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.
Selective citation and the shaping of scientific knowledge: Citation network analysis and the diet–heart debate

Rhodri Ivor Leng

Doctor of Philosophy
Science and Technology Studies
School of Social and Political Science
The University of Edinburgh
November 2019
Word count: 99,736
Abstract

Scientific knowledge is based, in part, on empirical evidence. Scientists contribute to a particular evidence-base by publishing the results of their experiments and observations, but in writing their papers they also use evidence produced by others. Scientists, however, are often selective in the evidence they use and can differ in their interpretations of that evidence. But how selective are scientists when evaluating specific scientific claims in the literature? And how does the use and interpretation of past studies shape what is eventually taken to be scientific knowledge?

This thesis develops an approach based on variants of Citation Network Analysis to study practices of selective citation and their consequences. I develop this approach around a specific case study that traces the emergence and evolution of the conjecture that dietary fats are causally related to the development of coronary heart disease (CHD), and the controversies that this provoked, in the scientific literature between 1900 and 1984.

First, I develop a novel method to detect citation bias, the act of selectively citing only evidence that supports a particular conclusion, by introducing sub-graph analysis to claim-specific citation network analysis. I apply this to trace how the results of the first four clinical trials examining the effectiveness of dietary fat modification in the secondary prevention of CHD were cited by reviews of literature prior to a major consensus conference in 1984. By this, I demonstrate that 82% of reviews that reached supportive conclusions regarding the effectiveness of dietary treatment cited only evidence that supported that conclusion.

Second, I adapt the claim-specific method to examine how the results of the first prospective cohort study to return findings related to the relationship between dietary fat and CHD were cited in the literature over a 7-year period. This study reported no association between dietary fat and CHD, but also reported on a plethora of other factors. Its findings were cited in many different ways that appeared to reflect the research interests of particular communities. I demonstrate that the dietary fat findings made the biggest impact not on those
interested in the relationship between dietary fat and CHD, but in a community advocating a link between sugar consumption and CHD.

Third, I develop a method for capturing a large body of literature representative of the ‘paradigm’ in which the diet–heart link arose and was evaluated via a layered Boolean search query methodology. From these retrieved data, I construct a large citation network that reconstructs the state of the literature prior to 1985, demonstrate network analytic methods for its validation, and construct a citation map of this research area that makes use of modularity analysis for community detection and temporal analysis. By this, I demonstrate that time played a major role in structuring the network – authors tended to cite recent literature. Authors also tended to cite other papers that shared similar research questions, and this appeared to cluster documents into sub-communities based on research focus. These results suggest the presence of a dynamic research front that, over time, focussed on different questions regarding the validity of the diet–heart hypothesis.

Finally, I apply Main Path Analysis to this network to understand how the link between dietary fat and CHD emerged and progressed through the literature. Main Path Analysis detects citation-paths that are most used in a body of literature – a method to understand how selective citation is involved in the development of the main research stream over time. I interpret the results by the reading of the identified documents and demonstrate that highly cited documents tended to be those that preserved the hypothesis rather than those that posed inconveniences to it.

The four major analyses performed throughout this thesis suggest that selective citation played an important role in the development of the diet–heart link. While I applied this approach to a specific case-study, its basic logic and methods ought to be applicable to other cases.
Lay Summary

Scientists contribute to a particular evidence-base by publishing the results of their experiments, but in writing their papers they also use evidence produced by others. However, scientists are often selective in the evidence they use and can differ in their interpretations of that evidence. But how selective are scientists when evaluating specific scientific claims in the literature? And how does the use and interpretation of past studies shape what is eventually taken to be scientific knowledge?

This thesis uses Citation Network Analysis to study citation practices and their consequences. Citation network analysis converts the bibliographies of scientific papers into datasets that record the references between papers. Papers are connected to other papers through their references, and this can be represented as a network of interaction between papers. While scientists must be selective in their referencing due to the volume of relevant papers, they are generally expected to reference a representative sample, when making particular claims. For example, when evaluating the results of clinical trials examining a particular treatment, a scientist would be expected to have read all relevant trials and to have evaluated their findings before making a claim regarding whether the evidence from clinical trials was supportive or unsupportive of the use of that particular treatment. In recent years, some scientists have expressed concern that studies that report findings that undermine popular hypotheses might be overlooked by scientists in their references, and that their exclusion from scientific discussion might lead to false conclusions.

To understand selective citation, I examined the emergence and development of the conjecture that dietary fats are causally related to the development of coronary heart disease (CHD), and the controversies that this provoked, in the scientific literature between 1900 and 1984.

I developed a novel method to detect citation bias, the act of selectively citing only evidence that supports a particular conclusion, by constructing a claim-specific citation network analysis. This records the content of references
by analysing the text that surrounds a particular reference within a citing paper. I apply this to trace how the results of the first clinical trials examining the effectiveness of dietary fat modification in the secondary prevention of CHD were cited by reviews of literature before a major consensus conference in 1984. I demonstrate that 82% of reviews that reached supportive conclusions regarding the effectiveness of dietary treatment cited only evidence that supported that conclusion.

I then adapted the claim-specific method to examine how the results of the first prospective cohort study on the relationship between dietary fat and CHD were cited in the literature over a 7-year period. This study found no relationship between dietary fat and CHD, but reported various incidental findings. Its findings were cited in many different ways that appeared to reflect the research interests of particular communities. The negative dietary fat findings made the biggest impact not on those interested in the relationship between dietary fat and CHD, but in a community proposing a link between sugar consumption and CHD.

I then developed a method for capturing a large body of literature representative of the ‘paradigm’ in which the diet–heart link arose. I used a search query methodology to searches for particular terms in the titles of scientific publications that are indicative of research on particular topics. Advocates of the hypothesis linking dietary fat to CHD, proposed that consumption of particular fats increased the level of cholesterol in the blood, which caused fatty material to be deposited in the arteries (atherosclerosis), increasing the risk of CHD. I captured all papers indexed in the Web of Science, a large bibliometric database, that contained terms relevant to (i) atherosclerosis; (ii) blood cholesterol and cholesterol metabolism; (iii) CHD; and then (iv) searched these for terms indicating a focus on diet and retrieved all publications and their reference lists.

From these data, I constructed a large citation network that reconstructs the state of the literature before 1985, demonstrated network analytic methods for its validation, and constructed a citation map of this research area using modularity analysis to detect communities and to study progression through
time. Time played a major role in structuring the network, as authors tended to cite recent literature. Authors also tended to cite other papers that shared similar research questions, and this clustered documents into sub-communities based on research focus. The map displays a dynamic research front that, over time, focussed on different questions regarding the validity of the diet–heart hypothesis.

Finally, I apply Main Path Analysis to understand how the link between dietary fat and CHD emerged and progressed through the literature. Main Path Analysis detects citation-paths that are most used in a body of literature – a method that shows how selective citation is involved in the development of the main research stream. I interpret the results by reading of the identified documents and demonstrate that highly cited documents tended to be those that preserved the hypothesis rather than those that challenged it.

The analyses performed throughout this thesis suggest that selective citation played an important role in the development of the diet–heart link. While I applied this approach to a specific case-study, its basic logic and methods ought to be applicable to other cases.
Acknowledgements

This thesis would not have been possible without the support of a great many people.

I would like to first thank to my supervisors, Steve Sturdy and Valeria Skafida, for helping me to secure funding, for their enthusiasm and faith in me, and for supporting me over the years. Their feedback on drafts and our many discussions shaped and improved my work considerably. I also thank Christine Knight, who supervised me during my MSc(R), who played an important role in securing me funding, for supporting me in my first year, and for her guidance. This thesis would not have been possible without their involvement, and their support and friendship has been invaluable.

I thank the Economic and Social Research Council (ESRC) for providing the funds to pursue my doctorate via the Advanced Quantitative Methods (AQM) studentship. I also thank the University of Edinburgh for providing me with the facilities and resources required to conduct my research.

I thank everyone in the STIS department, especially those who taught me over the years and listened to my ideas. In particular, I must thank Miguel Garcia-Sancho, James Lowe, Gil Viry, Niki Vermeulen, and Mark Wong for involving me in the TRANSGENE project. My involvement in this project helped me to develop the skills required to conduct the research that follows. I also wish to express my deep gratitude to Gil Viry for teaching me the fundamentals of network analysis and inspiring my passion for it. I also thank my examiners, Steven Yearley and Yves Gingras, for their careful reading of my work, constructive feedback, and for a very enjoyable viva.

Beyond STIS, I must thank Daniel Kenealy for his friendship and faith in me. Dan gave me the opportunity to tutor in the first year of my PhD, and we have worked together ever since. He has shaped my development more than he probably knows. Similarly, I thank Richard Brodie for his friendship over the years. Richard taught me how to teach, but also saw importance in my ideas and work. I thank Lindsay Paterson, with who I’ve also taught alongside, but also for his support and comments during my talks to the Q-step
group. Lindsay commented very positively on the first talk I ever gave on my research, and this gave me the confidence I needed.

Finally, I wish to convey my thanks to my family, though I cannot find the words to do this justice. I thank my brother and fellow network scientist, Trystan Leng, for his friendship, for our many long discussions, and for reading through drafts of my thesis. I thank my mother, Sharon Leng, for her unwavering support and for her comments on drafts. I thank my father, Gareth Leng, who saw importance in my research when few did, and for his support and comments. Finally, I thank my partner, Hayley Urquhart, for her love, kindness, and patience – I owe her the greatest debt.
Declar

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree or professional qualification. Some of the material used in this thesis has been published elsewhere:

- Limited material from Chapter 3 was adapted by myself into a small methods guide, ‘An introduction to citation network analysis’ (Leng 2019), provided to students attending the Economic and Social Research Council’s ‘Bibliometrics and Social Network Analysis’ workshop. I was co-organiser of this workshop and solely responsible for the teaching of citation network analysis and all associated teaching materials.
- The contents of Chapter 4, excluding the introduction and post-publication reflection, has been previously published in *PloS ONE* under the title ‘A network analysis of the propagation of evidence regarding the effectiveness of fat-controlled diets in the secondary prevention of coronary heart disease (CHD): Selective citation in reviews’ (Leng 2018).

Rhodri Ivor Leng
10/11/2019
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>I</td>
</tr>
<tr>
<td>LAY SUMMARY</td>
<td>III</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>VI</td>
</tr>
<tr>
<td>DECLARATION</td>
<td>VIII</td>
</tr>
<tr>
<td>CONTENTS</td>
<td>IX</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>XIII</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>XIV</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>XVII</td>
</tr>
<tr>
<td>SUPPLEMENTARY MATERIAL</td>
<td>XVIII</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1. OVERVIEW OF RESEARCH.</td>
<td>6</td>
</tr>
<tr>
<td>2. THESIS STRUCTURE</td>
<td>11</td>
</tr>
<tr>
<td><strong>CHAPTER 1: THE DIET–HEART DEBATE AND SCIENTIFIC CONTROVERSY</strong></td>
<td>15</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>15</td>
</tr>
<tr>
<td>1. DIET–HEART RESEARCH</td>
<td>16</td>
</tr>
<tr>
<td>Diet–heart and its early controversies</td>
<td>19</td>
</tr>
<tr>
<td>Observational studies</td>
<td>23</td>
</tr>
<tr>
<td>Intervention trials</td>
<td>26</td>
</tr>
<tr>
<td>Secondary prevention</td>
<td>26</td>
</tr>
<tr>
<td>Primary prevention</td>
<td>27</td>
</tr>
<tr>
<td>A conflicted evidence-base</td>
<td>29</td>
</tr>
<tr>
<td>State of the literature since 1984</td>
<td>33</td>
</tr>
<tr>
<td>Discussion</td>
<td>35</td>
</tr>
<tr>
<td>2. UNDERDETERMINATION, INTERPRETIVE FLEXIBILITY, AND EVIDENCE SELECTION</td>
<td>37</td>
</tr>
<tr>
<td>Kuhn and the sociology of scientific knowledge</td>
<td>40</td>
</tr>
<tr>
<td>3. DISCUSSION</td>
<td>45</td>
</tr>
<tr>
<td><strong>CHAPTER 2: SELECTIVE CITATION AND THE SHAPING OF SCIENTIFIC KNOWLEDGE</strong></td>
<td>47</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>47</td>
</tr>
<tr>
<td>1. SCIENCE AND THE CONVENTION OF CITATION</td>
<td>47</td>
</tr>
<tr>
<td>Theories of citation</td>
<td>49</td>
</tr>
<tr>
<td>2. DEREK DE Solla PRICE ON GROWTH, INEQUALITY AND CITATION NETWORKS</td>
<td>59</td>
</tr>
<tr>
<td>On growth of publication</td>
<td>59</td>
</tr>
<tr>
<td>Inequality in authorship</td>
<td>61</td>
</tr>
<tr>
<td>Inequality in citation</td>
<td>62</td>
</tr>
<tr>
<td>Research fronts</td>
<td>63</td>
</tr>
<tr>
<td>Cumulative advantage</td>
<td>66</td>
</tr>
<tr>
<td>Implications</td>
<td>68</td>
</tr>
<tr>
<td>3. WEBs OF BELIEF: CLAIM-SPECIFIC CITATION NETWORK ANALYSIS</td>
<td>70</td>
</tr>
<tr>
<td>Polarisation</td>
<td>73</td>
</tr>
<tr>
<td>Citation Distortions: Citation Bias, Amplification, and Invention</td>
<td>75</td>
</tr>
<tr>
<td>Citation bias</td>
<td>75</td>
</tr>
<tr>
<td>Research underutilisation</td>
<td>77</td>
</tr>
<tr>
<td>Citation diversion</td>
<td>78</td>
</tr>
<tr>
<td>Reference errors and dead-end citation</td>
<td>80</td>
</tr>
<tr>
<td>4. MAIN PATH ANALYSIS</td>
<td>82</td>
</tr>
<tr>
<td>5. DISCUSSION</td>
<td>87</td>
</tr>
<tr>
<td>Compatibility with SSK</td>
<td>90</td>
</tr>
</tbody>
</table>
List of Abbreviations

AHA – American Heart Association
CHD – Coronary Heart Disease
CI – Confidence Interval
CNA – Citation Network Analysis
CS-CNA – Claim-specific Citation Network Analysis
CVD – Cardiovascular Disease
HDL – High-Density Lipoproteins
HDL-C – High-Density Lipoprotein Cholesterol
IBM – Inclusion Body Myositis
LDL – Low-Density Lipoproteins
LDL-C – Low-Density Lipoprotein Cholesterol
MI – Myocardial Infarction
MPA – Main Path Analysis
MUFA – Monounsaturated Fatty Acids
PUFA – Polyunsaturated Fatty Acids
RCT – Randomised Controlled Trial
RR – Risk Ratio
SFA – Saturated Fatty Acids
SNA – Social Network Analysis
SSK – Sociology of Scientific Knowledge
STS – Science and Technology Studies
TC – Total Serum Cholesterol
UK – United Kingdom
US – United States of America
WoS – Web of Science
List of Figures

Fig 1: Number of publications on nitrogen fixation by plants published per year per contributing country, 1884–1931. Source: Wilson and Fred (1935 p.249) ..............60

Fig 2: Citation network representing the propagation of a phantom reference linked to claims regarding the clinical benefits of rutin .................................................................81

Fig 3: Citation Network Example 1 ..............................................................................93

Fig 4: Citation Network Example 2 ..............................................................................96

Fig 5: Citation Network Example 3 ..............................................................................99

Fig 6: Citation Network Example 4: Community detection ....................................101

Fig 7: Citation network mapping the citations from reviews to RCTs testing dietary fat modification in the secondary prevention of CHD ..................................................117

Fig 8: Sub-graph mapping the citations from neutral reviews to RCTs testing cholesterol lowering diets in the secondary prevention of CHD .............................................118

Fig 9: Sub-graph mapping the citations from unsupportive reviews to RCTs testing cholesterol lowering diets in the secondary prevention of CHD .............................................119

Fig 10: Sub-graph mapping the citations from supportive reviews to RCTs testing cholesterol lowering diets in the secondary prevention of CHD .............................................120

Fig 11: Citation Network representing citations between papers referencing Paul et al. 1963 .........................................................................................................................138

Fig 12: Citation network including diet and caffeine title classifications ..................139

Fig 13: Citation network displaying papers citing diet and caffeine findings ............140

Fig 14: Weakly-connected component showing interaction between citing papers classified by quotation analysis of referenced findings ..............................................141

Fig 15: Network displaying largest-component with vertices coloured by title classification .........................................................................................................................142

Fig 16: Network displaying largest-component with dietary carbohydrate and other dietary title classifications ...........................................................................................................144

Fig 17: Network displaying largest component with quotation classifications of references to serum cholesterol, diet, and caffeine findings ..................................................146

Fig 18: Number of indexed records from 1900 to 1984 in WoS and Scopus ..........171

Fig 19: Growth of indexed records related to atherosclerosis in WoS’ All Database and Core Collection. Top shows the distribution, 1900–1942; Bottom shows distribution, 1943–1984 ........................................................................................................176
Fig 20: Growth of indexed records related to atherosclerosis across WoS All Database search function and the Core Collection, 1943–1984

Fig 21: Growth of indexed records related to cholesterol metabolism and atherosclerosis in WoS Core Collection, 1900–1984. Top shows the period 1900–1942; Bottom shows distribution, 1943–1984.

