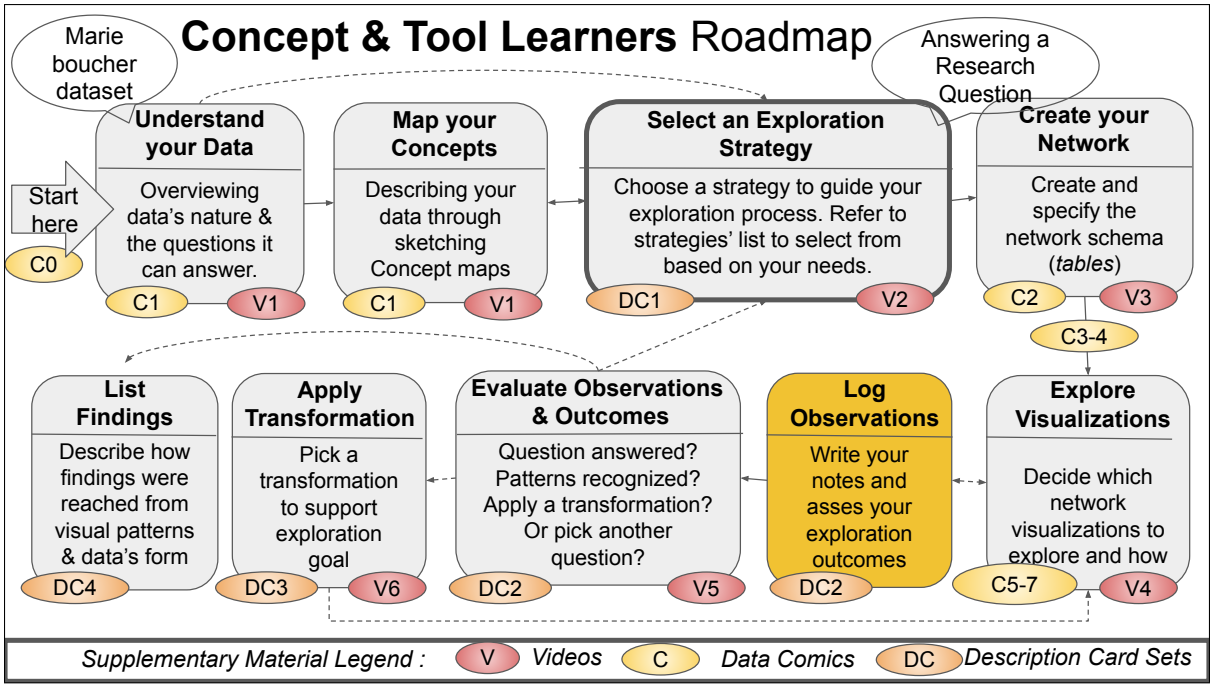

Visual Network Exploration
Coaching Program
Create a Network (Activity)
Network Specification
Create a Network with associated attributes
Who contacted whom? And in which context (e.g. commerce)?
Name 1
Name 2
Marie Boucher
Roze Boucher
Marie Boucher
Hubert Antheaume
Hubert Antheaume
Jacques Yvon
Vallet
Bec de Fort
Name
Sender's Place
Date
Sender
location of
Date sent
qualification
Type of
Receiver's Place
sent
location of
Receiver
Name





Name 1	Name 2
Marie Boucher	Roze Boucher
Marie Boucher	Hubert Antheaume
Hubert Antheaume	Jacques Yvon
Vallet	Bec de Fort


Visual Network Exploration
Coaching Program
What can you read from this node-link diagram? (Activity)
Explore Visualizations
Decide which network visualizations to explore and how
Who contacted whom?
Name 1
sent
Name 2
Thymen Van Schoonhoven
Cornelis Wils
René Cisson
Jehan Du Mesnil
Robert Bussiere
Charles Pocquelin
? Heron
Michel Germain sieur de Valecourt
Jean-Ory père
Louis Germain
Charles Cassard
Marie Antheaume
Marie Boucher et Hubert Antheaume Cie
Eloy Antheaume père
Jean Antheaume
Hubert Antheaume
Michel Villette
Euzèbe Foupault
Renée Maheur
? Bouchaud
Jacques Souchay
Euzèbe de Loynes
Marie-Anne de Lamarche
? Vallet
Jean Chesneau
Joseph Vallet
? Belliotte
Madeleine Gassau
Simon Chesneau
Renexent Vanarsen


Concept & Tool Learners Roadmap
Marie boucher dataset
Answering a Research Question
Start here
C0
Understand your Data
Overviewing data's nature & the questions it can answer.
C1
V1
Map your Concepts
Describing your data through sketching Concept maps
C1
V1
Select an Exploration Strategy
Choose a strategy to guide your exploration process. Refer to strategies' list to select from based on your needs.
DC1
V2
Create your Network
Create and specify the network schema (tables)
C2
V3
C3-4
List Findings
Describe how findings were reached from visual patterns & data's form
DC4
Apply Transformation
Pick a transformation to support exploration goal
DC3
V6
Evaluate Observations & Outcomes
Question answered? Patterns recognized? Apply a transformation? Or pick another question?
DC2
V5
Log Observations
Write your notes and asses your exploration outcomes
DC2
Explore Visualizations
Decide which network visualizations to explore and how
C5-7
V4
Supplementary Material Legend :
V Videos
C Data Comics
DC Description Card Sets

Exploration History, Journaling & Evaluation

	Dataset Name:	Network (1)	Network	
Selected Network Specification	Node 1:	Name 1		
	Node 2:	Name 2		
	Link Type:	-		
	Node Type(s):	-		
	Directed?	yes		
	Time ? *	-		
	Location? **	-		
	Questions you think this network can answer:	Who contacted who?		
Exploration	Network *Structure (e.g. Sparse, isolated nodes)*	Isolated nodes, why? clusters, why?		
	Identified Network *Clusters (e.g. count, type,patterns,..)*			
	Network *Temporal* Changes (e.g structure, clusters,..) *	-		
	Network *Geographical* Distribution (e.g. Density in areas) **	-		
	Any recognized *patterns by link type* filtering?			
	Any recognized *patterns by node type* filtering?			
Exploration Evaluation	Did the network *answered* to your RQs?	yes		
	Did the network *ignited* new questions?	But how		
	Which *transformations* can be applied to explore other aspects ? (e.g. use location as nodes, change link type)			

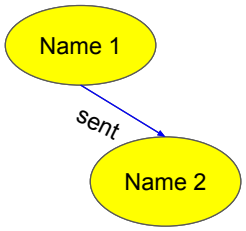
Visual Network Exploration Coaching Program

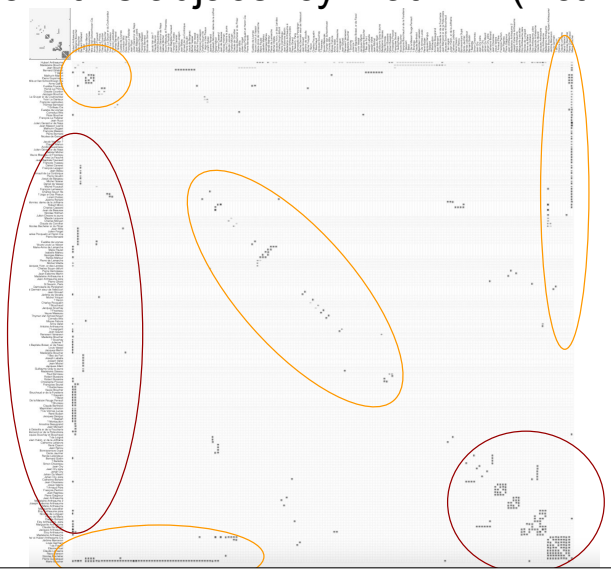
What can you read from this adjacency matrix? (Activity)

Explore Visualizations

Decide which network visualizations to explore and how

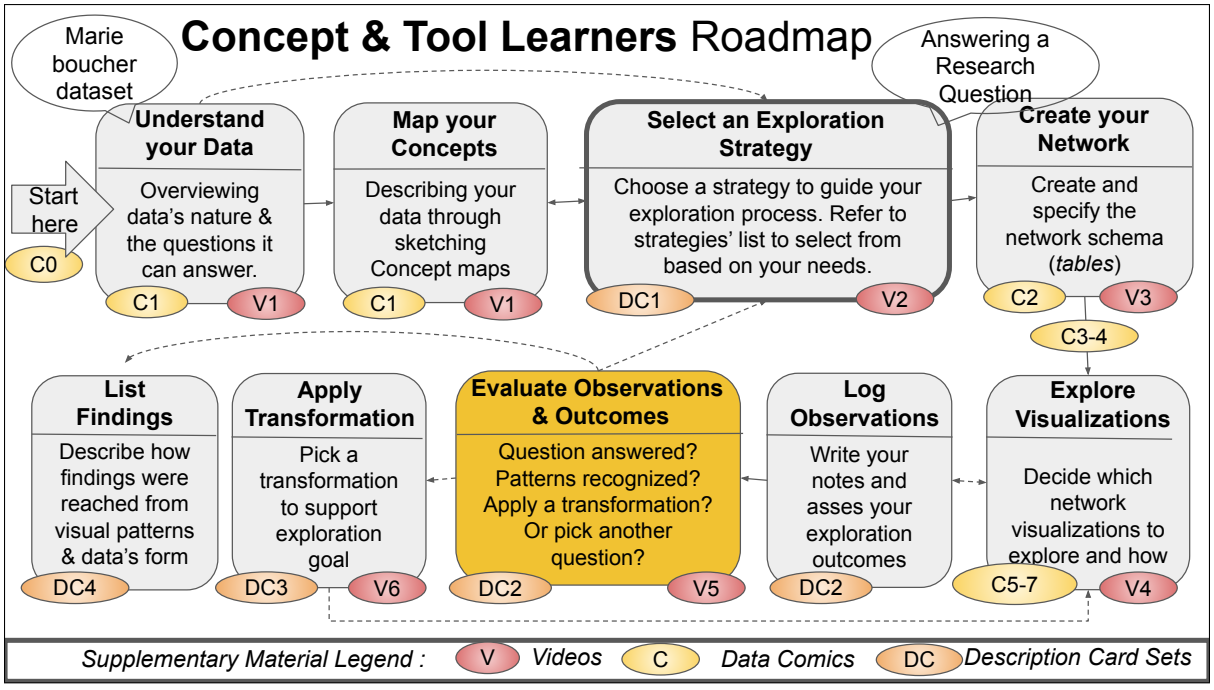
Who contacted whom?





Exploration History, Journaling & Evaluation

		Dataset Name:	Network (1)	Network (2)	…
Selected Network Specification		Node 1:	Name 1	Name 1	
		Node 2:	Name 2	Name 2	
		Link Type:	-	Qualification	
		Node Type(s):	-		
		Directed?	Yes	Yes	
		Time ? *	-		
		Location? **	-		
		Questions you think this network can answer:	Who contacted who? Isolated nodes, why? clusters, why?		
Exploration		Network *Structure (e.g. Sparse, isolated nodes)*			
		Identified Network *Clusters (e.g. count, type,patterns,..)*	Top clusters were 5		
		Network *Temporal* Changes (e.g structure, clusters,..) *	-		
		Network *Geographical* Distribution (e.g. Density in areas) **	-		
		Any recognized *patterns by link type* filtering?	-		
		Any recognized *patterns by node type* filtering?	-		
Exploration Evaluation		Did the network *answered* to your RQs?	yes		
		Did the network *ignited* new questions?			
		Which *transformations* can be applied to explore other aspects ? (e.g. use location as nodes, change link type)			



Exploration History, Journaling & Evaluation

		Dataset Name:	Network (1)	Network (2)	…
Selected Network Specification		Node 1:	Name 1	Name 1	
		Node 2:	Name 2	Name 2	
		Link Type:	-	Qualification	
		Node Type(s):	-		
		Directed?	Yes	Yes	
		Time ? *	-		
		Location? **	-		
		Questions you think this network can answer:	Who contacted who? Isolated nodes, why? clusters, why?		
Exploration		Network *Structure (e.g. Sparse, isolated nodes)*			
		Identified Network *Clusters (e.g. count, type,patterns,..)*	Top clusters were 5		
		Network *Temporal* Changes (e.g structure, clusters,..) *	-		
		Network *Geographical* Distribution (e.g. Density in areas) **	-		
		Any recognized *patterns by link type* filtering?	-		
		Any recognized *patterns by node type* filtering?	-		
Exploration Evaluation		Did the network *answered* to your RQs?	yes		
		Did the network *ignited* new questions?			
		Which *transformations* can be applied to explore other aspects ? (e.g. use location as nodes, change link type)			

How qualification can change our

- Add link type?
- Explore temporal changes?

Visual Network Exploration
Coaching Program

Summary: Steps to Form your Network Exploration

1. Identify *possible* nodes
2. Identity *possible* links
   a. Connect nodes through links
   b. Describe your data: What kind of links do you have?
3. Abstraction Decisions
4. Possible Exploration Paths
5. What are you interested in exploring: Ego network vs whole network

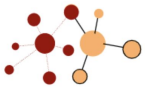
C.8.3 Extra Practice Example(Students Dataset)

VIS HUB Visual+ Interactive Data

THE UNIVERSITY of EDINBURGH

Extra Practice Example

(Students Dataset)



Visual Network Exploration

Coaching Program

Outline

- About the **Dataset**
- **Data Formation**
  - Activity (1): Creating Networks
  - Activity (2): Creating Possible Network Formations
  - Activity (3): Finding a Visual Answer for a Question
  - Activity (4): Creating Node/Link Tables

Alumni Dataset
(Members of the American University Club in Shanghai)