Fig 22: Growth of indexed records related to CHD, atherosclerosis, and cholesterol metabolism in WoS Core Collection, 1900–1984.


Fig 24: Growth of all indexed records in WoS’ Core Collection, 1900–1945.

Fig 25: Growth of all indexed records in WoS’ Core Collection, 1943–1984.

Fig 26: Growth of indexed records in WoS Core Collection regarding atherosclerosis, cholesterol, CHD, tuberculosis, syphilis, cholera, and diabetes research, 1900–1984.

Fig 27: Growth of indexed records in WoS Core Collection regarding dietary research related to atherosclerosis, CHD, or cholesterol metabolism, 1900–1984.

Fig 28: Growth of indexed records in WoS Core Collection related to overlapping of atherosclerosis, CHD, cholesterol metabolism, 1900–1984.

Fig 29: Growth of indexed records in WoS Core Collection related to atherosclerosis, CHD, cholesterol metabolism – excluding dietary terms 1900–1984.

Fig 30: Retrieved records from WoS Core Collection on diet, atherosclerosis, CHD, and cholesterol metabolism, 1900–1984.

Fig 31: Cumulative distribution of citations to documents in the FULL-largest-component.

Fig 32: Power-law fitted to log-log transformed frequency distribution of in-degree (k) of the FULL-largest-component.

Fig 33: Meta-vertex analysis of references from four sets of retrieved documents.

Fig 34: Weighted meta-vertex analysis of reference from the D and H literature sets.

Fig 35: Weighted meta-vertex analysis of references from the D and H literature sets.

Fig 36: Meta-vertex of the D literature set and its relationship to non-retrieved documents as established by weighted-in-degree.

Fig 37: Number of documents published per year in CUT-N1 and CUT-N2, 1900–1984.
Fig 38: Total citations to documents published in specific years in the CUT-N1 and CUT-N2 networks ................................................................. 232

Fig 39: Citation network. Right panel is the CUT-N1 network. Left panel is the CUT-N2 network ........................................................................................................... 238

Fig 40: Citation network of organic light-emitting diodes (OLEDs) by Kajikawa and Takeda (2009) ........................................................................................................ 239

Fig 41: Distribution of Edge (Reference) Age in the FULL, CUT-N1, and CUT-N2 networks ........................................................................................................ 241

Fig 42: Mean (top panel) and median (bottom panel) reference list age per year in FULL, CUT-N1, and CUT-N2 networks ........................................................................................................ 242

Fig 43: Proportion of references ≤5 years and ≤10 years old in the reference lists of retrieved publications per year ........................................................................................................ 244

Fig 44: Percentage of total citations to documents published in the 1950s that are older than five years of age ................................................................. 246

Fig 45: CUT-N1 citation network visualising the impact of year of publication on network structure ........................................................................................................ 247

Fig 46: Citation network of diet–heart research from 1923 to 1984 including references ≤5 years of age only ........................................................................................................ 249

Fig 47: Citation network of diet–heart research from 1847 to 1984 excluding references ≤5 years of age ........................................................................................................ 250

Fig 48: The CUT-N1 network clustered by modularity maximisation via the Leiden algorithm ........................................................................................................ 255

Fig 49: Number of documents in, and number of citations to, each cluster in the CUT-N1 network ........................................................................................................ 265

Fig 50: Citation network of the CUT-N2 network with documents coloured by terms contained within their titles indicating a focus on animal (blue), methodological (yellow), or human (red) research ........................................................................................................ 268

Fig 51: Citation network path example ................................................................................................................................. 272

Fig 52: Citation network traversal weights example ................................................................................................................................. 273

Fig 53: Combined results of local forward (F), backward(B), and global (G) main-path analyses of the CUT-N1 network ........................................................................................................ 278

Fig 54: Local and Global Key-route MPA analyses with 100 initiating edges ................................................................................................................................. 280

Fig 55: Keys' (1953 p.134) Six Country Graph ................................................................................................................................. 292

Fig 56: Distribution of documents per year related to diet–heart research contained in the CUT-N1, 1900–1961 ........................................................................................................ 313
List of Tables

Table 1: Example edge-list .................................................................103
Table 2: Example vertex-attribute-list..................................................104
Table 3: Contingency table of the distribution of citations from three different classifications of review to four secondary prevention RCTs.................................120
Table 4: Citations to ten prospective cohort studies by 1984 and 2018 as recorded by the Web of Science.................................................................130
Table 5: Citations to highly-cited papers in the dietary and caffeine title groups and citations contributed by each group..................................................143
Table 6: Dataset descriptors ................................................................213
Table 7: Network descriptors ................................................................214
Table 8: Total number of documents cited and uncited, and the total citations to documents from D, H, A and R literature sets ..............................................219
Table 9: Number of linked documents and isolates in D,H,A and R literature sets .........................................................................................220
Table 10: Total number of citations from one literature set to another, the proportion of these citations in relation to the sum of all citations between retrieved documents, and the proportion of these citations to the sum of all references ................................220
Table 11: Meta-vertex analysis results......................................................224
Table 12: Dataset descriptors CUT-N2 ....................................................230
Table 13: Description of CUT-N1 following modularity maximisation clustering via the Leiden algorithm .................................................................254
Table 14: Proportion of total references from within each cluster directed at other clusters of the CUT-N1 network.........................................................266
Table 15: Cluster membership of documents detected via MPA ...............279
Table 16: Cluster membership of documents detected via Key-route MPA ....281
Supplementary Material

All datasets used in this thesis are provided in the Supplementary Material file, which can be found either on the CD-ROM at the back of this thesis or as an electronic file. I list below the content of each folder, and have included an additional 'Data Note' that briefly describes the content of each dataset in the Supplement.

Chapter 2 Supplement
- Ch2.S1_vertex.csv. Vertex attribute list containing details of papers (vertices) citing a phantom reference in support of the claim that rutin has clinical benefits, including associated quotation data.
- Ch2.S2_edge.csv. Edge list containing all citations (edges) between papers in this analysis.
- Ch2.S3_workspace.gephi. A Gephi workspace containing a citation network of the propagation of a phantom reference associated with the claim that rutin has health benefits.

Chapter 4 Supplement
- Ch4.S1_table.docx. Study characteristics of four RCTs examining dietary fat restriction/modification in the secondary prevention of CHD. A comparison of the Oslo Diet–Heart Study, Rose Corn Study, Research Committee Low-fat Study, and Medical Research Council’s MRC Soya-bean Oil Trial.
- Ch4.S5_table.docx: A quotation analysis of four RCTs and 62 citing reviews.
- Ch4.S6_vertex.csv. Vertex attribute list containing data on all studies (vertices) included in this study.
- Ch4.S7_edge.csv. Edge-list containing data on all citations between reviews and RCTs included in this study.
- Ch4.S8_workspace.gephi. A Gephi workspace containing citation networks displaying the selective citation by reviews of secondary prevention dietary RCT data.

Chapter 5 Supplement
• Ch5.S1_table.csv. A quotation analysis of 110 citing studies of Paul et al. (1963).
• Ch5.S2_vertex.csv. Vertex-attribute-list including all studies (vertices) included in this study.
• Ch5.S3_edge.csv. Edge-list containing all citations between documents discussed in this study
• Ch5.S4_workspace.gephi. Gephi workspace containing a citation network of papers citing Paul et al. (1963)

Chapter 6 Supplement
• Ch6.S1_workbook.csv. An Excel workbook containing number of publications identified per year by four search queries designed to capture research on atherosclerosis, cholesterol, CHD, and their relationship to diet.

Chapter 7 Supplement
• Ch7.S1_data.txt. A .txt file containing the original bibliometric data retrieved from the Web of Science.
• Ch7.S2_A-search_classification.csv. Title classification data associated with 339 documents cited 20 or more times by documents identified during systematic literature retrieval.
• Ch7.S3_FULL_vertex.csv. Vertex attribute list containing attribute data for all vertices (documents) of the FULL-dataset.
• Ch7.S4_FULL_edge.csv. Edge list recording all citations between vertices (documents) contained in the FULL-dataset.
• Ch7.S5_BASE_vertex.csv. Vertex attribute list containing attribute data for all vertices (documents) of the BASE-dataset.
• Ch7.S6_BASE_edge.csv. Edge list recording all citations between vertices (documents) contained in the BASE-dataset.
• Ch7.S7_FULL_workspace.gephi. Gephi workspace containing the FULL-network.
• Ch7.S8_BASE_workspace.gephi. Gephi workspace containing the BASE network.
• Ch7.S9_metavertex_vertex.csv. Meta-vertex attribute list containing attribute data for all non-retrieved documents and meta-vertices.
• Ch7.S10_metavertex_edge.csv. Meta-vertex edge list recording all citations to non-retrieved documents from each meta-vertex.
• Ch7.S11_metavertex_workspace.gephi. Gephi workspace containing the meta-vertex analysis performed in Chapter 7.
• Ch7.S12_CUTN1_vertex.csv. Vertex attribute list containing attribute data for all vertices (documents) of the CUT-N1 network.
• Ch7.S13_CUTN1_edge.csv. Edge list recording all citations between vertices (documents) contained in the CUT-N2 network.
• Ch7.S14_CUTN1_workspace.gephi. Gephi workspace containing the CUT-N1-network.
• Ch7.S15_CUTN2_vertex.csv. Vertex attribute list containing attribute data for all vertices (documents) of the CUT-N2 network.
- Ch7.S16_\_CUTN2\_edge.csv. Edge list recording all citations between vertices (documents) contained in the CUT-N2 network.
- Ch7.S17_\_CUTN2\_workspace.gephi. Gephi workspace containing the CUT-N2 network.

**Chapter 8**
- Ch8.S1\_time\_analysis.csv. Excel workbook containing distribution of reference ages in the FULL, CUT-N1, and CUT-N2 datasets.
- Ch8.S2\_time\_CUTN1\_workspace.gephi. Gephi workspace containing the CUT-N1 network with time attribute data.
- Ch8.S3\_CUTN1\_cluster\_membership.csv. Excel workbook containing cluster membership of all vertices of the CUT-N1 network, as defined by a modularity maximisation analysis via the Leiden algorithm.
- Ch8.S4\_CUTN1\_cluster\_workspace.gephi. Gephi workspace containing CUT-N1 network with cluster membership data.
- Ch8.S5\_CUTN1\_title\_classification.csv. Excel workbook containing title classification data for all vertices of the CUT-N1 network.

**Chapter 9**
- Ch9.S1\_CUTN1\_MPA\_vertex.csv. Vertex attribute list containing attribute data for all vertices (documents) detected via Main Path Analysis of the CUT-N1 network.
- Ch9.S2\_CUTN1\_MPA\_edge.csv. Edge list recording all citations between vertices (documents) identified via Main Path Analysis of the CUT-N1 network.
- Ch9.S3\_CUTN1\_MPA\_workspace.gephi. Gephi workspace containing network displaying the results of a Main Path Analysis of the CUT-N1 network.
- Ch9.S4\_CUTN1\_keyrouteMPA\_vertex.csv. Vertex attribute list containing attribute data for all vertices (documents) detected via Key-route Main Path Analysis of the CUT-N1 network.
- Ch9.S5\_CUTN1\_keyrouteMPA\_edge.csv. Edge list recording all citations between vertices (documents) identified via Key-route Main Path Analysis of the CUT-N1 network.
- Ch9.S6\_CUTN1\_keyrouteMPA\_workspace.gephi. Gephi workspace containing network displaying the results of Key-route Main Path Analysis of the CUT-N1 network.
Introduction

The besetting danger is not so much of embracing falsehood for truth, as of mistaking a part of the truth for the whole


Scientific knowledge is based, in part, on empirical evidence. Scientists contribute to an evidence-base by publishing the results of their experiments and observations, but they also use evidence produced by others in their own works. Scientists, however, are often selective in the evidence they use and can differ in their interpretations of that evidence. But how selective are scientists when evaluating specific scientific claims in the literature? And how does the use and interpretation of past studies shape what is eventually taken to be scientific knowledge?

What scientists choose to cite forms the basis of modern work evaluating the impact and quality of research (Moed 2005). By this logic, papers that have accumulated many citations have made important contributions to the development of a research area. If a paper is cited it is being used by other researchers, it therefore must have some value, and if a paper has accumulated many citations it must have more value than a paper that has received few. A highly cited paper is cause for celebration, not only for the authors but also for the institutions that employ them, the funders that backed the research, and the publishing journal. Science has become an industry in which performance is measured by the number of papers produced and the citations that they accumulate.

In this environment, it easy to lose sight of the role that publication and citation play in the creation and evaluation of knowledge. By publishing their work, scientists communicate their ideas and observations to a broader community of scientists, and, through their references, scientists acknowledge the works explicitly used in those papers. This is a system in which ideas are publicly developed, endorsed, or rejected. An unpublished finding or theory is
not generally considered to be part of a knowledge-base, only after publication can the process of external scrutiny and validation begin, and only then can a work become cited.

Ideally, only works conducted to a high standard ought to make their way into scientific journals, and only those that make particularly important contributions ought to receive many citations. Scientific peers, in possessing relevant expert knowledge, ought to be able to judge the quality of scientific work reliably and should expose flaws or inaccuracies in the papers they read. On this account, the selectivity of scientists forms a self-correcting and quality enhancing system, identifying excellence, error, and fraud. Over the last decade, however, this story has become increasingly less tenable in light of many problems that have been detected by researchers.

There is an apparent ‘replication crisis’. Few studies are published that attempt to directly replicate previous experiments (Makel et al. 2012), one of the key elements believed necessary for effective external scrutiny and authentication; and many major findings have proved to be unreliable (Ioannidis 2005a; Begley and Ellis 2012; Open Science Collaboration 2015). This is joined by publication bias, the practice of preferentially publishing the results of experiments that have arrived at ‘positive’ statistically significant findings, while failing to publish ‘negative’ non-significant findings (Turner et al. 2008; Song et al. 2010; Nissen et al. 2016). Accompanying this is a widespread use of inappropriate statistical tests and flawed study designs (Simmons et al. 2011; Gelman and Loken 2014), leading some commentators to assert that most published findings in certain research areas are false (Ioannidis 2005b).

Awareness of these problems began in the psychology and medical literatures, but they are being documented in other areas. In 2012, Fanelli (2012) examined hypothesis-testing papers. In a sample of 2,434 papers derived from 20 disciplines between 2000 and 2007 and comparing these to 2,222 papers published in the 1990s, Fanelli found that ‘negative findings’ appeared to be disappearing from the literature in all disciplines. Of papers published between 1990 and 1991, 70% returned findings supportive to the hypothesis being tested in those papers, but by 2005 this rose to 89%. While
the physical sciences were more likely to report negative findings than the biomedical and social sciences, the rise in proportion of supportive findings since the 1990s appears to have affected all disciplines. Science, it is said, is suffering from ‘positive-outcome’ bias.

The last line of defence between these problems and the soundness of communally endorsed scientific knowledge are the choices that scientists make in deciding how to weigh evidence, and their decisions over what and how to reference. Collectively, if scientists critique irreproducible or non-reproduced results, if they recognise problems with statistical tests and study designs, and if they perform statistical tests to check whether publication bias is likely present in a body of literature or follow-up grant proposals, then these issues may not have caused major problems.

The evidence in regards to this, however, makes for bleak reading. Over the last few decades, evidence has emerged that citation biases and distortions are present in the scientific literature. Recently, a meta-analysis of studies of citation bias found that, on average, scientists tend to cite studies returning statistically significant ‘positive’ results more often than those reporting equivocal or negative findings (Duyx et al. 2017). Not only are ‘positive findings’ more likely to be published, they are also more likely to be cited. Fanelli (2013), in a study examining 2,545 hypothesis-testing papers published between 2000 and 2007, found that papers reporting a positive outcome had an average of 32% more citations than those that did not; however, this varied between disciplines.

While publications reporting findings relevant to these issues can be traced back to the 1970s and 80s (Rosenthal 1979, Sackett 1979, Gøtzsche 1987), recent scholarship has suggested that these problems are worsening. Baker (2016) reported the results of an anonymous survey sent to researchers in different disciplines to assess their views on whether there is a reproducibility crisis; 52% of respondents believed there to be a “significant crisis”, 38% a “slight crisis”, while only 3% believed there was “no crisis”. Other commentators, however, have raised a note of scepticism as to whether these
problems really are on the rise or whether increasing levels of research interest in them is responsible for this impression (Fanelli 2018).

As evidence accumulates about these apparent problems, certain widely treasured notions of how science functions and progresses are becoming less tenable. Most notably, the Popperian view – that science progresses through the posing of bold hypotheses and the determined attempts to refute them (Popper [1934]2000) – sits in stark contrast with findings that scientists tend to favour both publishing and citing studies that have reached ‘positive’ findings. While other views, particularly the Kuhnian view, appear better suited in terms of their descriptive accuracy. For Kuhn (1970), scientific work ‘progressed’ only when a particular scientific theory or ‘paradigm’ was adopted, and researchers devoted their efforts to developing and refining that theory, ironing out its flaws and expanding it to new cases. In this period of ‘normal science’, scientists actively ignored ‘anomalies’; instances that contradict or cannot be explained in reference to the existing theoretical perspective, “In a sense, to turn Sir Karl's view on its head, it is precisely the abandonment of critical discourse that marks the transition to a [normal] science” (p.6). Indeed, both Kuhn and later Feyerabend (1975), argued, in different ways, that however inefficient, irrational or absurd scientists might seem to be, some of those inefficiencies, irrationalities, and absurdities might nevertheless contribute to the progress of science. Today, whatever guides scientific progress remains a topic of dispute, but nevertheless scientists are reflecting on how science ought to function and how it deviates from ideals.

In response to these apparent problems, an entire field has arisen, commonly referred to as Meta-science or Science of Science, which has drawn a broad range of scholars from the natural and social sciences together to develop a novel understanding of how science operates (Ioannidis et al. 2015; Smaldino and McElreath 2016; Fortunato et al. 2018). This new wave of scholarship joins older disciplines in attempting to understand how science works and what causes it problems. Scientometrics, best known for developing quantitative measures of research performance, has, since the early 1960s, developed understanding of the dynamics of knowledge production at large
scales (Price 1965, 1976, 1986), developed high-quality databases that have indexed scientific publications and their references (Garfield 1955, 1979), and proposed a battery of quantitative methods for their analysis (Kessler 1963; Garfield et al. 1964; Small 1973; Hummon and Doreian 1989). The Sociology of Science and the Sociology of Scientific Knowledge (SSK) has produced a large body of primarily qualitative research on how scientists construct and interpret evidence, and a detailed understanding of the social mechanisms by which scientific knowledge is produced and the factors that influence its development (Merton 1968; Bloor [1976]1991; Gilbert 1977; Latour and Woolgar 1979; Barnes 1982; Collins 1981; Barnes et al. 1996). Moreover, philosophers have contributed important works regarding how science might progress, how science ought to function, and the limits of human reason (Duhem [1914]1991; Popper [1934]2000; Quine 1953; Kuhn [1962]1970; Feyerabend 1975).

These older disciplines have largely failed to systematically detect or propose solutions to the problems believed to be confronting science; at least, what work has been done on them has failed to alert the broader scientific community to their existence. This is not to say that these traditions have failed to produce work relevant to these apparent problems, but that there has been a failure to bring this research forward to address the concerns of scientists. Why this has occurred is complex. Scientometrics has primarily focussed on developing quantitative metrics for research evaluation (Mingers and Leydesdorff 2015). SSK, in its desire to separate itself from older forms of the Sociology of Science, adopted a largely non-evaluative perspective regarding the quality of research, focussing on describing the social processes involved in the construction of knowledge claims through ethnographic research and historical case studies (Latour and Woolgar 1979; Bloor [1976]1991).

With the arrival of Meta-science, with its explicit focus on detecting apparent problems and improving research (Ioannidis et al. 2015), this is an important time for interdisciplinary discussion not only about these questions, but also about how recent findings may alter understanding of how science functions. However, despite seemingly shared research interests, there is little
communication between these specialities; they use different methods, publish in different journals, and organise competing conferences. The literature, though rich, is fragmented and important findings, theories, and methods are failing to spread between disciplines.

It is against this backdrop that the following thesis was conceived. Its focus is on understanding selective citation practices and their consequences. It was influenced by, and hopes to contribute to, the empirical study of science, irrespective of current disciplinary boundaries.

1. Overview of Research

This thesis develops understanding of how scientists use and interpret evidence in the scientific literature and how this shapes scientific knowledge. To do this, I develop an approach that applies and modifies variants of Citation Network Analysis (CNA), an application of network analysis to bibliometric data, to examine different expressions of selective citation and their consequences in a particular research area.

Specifically, I examine the rise and eventual acceptance of the conjecture that the consumption of saturated fatty acids (SFA), by increasing serum cholesterol, contributes to coronary heart disease (CHD) via the development of atherosclerosis, while unsaturated fatty acids have an opposite or neutral effect (Keys et al. 1957c). Commonly referred to as the diet–heart hypothesis, this theory has dominated scientific discussion regarding the aetiology of atherosclerosis and CHD since the 1950s. Since the late 1970s and early 1980s, this theory has informed population-wide dietary advice to reduce SFA intake by avoiding foods such as full-fat dairy products and fatty meats (US Senate Select Committee on Nutrition and Human Needs 1977; UK Department of Health 1984). Following the National Institutes of Health's Consensus Conference in December 1984, raised serum cholesterol levels have been regarded as a cause of CHD, and it has been concluded that dietary treatment is an appropriate strategy to lower serum cholesterol in the United
States and beyond (NIH 1985). However, in the decades leading up to this consensus and the issuing of advice, the diet–heart hypothesis was controversial. Critics argued that the underpinning theory was erroneous, that the prevailing methodological approach to testing it was flawed, and that an ambiguous evidence-base was punctuated by unsupportive evidence that undermined dietary advice. Yet, despite a history strewn with controversy, the diet–heart idea exhibited a remarkable resilience.

The hypothesis was first clearly articulated in the 1950s and had apparently been accepted by the end of 1984. This provides an opportunity to study the trajectory of an idea through the scientific literature, from its first public articulation, through its controversies, to its eventual acceptance. This formed the rationale for selecting this particular case – here is a hypothesis with a short history, that made a large impact on research and policy, and that survived several major periods of controversy. Initially, I was guided by the following broad research questions:

(i) When did the proposition that dietary fats play a causal role in atherosclerosis and CHD become widely accepted? What evidence was this based on?

(ii) Did controversy polarise the scientific community into different camps that drew from different sources of evidence? Or was controversy the result of different theoretical interpretations of the same empirical evidence?

(iii) What role did different evidence-selection behaviours play in the dynamics of consensus formation and fragmentation?

I aspired to understand how this conjecture became a widely accepted ‘fact’ in the scientific community, and what influence the use of empirical evidence had on this process. The project started in 2015 during an MSc(R) at Edinburgh University. Before 2016, I focussed on writing historical case studies of important episodes of this history, and while I have not included these in the final thesis, they were vital in developing my knowledge of this area.
In 2016, I came across a study on citation bias. Greenberg (2009) had developed a methodology that he termed *claim-specific citation network analysis* (CS-CNA) to trace how a particular claim came to be accepted as a fact in the scientific literature despite an underlying ambiguous evidence-base. By recording the findings of all primary studies that contributed data towards this claim and recording how these were cited in later literature, Greenberg demonstrated that the belief in this proposition was based on supportive studies *alone* – unsupportive studies had failed to be cited by the vast majority of the citing literature. By constructing a citation network, Greenberg showed how the scientific community evaluating this particular claim had become apparently oblivious to the existence of unsupportive data.

This study had an immediate influence on me. I found a series of papers in which Ravnskov (1992, 1995, 1998, 2003), a critic of the diet–heart hypothesis, had claimed that the “diet-heart idea is kept alive by selective citation” (2003 p.1348). While Ravnskov, using a novel form of quotation analysis and citation analysis, had produced findings that looked plausible, they lacked the rigour and precision of Greenberg’s methodology. Furthermore, Ravnskov, a critic of the link between serum cholesterol and CHD, had collapsed the results of his analysis of dietary trials with cholesterol lowering drugs (1992). As a scientist clearly devoted to attempting to refute this hypothesis, his impartiality was questionable. Nevertheless, this was a claim that I could explore. Was the diet–heart idea influenced by selective citation, and, if so, in what manner?

In 2016, I won funding for a PhD from the Economic and Social Sciences Research Council’s Advanced Quantitative Methods studentship. My proposal set out to develop CNA further to understand selective citation practices and their consequences in the diet–heart case study. The aim was to use network analysis to trace how the diet–heart hypothesis emerged and developed through the literature by understanding how papers referenced other papers within that literature, and to use network analysis to assess whether citation bias or other distortions were present.
The approach that follows combines document analysis of the content of papers and their references with network analysis. This fusion of method is designed to produce an understanding of the interpretation of data by particular scientists, and how data and interpretation spread through the scientific literature – opening up the dynamics of the transition from experimentation and observation to knowledge held by a wider group of scientists.

Three variants of CNA are used. First, CS-CNA is performed to understand the prevalence of citation bias in segments of the literature and detect different interpretations of the meaning of particular findings. The rationale for applying CS-CNA is to understand whether the direction of findings reported in a study (e.g., supportive, unsupportive) affects how it is cited, and whether different communities of scientists use the available evidence in different manners.

Second, after systematically retrieving literature with terms relevant to diet–heart research in their titles, I construct a citation network representing the structure of the literature, as assembled through citations, prior to 1985. This allowed me to test whether these data cohere with:

(i) Kuhn’s ([1962]1970 p.20) claim that the birth of scientific ‘paradigm’ will be signified by a change in publication practices. Specifically, that documents focussing on a particular subject will begin to appear in scientific journals. To do this, I measure growth of publications per year.

(ii) Kuhn’s ([1962]1970 p.177) claim that the members of a paradigm will tend to communicate more frequently with other members of that paradigm. To test this, I perform a component analysis to measure the proportion of collected documents that form a connected network. I then examine whether the collected documents tend to cite other documents captured by my systematic literature search more often than those that were not. To do this, I construct a network containing all documents retrieved and all of their references, and measure the proportion of references directed at retrieved versus non-retrieved documents.
(iii) Price’s (1965) conjecture that a power-law describes the distribution of citations to documents in citation networks representing particular fields.

(iv) Price’s (1965) claim that a scientific field progresses through a dynamic research front that is composed of documents primarily referencing other documents published in the previous 5-years. To do this, I measure the distribution of the age of references.

(v) Kuhn’s ([1962]1970 p.10) claim that a community of scientists should be structured around a core-set of findings and methods if they share a paradigm. To do this, alongside understanding the skewness of citation distributions, I examine how papers cluster together. Modularity analysis for community detection (Newman 2006; Traag et al. 2019) is used to understand if sub-communities exist within this network, and whether these collect around a shared set of methods and findings.

(vi) Whether the controversy polarised the network into clusters of documents that relied on different segments of the literature. To do this, I use modularity analysis to understand if a supportive and critical community can be detected via this method.

Third, I apply Main Path Analysis (MPA) to this network to understand how the link between dietary fat and CHD emerged and progressed through the literature. Recently, it has been claimed that MPA detects citation-paths that are most used in a body of literature, and that this detects documents that played an important role in the development of a research area (Liu and Lu 2012, Batagelj et al. 2017). I examine whether this is the case, but also whether this method can be used to understand how selective citation is involved in the development of the main research stream over time. I interpret the results of this analysis by the reading of the identified documents and writing a narrative review.
2. Thesis Structure

In Chapter 1, I give an overview of the diet–heart hypotheses, its evidence-base, and its controversies. I argue that the history of this debate can be understood as a long period of underdetermination; scientists have come to conflicting positions as to what the evidence-base supports in the absence of a definitive study. Following this, I discuss perspectives from the philosophy and sociology of science that help explain scientific disagreement.

In Chapter 2, I discuss the convention of citation, CNA, and key empirical and theoretical perspectives. I argue that network analysis holds promise for understanding how publication and citation dynamics contribute to the formation and fragmentation of consensus in the scientific community. I discuss three variants of CNA – CS-CNA, CNA mapping and community detection, and MPA – and draw out the findings of recent applications of these methods for theories in the philosophy and sociology of science.

In Chapter 3, I provide a general introduction to CNA. As each chapter adapts CNA to fit particular questions, I outline network analysis; formally defining basic measures and terms, the software used, and the structure of the assembled datasets. In regards to materials and methods, each chapter that applies a particular analysis contains a full explanation of the specific methods and datasets used.

In Chapter 4, I perform CS-CNA to understand how four randomised controlled trials (RCTs) that evaluated the efficacy of fat-controlled diets in the secondary prevention of CHD were cited by reviews of the literature between 1969 and 1984. I introduce sub-graph analysis to the claim-specific approach to precisely detect citation bias and a new metric of research utilisation. Of the four RCTs, one supported dietary modification in the secondary prevention of CHD, while three were unsupportive. Of the 62 reviews, 28 suggested that data from existing trials supported dietary treatment, 17 were neutral in regards to whether the treatment was supported or not by the findings of trials, while 17 stated that the findings of trials did not support the use of dietary treatment. Supportive reviews underutilised the available RCTs to a greater degree than other reviews. Amongst the supportive group, citation bias was common – 23
(82%) reviews cited only the one RCT that was supportive. Most reviews that disseminated a supportive evaluation of the results of RCTs in the context of secondary prevention cited only data that supported this position.

In Chapter 5, I apply CS-CNA to understand how the results of the first prospective cohort study that reported findings on the relationship between dietary fat and CHD (Paul et al. 1963) were cited between 1963 and 1970. This study reported a lack of association between dietary fat, and particular classes of dietary fat, and the development of CHD, but also reported findings on many other factors. I demonstrate that these findings were cited in many different ways that appeared to reflect the research interests of particular communities. I identified all citing papers and classified them into research topics by terms within their titles; this classification helped to explain both community structure as judged by citation relationships, but also was predictive of what findings were cited within these citing papers. The three largest groups of papers focussed on CHD, diet, and caffeine. In a sample of 110 papers, I demonstrate that ~74% of studies cited Paul et al. for just a single finding. The most frequently cited findings were the significant associations between CHD and serum cholesterol, blood pressure, coffee consumption and smoking. The lack of association between all dietary variables and CHD was discussed in just 13 papers, but the dietary fat findings were only specifically mentioned in eight of these. Of concern, this lack of association was not reported in 22 of the 30 papers that cited the association between serum cholesterol and CHD. As serum cholesterol was significantly associated with CHD but dietary fat had no association with either serum cholesterol level or CHD, this finding required close scrutiny. The dietary fat findings appeared to make the biggest impact not on those interested in the diet–heart hypothesis, but in a community advocating a rival hypothesis linking sugar consumption to CHD.