About the Dataset

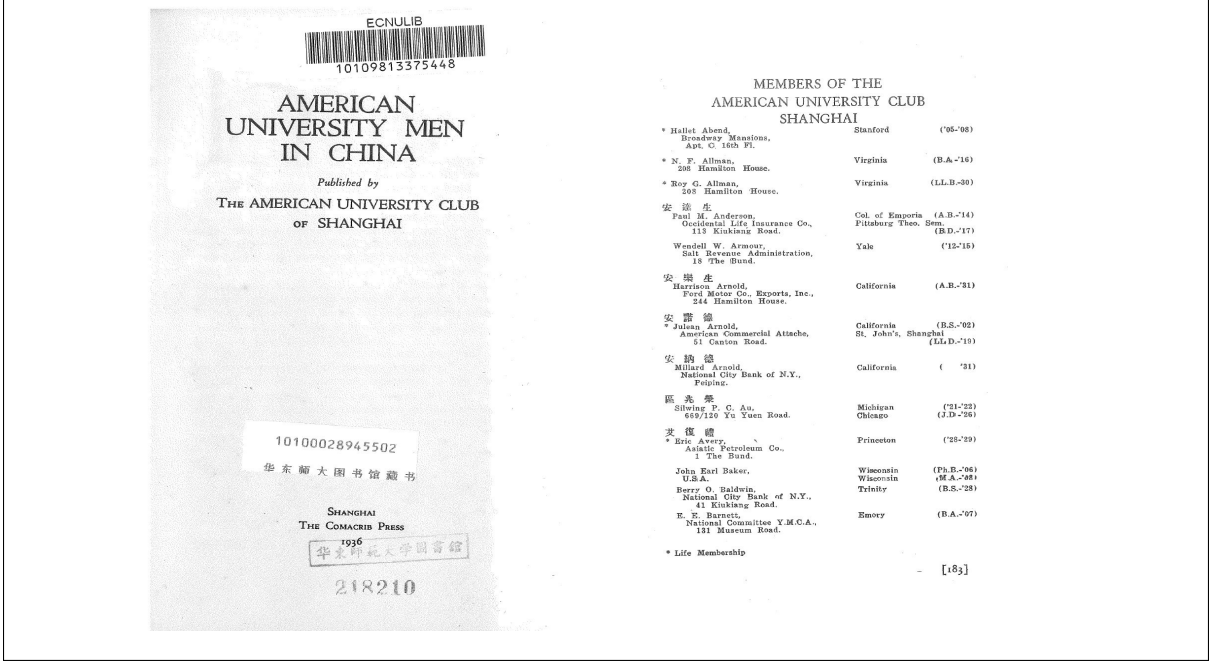
students_data

	Name_eng	Nationality	University	State	Uni_city	Degree_	Degree_level	Field_main	Field_2	Field_3	Year_s	Year_end	Honc	nDeg	start	end	duration	period	Degree_high
0	Wong_James	Chinese	New Bedford	Massachusetts	New Bedford			Unknown Major	Unknown Major	Unknown Major	1921	1925.0		2	1921	1926	5	1919-1935	
1	Wong_James	Chinese	New York University	New York	New York City			Unknown Major	Unknown Major	Unknown Major	1925	1926.0		2	1921	1926	5	1919-1935	
2	Hsia_Jui-Ching	Chinese	Chicago	Illinois	Chicago	M.A.	Master	Unknown Major	Unknown Major	Unknown Major	1925	1925.0		2	1925	1928	3	1919-1935	Doctorate
3	Hsia_Jui-Ching	Chinese	Columbia	New York	New York City	Ph.D.	Doctorate	Undetermined	Undetermined	Unknown Major	1928	1928.0		2	1925	1928	3	1919-1935	Doctorate
4	Lacy_Carleton	Western	Columbia	New York	New York City	M.A.	Master	Humanities	Arts	Theology	1914	1914.0		6	1911	1929	18	1909-1918	Doctorate
5	Lacy_Carleton	Western	Garrett Biblical Institute	Illinois	Evanston	D.D.	Doctorate	Theology	Theology	Theology	1928	1929.0	Hono	6	1911	1929	18	1909-1918	Doctorate
6	Lacy_Carleton	Western	Garrett Biblical Institute	Illinois	Evanston	S.T.B.	Bachelor	Theology	Theology	Theology	1913	1914.0		6	1911	1929	18	1909-1918	Doctorate
7	Lacy_Carleton	Western	Northwestern	Illinois	Evanston	M.A.	Master	Humanities	Arts	Theology	1914	1914.0		6	1911	1929	18	1909-1918	Doctorate
8	Lacy_Carleton	Western	Ohio Wesleyan	Ohio	Delaware	A.B.	Bachelor	Humanities	Arts	Theology	1911	1911.0		6	1911	1929	18	1909-1918	Doctorate
9	Lacy_Carleton	Western	Ohio Wesleyan	Ohio	Delaware	D.D.	Doctorate	Theology	Theology	Theology	1911	1911.0		6	1911	1929	18	1909-1918	Doctorate
10	Todnem_Lawrence	Western	South Dakota	South Dakota	Vermillion	B.A.	Bachelor	Humanities	Arts	Humanities	1912	1912.0		1	1912	1912	0	1909-1918	Bachelor
11	Howes_James Marvin	Western	Reed	Oregon	Portland	A.B.	Bachelor	Humanities	Arts	Humanities	1915	1915.0		1	1915	1915	0	1909-1918	Bachelor
12	Walker_M.P.	Western	Stevens Institute of Technology	New Jersey	Hoboken	M.E.	Master	Engineering	Engineering	Engineering	1898	1898.0		1	1898	1898	0	1883-1908	Master
13	Brown_Irving S.	Western	Drake	Iowa	Des Moines	A.B.	Bachelor	Humanities	Arts	Humanities	1920	1920.0		1	1920	1920	0	1919-1935	Bachelor
14	Butrick_Richard P.	Western	Georgetown	District of Columbia	Washington			Unknown Major	Unknown Major	Unknown Major	1918	1921.0		1	1918	1921	3	1909-1918	
15	Davis_Monnett B.	Western	Colorado	Colorado	Colorado Springs	A.B.	Bachelor	Humanities	Arts	Humanities	1917	1917.0		1	1917	1917	0	1909-1918	Bachelor
16	Drumright_E.F.	Western	Oklahoma	Oklahoma	Norman	B.S.	Bachelor	Sciences	Sciences	Sciences	1929	1929.0		1	1929	1929	0	1919-1935	Bachelor
17	Engdahl_Russell	Western	Georgetown	District of Columbia	Washington	B.S.	Bachelor	Sciences	Sciences	Sciences	1930	1930.0		2	1925	1930	5	1919-1935	Bachelor
18	Engdahl_Russell	Western	Washington	Washington	Seattle			Unknown Major	Unknown Major	Sciences	1925	1926.0		2	1925	1930	5	1919-1935	Bachelor
19	Pilcher_James B.	Western	Alabama Polytechnic Institute	Alabama	Auburn	B.S.	Bachelor	Sciences	Sciences	Sciences	1920	1920.0		2	1920	1921	1	1919-1935	Master

About the Dataset

About the Students Dataset

- Dataset was prepared and provided with all appreciation by Dr. Cécile Armand from Aix-Marseille Université
- Reference on Zenodo: https://zenodo.org/record/6370094#.Y8_wjuzMKkg
- Extracted from a roster of members of the American University Club in Shanghai published in 1936 .
- The dataset reflect the academic trajectories of the students in the United States individually.
- The Goal is to analyze the flows of students between universities overtime.



ECNULIB
10109813375448

AMERICAN UNIVERSITY MEN IN CHINA

Published by
THE AMERICAN UNIVERSITY CLUB OF SHANGHAI

10100028945502
华东师大图书馆藏书

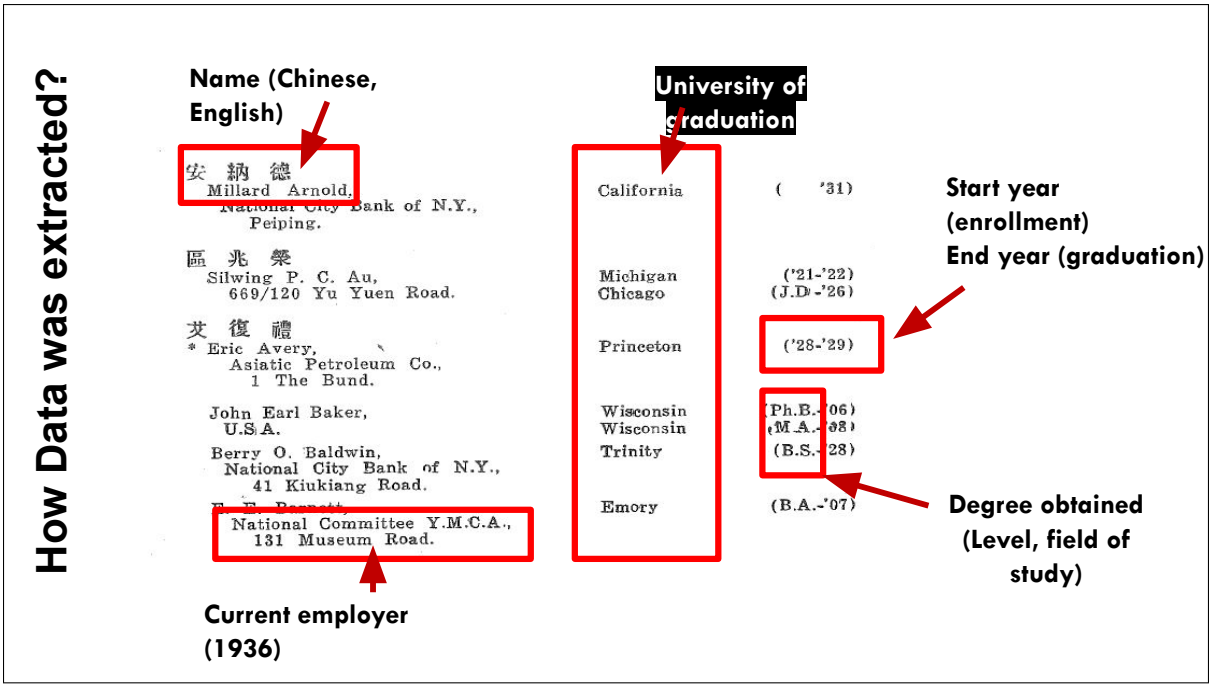
SHANGHAI
THE COMACRIB PRESS
1936

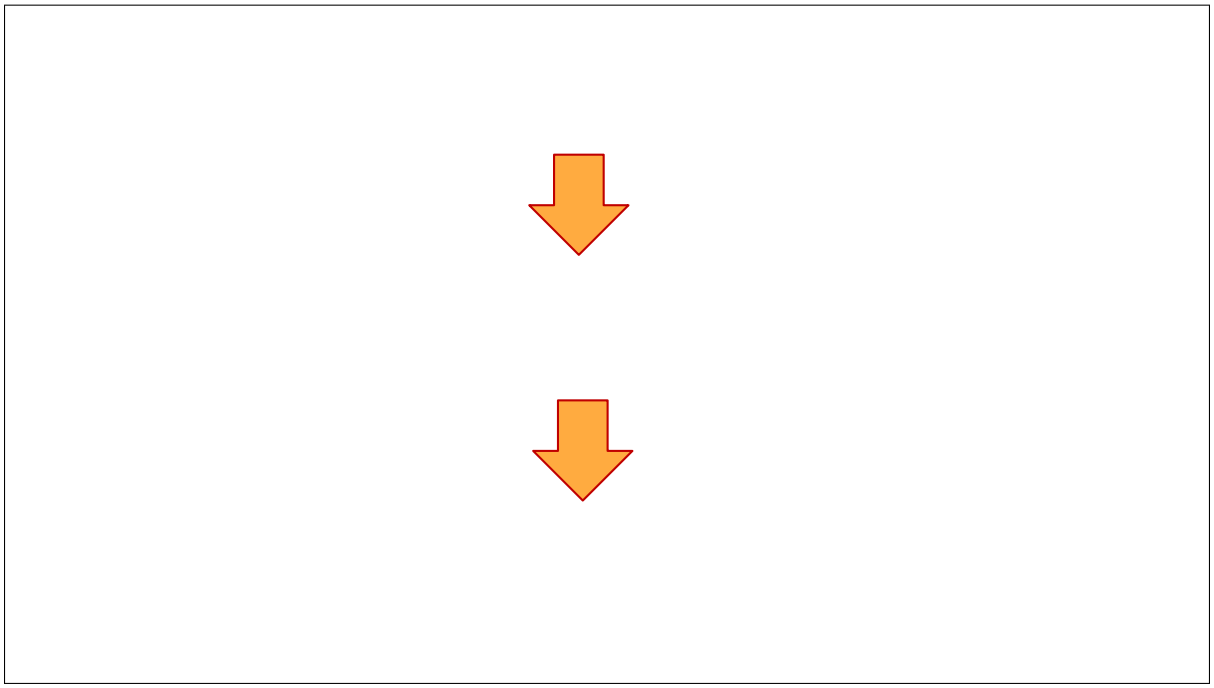
华东师范大学图书馆
218210

MEMBERS OF THE
AMERICAN UNIVERSITY CLUB
SHANGHAI

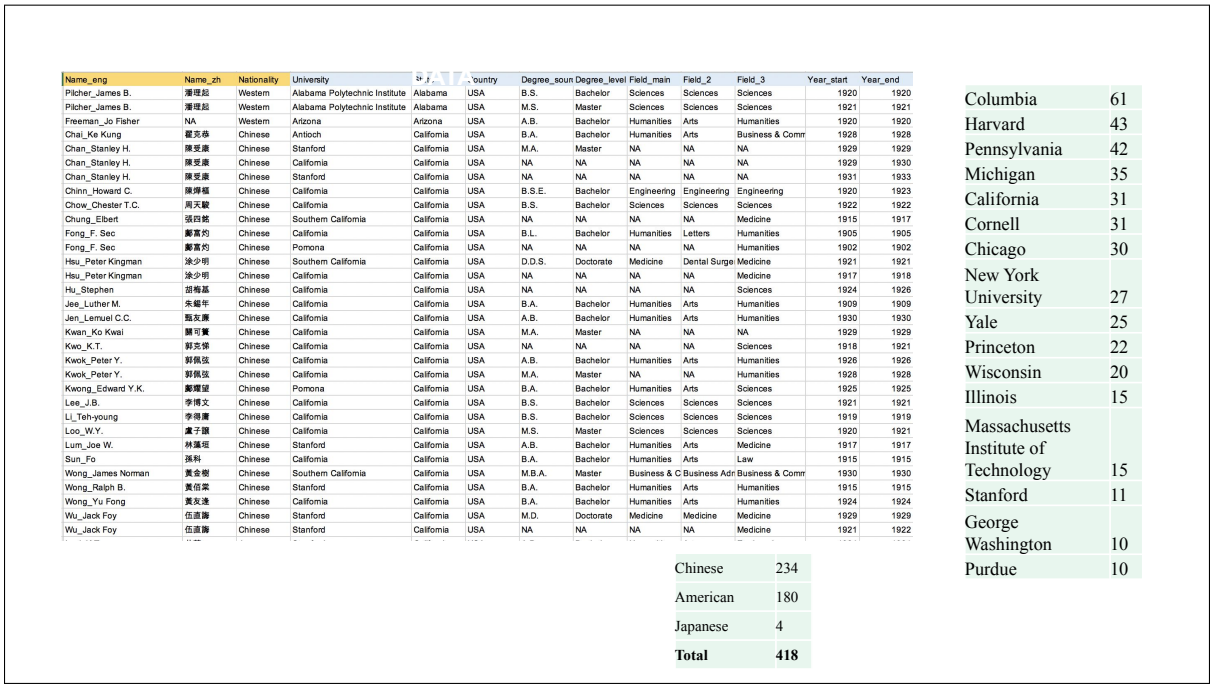
* Hallet Abend, Broadway Mansions, Apt. C 16th Fl.	Stanford	('05-'08)
* N. F. Allman, 208 Hamilton House.	Virginia	(B.A.-'16)
* Roy G. Allman, 208 Hamilton House.	Virginia	(LL.B.-'30)
安 德 生 Paul M. Anderson, Occidental Life Insurance Co., 113 Kiukiang Road.	Col. of Emporia Pittsburg Theo. Sem.	(A.B.-'14) (B.D.-'17)
Wendell W. Armour, Salt Revenue Administration, 18 The Bund.	Yale	('12-'15)
安 培 生 Harrison Arnold, Ford Motor Co., Exports, Inc., 244 Hamilton House.	California	(A.B.-'31)
安 露 德 * Julean Arnold, American Commercial Attache, 51 Canton Road.	California St. John's, Shanghai	(B.S.-'02) (LL.D.-'19)
安 納 德 Millard Arnold, National City Bank of N.Y., Peiping.	California	('31)
區 兆 榮 Silwing P. C. Au, 669/120 Yu Yuen Road.	Michigan Chicago	('21-'22) (J.D.-'26)
艾 復 禮 * Eric Avery, Asiatic Petroleum Co., 1 The Bund.	Princeton	('28-'29)
John Earl Baker, U.S.A.	Wisconsin Wisconsin	(Ph.B.-'06) (M.A.-'08)
Berry O. Baldwin, National City Bank of N.Y., 41 Kiukiang Road.	Trinity	(B.S.-'28)
E. E. Barnett, National Committee Y.M.C.A., 131 Museum Road.	Emory	(B.A.-'07)

* Life Membership





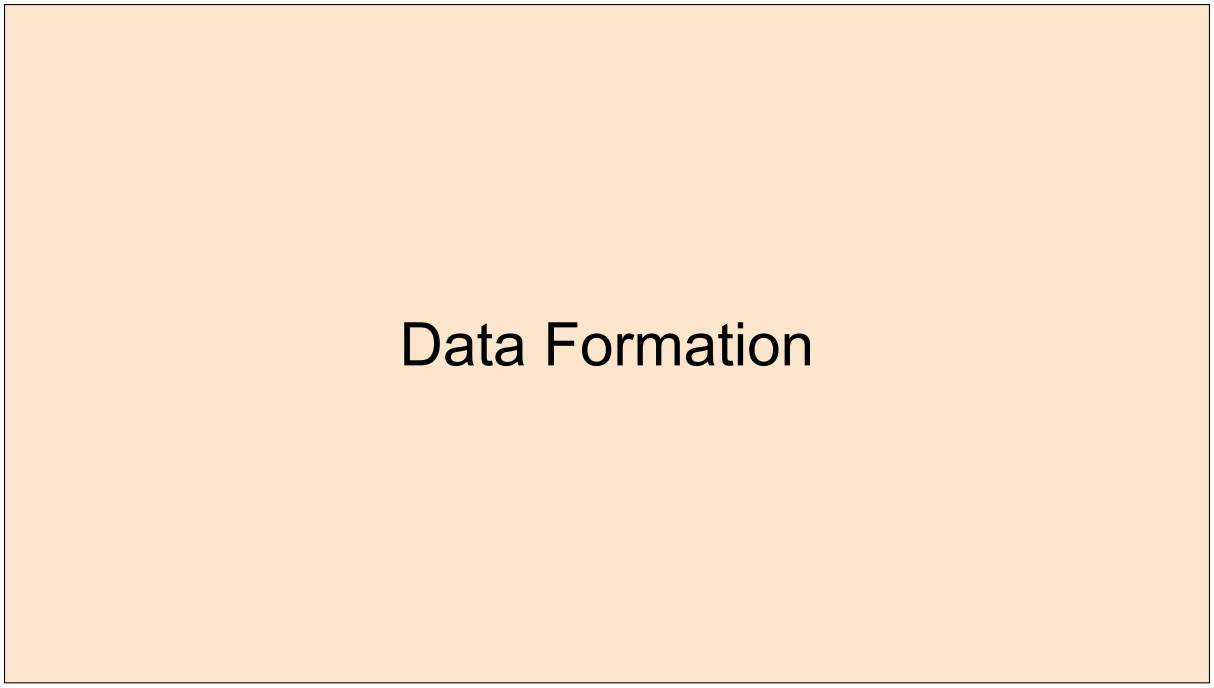
Two-mode edge list	Attributes
Source node (student)	Name of student
	Nationality of student
	Employer
	Sector of employment
	Current address
Target node (university)	University
	City
	State
Edge (curriculum)	Degree
	Field of study (2 levels)
	Year of enrollment
	Year of graduation