In Chapter 6, based on the findings from the two previous chapters, I capture a large body of literature related to the diet–heart debate. I develop a systematic literature retrieval strategy designed to capture diet–heart relevant documents published before 1985 from the Web of Science’s (WoS) Core Collection. This involved developing Boolean search queries to capture
literature containing specific title terms suggestive of a focus on cholesterol metabolism, atherosclerosis, or CHD, and then layering relevant dietary terms on top of these queries to retrieve relevant literature. By this, I retrieve a population of documents containing titles indicative of a focus on the diet–heart link. Prior to performing a network analysis, I measure the number of publication produced per year to understand periods of varying productivity, and show that periods of growth cohere with historical understanding of this research area. I interpret these results as consistent with Kuhn’s claim that the birth of a paradigm will be signified by an increase in journal publication output.

In Chapter 7, I detail how I converted and cleaned the retrieved publication data into datasets and citation networks. I further retrieved documents cited 20 times or more by the retrieved documents and included these in the dataset. I validated the retrieved literature in two ways. First, I used a computational method to compare the documents in my dataset with references from 19 reviews of the literature, all published at different time periods and written by different authors, and show that my search strategy captured 70% of the relevant literature discussed in these reviews. All relevant documents missed by my search strategy were retrieved and included in the network dataset. Second, I used network analysis to understand the structure of citation between the documents collected. This allowed me to demonstrate: (i) that the majority of documents formed a weakly-connected component and no other large weakly-connected component existed in these data; (ii) that citations to these documents follow a power-law distribution; (iii) and that the retrieved documents were more likely to reference another retrieved document than a non-retrieved document. From this, I refine the datasets down into two maps of the literature – one composed of the largest connected component formed from the retrieved literature, and another composed of the largest component formed from all the retrieved literature and any of their references cited by at least three of the retrieved documents. I interpret these results as confirming Price’s power-law conjecture, and cohering with Kuhn’s claim that citations will be primarily directed at other documents relevant to that paradigm.
In Chapter 8, I quantitatively analyse the constructed networks to understand the temporal dimension of citation. By ‘ageing’ references, I measure the ‘half-life’ of citation influence and examine how this relates to the development of this research area. On average, documents hit their peak citation rate at around 2 years following publication and collect half of all references within 5 years. However, I show that time’s impact is not uniform – highly cited literature not only receives more attention, but tends to be more resistant to the decaying effects of time. I interpret these findings as cohering with Price’s notion of a dynamic research front in specific fields. Further, I analyse community structure in these networks via modularity analysis and demonstrate that the network structure is also heavily influenced by research focus, defined as an adherence to general research questions. Here, however, I fail to detect a clear cluster of critical papers.

In Chapter 9, I perform a series of Main Path Analyses of my network dataset to understand the evolution of knowledge over these periods through the documents most used by researchers throughout the history of the diet–heart debate. I generate a narrative account of the development of this research area using only the documents identified. I reflect on its strengths and weaknesses, highlighting that this form of analysis largely excluded critical opinions and unsupportive data.

In Chapter 10, I draw together the findings to highlight the contributions to knowledge in two domains: what it adds to SSK/Scientometrics, and their methods and theoretical perspectives; and what it tells us about the development of the diet–heart hypothesis.
Chapter 1: The Diet–Heart Debate and Scientific Controversy

Introduction

This chapter introduces the diet–heart hypothesis, its evidence-base, and an overview of some of the important debates. The purpose is not to provide a comprehensive appraisal of the evidence-base, but to introduce the basic framework under which the diet–heart hypothesis and its controversies were evaluated. I then introduce theories from the philosophy and sociology of science that can help us to understand scientific disagreement.

In Section 1, I begin by introducing the hypothesis, the evidence that influenced its development, and the lines of criticism that were present during its initial construction. Second, I discuss the incorporation of this theory into the American Heart Association's (AHA) dietary guidelines in 1961, but highlight that neither randomised controlled trials (RCTs) nor prospective cohort studies had returned results on the relationship between dietary fat and the development of coronary heart disease (CHD) by this point. Third, I discuss the outcome of these forms of experiment and outline major debates that ensued. Fourth, although my thesis ends its analysis in 1984, I discuss modern scientific opinion regarding the relationship between dietary fat and CHD by examining the results of recent meta-analyses. As these include studies from before 1985, their results can help add perspective on the challenges in interpreting this evidence-base. I argue that the history of this debate can be understood as a long period of underdetermination; scientists have come to conflicting positions as to what the evidence-base supports.

In Section 2, I introduce perspectives in the philosophy and sociology of science on scientific disagreement. I distinguish between two classes of scientific disagreement – (i) disagreement based on different interpretations of the same empirical evidence by different groups; (ii) disagreement that stems from the use of different empirical evidence by different groups. These
categories are not mutually exclusive, but help to focus thinking on what is in dispute.

1. Diet–heart research

The ‘diet–heart hypothesis’ states that consumption of saturated fatty acids (SFA), by increasing serum cholesterol levels, contributes to CHD via the development of atherosclerosis, while polyunsaturated fatty acids (PUFA) modestly decrease serum cholesterol and lower the risk of CHD by reducing atherosclerosis (Keys et al. 1957c). The hypothesis was first proposed to explain why the incidence of CHD, particularly myocardial infarction (MI) [heart attack], had risen sharply in the Western, developed world.

Before the diet–heart hypothesis had been articulated, atherosclerosis, a condition where atheromatous plaques [fatty deposits] accumulate in the tunica intima [inner layer] of arteries, had been proposed as the major cause of CHD (Karsner 1933, Katz and Stamler 1953). CHD occurs when atheromatous plaque build-up inside the walls of the coronary arteries, which, over time, narrows the arterial lumen of affected arteries, restricting blood flow to the myocardium [heart muscle]. The restriction of oxygenated blood to the myocardium can provoke angina – pain in the chest, neck, or arms. In the event of total occlusion [obstruction] or plaque rupture leading to thrombosis, MI and sudden cardiac death can follow.

In the early 20th century, human atheromatous plaques were found to contain abnormally high concentration of cholesterol (Windaus 1910). Cholesterol, a type of lipid molecule, is an essential constituent of the membranes of all animal cells. Accordingly, researchers began investigating the possibility that the cholesterol that accumulated in the arterial walls was the result of excessive ingestion of animal foods. In 1913, Anitschkow and Chalatow (1913[1983]) showed that atherosclerosis-like lesions could be produced in some mammals, particularly rabbits, by altering the cholesterol content of the diet. Following previous reports that dietary cholesterol could enter the blood via the gut lumen of rabbits, Anitschkow and Chalatow
conjectured that consumption of cholesterol, by increasing serum cholesterol, led to the infiltration of cholesterol through the walls of arteries and the development of atherosclerosis. In 1933, after refining and reproducing his cholesterol-fed animal model, Anitschkow (1933) proposed the 'lipid-infiltration theory of atherosclerosis’, that a disturbance of cholesterol metabolism is a necessary, but not always sufficient, cause of human atherosclerosis. However, he noted that others factors, most notably raised blood pressure, were also probably implicated, particularly in damaging the layer of endothelial cells that form a protective arterial coating.

Anitschkow's ideas came up against stiff opposition. According to Katz and Stamler (1953), scientists and physicians were generally of the opinion that atherosclerosis was a consequence of senescence (ageing), because autopsy studies had consistently shown that arteries of most elderly people contained marked atherosclerosis. However, studies on soldiers who had died during the two World Wars soon demonstrated that atherosclerosis was also present in the coronary arteries of the young (Karsner 1933; Katz and Stamler 1953). French and Dock (1944), examining the autopsy records of 80 soldiers between the ages of 20 and 36 years who had died of suspected coronary incidents, found evidence of atherosclerosis in all cases and evidence of a recent thrombosis in 36% of cases. Both atherosclerosis and CHD, while more prevalent amongst the elderly, also affected the seemingly young and healthy. However, acceptance of this came uneasily in the medical community.

Before Herrick's (1912) studies of patients who had survived a MI, the medical community were of the opinion that MI was almost always fatal. Following Herrick's studies, MI became both diagnosable and as something recognised as requiring treatment. The most notable advance in diagnosis came from Herrick's (1919) observation that electrocardiographic (ECG) results of those with MI differed from those of healthy individuals and that this reflected damage to the myocardium. Advances in diagnosis, particularly the ability to diagnose acute MI via ECG following Pardee (1920), led to increased awareness of its incidence and prevalence. This was accompanied by Bloor's (1916) development of tests capable of quickly estimating the level of
cholesterol in blood serum (total cholesterol [TC]). As serum cholesterol became more easily measured, the number of studies investigating the relationship between CHD and serum cholesterol slowly increased. These attempted to ascertain whether there were any important differences between the average serum cholesterol of healthy people and those with CHD (Gorham and Meyer 1917, Davis et al. 1937, Steiner and Domanksi 1943), but these produced conflicting findings.

As awareness of CHD grew, it was joined, particularly in the United States (US), by an apparent increasing incidence. Heart diseases had been the biggest contributor to mortality in the US since 1910 (excluding the influenza epidemic years of 1917–1919), and had been rising steeply since around the 1920s (CDC 2015). In 1948, in response to the rising burden of cardiovascular disease (CVD), the US Congress passed the National Heart Act (1948). In a press release by the US Public Health service, the Act was announced alongside worrying figures; heart diseases alone were said to be responsible for more than a third of deaths, half of all deaths in those over the age of 45 years, and rates appeared to be increasing (US Public Health Service 1948). To launch “a full-scale attack on the Nation’s Number 1 destroyer of life” (p. 1059), the Act created the National Heart Institute (NHI), devoted to research, screening, diagnosis, and treatment of CVD. The purpose of this institution was to bring together the “limited and fragmentary” research into heart disease that had already been conducted, and to massively increase the funding, laboratory and clinical facilities, and the number of researchers specialising in heart diseases in the US (p.1060).

CHD, however, was not responsible for all heart disease deaths, which included deaths from rheumatic and congenital heart diseases. In 1946, the US National Office of Vital Statistics (1950) estimated that 11% of all deaths were caused by atherosclerosis of the coronary arteries, but by 1950 this figure had risen to 22% (Katz and Stamler 1953). This, particularly prominent in middle-aged men, caught the attention of Dock (1946), who called for the scientific community to make headway in understanding the causes of this disease and to find preventative strategies. In 1947, the US Public Health
Service funded two large prospective cohort studies, the Framingham Heart Study (Kannel et al. 1961) and the Minnesota Business and Professional Men Study (Keys et al. 1963), to investigate whether various lifestyle and environmental factors were associated with the development of CHD, including measuring serum cholesterol and blood pressure. Neither of these studies included dietary factors in their initial design, but both were capable of testing whether raised serum cholesterol was associated with CHD. However, these were not set to return findings until the early 1960s.

Diet–heart and its early controversies

Many scientists moved into CHD research. In 1949, Gofman et al. (1949), after developing a method to identify different classes of lipoprotein in the blood, demonstrated that cholesterol was predominately carried by low-density lipoproteins (LDL). In 1950, Gofman reported that, after feeding cholesterol to rabbits, levels of LDL rose in the blood, and this was associated with the development of atherosclerosis. After taking blood samples from people suffering a recent MI, he showed that these also had raised LDL levels.

For Keys (1951), however, Gofman’s typology unnecessarily complicated research into diet and CHD, and the technique required detection of lipoproteins, which was expensive and time-consuming. Generally, despite promising early results, the scientific community remained focussed on serum cholesterol, rather than concentrations of specific lipoproteins. In 1952, Abell et al. (1952) developed a cheap and standardised method for measuring the total concentration of serum cholesterol for the Framingham Study, which required such a method to sample large numbers of people, and this became particularly popular amongst researchers.

From 1950 onwards, many metabolic-ward trials – controlled feeding trials that alter the diet to understand how serum cholesterol related to dietary intake in humans – were conducted. In 1952, Groen et al. (1952) reported results suggesting that animal fats increased serum cholesterol, while vegetable fats appeared to have the opposite effect. Groen had been a student of Snapper (1941), who in Chinese Lessons for Western Medicine had linked the low consumption of animal fats and high consumption of a vegetable oil
high in linoleic acid, a PUFA, to the low rates of atherosclerosis in China (p.29-31). Snapper had been a student of De Langen who, following an expedition to Java, had found that rates of heart disease appeared to be particularly low in the Javanese, who also seemed to have low levels of serum cholesterol. De Langen attributed this to the Javanese diet, which was particularly low in dietary cholesterol and animal products (Blackburn 2012). However, De Langen’s results, published in Dutch in 1916 and 1922, failed to become widely known in the community. This was despite Snapper’s (1963) attempt to get the scientific community to recognise De Langen as the originator of the hypothesis linking diet to serum cholesterol levels and CHD.

However, in 1953, Keys (1953) reported a positive correlation between the consumption of total dietary fats and mortality from CHD in six countries by drawing data from national food consumption and vital statistics from 1948 and 1949. In Japan, where the average amount of fat eaten was below 10% of calories, less than 1 in every 1,000 men died from CHD. In England and Wales, with an average national consumption of dietary fat of 34% of calories, about four in 1,000 men died from CHD. In the US, a population eating on average about 40% of their calories from fat, about seven in 1,000 men died from CHD – a death rate about 700% higher than in Japan. Following this, Keys and colleagues found in several small epidemiological studies that fat consumption appeared to be associated with the serum cholesterol level in particular populations. Those that ate less fat had, on average, lower levels of serum cholesterol and lower rates of CHD. As early as 1953, Keys (1953) believed the evidence strong enough to warrant “a large extension of this type of epidemiological research”, and that public health programmes: “must take cognizance of the information already at hand” (p.137). For Keys, the appropriate response to rising population levels of CHD in the US was a population-wide shift in dietary habits to lower total fat consumption.

However, during this period, many metabolic-ward studies were returning findings that suggested that different types of dietary fat had different effects. By 1957, following independent findings from different laboratories (Bronte-Stewart et al. 1956, Ahrens et al. 1957, Keys et al. 1957), the scientific
community were generally of the opinion that SFA raise serum cholesterol, PUFA lowered it, while mono-unsaturated fatty acids (MUFA) appeared neutral. Fatty acids are compounds composed of a hydrocarbon chain, with carbon atoms connected to hydrogen atoms. SFA have no double bonds between the carbon atoms—they are “saturated” with hydrogen atoms. MUFAs have a single double bond between carbon atoms, and PUFAs have more than one double bond between carbon atoms—these are “unsaturated”. Generally, animal fats contain a high proportion of SFA, though certain non-animal products also contain high levels of SFA, such as coconuts and palm kernel oil. PUFA are found in high proportions in certain vegetable oils, particularly corn, soybean, and sunflower oil, but are also found in high concentrations in oily fish. MUFA are found in high concentrations in olive oil and other vegetable oil and products, but are also found in red meat and beef tallow. These terms are umbrella categories for individual fatty acids that are further classified by their carbon chain length.

That year, Keys et al. (1957c) proposed that lowering the amount of SFA in the population’s diet and moderately increasing that of PUFA ought to lower serum cholesterol and, in turn, lower the rate of CHD. They proposed a ‘predictive equation’ that described the effect of dietary fats on serum cholesterol in their metabolic-ward studies:

\[ \Delta TC = 2.74 \Delta SFA - 1.31 \Delta PUFA \]

where the change (\( \Delta \)) in TC is expressed in mg/100 ml, while changes in SFA and PUFA are expressed as changes in the % of total calories.

This suggested that serum cholesterol could be lowered more effectively by reducing the amount of SFA in the diet compared to increasing that of PUFA. Alongside evidence that the typical US citizen had one of the highest CHD risks globally, had a higher average TC level than most other populations, and who ate a diet of about 40% of calories from fat and >20% from SFA (Katz et al. 1958), these findings suggested that CHD in the US may be reduced by altering the diet.