Name_eng	Name_zh	Nationality	University	[illegible]	Country	Degree_soun	Degree_level	Field_main	Field_2	Field_3	Year_start	Year_end
Pilcher_James B.	潘理起	Western	Alabama Polytechnic Institute	Alabama	USA	B.S.	Bachelor	Sciences	Sciences	Sciences	1920	1920
Pilcher_James B.	潘理起	Western	Alabama Polytechnic Institute	Alabama	USA	M.S.	Master	Sciences	Sciences	Sciences	1921	1921
Freeman_Jo Fisher	NA	Western	Arizona	Arizona	USA	A.B.	Bachelor	Humanities	Arts	Humanities	1920	1920
Chai_Ke Kung	翟克恭	Chinese	Antioch	California	USA	B.A.	Bachelor	Humanities	Arts	Business & Comm	1928	1928
Chan_Stanley H.	陳受康	Chinese	Stanford	California	USA	M.A.	Master	NA	NA	NA	1929	1929
Chan_Stanley H.	陳受康	Chinese	California	California	USA	NA	NA	NA	NA	NA	1929	1930
Chan_Stanley H.	陳受康	Chinese	Stanford	California	USA	NA	NA	NA	NA	NA	1931	1933
Chinn_Howard C.	陳煒福	Chinese	California	California	USA	B.S.E.	Bachelor	Engineering	Engineering	Engineering	1920	1923
Chow_Chester T.C.	周天毅	Chinese	California	California	USA	B.S.	Bachelor	Sciences	Sciences	Sciences	1922	1922
Chung_Elbert	張四銘	Chinese	Southern California	California	USA	NA	NA	NA	NA	Medicine	1915	1917
Fong_F. Sec	鄺富灼	Chinese	California	California	USA	B.L.	Bachelor	Humanities	Letters	Humanities	1905	1905
Fong_F. Sec	鄺富灼	Chinese	Pomona	California	USA	NA	NA	NA	NA	Humanities	1902	1902
Hsu_Peter Kingman	徐少明	Chinese	Southern California	California	USA	D.D.S.	Doctorate	Medicine	Dental Surge	Medicine	1921	1921
Hsu_Peter Kingman	徐少明	Chinese	California	California	USA	NA	NA	NA	NA	Medicine	1917	1918
Hu_Stephen	胡梅基	Chinese	California	California	USA	NA	NA	NA	NA	Sciences	1924	1926
Jee_Luther M.	朱耀华	Chinese	California	California	USA	B.A.	Bachelor	Humanities	Arts	Humanities	1909	1909
Jen_Lemuel C.C.	甄友廉	Chinese	California	California	USA	A.B.	Bachelor	Humanities	Arts	Humanities	1930	1930
Kwan_Ko Kwai	關可貴	Chinese	California	California	USA	M.A.	Master	NA	NA	NA	1929	1929
Kwo_K.T.	郭克悌	Chinese	California	California	USA	NA	NA	NA	NA	Sciences	1918	1921
Kwok_Peter Y.	郭佩弦	Chinese	California	California	USA	A.B.	Bachelor	Humanities	Arts	Humanities	1926	1926
Kwok_Peter Y.	郭佩弦	Chinese	California	California	USA	M.A.	Master	NA	NA	Humanities	1928	1928
Kwong_Edward Y.K.	鄺耀望	Chinese	Pomona	California	USA	B.A.	Bachelor	Humanities	Arts	Sciences	1925	1925
Lee_J.B.	李博文	Chinese	California	California	USA	B.S.	Bachelor	Sciences	Sciences	Sciences	1921	1921
Li_Teh-young	李得庸	Chinese	California	California	USA	B.S.	Bachelor	Sciences	Sciences	Sciences	1919	1919
Loo_W.Y.	盧子頤	Chinese	California	California	USA	M.S.	Master	Sciences	Sciences	Sciences	1920	1921
Lum_Joe W.	林蕴垣	Chinese	Stanford	California	USA	A.B.	Bachelor	Humanities	Arts	Medicine	1917	1917
Sun_Fo	孫科	Chinese	California	California	USA	B.A.	Bachelor	Humanities	Arts	Law	1915	1915
Wong_James Norman	黃金樹	Chinese	Southern California	California	USA	M.B.A.	Master	Business & C	Business Adn	Business & Comm	1930	1930
Wong_Ralph B.	黃倍棠	Chinese	Stanford	California	USA	B.A.	Bachelor	Humanities	Arts	Humanities	1915	1915
Wong_Yu Fong	黃友逢	Chinese	California	California	USA	B.A.	Bachelor	Humanities	Arts	Humanities	1924	1924
Wu_Jack Foy	伍宣蔣	Chinese	Stanford	California	USA	M.D.	Doctorate	Medicine	Medicine	Medicine	1929	1929
Wu_Jack Foy	伍宣蔣	Chinese	Stanford	California	USA	NA	NA	NA	NA	Medicine	1921	1922

Chinese	234
American	180
Japanese	4
Total	**418**

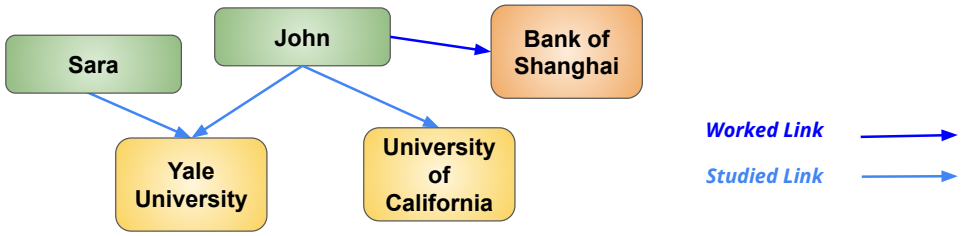
Columbia	61
Harvard	43
Pennsylvania	42
Michigan	35
California	31
Cornell	31
Chicago	30
New York University	27
Yale	25
Princeton	22
Wisconsin	20
Illinois	15
Massachusetts Institute of Technology	15
Stanford	11
George Washington	10
Purdue	10



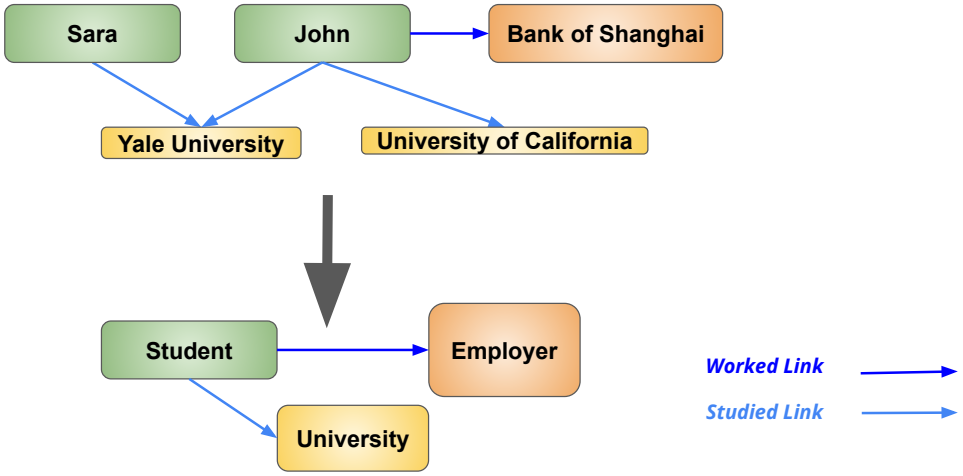
Activity (1): Creating Networks

Steps:

1. Extract ***objects (nodes)*** and their ***attributes*** from the dataset.
2. Select some/all of the nodes.
3. Sketch a network by connecting the selected nodes through possible links.
4. What kind of question/information can this network answer/tell?



Activity (1): Creating Networks



Activity (2): Creating Possible Network Formations

Share Sketches on Miro Board

1) Can we extract objects in our network?

Student	University	Degree	Major	Employer
• Nationality • Degrees	• Location	• Major(s) • Offering University • Date	• Degree(s)	• Location

2) What are the possible formations of the network?

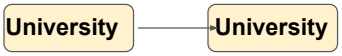
3) For each network form: write 1 or 2 questions it might answer.

Activity (2): Creating Possible Network Formations

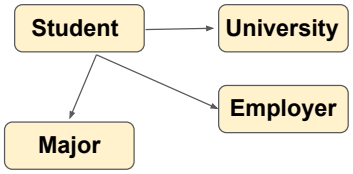
Network (1)



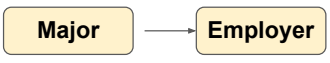
Network (2)



Network (3)



Network (4)



Activity (2): Creating Possible Network Formations

1) Can we extract objects in our network?

Student	**University**	**Degree**	**Major**	**Employer**
• Nationality	• Location	• Major(s) • Offering University • Date	• Degree(s)	• Location

2) What are the possible formations of the network?

3) For each network form: write 1 or 2 questions it might answer.

Decide on **what you want to visualize.**

If you visualize everything you won't see anything!

Activity (3): Finding a Visual Answer for a Question

For the following question:

How is the flow of students change over time?

1) Which of the created networks you find answers the question?

2) How can we track flows over or identify a pattern in the network? Use the following symbols to draw a path or a flow.

   Example: (S) ⇒ (U) ⇒ (U)

- **S** ⇒ Student
- **U** ⇒ University
- **D** ⇒ Degree
- **M** ⇒ Major
- **E** ⇒ Employer

Activity (3): Finding a Visual Answer for a Question

The following paths show examples of flows in a network:

- Interest in Universities: (S) ⇒ (U) ⇒ (U)
- Interest in Majors: (S) ⇒ (M) ⇒ (M)
- Employer's Interests: (S) ⇒ (U) ⇒ (U) ⇒(E)
- (U) [Major='Humanities'] ⇒(E)[Sector='Education']
- (Student) ⇒ Prog (University,Degree) ⇒ Job (Employer, Dep)
- (S) [Date=Between 1980 and 1984]

- **S** ⇒ Student
- **U** ⇒ University
- **D** ⇒ Degree
- **M** ⇒ Major
- **E** ⇒ Employer

Asking Questions

- What are the flows of students between universities over time?
- How can the flow differ based on nationality, academic/professional specialization?
- Let us add more ..

Activity (4): Creating Node/Link Tables

How can I decide the table type that I need?

1. **Link table**: if you have all of your data properties (attributes) placed on links only.
2. **Node table**: if you have your properties placed on nodes only.
3. **Both link and node tables**: if you have properties placed

Data Tables Used

1 studToUni

	Stud_Name	Nationality	University	Degree_level	Field_main	Year_start	Year_end
0	Wong_James	Chinese	New Bedford	Unknown Degree	Unknown Major	1921	1925
1	Wong_James	Chinese	New York University	Unknown Degree	Unknown Major	1925	1926
2	Hsia_Jui-Ching	Chinese	Chicago	Master	Unknown Major	1925	1925
3	Hsia_Jui-Ching	Chinese	Columbia	Doctorate	Undetermined	1928	1928

2 StudToEmp

	Stud_Name	University	UniversityLocation	Nationality	nDeg	Employer	Employer_Location	Sector	Degree_high	Degree_level	Field_main	Year_start	Year_end
0	Abend_Hallet	Stanford	Stanford,California	Western	1	Unknown Employer	Shanghai,China	Unknown Employment Sector	Unknown Degree	Unknown Degree	Unknown Major	1905	1908
1	Allman_N.F.	Virginia	Charlottesville,Virginia	Western	1	Unknown Employer	Shanghai,China	Unknown Employment Sector	Bachelor	Bachelor	Humanities	1916	1916
2	Allman_Roy G.	Virginia	Charlottesville,Virginia	Western	1	Unknown Employer	Shanghai,China	Unknown Employment Sector	Bachelor	Bachelor	Law	1930	1930
3	Anderson_Paul M.	Pittsburg Theological Seminar	Pittsburg,Pennsylvania	Western	2	Occidental Life Insurance Company	Shanghai,China	Company	Bachelor	Bachelor	Theology	1917	1917
4	Armour_Wendell	Yale	New Haven,Connecticut	Western	1	Salt Revenue Administration	Shanghai,China	Central Government	Unknown Degree	Unknown Degree	Unknown Major	1912	1915
5	Arnold_Harrison	California	Berkeley,California	Western	1	Ford Motor Company Exports, Inc.	Shanghai,China	Company	Bachelor	Bachelor	Humanities	1931	1931
6	Arnold_Julean	St. John's University	Shanghai,China	Western	2	U.S. Department of Commerce	Shanghai,China	Foreign Government	Bachelor	Doctorate	Law	1919	1919
7	Arnold_Millard	California	Berkeley,California	Western	1	National City Bank of New York	Peiping,China	Company	Unknown Degree	Unknown Degree	Unknown Major	1931	1931

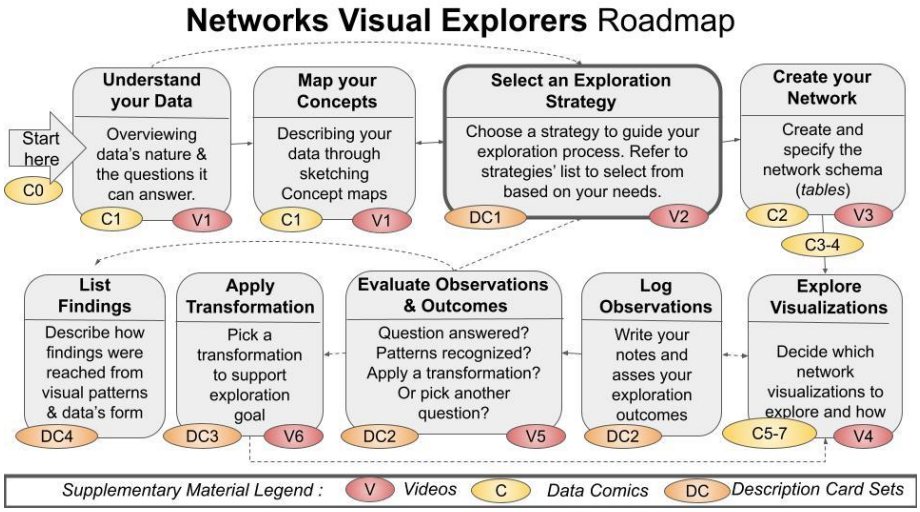
Appendix D

Coaching Program Feedback Questions

Coaching Program Participants Feedback & Reflection Form

Exploration Roadmap & Steps:

- What was your goal through the exploration process?
- Roadmap Steps:
  - Was the roadmap clear to you? What did you learn from it?
  - Which steps were helpful and clear?
  - Which of the steps were unclear to you and you find it needs further explanation? Were you able to follow on your own?
  - Were there some missing steps that weren't present?
- Can you describe/sketch the steps you have followed through the exploration roadmap?



Concept Maps & Network Schema:

- Did you find the concept map sketching clear in:
  - Was the idea of concept maps clear to you?
  - Were the symbols clear and easy to use?
  - Do you think some aren't clear or better symbols could be used (ex. Diamond for relationships)
  - Were some cases you thought of didn't exist in the example or weren't discussed in detail?
- What was hard or unable to decide on how to sketch in your concept map?

- Did the tutorials and examples provided help you in thinking about your data from different perspectives?
- Did the sketching exercise help you to understand more about your data?
- Were you able to transform it into a network schema?
- Were you able to proceed with creating visualizations?

Strategies:

- Did the strategies help you in exploring your data?
- Which strategy did you use?
- Was the description of the strategies clear to you?
- What are the obstacles you've faced?

Supporting Materials (e.g. videos, strategies description cards, data comics):

- What type of material was helpful? Do you have a preference ?
- What type of material have you used for different tasks?
- Were the materials complementing each other?
- Was the level of details in the material provided helpful or further was needed?

Summary:

- Will you be able to use the map on other datasets/problems?
- Was the decision-making and validation of observations helpful through the roadmap?
- Are you able to trust your findings?
- Did the roadmap help you to build a logical understanding of the visual exploration steps/journey?
- Where did most of the thinking or uncertainty take place?
- We have developed a new detailed roadmap, and will share it with you to have your opinion about.