In 1961, the AHA issued advice to those with, or at high risk of, CHD to alter their dietary fat consumption by reducing SFA intake, modestly increasing PUFA, and reducing their total fat consumption to under 35% of calories (Page
et al. 1961). This was despite a controversy in the 1950s over the validity and reliability of drawing causal inferences from, and basing dietary advice on, international statistics on food consumption and mortality. According to critics, these data were inaccurate and the relationship between dietary fats and CHD was confounded by almost all aspects of life in Western, developed nations (Yerushalmy and Hilleboe 1957, Mann 1957, Yudkin 1957). Critics proposed two competing hypotheses – physical inactivity and CHD (Mann 1957), and carbohydrate consumption, particularly sucrose, and CHD (Yudkin 1957).

A second controversy concerned whether the findings of metabolic-ward trials could be generalised beyond the walls of the laboratory. Ahrens (1957), one of the leading scientists in such trials, pointed out that no one currently understood what altering fat content of the diet in non-controlled settings would actually do. This was because all trials, typically conducted on no more than 20 or 30 men, placed individuals on specific controlled diets in which a single nutrient was changed. Nor was it clear whether the chain-length of particular fatty-acids might not have different effects on serum cholesterol, despite their level of saturation. Accordingly, Ahrens believed that scientists could not rule out that the number of calories and other nutrients consumed, along with a host of other factors such as exercise and stress, which had been controlled in these studies, might also impact serum cholesterol levels. Furthermore, to his eyes, it seemed that increasing the amount of unsaturated fats in the diet, particularly PUFA, appeared to be the most effective strategy. In one experiment, the largest fall in serum cholesterol came from a diet composed of over 70% of calories from unsaturated fats. Ahrens et al. (1957) and Sinclair (1956) proposed rival dietary fat hypotheses that placed greater emphasis on the protective role of unsaturated fats and essential fatty acids respectively, rather than on the harms of SFA.

Finally, Gofman challenged the prevailing method of measuring total cholesterol in serum. He championed instead his method of using the ultracentrifuge to measure the concentrations of particular lipoproteins in the blood. Gofman et al. (1949) had created a system of classification for lipoproteins that split these into particular bands based on their svedberg
flotation rate (Sf). Gofman et al. (1956) had previously shown that low-density lipoproteins of bands Sf 0–12, Sf 12-20, Sf 20-100, and Sf 100-400 were all positively and independently associated with MI\(^1\). For Gofman, these LDLs appeared to be the most important factor in CHD. If any of these specific bands were abnormally raised, then subjects were more likely to suffer a cardiac event, but these profiles were raised in unique manners from individual to individual. He then demonstrated that lipoprotein Sf 0–20 increased after eating fat and lipoprotein Sf 20–400 after eating carbohydrate. Sf 0–20 was the predominant carrier of cholesterol, so would explain a rise in serum TC, while a high Sf 20–400 appeared to lower TC. Yet, if either of these classes of lipoprotein were raised there appeared to be a greater risk of CHD. Accordingly, Gofman believed that dietary researchers must consider both the level of fat and carbohydrate in the diet and that they should abandon their favoured measure of total cholesterol.

In response to these controversies, two forms of experimentation were believed necessary by advocates and critics alike to prove dietary fats were causally related to CHD: (i) prospective observational studies that track the diets/serum cholesterol of a cohort over time to see who develops CHD; (ii) intervention trials where an experimental group is given a diet low in SFA and/or high in PUFA and compared to a control group to measure differences in serum cholesterol and CHD incidence. However, at the time of publication of the AHA’s advice, no data from these forms of experiment were available.

**Observational studies**

As discussed, in the early 1950s, major prospective studies were launched to test the ‘lipid-hypothesis’ – the link between serum cholesterol, atherosclerosis, and CHD. In the 1960s, these returned supportive evidence of a statistically significant association between raised serum TC levels and

---

\(^1\) Gofman uses the term LDLs to refer to what would today be recognised as low-density lipoproteins (LDL) [Sf 0–12], intermediate-density lipoproteins (IDL) [Sf 12–20], very-low-density lipoproteins (VLDL) [Sf 20–400], and chylomicrons (Williams et al. 1990). For Gofman, all of these were low density compared to another form of lipoprotein, high-density lipoproteins (HDL), which is retained in current terminology. These terms refer to the concertation of particular lipoproteins in the blood. These differ from the measurement of the amount of cholesterol bound in particular lipoproteins – e.g., LDL-C and HDL-C, popular measures in clinical medicine and epidemiology.
CHD, particularly in younger men (Kannel et al. 1961, Keys et al. 1963, Paul et al. 1963). However, the scientific community believed more evidence was required to understand whether dietary fat, particularly SFA, was also associated with CHD. To do this, a study must be able to track diet, serum cholesterol, and new cases of CHD. However, between the issuing of the AHA’s advice and the publication of population-wide dietary advice in the US and UK (among many others) in the 1970s and early 1980s, these forms of experiment produced conflicting findings.

In the 1970s, Keys (1970) published the findings of his Seven Countries Study, a cross-sectional study examining diet, serum cholesterol, and CHD mortality in 12,763 males between the ages of 40–59 years in seven countries. By further splitting this sample into 16 cohorts that represented separate communities within each country, Keys sought to understand whether diet could explain both variation between populations in regards to average serum cholesterol levels and CHD mortality, but also whether diet could explain variation between communities in particular countries. His findings suggested that calories from fat, but particularly from SFA, were significantly, positively associated with both serum cholesterol and CHD mortality between countries and communities. According to Keys, ~84% of the variance in CHD mortality could be explained by the average amount of SFA consumed.

Other studies returned similar findings when comparing different populations and communities. Most notably, the NI-HON-SAN study, which examined diet, serum cholesterol, and CHD in three groups of ethnic Japanese in Japan, Honolulu, and San Francisco – a total sample of 11,989 men aged between 45 and 69 years. This found that those in Japan had far lower average levels of TC – 20% lower – than those who had emigrated to Honolulu and San Francisco, who had similar average TC levels (Kato et al. 1973, Marmot et al. 1975). Furthermore, those in Japan had the lowest rate of CHD mortality across all ages, but the difference between these populations was only statistically significant between the ages of 55–59 years. In this age group, those in San Francisco were about 3.5-times more likely to die from CHD than those in Japan, while those in Honolulu had only a slightly higher mortality rate.
from CHD than those in Japan – 1.7 per 1,000 compared to 1.4 per 1,000 (Worth et al. 1975). Those in Japan also ate the smallest amount of total fat (15% of calories), SFA (6.7% of calories), and unsaturated fats (8.5% of calories). This was compared to an intake of 33% of calories from total fat, 23% from SFA, and 10.5% from unsaturated fats in Honolulu, and an intake of 37.5% from fat, 26% from SFA, and 11% from unsaturated fats in San Francisco (Kato et al. 1973). These intakes were held partly responsible for the death rate from CHD; however, the authors were puzzled that, while diet and TC levels appeared to explain well the differences between the Japan versus the San Francisco group, the predictive potential of both diet and TC became weaker when considering the Honolulu population (Syme et al. 1975).

However, prospective cohort studies that examined the relationship between diet, serum cholesterol, and CHD in a community, rather than between communities, produced generally unsupportive findings towards the diet–heart hypothesis. Before 1985, 10 such studies, covering 14 cohorts, had reported findings on the relationship between dietary fat and CHD mortality and morbidity within particular communities (Paul et al. 1963, Gordon and Kannel 1970, Medalie et al. 1973, Morris et al. 1977, Yano et al. 1978, Garcia-Palmieri et al. 1980, Gordon et al. 1981, Shekelle et al. 1981, McGee et al. 1984, Kromhout and Coulander 1984). In three of these cohorts (Yano et al. 1978, Garcia-Palmieri et al. 1980, and McGee et al. 1984), a significant difference in SFA consumption was observed between those with CHD and those without CHD, linking higher SFA intake with a higher risk of CHD. However, after controlling for other known risk-factors, this relationship was non-significant in two of them (Yano et al. 1978; Garcia-Palmieri et al. 1980). Thus, in 13 of these 14 cohorts no significant relationship was observed between SFA consumption and CHD after controlling for other known risk-factors.

Six of the cohorts showed a significant association between higher intake of PUFA (as % of total calories) and a lower risk of CHD (Yano et al. 1978, Garcia-Palmieri et al. 1980, Gordon et al. 1981 [Puerto Rico and Honolulu cohorts], Shekelle et al. 1981, and McGee et al. 1984). However, only in three
cohorts (Gordon et al. 1981 [Puerto Rico and Honolulu cohorts], Shekelle et al. 1981) was this relationship significant after controlling for other known risk-factors. Thus, in 11 of these 14 cohorts, no significant relationship was observed between PUFA consumption and CHD after controlling for other known-risk factors.

Thus, the evidence from prospective studies produced a paradoxical result – fat consumption appeared to predict the occurrence of CHD between countries or communities, but this predictive ability collapsed in specific populations. This raised concerns that confounding variables might be distorting the former findings (Stallones 1983). Alternatively, advocates of the diet–heart hypothesis argued that the limited population variance in dietary fat intake in particular communities meant that within-cohort studies were an inappropriate method. Rather, comparisons between populations ought to guide scientific opinion (Epstein 1965).

**Intervention trials**

Intervention trials were relied upon to isolate whether SFA really did cause CHD. Intervention trials tested fat modified (increased PUFA and reduced SFA) and fat-restricted (reduced SFA and total fat) diets in the contexts of primary prevention (before a clinical manifestation of CHD) and secondary prevention (after a clinical manifestation). The best controlled studies, RCTs, primarily examined the latter due to the ease in which patients who had recently suffered MI could be enrolled into a well-controlled experiment. However, these too produced inconclusive findings.

**Secondary prevention**

The Oslo study (Leren 1966) found a significant beneficial effect from fat modification, but three other RCTs found either equivocal (Research Committee 1965, Medical Research Council 1968) or harmful effects (Rose et al. 1965). These were the only published secondary prevention RCTs until the partial publication of the Sydney Diet–Heart Study in 1978, which only published results concerning all-cause mortality (Woodhill et al. 1978). The full results of that trial were recently recovered and re-analysed by Ramsden et al.
(2013), who found a significant harmful effect of replacing SFA with PUFA on CVD and CHD despite significantly reducing TC compared to controls. The Bierenbaum et al. (1973) study, a non-randomised intervention trial, found a beneficial effect of SFA restriction in secondary prevention, but only in those aged under 60 years. All trials, however, successful lowered serum TC levels by 6–14% in the intervention group compared to controls.

While some interpreted the results as evidence against the diet–heart hypothesis, others believed they demonstrated that dietary intervention needed to be began before a clinical manifestation of atherosclerosis (Stamler 1967).

**Primary prevention**

In primary prevention the picture looked more positive. The only RCT (Dayton et al. 1969) fully available before 1985 examined fat modification in 424 men aged between 54 and 88 years compared to a matched control group of 422 men with a mean follow-up of 3 years. A 12.7% decrease in TC was reported in the intervention group compared to controls, which was held responsible for a statistically significant 24% decrease in combined non-fatal and fatal CHD/CVD events (including stroke, amputation and aneurism) in the intervention group. This combined end-point analysis, however, obscured the non-significant results of treatment on the original primary outcomes – MI mortality, MI incidence, and sudden coronary death. In terms of MI mortality, 23 deaths were recorded in both groups. For non-fatal MI, there were 27 events in the control group and 19 in the intervention – a non-significant difference. For sudden coronary death, a category used when patients die rapidly and not under direct observation, there were 27 in the control group and 18 in the intervention – a non-significant difference. While hinting at a possible beneficial effect, the study failed to demonstrate that dietary treatment was an effective treatment for CHD, at least by conventional statistical thinking. In regards to all-cause mortality, 177 deaths were recorded in the intervention versus 174 in the control group.

Another primary prevention trial (Frantz et al. 1975) failed to fully publish its results. Frantz et al. partially reported their findings suggesting that their trial
had found a beneficial effect of fat-modified diets in reducing the combined rate of MI, sudden death, and stroke in men. However, 14 years later, Frantz et al. (1989) published their results in a peer-reviewed journal. Now, the authors concluded that there was no difference between the intervention and control groups in terms of CHD events, CHD mortality, or all-cause mortality for the entire population studied, which included males and females between the 30 and 70 years. This was despite lowering TC by 14%, while the control group experienced only a 1% reduction. In 2016, following recovery of the original unpublished data, Ramsden et al. (2016) found that, for those participants in the trial for a year or more, there was a “22% higher risk of death for each 30 mg/dL (0.78 mmol/L) reduction in serum cholesterol... (hazard ratio 1.22, 95% confidence interval [CI] 1.14 to 1.32; P<0.001)”. Additionally, there was “no evidence of benefit in the intervention group for coronary atherosclerosis or myocardial infarcts” (p.1).

Other forms of intervention trials arrived at apparently supportive findings. Christakis et al. (1966) reported a statistically significant reduction in CHD events in an intervention group (n=814 men, aged 40–59 years) on a diet in which calories from fat were reduced to 30%, with a PUFA:SFA ratio of 1.25–1.50:1. However, this study enrolled a non-randomised control (n=463 men, aged 40–59 years) 2 years after its initiation. The CHD results included a total of eight men in the intervention group who had suffered a CHD event compared to 12 in the control group. The intervention group had experienced about a 13% fall in TC level while the control groups level remained unchanged, which the authors attributed as the cause for the difference in CHD events. To arrive at their statistically significant finding, the authors measured the number of person-years of experience (the number of years on the trial) and then estimated the CHD incidence rate per 100,000 person-years of experience. By this, they found, in men aged 40–49 years that the intervention group had an incidence rate of 196 per 100,000 person-years versus the control group that had a rate of 642 per 100,000 person-years. This was based on one CHD event in the intervention group versus four in the control, but even by this method no significant difference was found between the intervention
and control for those aged 50–59 years. The authors concluded “A significantly lower incidence of coronary heart disease was observed in the experimental group consuming the study diet” and highlighted “The public health significance of our salutary effect on coronary heart disease morbidity” (pp.311–313). In 1968, Rinzler (1968), in a follow-up using the same methods of analysis, found a further 20 CHD events in the control group compared to only eight additional events in the intervention group. Thus, by the close of this study, a total of 32 men in the control group developed CHD versus 17 in the intervention.

Other trials reported similar decreases in CHD following dietary treatment. The Finnish Mental Hospital Study (Turpeinen et al. 1979), a crossover study between two mental hospitals, found CHD events were reduced by half in both the first and second treatment groups after successfully lowering TC by between 12–20%. Hjermann et al. (1981), a multifactorial intervention trial, provided probably the strongest evidence that dietary intervention and cessation of smoking could lead to a major reduction in CHD in those with exceptionally high serum cholesterol levels – with a 47% reduction in CHD in the experimental group compared to control. However, due to multiple interventions, the effect of SFA could not be precisely isolated. Other studies, notably the Multiple Risk Factor Intervention Trial (MRFIT 1982), a major multifactorial trial, found no benefit for either CHD mortality or all-cause mortality from dietary and other lifestyle interventions.

**A conflicted evidence-base**

By the 1980s, there were still unanswered questions over the validity of the diet–heart hypothesis. The findings from RCTs appeared, on balance, to undermine the hypothesis, at least in the context of secondary prevention, as too did the lack of an association between SFA and CHD from prospective cohort studies. However, evidence from cross-country and community comparisons appeared to be strongly supportive, and the less rigorously controlled primary prevention trials returned promising findings. Importantly, the evidence-base was also impacted by publication bias – two of the seven major trials that could have published unsupportive results did not. While today
we know the results of these trials, only five of these were available to researchers until recently.

It was in this context that the dietary guidelines were developed and they came down heavily on the side of the diet–heart hypothesis. In 1977, the US issued population-wide dietary advice to reduce SFA consumption to 10% of food energy (US Senate Select Committee 1977), and the UK, via the COMA report (UK Department of Health 1984), issued similar advice to reduce SFA to 15% of food energy. These guidelines suggested that SFA should be restricted and partially replaced by foods rich in PUFA. Major supporters hailed it as a historic achievement for public health (Stamler 1978; Walker and Canon, 1985) but some warned that the evidence was not strong enough (Ahrens 1976, 1979), while others strongly criticised the decision (Mann 1977; McMichael 1979). However, according to Norum (1978), by the late 1970s these critics were a minority – 92% of 200 scientific experts from 22 countries that he surveyed agreed that SFA should be restricted in the diet.