Thank you for your time.

Appendix E

Earlier Sketches of Roadmap

We add here some of the early and abstract versions of the roadmap.

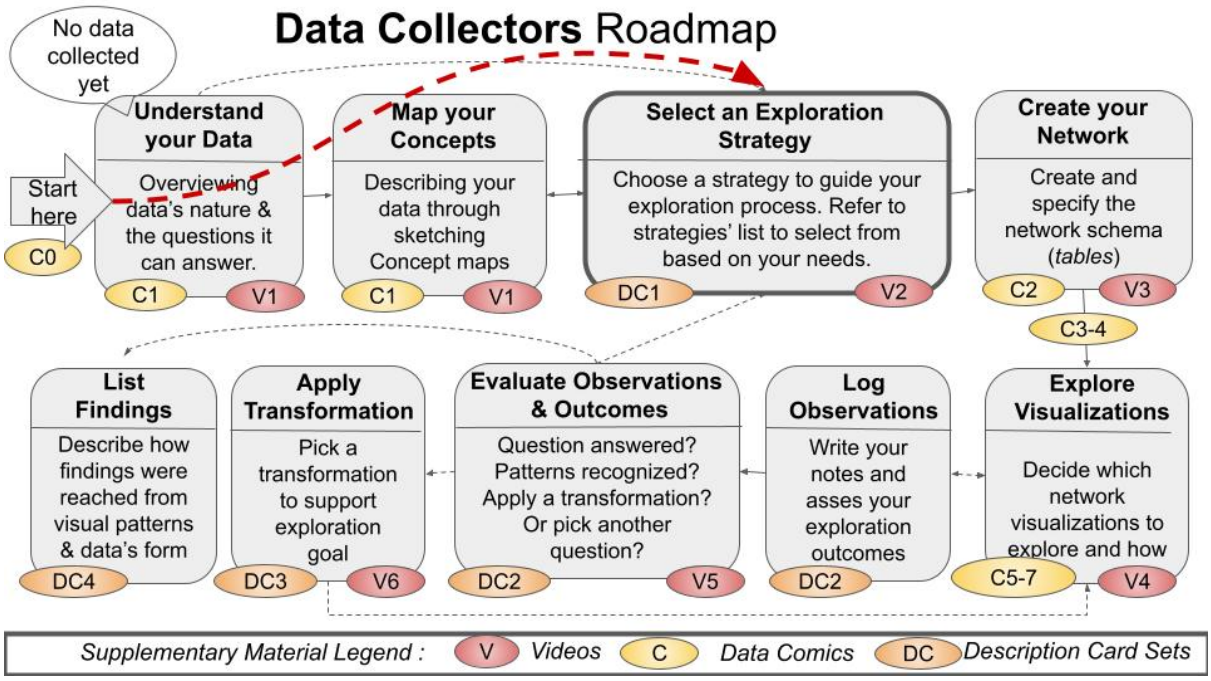


Figure E.1: Early Version of: Informing Data Collection Roadmap

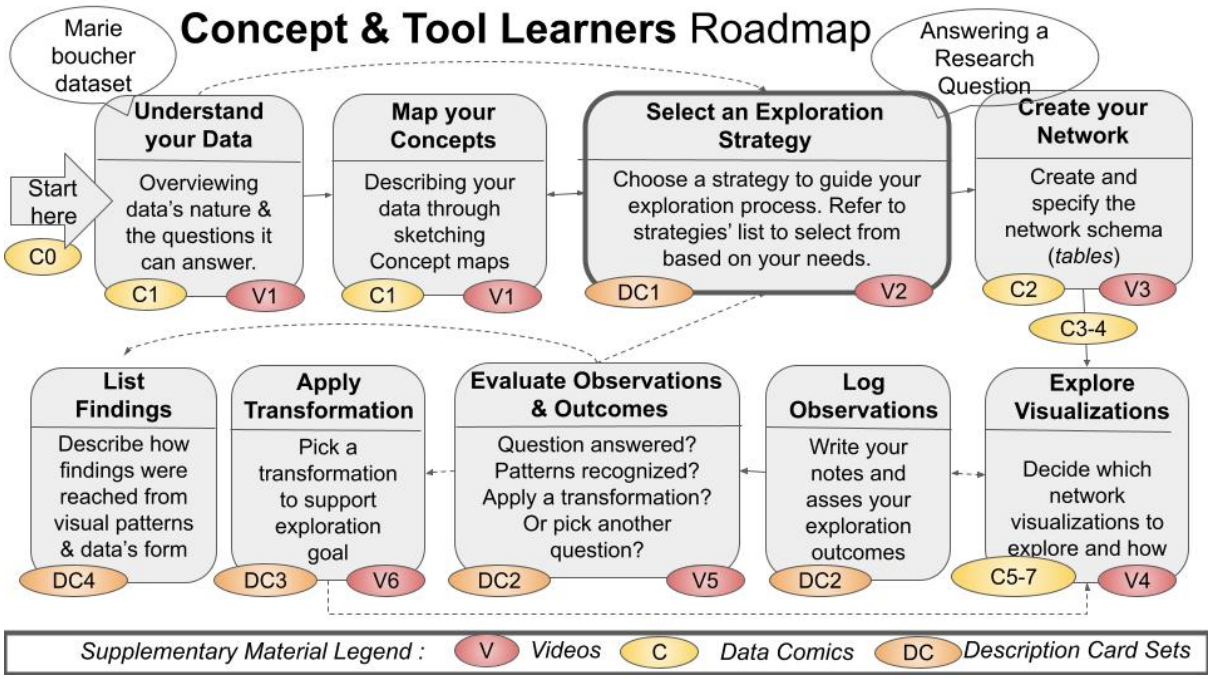


Figure E.2: Early Version of: Learning about Concepts and Tools Roadmap

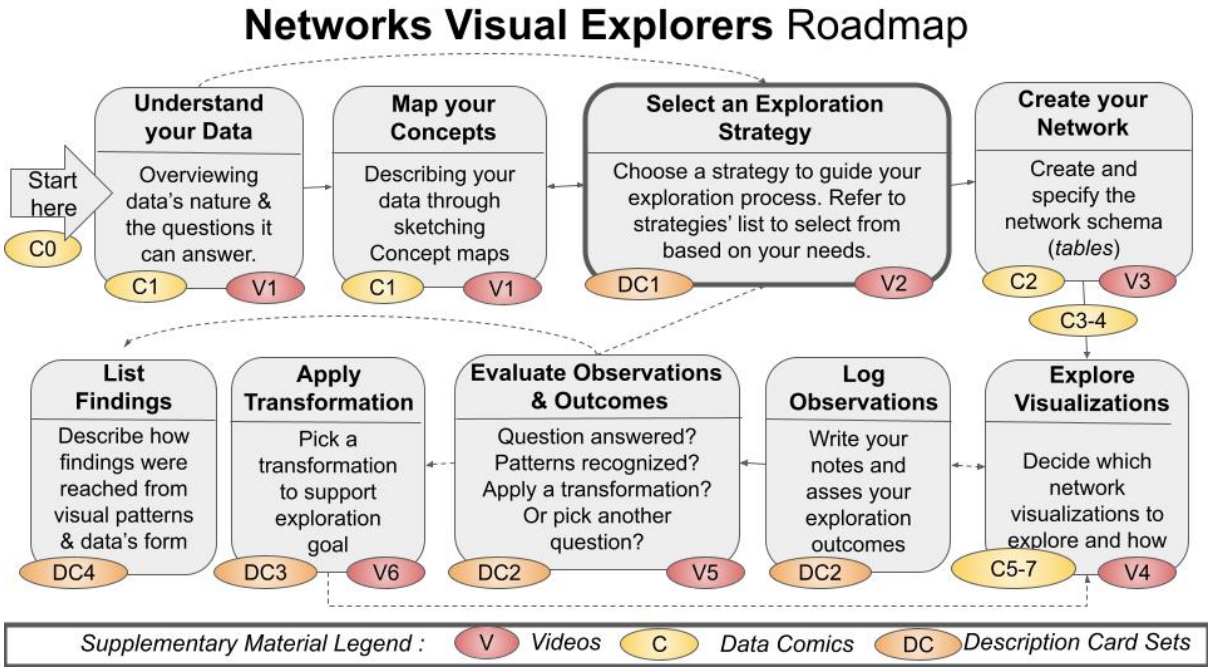


Figure E.3: Early Version of: Explorers Roadmap

Appendix F

Designing Educational Material for Networks Visual Exploration

Selecting an appropriate learning style is an important task when designing material for visual exploration. Roadmap's design can support different **learning styles** [257] of learners (active or passive) and teaching methods (top-down or bottom-up). Using such a roadmap is effective by emphasizing the analyst's involvement, interaction, and activeness within the exploration process. The design of the roadmap and training material should include reflection questions[257] that can enhance user engagement and understanding. However, Sperrle et al. [250] describe guidance based on a bottom-up learning style. Thus, we list the factors we found affecting the learning process of network visual exploration which are:

- Learning Style: (e.g. active vs. passive)
- Exploration Pathways: (tools learners pace differs from that goal-oriented exploration pathway)
- Levels of Guidance: from overview (Roadmap) to detailed (Steps)
- Nature of Supportive Material: data comics, videos, step-by-step tutorials, and others.
- Existence and Nature of Data: Having real-world data to explore offers a different experience than those learning with demo data or those who aim to collect data.

Selecting criteria that define the exploration or learning pathways is a major step in the roadmap design. The 4 **pathways** restructures how we can look into the learning experience from the analyst's goals' perspective, as Andrienko et al. [7] suggest. For example, the *training goal* is applicable more to learners and data collectors. Whereas explorers go under the goals of *informing analysts* and *verifying conclusions*. Some of the goals can be achieved through several steps. It is important to break the process into steps *reduce the cognitive load*, *mitigate bias*,

and *higher levels of user engagement* These 4 pathways, including the lack of data and/or goals, are encountered in real industrial settings, which makes the proposed structure reasonable and useful. This kind of roadmap is a useful resource for teaching network analysis. ARoadmaps can be studied further to customize them based on industry stakeholders and evaluate the suitability of network visualization tools in their own domain. Pathways not only help analysts in their visual exploration of networks but also help tool designers in addressing different needs.

Understanding more about trajectories and personal learning is a must. A **trajectory** is the actual set of steps followed by the analyst in their exploration in comparison to the expected pathway. Several factors steer the learning experience through those trajectories, which are: *the nature of the dataset, the pathway, and the strategies selected*. Analysts create their own trajectories when exploring by combining their pathway with one or more exploration strategies. Further investigation of commonalities within trajectories can assist in learning more about analysts' preferences and practices. In addition, most points where analysts are trapped or end their exploration suddenly can promote further investigation into uncertainty in exploration. In other words, where do analysts need the roadmap to be more enriched?

F.1 How to Design a Roadmap for Visual Exploration Purposes?

The design of roadmaps for visual exploration purposes should offer a systematic approach to apply the steps of the exploration process in a self-regulated manner as much as possible. The following list compiles the main criteria to consider for and in the design of roadmaps:

- **Mirror** the exploration process: step-by-step through the roadmap design.
- **Inform** the analysts: about data's underlying structure and semantics (e.g., relational data, hierarchical data) and how it is mapped into visualizations.
- **Ignite** ideas of how to think about data through providing examples, discussing implementation, and deepening the understanding of how the exploration works. (e.g. exploration strategies, supportive material design, data transformation operations).
- Provoke analysts to **Iterate**: by promoting repetitive exploration and encouraging analysts to practice exploration through different network structures and formations of the same/different datasets.
- **Establish Trust**: in the exploration process through observations logging and validation.

- **Capture** exploration goals: through the design of adaptive means to understand analysts' decisions and actions (e.g. strategies).

- **Customize/Personalize** exploration experiences: through offering pathways and opportunities to support personal learning (e.g. decision points, branching through questions, variety of exploration strategies).

- Provide Alternatives:

- **Ensure Neutrality and generalizability** of the roadmap and resources design: does not limit analysts' thinking and sensemaking process to what specific tool(s) offer (e.g. imposing implicit assumptions about how date or location is used). Roadmap's neutrality can be examined with its ability to be used it to guide along with other tools.

Identifying symbols that represent characteristics of networks' components is crucial, as network sketching. For example, showing where missing data is expected[259] vs. symbols representing absent unknown values. In addition, the reflection step is a crucial step as it validates findings and observations in an accumulative manner. As the analyst explores, the accumulated knowledge enriches the analyst to make better decisions in the future on how to explore data sets with similar structure or nature. For example, sometimes it is better to compare different visualizations of the same network schema. Whereas in some cases, it is better to compare the same visualizations of different schemes of the data[259].

F.2 How can we support the Knowledge Generation Model when Designing Network Visualization Tools and Guides?

The **sensemaking** process through the network's visual exploration is revealed through each step of the roadmap. It shows how the analysts describe their data, which strategies they choose to apply, on what basis they create the network schema and even which type of data they fill the table with. Those should be captured and used to provide a better understanding and support analysts of all trajectories based on the different steps, pathways, and strategies they follow in both tools and roadmap design. Thus, self-regulated Learning should be promoted through roadmaps and visualization tools designed to enable a better understanding of analysts' choices.

In addition, to support the learning process. **Analysts should be exposed to recommended visualization only after the analyst is able to create at least a single visualization, not earlier.** Recommending visualizations at an early stage may overwhelm the analyst through the exploration process. In addition, it hinders the analyst from understanding more about the data's

How to decide what **Visualizations to Explore**?

	Questions to Ask	Explanation & Implementation
1	**What is your strategy(ies)?**	Some strategies can directly point you to a specific visualization to explore, whereas others might involve multiple visualizations to explore and get insights through visual patterns.
2	**What is your chosen network schema?**	Data types of attributes in your schema plays a role in the decision of which visualization to choose (ex. Including time then consider temporal exploration).
3	**Is it your first acquaintance with the dataset?**	Whether you are a beginner or not, if this your first time to explore your network start with node-link diagram to have an overview about your data. Otherwise, If you are looking for other facets to explore your data, choose visualizations supporting your strategy.
4	**Is it your first time to explore networks?**	For beginners, it's recommended to explore using node-link diagram to understand more about networks exploration and analysis concept.
5	**Are you looking for more analytical operations (ex. comparing and contrasting?)**	• If you were comparing/contrasting between different sub-/datasets (ex. ego-networks), compare between same visualization type of each network. • If you were complementing/interpreting different visualizations of the same network, make use of coordinated views or even create your own side-to-side view. • If you were comparing/contrasting between different networks (different schemas created from the same dataset) log your observations from each visualizations and build upon them.

Figure F.1: A question-based guide to assist analysts in selecting the suitable visualization(s) for their exploratory tasks.

nature, how the visualization was developed, and what better presentations could be created of the data as a network.

It is important for each type of data model and its visualization (network visualizations in our case) to create an initial guide on which visualization to explore. As (Figure F.1) shows, a set of questions can guide analysts in their visual exploration process and assist in choosing which visualization to explore. In network visualizations, exploring node-link diagrams as a first step contributes to the analyst's knowledge of the data structure and connectivity. This establishes an initial understanding before moving forward to explore other visualizations. Some might be compacted or hard to interpret in patterns (e.g., adjacency matrix). Thus, guides or sets of questions that will help analysts choose which visualization to start with and what to explore afterward are very crucial as Figure F.2.

F.3 Bloom's Taxonomy and Knowledge Acquisition in Networks Visualization

Bloom's taxonomy represents the advancement in the learning process and knowledge acquisition. The network visual exploration process involves several tasks, starting with network modeling, creating network(s) visualization(s), and making sense of the findings. In Figure F.3, we distribute network visual exploration tasks among Bloom's taxonomy levels. From our

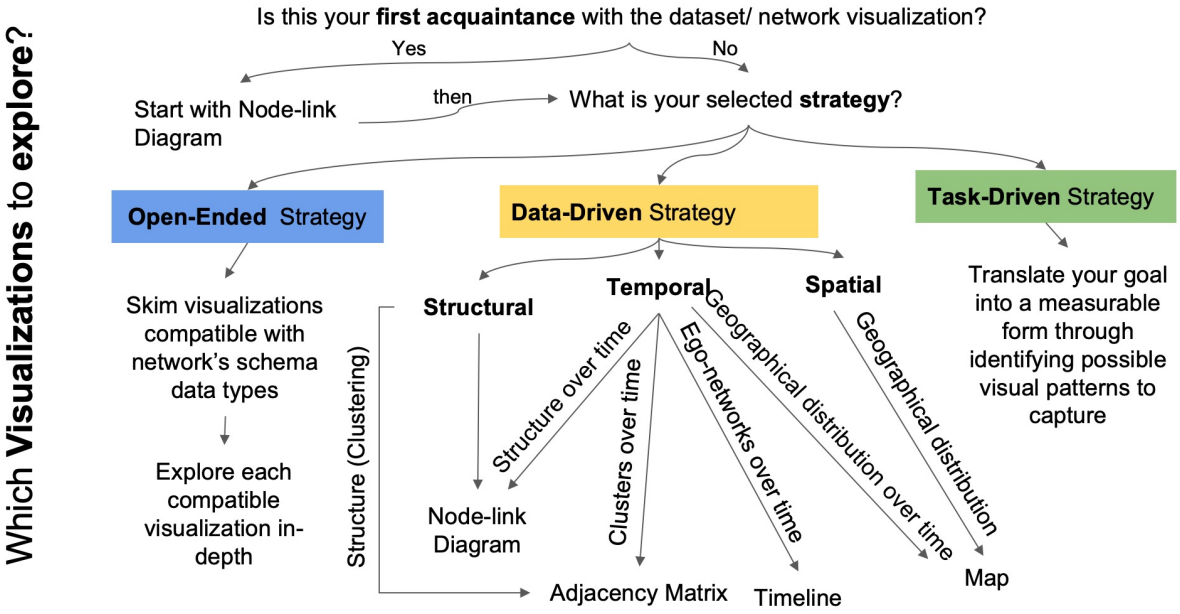


Figure F.2: A partial part of the roadmap flow-chart guiding analysts on choosing which network visualization to explore and in which sequence.

studies, we can describe the user types and their advancement in network visual exploration:

- **Demo Users**: are users whose intention is to learn the concept or tool of network visualization. They experiment only the first two levels of Bloom's taxonomy which are:
  - The ability to **remember** basic facts about network visualizations (e.g. what does a network consist of, mention number of network visualization).
  - The ability to **understand** the demo examples and how they work as explained to them through educational material (e.g. tutorials). Even if they are given a demo data set to create a network and explore visually, they are not able to reach the same level of in-depth understanding as if they work with their own data.
- **Data Strugglers**: are users who intend to explore their data but are not able to create even a single network. Such users are not able to go from (*understand*) to (*apply*) in Bloom's taxonomy. Thus, exercises and activities should be designed to assist them in gaining such skills.
- The **Problem-Solving Cycle**: including both the foraging loop and sensemaking process. Thus, multi-session explorers show stronger signs in their ability to explore data visually by applying, analyzing, and evaluating data and visualization *iteratively* further than single-session explorers. However, although explorers are able to make sense of data,

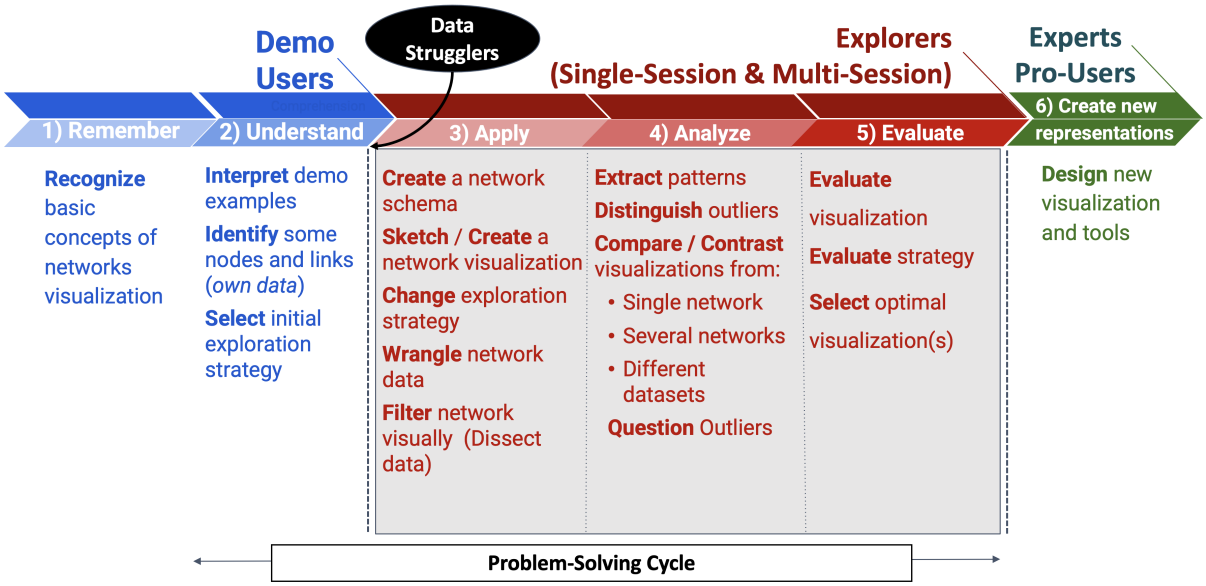


Figure F.3: Network Visual Explorations Tasks Across Bloom's Taxonomy Levels

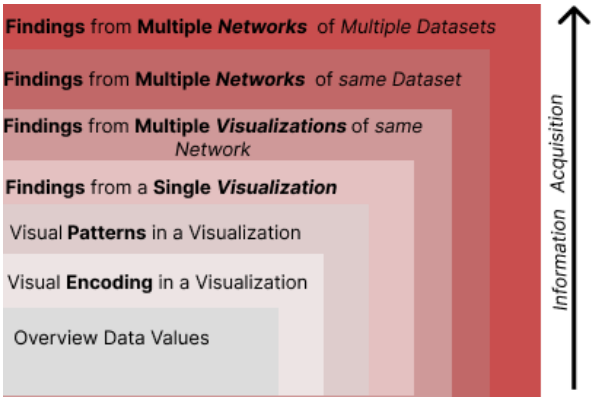


Figure F.4: Knowledge Acquisition through Several Iterations of Network Modeling

not all of them are able to create new designs of visualizations, as this shows one of the highest cognitive levels. As analysts explore, they can reflect on the insights they capture as they explore ranging from a single visualization of a single dataset to multiple visualizations of several datasets. Thus, further studies should investigate how to assist analysts from one level to the following in Bloom's taxonomies. Such activities should assist in sensemaking, problem-solving, and decision-making.

F.4 How do our Findings fit with the Knowledge Generation Model?

In this part, we describe our findings in network visual exploration in the context of the knowledge generation model. Through the network's visual exploration process, the analyst goes through one or more iterations over the exploration steps. For a specific exploration goal, the resulting network visualization can be influenced by any of the 4 factors which are:

1. an initial **mental image** of how a network visualization should look like,
2. a selected **exploration strategy** that decide upon nodes and links,
3. the **design of the network visualization tool** interface and workflow,
4. and the **training and educational material** such as the demo network visualizations and example data sets.

As some analysts initiate their exploration process, their *mental image* of the expected network visualizations might change based on the exploration strategy. However, the last two factors, which are the design of the visualization tool and educational material, can either ignite ideas on how to explore the network data or limit it. Thus, we can illustrate using Figure F.5 that some analysts might start with an initial image, and as they interact with their own data and the demo examples, the resulting network visualization changes. As the analysts advance and become more experienced in network visual exploration, the gap between the resulting network visualization and the initial mental image gets reduced. In other words, the network schema can be easily created and lead to the targeted network visualization.

Analysts initiate the visual exploration process (Figure F.6) with an expected visual image of their network and how they will measure or analyze what they can find through their visual exploration process. By comparing expected visual patterns to those captured, analysts understand more about their decisions, the data they are exploring, and the impact of the chosen strategy.

The knowledge generation model (section 2.4) is influenced by the selected exploration strategy. It defines how the strategy impacts the way the user reaches their findings. In network visual exploration, analysts start generating knowledge even before interacting with the visualization. That is when the analyst starts deciding on the selection of nodes and links and the creation of the network schema, analysts start to understand about data. Then as the user chooses which visualization to create, the user can take any of the following actions: **explore**: by interacting with the visualization interface, **wrangle** data: by transforming data to reach the targeted network visualization, and **select a strategy**: by selecting an initial strategy or changing the current one used.

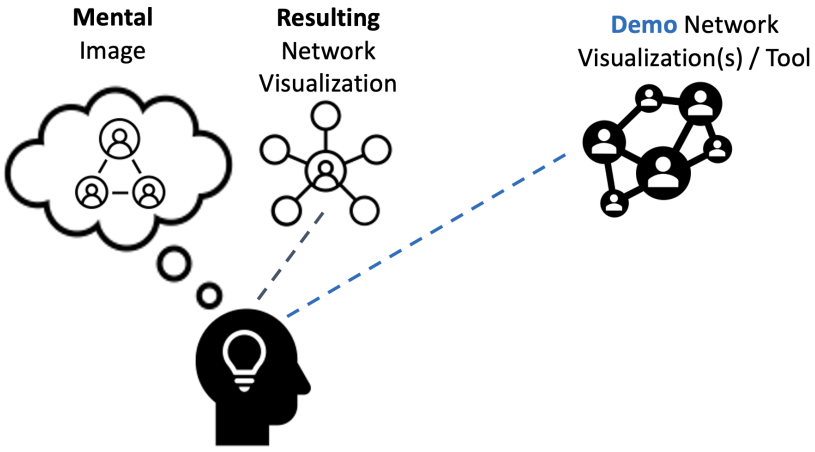


Figure F.5: Networks Visualizations Between Expected *Mental Image* and Actual Resulting Visualization

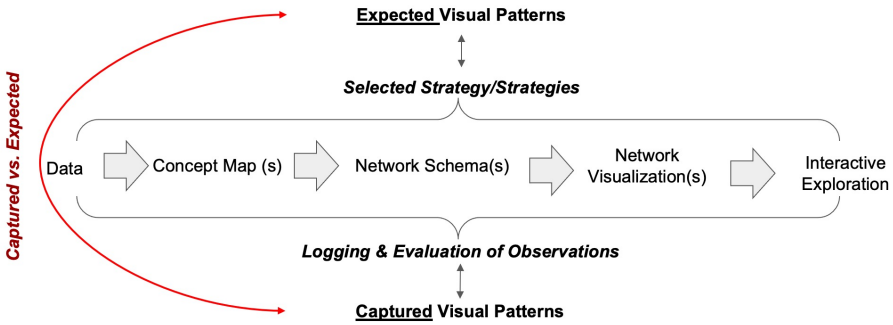


Figure F.6: Comparing Captured Visual Patterns to Expected Visual Patterns Contributes to Knowledge Generation

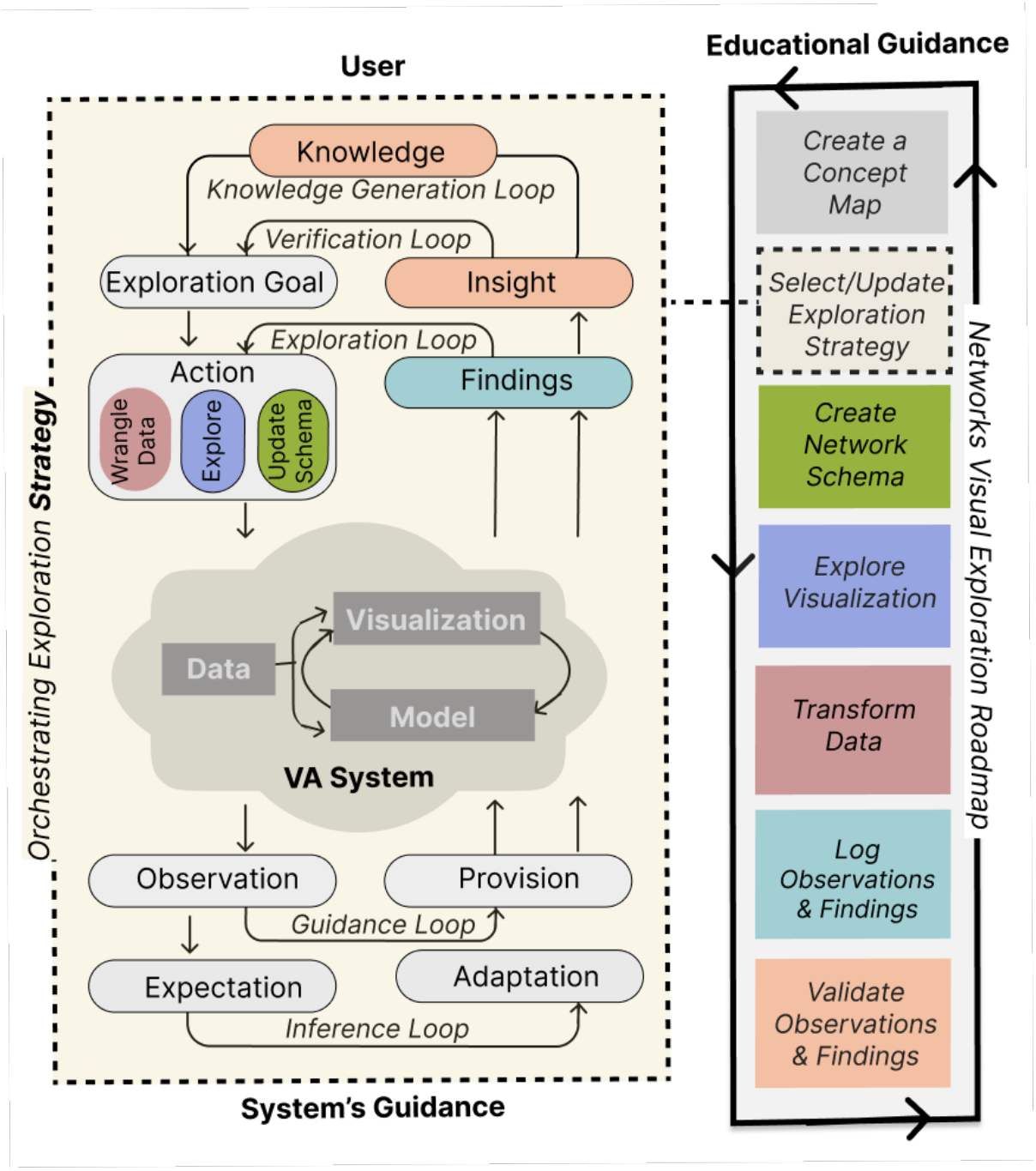


Figure F.7: Knowledge Generation Model in Guided Visual Analytics

As the analyst explores, the captured findings can be used as the basis for further actions, as mentioned above. The captured findings can be verified as insights and used to evaluate the goal (e.g., validate the hypothesis). The goal is defined based on the strategy and data context. This accumulative process contributes to the knowledge about the data itself, not only findings and insight about the data.

Bibliography

[1] N. Aggrawal and A. Arora. Visualization, analysis and structural pattern infusion of dblp co-authorship network using gephi. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, pp. 494–500. IEEE, 2016.

[2] M. AlKadi, V. Serrano, J. Scott-Brown, C. Plaisant, J.-D. Fekete, U. Hinrichs, and B. Bach. Understanding barriers to network exploration with visualization: A report from the trenches. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):907–917, 2022.

[3] B. Alper, B. Bach, N. Henry Riche, T. Isenberg, and J.-D. Fekete. Weighted graph comparison techniques for brain connectivity analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 483–492, 2013.

[4] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 111–117. IEEE, 2005.

[5] R. Amar and J. Stasko. Best paper: A knowledge task-based framework for design and evaluation of information visualizations. In *IEEE Symposium on Information Visualization*, pp. 143–150, 2004. doi: 10.1109/INFVIS.2004.10

[6] M. H. Amini, A. Imteaj, and P. M. Pardalos. Interdependent networks: A data science perspective. *Patterns*, 1(1):100003, 2020.

[7] N. Andrienko, T. Lammarsch, G. Andrienko, G. Fuchs, D. Keim, S. Miksch, and A. Rind. Viewing visual analytics as model building. In *Computer graphics forum*, vol. 37, pp. 275–299. Wiley Online Library, 2018.

[8] D. Archambault, T. Munzner, and D. Auber. Topolayout: Multilevel graph layout by topological features. *IEEE transactions on visualization and computer graphics*, 13(2):305–317, 2007.

[9] D. Archambault, H. Purchase, and B. Pinaud. Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):539–552, 2010.

[10] B. Bach, M. AlKadi, J.-D. Fekete, C. Plaisant, V. Serrano, N. H. Riche, and N. Dufournaud. The vistorian. `https://vistorian.github.io/`. Accessed on July, 1st, 2022.

[11] B. Bach, M. AlKadi, U. Hinrichs, and J. Scott-Brown. Free 6-Weeks Online Course on Interactive Visual Network Exploration. `https://vistorian.github.io/courses.html`. Accessed on July, 1st, 2022.