Nevertheless, Stallones (1983), while accepting that many believed in the validity of the diet–heart hypothesis, argued that the evidence-base was ultimately ambiguous. For Stallones, this ambiguity was the result of a failure to fund a long-term intervention trial to adequately test the hypothesis. He berated the decision not to extend the National Diet-Heart Feasibility Study (1968), which, despite demonstrating that serum cholesterol levels could be reduced by about 13% in a large experimental group and maintained adherence levels at close to 90%, did not receive funding to continue the trial.

The report was treated as though it were contaminated; the proposed sample size, nearly 60,000 persons, and the price tag, from 50 to 250 million dollars, were judged to render the feasible unfeasible. Since that judgment was made, a number of studies of less definitive character have been reported and others are still under way...In any case, we are left with what we have, and what we have contains the seeds of conflict. (pp.169–170, italicised for emphasis)

In 1984, in response to what appeared to be a mounting controversy, the National Institutes for Health (NIH 1985) gathered a panel of experts to resolve whether the link between diet, raised serum cholesterol, and CHD was causal.
After reviewing the evidence, the panel concluded that elevated serum cholesterol is a cause of CHD and that lowering by dietary means is protective:

The evidence justifies for men, women, and children ages 2 years and older the reduction of calories from fat from the present average level of 40 percent to 30 percent, calories from saturated fat to 10 percent or less, and dietary cholesterol to no more than 250 to 300 mg daily. (NIH 1985)

The conference placed considerable weight on the Lipid Research Clinics Coronary Primary Prevention Trial findings (LRC-CPPT 1984). This was a large, multicentre double-blind RCT of 3,806 men aged between 35 and 59 years with high serum cholesterol levels (clinical hypercholesterolaemia) but free of CHD at the start of the trial. Both groups received the same dietary advice to lower their consumption of SFA and dietary cholesterol, but the intervention group was given cholestyramine, a bile acid sequestrant, a drug demonstrated to lower TC and LDL-C levels (cholesterol bound in LDL), while the control was provided with a placebo. After 7.4 years of follow-up, 44 men in the control group had died of CHD compared to 32 in the intervention group. Regarding non-fatal MI, 225 men experienced at least one event in the control group versus 195 in the intervention. Combining the results of these two primary end-points, the intervention group experienced a statistically significant 19% reduction in CHD risk compared to the control group. This was attributed to a significant reduction of serum cholesterol by 8.5–12.6% compared to the control group. The authors concluded “This clinical trial provides strong evidence for a causal role for these lipids in the pathogenesis of CHD” (p.351), and while the trial was “not designed to assess directly whether cholesterol lowering by diet prevents CHD. Nevertheless, its findings, taken in conjunction with the large volume of evidence relating diet, plasma cholesterol levels, and CHD, support the view that cholesterol lowering by diet also would be beneficial” (p.360).

For many, this was the end of the debate. Yet, while the LRC-CPPT provided support for the lipid-hypothesis whether it also supported the diet-heart hypothesis was questionable. First, this was a study in a particular group of men with established hypercholesterolaemia, and whether its findings could
be generalised beyond this specific group were questioned (Oliver 1984). Second, as this was not a test of diet, it remained unclear as to whether diet really did alter serum cholesterol in a manner that was dangerous (Oliver 1984).

Complicating matters further, one difficulty that arose prior to the results of this trial was the arrival of unexpected findings regarding the relationship between high-density lipoproteins (HDL) and CHD. Several large epidemiological studies reported statistically significant inverse associations between HDL-C levels and CHD risk (Miller and Miller 1975, Rhoads et al. 1976, Miller et al. 1977, Gordon et al. 1977, Castelli et al. 1977), leading some to claim that a low HDL-C level was a better predictor of CHD risk than either TC or LDL-C.

In the body, LDL transports cholesterol and other fat molecules to cells, while HDL removes cholesterol by transporting it to the liver to be excreted. As both the lipid and dietary hypotheses focussed on the harmful effect of serum cholesterol, and devoted most of their attention to how TC or LDL-C contributed to atherosclerosis and CHD, the HDL-C findings were unexpected and posed problems for how to interpret previous studies that had not measured this (Levy 1981). Further, Schaefer et al. (1981) reported that, while reducing the amount of fat in the diet, particularly SFA, reduced both TC and LDL-C levels in the blood, it simultaneously lowered HDL-C. According to the authors, the ratio of LDL-C to HDL-C remained unchanged following serve restriction of dietary fats. Whether dietary fats raised or lowered lipoprotein profiles in a beneficial or harmful manner became unclear. While Shepherd et al. (1978), in a metabolic-ward study, reported that PUFA feeding appeared to lower HDL-C to a greater extent than it did LDL-C.

As Blackburn (2012) recalls, this provoked a reflection by some scientists regarding the basis of their approach to understanding the relationship between diet and CHD, specifically whether to adopt a measure of lipoprotein concentrations rather than serum cholesterol. According to Krauss, this led to the rediscovery of Gofman’s work on lipoproteins published more than 20 years' prior (Krauss quoted in Blackburn 2012).
Despite renewed controversy in the early 1990s (Ravnskov 1992,1995; Atrens 1994) in which it was claimed that the belief in the hypothesis stemmed from a biased selection of evidence, the general opinion of the scientific community has been that the diet–heart hypothesis is supported by the available evidence. Recently, however, controversy has re-erupted due to the publication of several meta-analyses re-analysing that evidence-base.

In 2010, Siri-Tarino et al. (2010) published a meta-analysis of prospective cohort studies that examined the relationship between fatty acids and CHD/CVD. After pooling the results from 21 trials they concluded “prospective epidemiologic studies showed that there is no significant evidence for concluding that dietary saturated fat is associated with an increased risk of CHD or CVD” (p.535).

This was followed by Chowdhury et al. (2014) systematic review and meta-analysis of 32 prospective studies and 27 RCTs, which found no significant association between SFA, PUFA, or MUFA and CHD in a pooled risk-ratio (RR) analysis. However, they did find a significant adverse effect of dietary trans-fatty acids – a fatty acid formed from the hydrogenation of vegetable oils. From this, the authors concluded “Current evidence does not clearly support cardiovascular guidelines that encourage high consumption of polyunsaturated fatty acids and low consumption of total saturated fats” (p.398).

de Souza et al. (2015) found via meta-analyses of 12 prospective studies that SFA was not associated with all-cause mortality (RR 0.99, 95% CI 0.91–1.09), CVD mortality (RR 0.97, 95% CI 0.84–1.12), CHD mortality (RR 1.15, 95% CI 0.97–1.36) or CHD events (RR 1.06, 95% CI 0.95–1.17). The authors concluded that “Saturated fats are not associated with all-cause mortality, CVD, CHD…but the evidence is heterogeneous with methodological limitations” (p.1).

However, Hooper et al. (2015) meta-analysis of 17 RCTs found SFA reduction was not significantly associated with all-cause mortality (RR 0.97, 95% CI 0.72–0.96), CVD mortality (RR 0.95, 95% CI 0.80–1.12), or combined
fatal and non-fatal MI events (RR 0.90, 95% CI 0.80–1.01). However, there was a significant association between the reduction of SFA and the number of CVD events (RR 0.83, 95% CI 0.72–0.96). From this, the authors claim that reducing SFA reduces CVD events by 17% and concluded that their findings were “suggestive of a small but potentially important reduction in cardiovascular risk on reduction of saturated fat intake” (p.2).

Harcombe et al. (2015), after subjecting evidence derived from six RCTs conducted before 1984 to meta-analysis, found no significant difference between all-cause mortality (RR 0.996, 95% CI 0.865–1.147) and CHD mortality between interventions and controls (RR 0.989, 95% CI 0.78–1.25). This led the authors to conclude that dietary guidelines were introduced “in the absence of supporting evidence from RCTs” (p.1). In 2016, Harcombe et al. (2016) expanded this study to cover RCTs published before 2015, which included an additional four trials, for all-cause mortality RR 0.991 (95% CI 0.94–1.05), for CHD mortality RR 0.976 (95% CI 0.88–1.08). Similarly, in 2017, Harcombe et al. (2017) examined via meta-analysis the available prospective cohort studies at the time of devising dietary guidelines. They concluded “that the prospective cohort study evidence available at the time did not support the introduction of dietary guidelines” (p.1740). These authors went on to argue that, not only did the available evidence at the time of writing not support the introduction of dietary guidelines, but that the evidence still does not support their use.

In 2017, the AHA (Sacks et al. 2017) in response to mounting controversy published a Presidential Advisory Statement that analysed a selection of what they considered the most robust studies. Specifically, they analysed what they deemed to be the core early trials (Dayton et al. 1969, Leren 1966, Medical Research Council 1968, Turpeinen et al. 1979), and came to a rather different position to other meta-analyses, “The results showed that lowering saturated fat and replacing it with vegetable oil rich in polyunsaturated fat, primarily soybean oil, lowered CHD by 29%” (p.e5). For the authors, “we conclude strongly that lowering intake of saturated fat and replacing it with unsaturated fats, especially polyunsaturated fats, will lower the incidence of CVD” (p.e1).
A few months later, the Prospective Urban Rural Epidemiology study (Dehghan et al. 2017), a large, multi-country prospective cohort study examining 135,335 individuals between the ages of 30 and 70 years in 18 different countries over a median of 7.4 years, concluded:

Total fat and types of fat were not associated with cardiovascular disease, myocardial infarction, or cardiovascular disease mortality, whereas saturated fat had an inverse association with stroke. Global dietary guidelines should be reconsidered in light of these findings (p.2050).

Discussion

After more than 60 years of investigation, the diet–heart hypothesis remains controversial. The history of the debate appears to be one of a long period of underdetermination; scientists have drawn dramatically different conclusions from the available evidence in the absence of a definitive study. Indeed, in the few sociological studies of this controversy, Garrety (1998) concluded that:

Although scientists have produced a vast amount of information about fats, cholesterol, and CHD, they have never been able to produce a “definitive experiment,” that is a convincing demonstration that reductions in saturated fat and cholesterol intakes will, through a decrease in serum cholesterol levels, actually result in a lower incidence of CHD. The science which is used to justify dietary recommendations draws on a great deal of circumstantial evidence…However, there are also many anomalies and contradictions within this knowledge, as skeptics have frequently pointed out (p.403)

Yet, the different sides of this argument have rarely acknowledged such uncertainty. Those in support claim their opponents are capitalising on populist prejudices and are advocating dangerous advice, ignorant to the apparent mass of supportive evidence (Stamler 1980; Walker and Canon 1985; Steinberg 2007; Williams 2015). Those in opposition accuse their opponents of following a populist theory that produced inaccurate advice that was, at best, a pointless intervention, or, at worst, dangerous – apparently blind to the wealth of unsupportive data (Mann, 1977, Ravnskov 1998, Teicholz 2015). While several researchers have come to the conclusion that the evidence-base
leads to no clear conclusions as to whether the hypothesis is supported or refuted (Ahrens 1976; Stallones 1983)

Writing in 1957, Yudkin (1957), in a survey of the literature, proclaimed “one begins to have the uneasy feeling that both the proponents and opponents of the dietary hypothesis are quoting only those data which support their view” (p.6987). In 1975, Blackburn (1975), a major figure in dietary epidemiology, claimed “Two strikingly polar attitudes persist on this subject, with much talk from each and little listening between” (pp.105–106). In 1983, Stallones observed that “results of research aimed at testing the [diet–heart] hypothesis experimentally have been used to support opposed positions” (p.168). In 1996, Rieger (1996) concluded “definitive answers regarding the diet-heart disease hypothesis are premature and that the polarized positions of acceptance vs rejection of the hypothesis fail to account for the full range of results” (p. 1227). In 2018, Forouhi et al. (2018) noted that, while the “The public health debate about dietary fats and health has been ongoing for over 60 years…the medical literature is still full of articles arguing opposing positions” (p.1).

As a long-running controversy with an evidence-base widely recognised to be incomplete in parts, this represents an important case study to understand how, in the absence of definitive evidence, scientific opinion is formed and the factors that lead to disagreement between scientists. This thesis explores one potential contributing factor – whether the way in which scientists used the evidence that was available to them shaped consensus and disagreement over time. For some scientists, the evidence was strong enough to support the issuing of population-wide dietary guidelines. For others, however, important findings regarding whether reducing SFA in the diet would lower CHD rates were still missing. For others still, the existing evidence was enough to refute the hypothesis. Were these scientists using the same evidence but interpreting it differently? Or were they using different segments of evidence-base?
2. Underdetermination, Interpretative Flexibility, and Evidence Selection

Why can’t scientists agree on these matters? This is a matter of public and political concern, as the diet–heart hypothesis is intertwined with dietary guidelines. There is pressure on scientists to reach a definitive position, but to expect certainty is unreasonable due to the fallibility of scientific knowledge. As Popper ([1934]2000) stated:

> [E]very scientific statement must remain tentative forever. It may be corroborated, but every corroboration is relative to other statements which, again, are tentative. Only in our subjective experiences of conviction, in our subjective faith, can we be absolutely certain (p.280).

Science is often inaccurate, always incomplete, and, typically, there are different interpretations of evidence along with several reasonable rival explanations of any topic. These characteristics are widely accepted as a feature of science, and many believe that it is through the holding and comparison of different, empirically testable, theoretical perspectives that science progresses (Popper [1934]2000; Kuhn [1962]1970; Feyerabend 1975).

When these disputes cannot be resolved easily, a fissure in the scientific community emerges (a scientific controversy) – with different scientists arguing from different perspectives with no obvious way of resolving the argument. For Popper ([1934]2000), disagreements end when a conjecture is falsified. However, scientific controversies are generally not settled so easily. Scientists often differ about how data should be interpreted, what methods are appropriate for testing a hypothesis, whether the phenomenon has been measured or categorised correctly, and whether theories are actually supported by the evidence. Kuhn ([1962]1970) argued that scientific disagreements stem from the holding of different theoretical edifices, standards of evaluation, and research problems by different communities. Controversies revolve around scientists holding different epistemological and
ontological assumptions and methodological preferences, and corroboration and falsification requires different standards for different groups.

But how can scientists hold different interpretations of empirical evidence? Mill ([1843]2011), in *A System of Logic*, gives an insight into the complexities of scientific interpretation:

> [A]n hypothesis is not to be received as probably true because it accounts for all the known phenomena, since this is a condition sometimes fulfilled tolerably well by two conflicting hypotheses; while there are probably a thousand more which are equally possible, but which, for want of anything analogous in our experience, our minds are unfitted to conceive (p.503).

Here, Mill suggests that there are typically several feasible hypotheses for any given event, but these are generally not seriously considered or even conjectured. Rather, people tend to back a particular hypothesis despite the fact that it is always logically possible that other hypotheses could explain a particular event. This alludes to one of the most influential ideas in the philosophy of science – the underdetermination of theory by empirical evidence (Duhem [1914]1991, Quine 1953). Quine proposed that, at any time, empirical evidence available is, in itself, insufficient to logically justify our belief in the truth of our theoretical explanations. According to Quine, abstract reasoning from evidence cannot be the decisive factor in the selection of one hypothesis over another, if other hypotheses are equally able to explain this evidence.

To understand this, imagine we find a correlation between the consumption of dietary fat and the prevalence of CHD. We can conceive of several different explanations: (i) the consumption of fat causes CHD; (ii) the consumption of fat is an effect of CHD; (iii) the consumption of fat is neither a cause or effect of CHD, and the correlation is the result of error or chance; (iv) that the correlation is true, but that the consumption of fat and CHD are both independently affected by some third, unknown factor – a confounded relationship. The evidence here is consistent with different theoretical explanations.