[12] B. Bach, M. AlKadi, U. Hinrichs, and J. Scott-Brown. Vistorian Data Formats Course. `https://vistorian.github.io/formattingdata.html`. Accessed on July, 1st, 2022.

[13] B. Bach, E. Freeman, A. Abdul-Rahman, C. Turkay, S. Khan, Y. Fan, and M. Chen. Dashboard Design Patterns. *Transactions of Visualization and Computer Graphics (TVCG)*, p. 11, 2023.

[14] B. Bach, M. Keck, F. Rajabiyazdi, T. Losev, I. Meirelles, J. Dykes, R. S. Laramee, M. AlKadi, C. Stoiber, S. Huron, C. Perin, L. Morais, W. Aigner, D. Kosminsky, M. Boucher, S. Knudsen, A. Manataki, J. Aerts, U. Hinrichs, J. C. Roberts, and S. Carpendale. Challenges and Opportunities in Data Visualization Education: A Call to Action. In *IEEE TVCG, Proc VIS 2023*. IEEE, 2024.

[15] B. Bach, E. Pietriga, and J.-D. Fekete. Graphdiaries: Animated transitions and temporal navigation for dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, 20(5):740–754, 2013.

[16] B. Bach, N. H. Riche, S. Carpendale, and H. Pfister. The emerging genre of data comics. *IEEE computer graphics and applications*, 37(3):6–13, 2017.

[17] B. Bach, N. H. Riche, R. Fernandez, E. Giannisakis, B. Lee, and J.-D. Fekete. NetworkCube: bringing dynamic network visualizations to domain scientists. In *Posters of the Conference on Information Visualization (InfoVis)*, 2015.

[18] L. Bartram, M. Correll, and M. Tory. Untidy data: The unreasonable effectiveness of tables. *arXiv preprint arXiv:2106.15005*, 2021.

[19] R. C. Basole, A. Srinivasan, H. Park, and S. Patel. ecoxight: Discovery, exploration, and analysis of business ecosystems using interactive visualization. *ACM Transactions on Management Information Systems (TMIS)*, 9(2):1–26, 2018.

[20] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *International AAAI Conference on Weblogs and Social Media*, 2009.

[21] V. Batagelj and A. Mrvar. Pajek-program for large network analysis. *Connections*, 21(2):47–57, 1998.

[22] A. Batch and N. Elmqvist. ori. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):278–287, 2017. doi: 10/gcp8bz

[23] G. D. Battista, P. Eades, R. Tamassia, and I. G. Tollis. *Graph drawing: algorithms for the visualization of graphs*. Prentice Hall PTR, 1998.

[24] L. Battle, R. Chang, and M. Stonebraker. Dynamic prefetching of data tiles for interactive visualization. In *Proceedings of the 2016 International Conference on Management of Data*, pp. 1363–1375, 2016.

[25] L. Battle and J. Heer. Characterizing Exploratory Visual Analysis: A Literature Review and Evaluation of Analytic Provenance in Tableau. *Computer Graphics Forum*, 38(3):145–159, June 2019. doi: 10.1111/cgf.13678

[26] M. Baur. *visone - Software for the Analysis and Visualization of Social Networks*. PhD thesis, Karlsruhe Institute of Technology, 2008.

[27] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. The state of the art in visualizing dynamic graphs. *EuroVis (STARs)*, 2014.

[28] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. A taxonomy and survey of dynamic graph visualization. In *Computer graphics forum*, vol. 36, pp. 133–159. Wiley Online Library, 2017.

[29] S. S. Beheshitha, M. Hatala, D. Gašević, and S. Joksimović. The role of achievement goal orientations when studying effect of learning analytics visualizations. In *Proceedings of the sixth international conference on learning analytics & knowledge*, pp. 54–63, 2016.

[30] M. Behrisch, B. Bach, N. Henry Riche, T. Schreck, and J.-D. Fekete. Matrix reordering methods for table and network visualization. In *Computer Graphics Forum*, vol. 35, pp. 693–716. Wiley Online Library, 2016.

[31] S. Bender-deMoll and D. A. McFarland. The art and science of dynamic network visualization. *Journal of Social Structure*, 7(2):1–38, 2006.

[32] S. Bender-deMoll and D. A. McFarland. The art and science of dynamic network visualization. *Journal of Social Structure*, 7, 2006.

[33] E. Bertini, A. Perer, C. Plaisant, and G. Santucci. Beliv'08: Beyond time and errors: novel evaluation methods for information visualization. In *CHI'08 extended abstracts on Human factors in computing systems*, pp. 3913–3916. 2008.

[34] A. Bigelow, C. Nobre, A. Lex, and M. Meyer. Mure.js: Toward flexible authoring and reshaping of networks. In *Proceedings of the IEEE Information Visualization Conference - Posters (InfoVis)*, 2018.

[35] A. Bigelow, C. Nobre, M. Meyer, and A. Lex. Origraph: Interactive network wrangling. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, Oct. 2019. doi: 10/gpsvww

[36] A. Bigelow, K. Williams, and K. E. Isaacs. Guidelines for pursuing and revealing data abstractions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1503–1513, 2020.

[37] M. Bilgic, L. Licamele, L. Getoor, and B. Shneiderman. D-dupe: An interactive tool for entity resolution in social networks. In *2006 IEEE Symposium On Visual Analytics Science And Technology*, pp. 43–50. IEEE, 2006.

[38] T. Blascheck, M. John, K. Kurzhals, S. Koch, and T. Ertl. Va 2: a visual analytics approach for evaluating visual analytics applications. *IEEE transactions on visualization and computer graphics*, 22(1):61–70, 2015.

[39] T. Blascheck, S. Koch, and T. Ertl. Logging Interactions to Learn About Visual Data Coverage. *LIVVIL: Logging Interactive Visualizations & Visualizing Interaction Logs*, p. 4, 2016.

[40] J. E. Block and E. D. Ragan. Micro-entries: Encouraging deeper evaluation of mental models over time for interactive data systems. In *2020 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)*, pp. 38–47. IEEE, 2020.

[41] J. E. Block and E. D. Ragan. Micro-entries: Encouraging deeper evaluation of mental models over time for interactive data systems. In *Proceedings of the 2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, pp. 38–47, 2020. doi: 10/gpsvws

[42] A. Bokhare and P. Metkewar. Visualization and interpretation of gephi and tableau: a comparative study. In *Advances in Electrical and Computer Technologies*, pp. 11–23. Springer, 2021.

[43] E. M. Bonsignore, C. Dunne, D. Rotman, M. Smith, T. Capone, D. L. Hansen, and B. Shneiderman. First steps to netviz nirvana: Evaluating social network analysis with NodeXL. In *Proceedings of the 2009 International Conference on Computational Science and Engineering*. IEEE, 2009. doi: 10/dx2p6n

[44] K. Börner, A. Maltese, R. N. Balliet, and J. Heimlich. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3):198–213, 2016.

[45] K. Borner and D. E. Polley. *Visual insights: A practical guide to making sense of data*. MIT Press, 2014.

[46] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, dec 2011. doi: 10/b7bhhf

[47] L. Bounegru, T. Venturini, J. Gray, and M. Jacomy. Narrating networks: Exploring the affordances of networks as storytelling devices in journalism. *Digital Journalism*, 5(6):699–730, 2017.

[48] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1963–1972, 2014.

[49] R. E. Boyatzis. *Transforming Qualitative Information. Thematic Analysis and Code Development*. Sage Publishing, 1998.

[50] R. E. Boyatzis. *Transforming Qualitative Information. Thematic Analysis and Code Development*. Sage Publishing, 1998.

[51] U. Brandes, P. Kenis, and J. Raab. Explanation through network visualization. *Methodology*, 2(1):16–23, 2006.

[52] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics*, 19(12):2376–2385, 2013.

[53] N. Bressa, K. Wannamaker, H. Korsgaard, W. Willett, and J. Vermeulen. Sketching and ideation activities for situated visualization design. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, pp. 173–185, 2019.

[54] V. J. Bristor and S. V. Drake. Linking the language arts and content areas through visual technology. *THE journal*, 22(2):74–77, 1994.

[55] E. T. Brown, A. Ottley, H. Zhao, Q. Lin, R. Souvenir, A. Endert, and R. Chang. Finding waldo: Learning about users from their interactions. *IEEE Transactions on visualization and computer graphics*, 20(12):1663–1672, 2014.

[56] A. Burns, C. Lee, R. Chawla, E. Peck, and N. Mahyar. Who do we mean when we talk about visualization novices? In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2023.

[57] V. L. Byrd and N. Dwenger. Activity worksheets for teaching and learning data visualization. *IEEE Computer Graphics and Applications*, 41(6):25–36, 2021. doi: 10.1109/MCG.2021.3115396

[58] D. Camacho, Á. Panizo-LLedot, G. Bello-Orgaz, A. Gonzalez-Pardo, and E. Cambria. The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion*, 63:88–120, 2020.

[59] Y. Carbonneaux, J.-M. Laborde, and R. M. Madani. Cabri-graph: A tool for research and teaching in graph theory. In *International Symposium on Graph Drawing*, pp. 123–126. Springer, 1995.

[60] S. K. Card, J. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[61] D. Ceneda, N. Andrienko, G. Andrienko, T. Gschwandtner, S. Miksch, N. Piccolotto, T. Schreck, M. Streit, J. Suschnigg, and C. Tominski. Guide me in analysis: A framework for guidance designers. In *Computer Graphics Forum*, vol. 39, pp. 269–288. Wiley Online Library, 2020.

[62] D. Ceneda, T. Gschwandtner, T. May, S. Miksch, M. Streit, and C. Tominski. Guidance or no guidance? a decision tree can help. In *EuroVA@ EuroVis*, pp. 19–23, 2018.

[63] D. Ceneda, T. Gschwandtner, and S. Miksch. A review of guidance approaches in visual data analysis: A multifocal perspective. In *Computer Graphics Forum*, vol. 38, pp. 861–879. Wiley Online Library, 2019.

[64] A. Chamberlain, A. Crabtree, T. Rodden, M. Jones, and Y. Rogers. Research in the wild: understanding'in the wild'approaches to design and development. In *Proceedings of the Designing Interactive Systems Conference*, pp. 795–796, 2012.

[65] C. Chang, B. Bach, T. Dwyer, and K. Marriott. Evaluating perceptually complementary views for network exploration tasks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1397–1407, 2017.

[66] R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky. Defining insight for visual analytics. *IEEE Computer Graphics and Applications*, 29(2):14–17, 2009.

[67] D. H. Chau. Data mining meets hci: Making sense of large graphs. *CMU Pittsburgh PA Machine Learning Dept, Tech. Rep*, 2012.

[68] C. Chen. Top 10 unsolved information visualization problems. *IEEE computer graphics and applications*, 25(4):12–16, 2005.

[69] H. Chen, U. Soni, Y. Lu, R. Maciejewski, and S. Kobourov. Same stats, different graphs. In *International Symposium on Graph Drawing and Network Visualization*, pp. 463–477. Springer, 2018.

[70] K.-T. Chen, T. Dwyer, B. Bach, and K. Marriott. It's a wrap: Toroidal wrapping of network visualisations supports cluster understanding tasks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2021.

[71] B. chul Kwon, B. Fisher, and J. S. Yi. Visual analytic roadblocks for novice investigators. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 3–11. IEEE, 2011.

[72] C. Collins, N. Andrienko, T. Schreck, J. Yang, J. Choo, U. Engelke, A. Jena, and T. Dwyer. Guidance in the human–machine analytics process. *Visual Informatics*, 2(3):166–180, 2018.

[73] M. Conroy. Networks, maps, and time: Visualizing historical networks using palladio. *DHQ: Digital Humanities Quarterly*, 15(1), 2021.

[74] A. Conzemius and J. O'Neill. *The power of SMART goals: Using goals to improve student learning*. Solution Tree Press, 2009.

[75] K. Cook, N. Cramer, D. Israel, M. Wolverton, J. Bruce, R. Burtner, and A. Endert. Mixed-initiative visual analytics using task-driven recommendations. In *2015 IEEE conference on visual analytics science and technology (VAST)*, pp. 9–16. IEEE, 2015.

[76] C. J. Costa and M. Aparicio. Supporting the decision on dashboard design charts. In *Proceedings of 254th The IIER International Conference 2019*, pp. 10–15, Sept. 2019. doi: 10362/85057

[77] C. Cramer, L. Sheetz, H. Sayama, P. Trunfio, H. E. Stanley, and S. Uzzo. NetSci high: Bringing network science research to high schools. In *Studies in Computational Intelligence*, pp. 209–218. Springer International Publishing, 2015. doi: 10/gpb58v

[78] Cyram Inc. Netminer - social network analysis software.

[79] N. Dabbagh and A. Kitsantas. Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *The Internet and higher education*, 15(1):3–8, 2012.

[80] F. Dabek and J. J. Caban. A grammar-based approach for modeling user interactions and generating suggestions during the data exploration process. *IEEE transactions on visualization and computer graphics*, 23(1):41–50, 2016.

[81] T. N. Dang, N. Pendar, and A. G. Forbes. TimeArcs: Visualizing Fluctuations in Dynamic Networks. *Computer Graphics Forum*, 2016. doi: 10/f8wcjk

[82] W. de Nooy, A. Mrvar, and V. Batagelj. *Exploratory social network analysis with Pajek*. Cambridge University Press, Cambridge, 2018.

[83] A. Diehl, A. Abdul-Rahman, M. El-Assady, B. Bach, D. A. Keim, and M. Chen. Visguides: A forum for discussing visualization guidelines. In *EuroVis (Short Papers)*, pp. 61–65, 2018.

[84] F. Du, C. Plaisant, B. Shneiderman, et al. EventFlow Visualization Tool for Categorical Temporal Event Data Analysis. `https://hcil.umd.edu/eventflow/`. Accessed on April, 30th, 2023.

[85] J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public interest*, 14(1):4–58, 2013.

[86] E. Duval. Attention please! learning analytics for visualization and recommendationåa. In *Proceedings of the 1st international conference on learning analytics and knowledge*, pp. 9–17, 2011.

[87] C. D'Ignazio and R. Bhargava. 13. data visualization literacy: A feminist starting point. *Data visualization in society*, p. 207, 2020.

[88] V. Echeverria, R. Martinez-Maldonado, S. B. Shum, K. Chiluiza, R. Granda, and C. Conati. Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling. *Journal of Learning Analytics*, 5(3):73–97, 2018.

[89] D. Edge, N. H. Riche, J. Larson, and C. White. Beyond tasks: An activity typology for visual analytics. *IEEE transactions on visualization and computer graphics*, 24(1):267–277, 2017.

[90] M. Eiglsperger, S. P. Fekete, and G. W. Klau. Orthogonal graph drawing. In *Drawing Graphs: Methods and Models*, pp. 121–171. Springer, 2001.

[91] O. ElTayeby and W. Dou. A survey on interaction log analysis for evaluating exploratory visualizations. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV)*, pp. 62–69, 2016.

[92] R. Engels. Planning tasks for knowledge discovery in databases; performing task-oriented user-guidance. In *KDD*, pp. 170–175. Citeseer, 1996.

[93] J.-D. Fekete. Interaction Trace Manager. `https://github.com/INRIA/intertrace`. Accessed on July, 1st, 2022.

[94] D. Fernandes and J. Bernardino. Graph databases comparison: Allegrograph, arangodb, infinitegraph, neo4j, and orientdb. In *Data*, pp. 373–380, 2018.

[95] E. E. Firat, A. Joshi, and R. S. Laramee. Interactive visualization literacy: The state-of-the-art. *Information Visualization*, p. 14738716221081831, 2021.

[96] A. Fortenbacher, L. Beuster, M. Elkina, L. Kappe, A. Merceron, A. Pursian, S. Schwarzrock, and B. Wenzlaff. Lemo: A learning analytics application focussing on user path analysis and interactive visualization. In *2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)*, vol. 02, pp. 748–753, 2013. doi: 10.1109/IDAACS.2013.6663025

[97] E. G. Freedman and P. Shah. Toward a model of knowledge-based graph comprehension. In *International conference on theory and application of diagrams*, pp. 18–30. Springer, 2002.

[98] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the national academy of sciences*, 111(23):8410–8415, 2014.

[99] T. M. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

[100] R. Garner. *Metacognition and reading comprehension.* Ablex Publishing, 1987.

[101] S. Gathani, S. Monadjemi, A. Ottley, and L. Battle. A grammar-based approach for applying visualization taxonomies to interaction logs. In *Computer Graphics Forum*, vol. 41, pp. 489–500. Wiley Online Library, 2022.

[102] M. Ghasemi and D. Amyot. From event logs to goals: a systematic literature review of goal-oriented process mining. *Requirements Engineering*, 25(1):67–93, 2020.

[103] M. Ghoniem, J.-D. Fekete, and P. Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *IEEE Symposium on Information Visualization*, pp. 17–24, 2004. doi: 10.1109/INFVIS.2004.1

[104] F. Gilbert and D. Auber. From databases to graph visualization. In *2010 14th International Conference Information Visualisation*, pp. 128–133. IEEE, 2010.

[105] S. Gomez and D. Laidlaw. Modeling task performance for a crowd of users from interaction histories. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2465–2468, 2012.

[106] D. Gotz and Z. Wen. Behavior-driven visualization recommendation. In *Proceedings of the 14th international conference on Intelligent user interfaces*, pp. 315–324, 2009.

[107] D. Gotz and M. X. Zhou. Characterizing users' visual analytic activity for insight provenance. *Information Visualization*, 8(1):42–55, 2009.

[108] L. Grammel, M. Tory, and M.-A. Storey. How information visualization novices construct visualizations. *IEEE transactions on visualization and computer graphics*, 16(6):943–952, 2010.

[109] S. Gratzl, A. Lex, N. Gehlenborg, N. Cosgrove, and M. Streit. From visual exploration to storytelling and back again. In *Computer Graphics Forum*, vol. 35, pp. 491–500. Wiley Online Library, 2016.

[110] W. D. Gray, B. E. John, and M. E. Atwood. Project ernestine: Validating a goms analysis for predicting and explaining real-world task performance. *Human-computer interaction*, 8(3):237–309, 1993.

[111] T. R. Green. Cognitive dimensions of notations. *People and computers V*, pp. 443–460, 1989.

[112] C. Grimes, D. Tang, and D. Russell. Query logs alone are not enough. In *WWW 2007 Workshop on Query Log Analysis: Social and Technological Changes*, 2007.

[113] G. Guest, K. M. MacQueen, and E. E. Namey. *Introduction to Applied Thematic Analysis*. Sage Publishing, 2012.

[114] H. Guo, S. R. Gomez, C. Ziemkiewicz, and D. H. Laidlaw. A case study using visualization interaction logs and insight metrics to understand how analysts arrive at insights. *IEEE transactions on visualization and computer graphics*, 22(1):51–60, 2015.

[115] A. A. Hagberg, D. A. Schult, and P. J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pp. 11–15. Pasadena, CA USA, Aug. 2008.

[116] P. Hanrahan. Vizql: a language for query, analysis and visualization. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pp. 721–721, 2006.

[117] D. L. Hansen, D. Rotman, E. Bonsignore, N. Milic-Frayling, E. M. Rodrigues, M. Smith, and B. Shneiderman. Do you know the way to SNA?: A process model for analyzing and visualizing social media network data. In *Proceedings of the 2012 International Conference on Social Informatics*. IEEE, Dec. 2012. doi: 10/ggb6g3

[118] S. M. Harrison. A comparison of still, animated, or nonillustrated on-line help with written or spoken instructions in a graphical user interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 82–89, 1995.

[119] H. J. Hartman. *Metacognition in learning and instruction: Theory, research and practice*, vol. 19. Springer Science & Business Media, 2001.

[120] S. He and E. Adar. V iz i t c ards: A card-based toolkit for infovis design education. *IEEE transactions on visualization and computer graphics*, 23(1):561–570, 2016.

[121] J. Heer, M. Bostock, and V. Ogievetsky. A tour through the visualization zoo. *Communications of the ACM*, 53(6):59–67, 2010.

[122] J. Heer and D. Boyd. Vizster: visualizing online social networks. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 32–39, 2005. doi: 10.1109/INFVIS.2005.1532126

[123] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE transactions on visualization and computer graphics*, 14(6):1189–1196, 2008.

[124] J. Heer and A. Perer. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. *Information Visualization*, 13(2):111–133, dec 2012. doi: 10/f5wggh

[125] N. Henry. Visually exploring large social networks. In *IFIP Conference on Human-Computer Interaction*, pp. 604–610. Springer, 2007.

[126] N. Henry and J.-D. Fekete. Matlink: Enhanced matrix visualization for analyzing social networks. In *Proceedings of the IFIP Conference on Human-Computer Interaction*, pp. 288–302. Springer, 2007.

[127] I. Herman, G. Melançon, and M. S. Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on visualization and computer graphics*, 6(1):24–43, 2000.

[128] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al. Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37, 2021.

[129] D. Holten, P. Isenberg, J. J. van Wijk, and J.-D. Fekete. An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs. In *Proceedings of the IEEE Pacific Visualization Symposium*. IEEE, Mar. 2011. doi: 10/cn3tm2

[130] M. Huisman and M. A. Van Duijn. Software for social network analysis. *Models and methods in social network analysis*, 270:e316, 2005.

[131] Humanities + Design a Research Lab at Stanford University. Palladio visualization tool of complex historical data. Accessed on July 29, 2024.

[132] Y.-H. Hung and P. Parsons. Assessing user engagement in information visualization. In *Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1708–1717, 2017.

[133] S. Huron, S. Carpendale, J. Boy, and J.-D. Fekete. Using viskit: A manual for running a constructive visualization workshop. In *Pedagogy of Data Visualization Workshop at IEEE VIS 2016*, 2016.

[134] S. Huron, S. Carpendale, A. Thudt, A. Tang, and M. Mauerer. Constructive visualization. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14, p. 433–442. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10/gfspd2

[135] M. Imre, W. Chang, S. Wang, C. P. Trinter, and C. Wang. Graphvisual: Design and evaluation of a web-based visualization tool for teaching and learning graph visualization. In *Proceedings of American Society for Engineering Education Annual Conference*, 2020.

[136] S. Jänicke. Teaching on the intersection of visualization and digital humanities. In *VISIGRAPP (3: IVAPP)*, pp. 100–109, 2020.

[137] S. Kandel, J. Heer, C. Plaisant, J. Kennedy, F. van Ham, N. H. Riche, C. Weaver, B. Lee, D. Brodbeck, and P. Buono. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, 10(4):271–288, Sept. 2011. doi: 10/bs2nc2

[138] S. Kandel, A. Paepcke, J. Hellerstein, and J. Heer. Wrangler: Interactive visual specification of data transformation scripts. In *Proceedings of the sigchi conference on human factors in computing systems*, pp. 3363–3372, 2011.

[139] P. Karampelas. *Visual Methods and Tools for Social Network Analysis*, pp. 2314–2327. Springer New York, New York, NY, 2014. doi: 10.1007/978-1-4614-6170-8_268

[140] S. Kasica, C. Berret, and T. Munzner. Table scraps: an actionable framework for multi-table data wrangling from an artifact study of computational journalism. *IEEE Transactions on visualization and computer graphics*, 27(2):957–966, 2020.

[141] T. Kauer, M. Dörk, A. L. Ridley, and B. Bach. The public life of data: Investigating reactions to visualizations on reddit. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2021.

[142] N. Kerracher, J. Kennedy, and K. Chalmers. A task taxonomy for temporal graph visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 21(10):1160–1172, Oct. 2015. doi: 10/f7qbxq

[143] H. Khosravi, S. Shabaninejad, A. Bakharia, S. Sadiq, M. Indulska, and D. Gasevic. Intelligent learning analytics dashboards: Automated drill-down recommendations to support teacher data exploration. *Journal of Learning Analytics*, 8(3):133–154, 2021.

[144] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.

[145] G. Klein, J. K. Phillips, E. L. Rall, and D. A. Peluso. A data-frame theory of sensemaking. In *Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making*, vol. 113, 2007.

[146] KNIME. Knime data analytics platform. Accessed on July 29, 2024.

[147] C. Konold and T. Higgins. Reasoning about data. *A research companion to principles and standards for school mathematics*, p. 193–215, 2003.

[148] D. Krackhardt, J. Blythe, and C. McGrath. An improved network drawing program. *Connections*, 17(2):53–55, 1994.

[149] A. Krekhov, M. Michalski, and J. Krüger. Integrating visualization literacy into computer graphics education using the example of dear data. *arXiv preprint arXiv:1907.04730*, 2019.

[150] L. Krempel. Network visualization. *The SAGE handbook of social network analysis*, pp. 558–577, 2011.

[151] M. Krzywinski, I. Birol, S. J. Jones, and M. A. Marra. Hive plots—rational approach to visualizing networks. *Briefings in bioinformatics*, 13(5):627–644, 2012.

[152] K. Kuosa, D. Distante, A. Tervakari, L. Cerulo, A. Fernández, J. Koro, and M. Kailanto. Interactive visualization tools to improve learning and teaching in online learning environments. *International journal of distance education technologies (IJDET)*, 14(1):1–21, 2016.

[153] M. Lal. *Neo4j graph data modeling*. Packt Publishing Ltd, 2015.

[154] H. Lam. A framework of interaction costs in information visualization. *IEEE transactions on visualization and computer graphics*, 14(6):1149–1156, 2008.

[155] C. Lang, G. Siemens, A. Wise, and D. Gasevic. *Handbook of learning analytics*. SOLAR, Society for Learning Analytics and Research New York, 2017.

[156] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI workshop on BEyond time and errors novel evaluation methods for information visualization - BELIV '06*. ACM Press, 2006. doi: 10/b2pm7w

[157] D. J. L. Lee, H. Dev, H. Hu, H. Elmeleegy, and A. G. Parameswaran. Avoiding drill-down fallacies with *VisPilot*: assisted exploration of data subsets. In *Proceedings of the 24th International Conference on Intelligent User Interfaces, IUI 2019, Marina del Ray, CA, USA, March 17-20, 2019*, pp. 186–196. ACM, 2019. doi: 10/gpq2rd

[158] S. Lee, S.-H. Kim, Y.-H. Hung, H. Lam, Y.-a. Kang, and J. S. Yi. How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking. *IEEE transactions on visualization and computer graphics*, 22(1):499–508, 2015.

[159] S. Lee, S.-H. Kim, and B. C. Kwon. Vlat: Development of a visualization literacy assessment test. *IEEE transactions on visualization and computer graphics*, 23(1):551–560, 2016.

[160] N. Li, Ł. Kidziński, P. Jermann, and P. Dillenbourg. Mooc video interaction patterns: What do they tell us? In *Design for Teaching and Learning in a Networked World: 10th European Conference on Technology Enhanced Learning, EC-TEL 2015, Toledo, Spain, September 15-18, 2015, Proceedings 10*, pp. 197–210. Springer, 2015.

[161] W. Li, S. Schöttler, J. Scott-Brown, Y. Wang, S. Chen, H. Qu, and B. Bach. Networknarratives: Data tours for visual network exploration and analysis. *Proc. of ACM SIGCHI*, 2023.

[162] Z. Liu and J. Heer. The effects of interactive latency on exploratory visual analysis. *IEEE transactions on visualization and computer graphics*, 20(12):2122–2131, 2014.

[163] Z. Liu, S. B. Navathe, and J. T. Stasko. Ploceus: Modeling, visualizing, and analyzing tabular data as networks. *Information Visualization*, 13(1):59–89, jun 2013. doi: 10/f5t2qt

[164] Y. Livnat, J. Agutter, S. Moon, R. F. Erbacher, and S. Foresti. A visualization paradigm for network intrusion detection. In *Proceedings from the sixth annual IEEE SMC information assurance workshop*, pp. 92–99. IEEE, 2005.

[165] H.-Y. Lu, T. Fujiwara, M.-Y. Chang, Y.-c. Fu, A. Ynnerman, and K.-L. Ma. Visual analytics of multivariate networks with representation learning and composite variable construction. *arXiv preprint arXiv:2303.09590*, 2023.

[166] O. Lucaciu, M. Băciuţ, G. Băciuţ, R. S. Câmpian, and A. Ionel. The gephi network visualisation of the guided bone regeneration process induced with tissue engineered grafts. In *Biomaterials and Regenerative Biomedicine*, vol. 695 of *Key Engineering Materials*, pp. 215–221. Trans Tech Publications Ltd, 6 2016. doi: 10.4028/www.scientific.net/KEM.695.215

[167] N. Mahyar, S.-H. Kim, and B. C. Kwon. Towards a taxonomy for evaluating user engagement in information visualization. In *Workshop on Personal Visualization: Exploring Everyday Life*, vol. 3, p. 2, 2015.

[168] Maptiler: Maps for developers. https://www.maptiler.com/. Accessed on July, 1st, 2022.

[169] H. M. Marin-Castro and E. Tello-Leal. Event log preprocessing for process mining: a review. *Applied Sciences*, 11(22):10556, 2021.

[170] W. Matcha, D. Gašević, A. Pardo, et al. A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies*, 13(2):226–245, 2019.

[171] W. Matcha, D. Gašević, N. A. Uzir, J. Jovanović, and A. Pardo. Analytics of learning strategies: Associations with academic performance and feedback. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pp. 461–470, 2019.

[172] J. Matejka and G. Fitzmaurice. Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1290–1294, 2017.

[173] E. Mayr, N. Hynek, S. Salisu, and F. Windhager. Trust in Information Visualization. In *Proceedings of the EuroVis Workshop on Trustworthy Visualization (TrustVis)*. The Eurographics Association, 2019. doi: 10.2312/trvis.20191187

[174] S. McKenna, A. Lex, and M. Meyer. Worksheets for guiding novices through the visualization design process. *arXiv preprint arXiv:1709.05723*, 2017.

[175] S. McKenna, D. Mazur, J. Agutter, and M. Meyer. Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2191–2200, 2014. doi: 10.1109/TVCG.2014.2346331

[176] M. Meyer, M. Sedlmair, P. S. Quinan, and T. Munzner. The nested blocks and guidelines model. *Information Visualization*, 14(3):234–249, 2015.

[177] P. Mishra, S. Kumar, and M. K. Chaube. Graph interpretation, summarization and visualization techniques: A review and open research issues. *Multimedia Tools and Applications*, 82(6):8729–8771, 2023.

[178] A. V. Moere, M. Tomitsch, C. Wimmer, B. Christoph, and T. Grechenig. Evaluating the effect of style in information visualization. *IEEE transactions on visualization and computer graphics*, 18(12):2739–2748, 2012.

[179] M. Muller, I. Lange, D. Wang, D. Piorkowski, J. Tsay, Q. V. Liao, C. Dugan, and T. Erickson. How data science workers work with data: Discovery, capture, curation,

design, creation. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15, 2019.

[180] T. Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009.

[181] T. Munzner. *Visualization analysis and design*. CRC press, 2014.

[182] S. Murray. *Interactive data visualization for the web: an introduction to designing with D3*. " O'Reilly Media, Inc.", 2017.

[183] B. Mutlu, M. Gashi, and V. Sabol. Towards a task-based guidance in exploratory visual analytics. In *54th Annual Hawaii International Conference on System Sciences: HICSS 2021*, pp. 1466–1474. IEEE Computer Society, 2021.

[184] A. Noack and C. Lewerentz. A space of layout styles for hierarchical graph models of software systems. In *Proceedings of the 2005 ACM symposium on Software visualization*, pp. 155–164, 2005.

[185] C. Nobre, M. Meyer, M. Streit, and A. Lex. The state of the art in visualizing multivariate networks. In *Computer Graphics Forum*, vol. 38, pp. 807–832. Wiley Online Library, 2019.

[186] C. Nobre, D. Wootton, L. Harrison, and A. Lex. Evaluating multivariate network visualization techniques using a validated design and crowdsourcing approach. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, 2020. doi: 10/gpsvwt

[187] C. Nobre, K. Zhu, E. Mörth, H. Pfister, and J. Beyer. Reading between the pixels: Investigating the barriers to visualization literacy. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2024.

[188] D. A. Norman. *The Design of Everyday Things*. Basic Books, Inc., USA, 2002.

[189] C. North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3):6–9, 2006. doi: 10.1109/MCG.2006.70

[190] NWB Team. Indiana University, Northeastern University, and University of Michigan. Network workbench tool., 2006.

[191] M. Okoe, R. Jianu, and S. Kobourov. Revisited experimental comparison of node-link and matrix representations. In *Lecture Notes in Computer Science*, pp. 287–302. Springer International Publishing, 2018. doi: 10/gpsvwz

[192] J. O'Neill. Smart goals, smart schools. *Educational Leadership*, 57(5):46–50, 2000.

[193] J. O'Madadhain, D. Fisher, S. White, and Y. Boey. The jung (java universal network/graph) framework. *University of California, Irvine, California*, 43, 2003.

[194] A. Pachuilo, E. D. Ragan, and J. R. Goodall. Leveraging Interaction History for Intelligent Configuration of Multiple Coordinated Views in Visualization Tools. *LIVVIL: Logging Interactive Visualizations & Visualizing Interaction Logs 1*, p. 3, 2016.