However, determining whether evidence supports or refutes a particular
proposition is complex. Duhem ([1914]1991) proposed that any test of a hypothesis always implies acceptance of many other beliefs. Duhem argued that no experiment, however well designed, tests a single hypothesis in isolation. It is always a test of an entire interlocking body of assumptions and hypotheses: “when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses constituting this group is unacceptable and ought to be modified; but the experiment does not designate which one should be changed” (p.187). While a falsifying result can tell us that there is a problem with a system of beliefs, it cannot establish exactly which belief has been falsified.

For example, we might test whether eating a low-fat diet affects the risk of CHD in an RCT. Our theory might lead us to predict a reduced risk. But if the trial shows no effect, has our hypothesis been falsified? Maybe there was a problem with the experiment: perhaps the groups were not properly randomised, or perhaps the randomisation produced, by the chance inherent in randomisation, some imbalance between groups that was overlooked as irrelevant but for unknown reasons was critical. Was the 'low-fat diet' really low enough in the sort of fat that is dangerous? Was CHD reliably detected – did different physicians at different hospitals use exactly the same criteria? Maybe the control subjects also changed their diet, just because they were part of a trial and so more aware of their health, or perhaps they changed their behaviour in other, unrecorded ways. Perhaps people who agree to be part of a trial are not typical of the whole population? Were the statistical tests appropriate – was the study adequately powered? Were there errors in recording CHD or food consumption? Perhaps another, unknown factor distorted the results? Whether our experiment has falsified our hypothesis is underdetermined by the empirical evidence alone.

Duhem and Quine’s theses, when considered together, imply that an isolated scientific hypothesis is not open to empirical refutation (Harding 1976). Predictions depend on many background hypotheses and assumptions. Any prediction from our hypothesis that ‘fat causes CHD’, depends on the belief in the appropriateness of the procedure used in intervention trials, the statistical
methods used to assess the correlation, the reliability of data, on the underlying mechanistic theory that warrants such a hypothesis, and on the belief that ‘fat’ and ‘CHD’ are correctly classified. The acceptance of a hypothesis depends on belief in other hypotheses. Belief in a particular scientific fact depends on a broader belief set – unsupportive evidence never simply refutes a single claim.

Generally, in the biomedical sciences, it is believed that to establish a level of certainty regarding causation of disease there must be three bodies of evidence that cohere with one another – evidence of mechanism, association, and intervention. Where there is discordance between these three branches, reasoned scientific disagreement can ensue. Furthermore, if a rival theory can equally well explain the findings from these, then a reasonable scientific disagreement can follow because, in these instances, there is no clear solution to whether a particular hypothesis should stand or fall.

Yet, in cases where there are empirically equivalent alternatives, it is not typical to hold that they are equally valid or important. What then influences the selection of one over another? If, as underdetermination implies, empirical reality plays a non-decisive role in the way we interpret evidence and evaluate particular theories, then other factors beyond observation and reasoning must also be involved in knowledge construction. Hesse (1980) believes “it is only a short step from this philosophy of science to the suggestion that adoption of such criteria…should be explicable by social rather than logical factors” (1980, p.33).

Kuhn and the sociology of scientific knowledge

Since the 1960s, particularly following Kuhn ([1962]1970), sociologists of scientific knowledge (SSK) have been interested in how scientists construct understanding following the implications of underdetermination. To understand how scientists come to agree and disagree on matters, SSK researchers set out to understand the social process involved in shaping scientific knowledge (Bloor [1976]1991).

Kuhn ([1962]1970) argued that scientists in a given field work within an accepted paradigm, sharing rules that determine what questions are important and how they should be answered. Scientists work within these ‘paradigms’
and knowledge advances only when experimentation follows a common set of agreed upon principles. He proposed that scientific research went through three stages of development. In its early (pre-paradigm) stage, research is often chaotic, with several rival theories available, but that once a particular position gains enough interest and support the scientific field exhibits *path dependency* (normal science):

Philosophers of science have repeatedly demonstrated that more than one theoretical construction can always be placed upon a given collection of data... But that invention of alternates is just what scientists seldom undertake except during the pre-paradigm stage of their science’s development and at very special occasions during its subsequent evolution. So long as the tools a paradigm supplies continue to prove capable of solving the problems it defines, science moves fastest and penetrates most deeply through confident employment of those tools. The reason is clear. As in manufacture so in science—retooling is an extravagance to be reserved for the occasion that demands it (p. 76).

Here, Kuhn drew attention to how the equipment and expertise needed to test a theory and the access to funds and personnel constrained the scientific community to follow particular lines of enquiry.

Kuhn’s perspective sat in direct contrast to Popper’s ([1934]2000). Popper held that progress in science depended on rigorous testing of theories and the critical interaction between scientists over the findings. By contrast, Kuhn described a dogmatic science; one where scientists actively avoided criticising their respective theoretical frameworks, and he argued that this dogmatic acceptance of theory was integral to progress in science:

Normal science...is predicated on the assumption that the scientific community knows what the world is like. Much of the success of the enterprise derives from the community’s willingness to defend that assumption, if necessary at considerable cost. Normal science, for example, often suppresses fundamental novelties because they are necessarily subversive of its basic commitments (p.5).

For Kuhn, a scientific controversy (crisis) arises when an existing paradigm encounters ‘anomalies’; instances that contradict or cannot be explained in reference to the existing theoretical perspective. However, he argued that the existence of anomalies is never enough in itself to explain the
existence of controversies because advocates of a perspective “devise numerous articulations and ad hoc modifications of their theory in order to eliminate any apparent conflict”, while scientists “may begin to lose faith and then to consider alternatives, they do not renounce the paradigm that has led them into crisis” (pp.77–78).

How then are scientific crises brought to a close? Kuhn ([1962]1970) proposed that “a scientific theory is declared invalid only if an alternate candidate is available to take its place” (p.77). This alternative must be able to “attract an enduring group of adherents away from competing modes of scientific activity”, while also being “sufficiently open-ended to leave all sorts of problems for the redefined group of practitioners to resolve”, and capable of sufficiently explaining anomalies (p.10). Thus, for Kuhn, controversies and eventual refutations were provoked by social revolutions that transformed the structure of particular communities; of personnel and their networks of collaboration and citation.

For Kuhn, the growth of scientific paradigms is based on the successful propagation of scientific theories, evidence, and methods that shape what scientists eventually hold to be true. By this account, to understand why certain scientists pursue certain research objectives, conform to particular theories, use certain methods, or discuss strands of the evidence, one must look to what other scientists have done and the ideas that became popular in the scientific community.

Kuhn provided an overarching framework by which to understand how science functioned, including how different interpretations and evidence behaviours might help understand the content of its knowledge. Following Kuhn, the Strong Programme in SSK sought to understand in more empirical detail the social process involved in the construction of scientific knowledge and its controversies (Bloor [1976]1991). The Strong Programme proposed four tenets to guide researchers in their analysis of science: causality, impartiality, symmetry, and reflexivity. To seek for causal explanations of scientific belief, one looks at the practices of scientists in the methods they apply, the theories they hold, the evidence they use, the justifications they
provide for these, and broader influences such as funding structures or policy interest. These accounts should be produced from an impartial perspective that treats opposing positions equally – that is, by applying the same standards of evaluation to both – in the same sense that a jury and judge are impartial at the outset of a trial. In these accounts, causal explanations for scientific belief should by symmetrical in the sense that the causes of belief are to be explained by applying the same methods to generate findings and theories of belief formation. Finally, the methods used to understand the formation of scientific belief should be capable of being applied to your own analysis to understand how you reached your position.

Initially, the different interpretations that scientist came to in regards to the meaning of different findings and theories attracted attention. Following a desire to separate themselves from the Sociology of Error (a view attributed to Merton that assumed that scientific disagreement stemmed from some irrationality or deviance of one side of a debate), SSK researchers tried explain the factors shaping scientific opinion without recourse to evaluating the strengths and weaknesses of different arguments. Their attention focussed on the social processes involved in the endorsement of particular knowledge claims.

For Barnes (1983), the very meaning of concepts is established through *Bootstrapped Induction* – a process where terms used to describe particular things in particular ways take on stability through their repeated usage by social groups. Barnes argued that the meaning of a term is not simply a by-product of pattern recognition of sensory data, but contains a substantial self-referential component – a conventional meaning upheld and adapted to new instances by a community. Generalising this argument, the meaning of a particular finding or theory is established by the repetition of particular interpretations by members of a particular scientific community.

In *Laboratory Life*, Latour and Woolgar (1979) focussed on understanding how knowledge is constructed by observing scientists in a laboratory. From ethnography of Guillemin’s laboratory, the authors drew two major conclusions: (i) scientists negotiated the meaning of terms, experimental
results, and the utility and precision of instruments. Scientific knowledge gained its stability and truth-likeness in this laboratory not solely from logical thought and empirical observation, but through this process of social negotiation; and (ii), scientific work was directed at publication, in which the main goal was transforming statements into statements of lesser or greater credibility by persuading a broader audience beyond that laboratory. Thus, Latour and Woolgar identified that scientific knowledge construction involved two phases. The construction of a particular position within a laboratory involved agreement amongst a small number of relevant individuals, but then the goal was to persuade a much larger audience of the credibility of this position. This was not presumed to be a linear process; rather scientists were both in the business of persuading, and being persuaded by, other scientists. Accordingly, scientific knowledge is constructed through a process of persuasion.

Collins (1981) proposed that to understand the functioning of science, analysts ought to focus their attention on controversies. To do this, he proposed that a researcher first tries to capture the ‘interpretive flexibility’ of the theories, concepts, and data that are involved in a particular dispute. This, a sociological interpretation of underdetermination, sought first to establish that different interpretations actually exist in the scientific community, and second to understand why scientists come to different positions and how such disagreements are eventually resolved. If the meaning of findings and theories are constructed through social use, then divergent interpretations must stem from different social processes in different communities.

Instances where scientists have come to conflicting positions regarding the meaning of particular findings have drawn the attention of philosophers because of their implication for logical reasoning, and sociologists because of the need for extra-scientific reasons to explain the holding of one interpretation rather than another. But this is not the only form of scientific disagreement.

Scientists, at present, are deeply concerned at a wide range of factors that they understand to be distorting science. Some have proposed that systemic bias can lead to distortions in scientific knowledge – citation bias
(Jannot et al. 2013, Fanelli 2013, Duyx et al. 2017), publication bias, failure to conduct replication studies, and the widespread use of inappropriate statistical tests (Ioannidis, 2005). Citation bias, the act of selective referencing studies of a particular outcome, is thought to lead to distortions of the evidence-base, whereby a group of scientists do not interact with, and in some cases are completely unaware of, rival perspectives and unsupportive data (Greenberg 2009, 2011). This can lead to polarisation, where different scientists draw from different sources of evidence and arrive at opposite conclusions (Trinquart et al. 2016). This form of controversy may be pervasive, but few philosophers or sociologists have analysed it. In the following chapter, I discuss such findings and their implication for our understanding of science.

3. Discussion

However, for now, we may say that there are at least two forms of scientific disagreement:

(i) Disagreements that stem from different interpretations of the same data.

(ii) Disagreements that stem from the use of different data.

These are ideal types, and it is likely that they overlap in interesting ways. The history of diet–heart research represents a particularly rich body of work in which to explore both of these forms of disagreement and their resolution. To understand the diet–heart debate, we need to understand both how scientists selected evidence and how they interpreted it.

From SSK, we might expect that the meaning of terms, findings, and theories are established through a process of persuasion and use in particular communities. Interpretations that spread and which are communally endorsed and used are, by SSK’s definition, what we mean by scientific knowledge (Bloor [1976]1991):

Instead of defining it as true belief - or perhaps, justified true belief - knowledge for the sociologist is whatever people take to be knowledge…Of course knowledge must be distinguished from mere belief. This can be done by reserving the word ‘knowledge’ for what
is collectively endorsed, leaving the individual and idiosyncratic to count as mere belief (p.5).

In light of findings regarding the prevalence of citation bias in the scientific literature, the selective use of the available evidence also requires attention. One way of studying how groups of scientists come to agree or disagree on an issue is to trace the propagation of findings and their interpretations within scientific communities, and the factors that lead to their endorsement or rejection.

To do this, we may, like Latour and Woolgar (1979), examine how scientists negotiate the meaning of results in particular laboratories and the process of authoring scientific publications. However, this would be only part of the story of how, over time, groups of scientist come to agree or disagree on an issue. And, as Latour and Woolgar stressed, the work conducted within the laboratory, and the publication of particular findings from it, was aimed to persuade those beyond that laboratory of the credibility of particular perspectives. Or, we might, following this observation, try to understand how particular findings and interpretations are proposed, propagated, and evaluated in the scientific literature. If scientists are in the business of persuading and being persuaded by others, then understanding this process of persuasion in the scientific literature is important. It is this latter form of analysis that I pursue in this thesis, and, in the following chapter, I argue that Citation Network Analysis is particular suited to this aim.
Chapter 2: Selective Citation and the Shaping of Scientific Knowledge

Introduction

The success of a scientific work depends, in part, on how it is communicated and received. In this chapter, I focus on the conventions of publication and citation, which, I argue, can be used to study how, over time, findings produced by particular individuals and groups spread amongst a wider community. By understanding the content of papers and particular references, it is possible to evaluate whether particular claims are supported by the unbiased or biased selection of evidence. Further, it is possible to understand how the diffusion of knowledge through the literature, as documented by citations, shapes that knowledge; whether particular findings become widely endorsed or rejected, whether they are simplified, whether they are distorted by misinterpretation, and whether certain interpretations of the meaning of particular findings become popular.

1. Science and the convention of citation

Reflecting on his achievements, Newton wrote “If I have seen further, it is by standing on the shoulders of giants” (1675 p.1). While the origin and meaning of the phrase are disputed (Merton 1965), it is now ingrained in our understanding of the nature of science. The popular interpretation suggests that Newton was paying homage to scientists who preceded him. It is an assertion of humility – a reflection that science is not conducted in isolation, that any scientific position is not solely the product of the individual but relies on the work of many. Further, it encapsulates the idea that science is an accumulative process; while many doubt that this is an unproblematic path to truth, few deny that science is a collective process of knowledge creation that relies, in some way, on the work of those who came before.
It is this reliance of scientists on the work of others that the convention of citation intends to capture – to make manifest the flow of information from one scientific paper to another. Any given paper will, of course, be influenced by a plethora of other papers and experiences that are not explicitly acknowledged by an author. For example, to understand all of the influences on this thesis would be impossible – I can’t account for how the entirety of my experience has shaped my work. As Bacon ([1605]2008) observed, “no man knoweth how he came to the knowledge he hath obtained” (p.234). Nevertheless, the references remain an important signifier of the works that I have read and believed important, and they provide a way of scrutinising the claims I have made.

Typically, the word ‘citation’ is reserved for describing the acknowledgement one paper receives from another, while ‘reference’ describes the acknowledgement one paper gives to another (Smith 1981). The convention of citation creates formal links between papers – links that go from the citing papers to the cited. A formal acknowledgement of the shoulders that one has, metaphorically speaking, stood upon.

This convention gives an opportunity to study several features of science. Citations have been subject to academic study since Gross and Gross’ (1927) measurement of the use of journals in particular disciplines by using citation counts to gauge which journals libraries should purchase. In the 1950s, Garfield (1955) developed this into a distinctive approach, known today as Scientometrics, that uses citation data to evaluate the quality and influence of particular scientific papers and journals. In 1960, he founded the Institute for Scientific Information (ISI), which developed the Science Citation Index – a bibliographic database of published scientific work, which includes a record of the citations between these works. Today, ISI’s Web of Science (WoS) holds more than 72 million records on the scientific literature, including articles, reviews, notes and letters, discussion and opinion pieces, revision and retractions, and meeting abstracts, and over a 1.4 billion references from these documents.
Theories of citation

As citation data has become readily available, the use of it to evaluate features of scientific knowledge has increased dramatically. Most work using citation data is aimed at evaluating research performance (Moed 2005). Advocates of evaluative citation analysis subscribe to a normative theory that posits that citations can be used to assess the influence/quality of particular papers. Following Merton’s ([1942]1973) norms, scientists are believed to publish all findings in the spirit of communism and to reference their publications in a manner that allows other researchers to find the relevant works that support particular claims. In their referencing, scientists are believed to uphold universal standards that require scientists to judge papers by some objective criteria, and not on the apparent authority of the scientists who produced that work. In the spirit of disinterestedness, a scientist’s judgement is neither shaped by personal ambition or subjective opinion – studies are not evaluated in such a way as to support treasured ideas or to advance particular interests. Finally, scientists, in their commitment to organised skepticism, subject claims to critical scrutiny.