[195] M. Palacin-Silva, J. Khakurel, A. Happonen, T. Hynninen, and J. Porras. Infusing design thinking into a software engineering capstone course. In *2017 IEEE 30th Conference on Software Engineering Education and Training (CSEE&T)*, pp. 212–221. IEEE, 2017.

[196] K. Peppler, A. Keune, and A. Han. Cultivating data visualization literacy in museums. *Information and Learning Sciences*, 122(1/2):1–16, 2021.

[197] A. Perer and B. Shneiderman. Balancing systematic and flexible exploration of social networks. *IEEE transactions on visualization and computer graphics*, 12(5):693–700, 2006.

[198] A. Perer and B. Shneiderman. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 265–274, 2008.

[199] A. Perer and B. Shneiderman. Systematic yet flexible discovery. In *Proceedings of the 13th International Conference on Intelligent User interfaces - IUI '08*. ACM Press, 2008. doi: 10/cpvcjz

[200] A. Perer and B. Shneiderman. Integrating statistics and visualization for exploratory power: From long-term case studies to design guidelines. *IEEE Computer Graphics and Applications*, 29(3):39–51, may 2009. doi: 10/dh4fgc

[201] I. Pérez-Messina, D. Ceneda, M. El-Assady, S. Miksch, and F. Sperrle. A typology of guidance tasks in mixed-initiative visual analytics environments. In *Computer Graphics Forum*, vol. 41, pp. 465–476. Wiley Online Library, 2022.

[202] R. Pienta, J. Abello, M. Kahng, and D. H. Chau. Scalable graph exploration and visualization: Sensemaking challenges and opportunities. In *2015 International Conference on Big Data and Smart Computing (BIGCOMP)*, pp. 271–278, 2015. doi: 10.1109/35021BIGCOMP.2015.7072812

[203] R. Pienta, A. Tamersoy, A. Endert, S. Navathe, H. Tong, and D. H. Chau. Visage: Interactive visual graph querying. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 272–279, 2016.

[204] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, vol. 5, pp. 2–4. McLean, VA, USA, 2005.

[205] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces*, pp. 109–116, 2004.

[206] C. Plaisant, J.-D. Fekete, and G. Grinstein. Promoting insight-based evaluation of visualizations: From contest to benchmark repository. *IEEE transactions on visualization and computer graphics*, 14(1):120–134, 2007.

[207] J. R. Ponciano, C. D. Linhares, E. R. Faria, and B. A. Travençolo. An online and nonuniform timeslicing method for network visualisation. *Computers & Graphics*, 97:170–182, 2021.

[208] J. Pretorius, H. C. Purchase, and J. T. Stasko. Tasks for multivariate network analysis. In *Multivariate Network Visualization: Dagstuhl Seminar# 13201, Dagstuhl Castle, Germany, May 12-17, 2013, Revised Discussions*, pp. 77–95. Springer, 2014.

[209] K. Puolamäki, E. Oikarinen, B. Kang, J. Lijffijt, and T. De Bie. Interactive visual data exploration with subjective feedback: an information-theoretic approach. *Data Mining and Knowledge Discovery*, 34(1):21–49, 2020.

[210] H. C. Purchase, R. F. Cohen, and M. I. James. An experimental study of the basis for graph drawing algorithms. *ACM Journal of Experimental Algorithmics*, 2:4, jan 1997. doi: 10/fkqn85

[211] Z. Qu and J. Hullman. Keeping Multiple Views Consistent: Constraints, Validations, and Exceptions in Visualization Authoring. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):468–477, Jan. 2018. doi: 10.1109/TVCG.2017.2744198

[212] I. Rae, F. Zhou, M. Bilsing, and P. Bunge. Understanding visual investigation patterns through digital "field" observations. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2022.

[213] K. Reda, A. E. Johnson, J. Leigh, and M. E. Papka. Evaluating user behavior and strategy during visual exploration. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pp. 41–45, 2014.

[214] K. Reda, A. E. Johnson, M. E. Papka, and J. Leigh. Modeling and evaluating user behavior in exploratory visual analysis. *Information Visualization*, 15(4):325–339, 2016.

[215] D. Ren, L. R. Marusich, J. O'Donovan, J. Z. Bakdash, J. A. Schaffer, D. N. Cassenti, S. E. Kase, H. E. Roy, W. yi (Sabrina) Lin, and T. Höllerer. Understanding node-link and matrix visualizations of networks: A large-scale online experiment. *Network Science*, 7(2):242–264, June 2019. doi: 10/gpsvw2

[216] B. Renoust, V. Claver, and J.-F. Baffier. Multiplex flows in citation networks. *Applied network science*, 2(1):1–34, 2017.

[217] A. Rind, W. Aigner, M. Wagner, S. Miksch, and T. Lammarsch. Task cube: A three-dimensional conceptual space of user tasks in visualization design and evaluation. *Information Visualization*, 15(4):288–300, 2016.

[218] J. C. Roberts, C. J. Headleand, and P. D. Ritsos. *Five design-sheets: creative design and sketching for computing and visualisation*. Springer, 2017.

[219] J. C. Roberts, P. D. Ritsos, J. R. Jackson, and C. Headleand. The explanatory visualization framework: An active learning framework for teaching creative computing using explanatory visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):791–801, 2018. doi: 10.1109/TVCG.2017.2745878

[220] A. M. B. Rodrigues, G. D. J. Barbosa, H. C. V. Lopes, and S. D. J. Barbosa. What questions reveal about novices' attempts to make sense of data visualizations: Patterns and misconceptions. *Computers & Graphics*, 94:32–42, 2021.

[221] Y. Rogers, P. Marshall, and J. M. Carroll. *Research in the Wild*. Springer, 2017.

[222] G. Rossetti and R. Cazabet. Community discovery in dynamic networks: a survey. *ACM computing surveys (CSUR)*, 51(2):1–37, 2018.

[223] M. Roughan and J. Tuke. Unravelling graph-exchange file formats, 2015. doi: 10.48550/ARXIV.1503.02781

[224] P. Ruchikachorn and K. Mueller. Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE transactions on visualization and computer graphics*, 21(9):1028–1044, 2015.

[225] S. P. Borgatti and M.G. Everett and L.C. Freeman (Analytic Technologies). UCINET for windows: Software for social network analysis., 2002.

[226] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE transactions on visualization and computer graphics*, 20(12):1604–1613, 2014.

[227] B. Saket and A. Endert. Investigating the manual view specification and visualization by demonstration paradigms for visualization construction. In *Computer Graphics Forum*, vol. 38, pp. 663–674. Wiley Online Library, 2019.

[228] P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE transactions on visualization and computer graphics*, 11(4):443–456, 2005.

[229] P. Saraiya, C. North, V. Lam, and K. A. Duca. An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1511–1522, 2006.

[230] A. Sarikaya, M. Correll, L. Bartram, M. Tory, and D. Fisher. What Do We Talk About When We Talk About Dashboards? *IEEE Transactions on Visualization and Computer Graphics*, 25(1):682–692, Jan. 2019. doi: 10.1109/TVCG.2018.2864903

[231] H. Sayama, C. Cramer, M. A. Porter, L. Sheetz, and S. Uzzo. What are essential concepts about networks? *Journal of Complex Networks*, 4(3):457–474, 2016.

[232] S. Schoffel, J. Schwank, and A. Ebert. A user study on multivariate edge visualizations for graph-based visual analysis tasks. In *Proceedings of the 2016 20th International Conference Information Visualisation (IV)*. IEEE, July 2016. doi: 10/gpsvw3

[233] S. Schöttler, Y. Yang, H. Pfister, and B. Bach. Visualizing and interacting with geospatial networks: A survey and design space. In *Computer Graphics Forum*, vol. 40, pp. 5–33. Wiley Online Library, 2021.

[234] G. Schraw. Promoting general metacognitive awareness. In *Metacognition in learning and instruction: Theory, research and practice*, pp. 3–16. Springer, 2001.

[235] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2366–2375, 2013.

[236] H.-J. Schulz, M. Streit, T. May, and C. Tominski. Towards a characterization of guidance in visualization. In *Poster at IEEE Conference on Information Visualization (InfoVis)*, vol. 2, 2013.

[237] B. A. Schwendimann, M. J. Rodriguez-Triana, A. Vozniuk, L. P. Prieto, M. S. Boroujeni, A. Holzer, D. Gillet, and P. Dillenbourg. Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, 10(1):30–41, 2016.

[238] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012. doi: 10/f4fv7x

[239] V. Serrano Molinero, B. Bach, C. Plaisant, N. Dufournaud, and J.-D. Fekete. Understanding the Use of The Vistorian: Complementing Logs with Context Mini-Questionnaires. In *Proceedings of the Visualization for the Digital Humanities Workshop*. Phoenix, United States, Oct. 2017.

[240] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, et al. Cytoscape: a software environment for integrated models. *Genome Res*, 13(22):426, 1971.

[241] L. Shi, C. Wang, and Z. Wen. Dynamic network visualization in 1.5 d. In *2011 IEEE Pacific Visualization Symposium*, pp. 179–186. IEEE, 2011.

[242] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*, pp. 364–371. Elsevier, 2003.

[243] B. Shneiderman, G. Fischer, M. Czerwinski, M. Resnick, B. Myers, L. Candy, E. Edmonds, M. Eisenberg, E. Giaccardi, T. Hewett, et al. Creativity support tools: Report from a us national science foundation sponsored workshop. *International Journal of Human-Computer Interaction*, 20(2):61–77, 2006.

[244] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pp. 1–7, 2006.

[245] T. J. Shuell. Phases of meaningful learning. *Review of educational research*, 60(4):531–547, 1990.

[246] G. Siemens and D. Gasevic. Guest editorial-learning and knowledge analytics. *Journal of Educational Technology & Society*, 15(3):1–2, 2012.

[247] M. Smith, A. Ceni, N. Milic-Frayling, B. Shneiderman, E. Mendes Rodrigues, J. Leskovec, and C. Dunne. NodeXL: a free and open network overview, discovery and exploration add-in for excel 2007/2010/2013/2016 from the social media research foundation, 2010.

[248] M. A. Smith, B. Shneiderman, N. Milic-Frayling, E. Mendes Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, and E. Gleave. Analyzing (social media) networks with nodexl. In *Proceedings of the fourth international conference on Communities and technologies*, pp. 255–264, 2009.

[249] F. Sperrle, J. Bernard, M. Sedlmair, D. Keim, and M. El-Assady. Speculative execution for guided visual analytics. In *2018 IEEE Workshop on Machine Learning from User Interaction for Visualization and Analytics (MLUI)*, pp. 1–8. IEEE, 2018.

[250] F. Sperrle, D. Ceneda, and M. El-Assady. Lotse: A practical framework for guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1124–1134, 2022.

[251] F. Sperrle, A. Jeitler, J. Bernard, D. Keim, and M. El-Assady. Co-adaptive visual data analysis and guidance processes. *Computers & Graphics*, 100:93–105, 2021.

[252] R. R. Springmeyer, M. M. Blattner, and N. L. Max. A characterization of the scientific data analysis process. In *IEEE Visualization*, vol. 92, pp. 235–242, 1992.

[253] A. Srinivasan, H. Park, A. Endert, and R. C. Basole. Graphiti: Interactive specification of attribute-based edges for network modeling and visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):226–235, 2017.

[254] D. Steinbock, E. Groller, and M. Waldner. Casual visual exploration of large bipartite graphs using hierarchical aggregation and filtering. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*, pp. 1–10. IEEE, 2018.

[255] P. T. Sukumar, G. J. Martinez, T. Grover, G. Mark, S. K. D'Mello, N. V. Chawla, S. M. Mattingly, and A. D. Striegel. Characterizing exploratory behaviors on a personal visualization interface using interaction logs. *EuroVis 2020-Short Papers*, 2020.

[256] G. Summer, T. Kelder, K. Ono, M. Radonjic, S. Heymans, and B. Demchak. cyneo4j: connecting neo4j and cytoscape. *Bioinformatics*, 31(23):3868–3869, 2015.

[257] Y. Tanahashi, N. Leaf, and K.-L. Ma. A study on designing effective introductory materials for information visualization. In *Computer Graphics Forum*, vol. 35, pp. 117–126. Wiley Online Library, 2016.

[258] J. W. Tukey et al. *Exploratory data analysis*, vol. 2. Reading, MA, 1977.

[259] F. Van Ham, H.-J. Schulz, and J. M. Dimicco. Honeycomb: Visual analysis of large scale social networks. In *Human-Computer Interaction–INTERACT 2009: 12th IFIP TC 13 International Conference, Uppsala, Sweden, August 24-28, 2009, Proceedings, Part II 12*, pp. 429–442. Springer, 2009.

[260] K. Verbert, E. Duval, J. Klerkx, S. Govaerts, and J. L. Santos. Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10):1500–1509, 2013.

[261] K. Verbert, S. Govaerts, E. Duval, J. L. Santos, F. Van Assche, G. Parra, and J. Klerkx. Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, 18:1499–1514, 2014.

[262] A. Vogogias, D. Archambault, B. Bach, and J. Kennedy. Visual encodings for networks with multiple edge types. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pp. 1–9, 2020.

[263] T. von Landesberger, S. Fiebig, S. Bremm, A. Kuijper, and D. W. Fellner. Interaction taxonomy for tracking of user actions in visual analytics applications. *Handbook of Human Centric Visualization*, pp. 653–670, 2014.

[264] T. Von Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J. J. van Wijk, J.-D. Fekete, and D. W. Fellner. Visual analysis of large graphs: state-of-the-art and future research challenges. In *Computer graphics forum*, vol. 30, pp. 1719–1749. Wiley Online Library, 2011.

[265] E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):491–500, 2019. doi: 10.1109/TVCG.2018.2865146

[266] J. Walny, C. Frisson, M. West, D. Kosminsky, S. Knudsen, S. Carpendale, and W. Willett. Data Changes Everything: Challenges and Opportunities in Data Visualization Design Handoff. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):12–22, 2019.

[267] C. Wang. Visvisual: A toolkit for teaching and learning data visualization. *IEEE computer graphics and applications*, 42(4):20–26, 2022.

[268] Z. Wang, L. Sundin, D. Murray-Rust, and B. Bach. Cheat sheets for data visualization techniques. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Apr. 2020. doi: 10/gnrq7t

[269] C. Ware. *Visual thinking for design*. Elsevier, 2010.

[270] S. Wasserman and K. Faust. Social network analysis: Methods and applications. 1994.

[271] M. Wattenberg. Visual exploration of multivariate graphs. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. ACM, apr 2006. doi: 10/fknmzs

[272] C. Weaver. Multidimensional data dissection using attribute relationship graphs. In *2010 IEEE Symposium on Visual Analytics Science and Technology*, pp. 75–82. IEEE, 2010.

[273] P. H. Winne. Bootstrapping learner's self-regulated learning. *Psychological Test and Assessment Modeling*, 52(4):472, 2010.

[274] M. A. Wojton, D. Hayde, J. E. Heimlich, and K. Börner. Begin at the beginning: a constructionist model for interpreting data visualizations. *Curator: The Museum Journal*, 61(4):559–574, 2018.

[275] J. wook Ahn, C. Plaisant, and B. Shneiderman. A task taxonomy for network evolution analysis. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):365–376, Mar. 2014. doi: 10/f58vqt

[276] J. S. Yi, Y. ah Kang, J. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics*, 13(6):1224–1231, 2007.

[277] J. S. Yi, Y.-a. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: how do people gain insights using information visualization? In *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*, pp. 1–6, 2008.

[278] V. Yoghourdjian, D. Archambault, S. Diehl, T. Dwyer, K. Klein, H. C. Purchase, and H.-Y. Wu. Exploring the limits of complexity: A survey of empirical studies on graph visualisation. *Visual Informatics*, 2(4):264–282, Dec. 2018. doi: 10/ggjnn5

[279] A. Zeid, S. Kamarthi, C. Duggan, and J. Chin. Capsule: An innovative capstone-based pedagogical approach to engage high school students in stem learning. In *ASME International Mechanical Engineering Congress and Exposition*, vol. 54914, pp. 305–314, 2011.

[280] E. Zgraggen, S. M. Drucker, D. Fisher, and R. DeLine. (s—qu)eries: Visual regular expressions for querying and exploring event sequences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, p. 2683–2692. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2702123.2702262

[281] A. Zoss. Challenges and solutions for short-form data visualization instruction. In *Pedagogy Data Visualization, IEEE VIS Workshop*, 2016.

[282] A. Zoss, A. Maltese, S. M. Uzzo, and K. Börner. Network visualization literacy: Novel approaches to measurement and instruction. In *Network Science In Education*, pp. 169–187. Springer International Publishing, 2018. doi: 10/gpsvw4