Merton’s early views on scientific behaviour appear to have been shaped by the philosophy of Logical Positivism (Ayer 2001). This saw science as a progressive activity, building from empirically demonstrable “facts”, drawing inferences from them, and verifying those inferences by further empirical studies. Merton saw the institutional goal of science as extending “certified knowledge”:

The technical methods employed toward this end provide the relevant definition of knowledge: empirically confirmed and logically consistent statements of regularities…the technical norm of empirical evidence, adequate and reliable, is a prerequisite for sustained true prediction; the technical norm of logical consistency, a prerequisite for systematic and valid prediction ([1942]1973) p.270).

By this thinking, the practices of publishing and referencing spread awareness of findings and ensured their robustness. Findings were replicated, and, if
verified, were accepted as knowledge. For Merton, however, this system was also part of the scientific reward system. Scientists were said to provide though their references, “pellets of peer-recognition of the knowledge claim, accepted or expressly rejected, that was made in that source” (1988 p. 622). Thus, both publication and citation played a social function. But, by Merton’s norms, peer-recognition was supposed to be guided by the same standards used to evaluate the quality of work – and so sound science worked in synchrony with its social system. Merton came to see problems with this argument, but, for now, let us retain this simplified account of his views.

By this account, it might seem that citations can be used to judge the quality of scientific work. Only those works conducted to a high standard and which make important contributions to the literature should remain highly cited over time. This is because scientific peers ought to be able to judge the quality of scientific work reliably and should expose flaws or inaccuracies undetected by pre-publication peer review. Through an obligation to be balanced in writing a paper, a scientist ought to reference relevant discrepant data and explain why they are discounted. Citations to weak papers might fall steeply as scientists disregard them because of their perceived unreliability. However, initially, all relevant papers ought to receive a balanced coverage in the literature.

A variant of this view has also been proposed by those who rejected logical positivism. In Popper’s ([1945]2013) philosophy it was the publication of scientific work, and the critical scrutiny that it was subjected to, that imbued scientific knowledge with its objectivity. To illustrate, Popper asks us to imagine Robinson Crusoe, abandoned on a desert island with all the equipment and resources necessary to conduct a scientific investigation, and asks if anything Crusoe produces can be called science? According to Popper, it cannot:
For there is nobody but himself to check his results; nobody but himself to correct those prejudices which are the unavoidable consequence of his peculiar mental history; nobody to help him get rid of that strange blindness concerning inherent possibilities of our results which is a consequence of the fact that most of them research through comparatively irrelevant approaches. And concerning his scientific papers, it is only in attempts to explain his work to somebody who has not done it that he can acquire the discipline of clear and reasoned communication which too is part of the scientific method (p. 426).

For Popper, if scientists upheld a critical attitude inspired by an ideal of falsification – that claims once conjectured are subjected to the most severe scrutiny possible – then peer judgement ought to form a self-correcting and quality-enhancing system which, over time, detects errors and strips these from scientific thinking. Popper, in viewing humans as fallible and imperfect, recognised that this system would not work perfectly nor would it strip all errors, but that it would nevertheless capture many problems.

Most academics interested in citation data believe that citations indicate both influence and quality. However, many studies have shown that citation rates of papers are associated with many factors beyond methodological quality (Bornmann and Daniel 2008). Most studies reporting such findings have used large datasets in which particular variables, such as language of publication, author prestige, and gender, are correlated with the number of citations received. By this, it can be shown that many attributes are significantly related to citation rates (Tahamtan et al. 2016). Such studies generally assume that it is some property of the cited article that influences its citation rate – but is this a sound assumption?

Two broad classes of criticisms have underpinned skepticism of evaluative citation analysis. First, many have raised concerns that citation data is too incomplete to be used to understand the influence of a work. Edge (1979) highlighted that many influences on scientific work are not formally referenced in-text, and this undermined studies simply using the number of citations a paper receives to infer its impact. MacRoberts and MacRoberts (1989, 1996, 2018) have repeatedly argued that “the vast majority of influence is not cited.
Period.” (2018 p.478). Many ‘facts’ and ideas used within a text fail to be adequately referenced, and knowledge of these likely derive from a large variety of potentially unmeasurable sources, from informal discussion with colleagues and conferences to cultural influences on specific groups (Hick and Potter 1991; MacRoberts and MacRoberts 1997). Some highly influential papers experience ‘obliteration by incorporation’, a term used to describe the tendency of well-known ‘classic’ works being referred to in-text but without being formally referenced (Merton 1968b; McCain 2011). As will be discussed later, many references also appear to be erroneous and it is not uncommon for authors to misinterpret the studies they reference (Jergas and Baethge 2015) Recent studies also suggest that papers may even be referenced without having been read by an author (Simkin and Roychowdhury 2003)².

Second, there are a plethora of different functions that references can play within a text. Some references point to specific ideas, methods, or data in a cited paper, and a citing author might use a reference to support a particular claim, to critique or elaborate on a point, or to establish a background for their own study (Garfield 1962). As references have varied functions, citation counts that collapse these different roles into a single number can give a false impression of an article having a specific homogenous influence on the literature.

Since the 1970s, scientometricians have applied a form of content analysis – citation context analysis – to classify the role particular references play within a text in order to ascertain which citations ought to be counted as ‘important’ or ‘influential’ and which citations should be regarded as ‘redundant’ or ‘perfunctory’. While many studies have used quite different classification systems to guide their analysis, making generalisations from this body of work difficult, one consistent finding from such attempts has been that the ‘negative’ or ‘critiquing’ citations are seemingly rare in this scientific literature (Bornmann and Daniel 2008; Tahamtan and Bornmann 2019).

² Garfield (1955) originally conceived of a bibliometric index as having utility specifically in enabling researchers to “eliminate the uncritical citation of fraudulent, incomplete, or obsolete data by making it possible for the conscientious scholar to be aware of criticisms of earlier papers” (p.108). Citation indexing, however, does not appear to have addressed this problem.
Moravcsik and Murugesan (1975) examined 575 references from 30 papers on theoretical high energy physics published in *Physical Review* between 1968 and 1972. Only 14% of references were ‘negational’ – a term used to describe author’s directly challenging previous findings or theories. Spiegel-Rösing (1977), examining 2,309 references in 66 papers published in *Science Studies* between 1971 and 1974, found that the “the incidence of ‘negative’ or ‘critical’ citations is very small (0.8%)” (p.111). More recently, Lin *et al.* (2018), in a study of 15,875 references in 180 papers in the sociology, psychology, and economics found negative references rare – making up only 0.9%, 0.1%, and 0.6% of references in each respective sample. However, the findings from such studies are hard to interpret. As authors attempt to classify the role every reference plays in an article, these findings only tell us that ‘negative’ references make up a small proportion of most articles reference lists – the majority of articles are not devoted to attacking the works of others. Importantly, authors performing such studies have not taken their results indicating a lack of ‘negative’ references as suggesting that bias is occurring in the literature, and these results should not confused with findings related to citation bias. However, such findings do highlight that the motivation for citing previous works tends to be for supporting, elaborating, or contextualising an argument an author wishes to make or justifying the selection of particular methods. This view has been echoed by scientists, such as Goudsmit (1974), who claimed “The reason for citing a paper is *primarily* for possible support of the author’s contentions and only secondarily in recognition of previous work” (p. 28). If this is an accurate description of the motivation for referencing work, then this causes problems for the Popperian view that scientists form a rigirous error-detecting and self-correcting mechanism through their communications, which would depend on scientists dedicating time to critique.

---

3 Citation context analysis, as it is currently performed in the literature, cannot be used to establish citation bias – as authors might cite negative studies in a supportive manner. As will be discussed later, citation context analysis is quite unlike claim-specific citation network analysis (Greenberg 2009) or quotation analysis (Ravnskov 1995) beyond its focus on reference contexts. These begin their analysis by identifying a set of primary data trials, classifying them by their findings as supportive, negative, or inconclusive towards some hypothesis. Then authors search through the citing literature for references to these studies and classify these by whether the citing author is using that reference to support or challenge that hypothesis or remains neutral to it.
Constructivist understanding of citation

Social constructivist scholars have raised further problems with the way evaluative citation analysts view scientific writing and referencing. Constructivists view scientific texts not as objective technical accounts of experiments or findings, but as tools used by scientists to advance particular arguments and perspectives (Gilbert 1977; Latour and Fabbri [1977]2000; Latour and Woolgar 1979; Gilbert and Mulkay 1980, 1984; Yearley 1981; Callon et al. 1986; Bazerman 1988). Accordingly, referencing practices ought to be understood in their appropriate social and rhetorical context.

Gilbert (1977) rejected the idea that referencing behaviour was guided purely by considerations of the scientific quality of papers. Instead referencing should be understood as a rhetorical tool employed to persuade an audience of the importance and novelty of a particular work.

This involved referencing well-known papers in order to frame a work against a backdrop of literature familiar to those in a specific field and to highlight its unique contribution to that literature. To do this, scientists favoured referencing studies already well-known and liked in a community. This played two functions. First, it allowed an author to present an argument against a backdrop of works that were already widely endorsed; authors tried to “shine in their reflected glory” (p.116). Second, it played a social function in demonstrating to a particular community that an author knew of and respected the apparently important works of the field.

In doing so, however, scientists had to make educated guesses as to what the important and liked papers were, and to do this they scanned through recently published papers to find papers that were regularly cited in a positive manner. Gilbert suggested that it was through this mechanism that certain studies became recognised as ‘exemplars’4. However, he went on to suggest that reference lists rarely showed distinct signs that authors were using past papers to find relevant papers. Rather, because scientists had to present their

---

4 Gilbert references Kuhn directly for this concept. According to Gilbert, many highly cited papers were ‘regarded as 'exemplars' in the Kuhnian sense of the term” (p.122) by the communities that direct many citations to them.
work as novel, most reference list would introduce novel references – either unique methods, findings, or theories – that distinguished the novelty of a current work from existing scholarship in a field.

For Gilbert, these referencing behaviours not only impacted the reception of the authors work, but also impacted how the literature that was referenced was viewed by other scientists. By referencing some favourably, some critically, and many not at all, scientists pinned their allegiance to particular strands of the literature and communities advocating particular positions. According to Gilbert, this complex process was largely responsible for generating consensus in specific areas:

[I]n citing certain papers, the author can be seen to be making an assertion about his own opinion concerning the validity of the findings of the cited papers, and is thus contributing, albeit only in small measure, to the overall consensus of his research area, a consensus which is also continually being re-established through the choices of references in his fellow participants’ own publications. (p.117)

Advocates of evaluative citation analysis have often interpreted the constructivist position as claiming that non-scientific factors of articles, particularly author prestige, has a larger influence on citation rates than their methodological quality (Baldi 1998; White 2004). However, this interpretation, that scientists cite those in particular positions of authority to inflate their claims, has obscured important differences between the normative and constructivist positions. Where normative theorists assume that it some quality of the literature that determines how it is cited, the constructivists reject this as simplistic and stress that the role citations play in citing articles must also be considered. This is a disagreement, in part, over what should be the independent or dependent variable in analysis.

For Latour and Woolgar (1979), scientific activity is organised around publishing articles, scientists are professional “writers and readers in the business of being convinced and convincing others” (p.88), and scientific work is considered successful when it persuades others. Scientists gauge the success of their articles by tracing how other articles reference them. Scientists
carefully examine how articles are being referenced, and direct their efforts, both in the laboratory and in writing publications, to attempting to transform particular claims to higher or lower levels of credibility. On this account, scientists are advocates of particular perspectives, and, in their advocacy, they direct their work to persuading others of the merits of their favoured positions. For Latour (1987), it was through the act of referencing past literature that scientists built stories that could persuade their audiences:

Whatever the tactics, the general strategy is easy to grasp: do whatever you need to the former literature to render it as helpful as possible for the claims you are going to make. The rules are simple enough: weaken your enemies, paralyse those you cannot weaken, help your allies if they are attacked, ensure safe communications, with those who supply you with indisputable instruments, oblige your enemies to fight one another; if you are not sure of winning, be humble and understate (pp.37-38)

According to Latour (1987), this resulted in the scientific literature being replete with diverse interpretations that stemmed, in part, from the literature being:

[C]ited without being read, that is perfunctorily; or to support a claim which is exactly the opposite of what its author intended; or for technical details so minute that they escaped their author's attention (p.40)

As scientists are in the business of both persuading and being persuaded by others, exactly which works are influencing who becomes a complex question.

To elaborate, take a situation where an author, Bob, references a study, “A”, as implying “Y”. Now, assume study A contains findings that could be read in several different ways, but that some of its findings imply “Z” – a position that undermines the Y case. Bob publishes his paper, “B”, in which it is claimed that A is evidence for Y, but includes only those findings that help him in this, ignoring those that might suggest Z. Maybe he purposefully neglected them, or, perhaps, due to the quirks of human cognition, only paid attention to the findings that interested him. Now, Sandra reads Bob’s paper, a beautifully written and strongly supportive account of why we should accept Y, and she is convinced. Sandra now publishes a paper “C” echoing the position of Bob’s, but also repeating the claim that A implied Y. Perhaps Sandra didn’t read A.
Perhaps she did, but, following Bob’s argument, only skimmed A to check the finding that Bob had quoted. For Bob and Sandra, paper A means Y – and paper A is cited because these authors believe Y is true.

Now, in a separate community, the Z theorists have also read A and they take it as implying Z. Rachel, the originator of the Z theory, authors a paper in which she claims that A, in implying Z, is actually a refutation of Y when the results are considered in their totality. Rachel’s views spread amongst the Z community, and for this group A implies Z and refutes Y.

Simultaneously, Jim, a skeptic of the type of experiment performed in A, authors a paper claiming that A cannot be used as evidence for either Y or Z! Jim, however, because of his aggressive style, finds himself detached from both the Y and Z groups, who decided his views are not worth paying attention to. No one reads his paper. Jim, dejected with his low citation rate, leaves science.

Here, A’s citation rate is influenced by the beliefs of those who reference it, but also by persuasiveness of particular interpretations of A in other papers and the credibility of particular authors5.

For the constructivists, citation counts and terms like ‘influence’ obscure the complex reality of how scientist actually reference and the environment in which they do so. Citation counts can only be assumed to mean that some reference plays some rhetorical role in some argument.

According to Scientometricians, empirical tests of the constructivist position have largely failed to produce supportive evidence (Baldi 1998; White 2004). For White (2004) citation behaviour is “better explained by Merton’s norm of universalism, which holds that citers are rewarding use of relevant

---

5 For evidence that scientists may discount the views of scientists that have lost credibility or fallen out of favour in specific communities, see Collins’ account of the gradual exclusion of scientists proposing the existence of high-visibility gravitational waves from the main-stream of gravitational wave research (1999, 2000). The major proponents of the existence of high-visibility gravitational waves was a group run by Weber, and this group faced difficulties in getting their results taken seriously over time as consensus in gravitational wave community shifted against them. This included some journals refusing to publish certain claims, and Weber’s later articles being largely ignored by the core-group in the gravitational wave field. According to Collins, “Each of these articles presented a hugely important result if taken at face value, but each article was either ignored or attacked by the core-group. The 1996 article, as far as I can ascertain, has not been read by a single member of the core-group” (2000, p.828)
intellectual property, than by the constructivists' particularism, which holds that citers are trying to persuade through manipulative rhetoric” (p.93). White (2004) tested whether scholars tended to cite papers based on the author's citation reputation (the number of previous citations received), and presented results suggesting that most citations are to authors of “middling reputation”. White claimed this result undermined the core claim of the constructivists, that “that the main function of citation is to marshal big-name support for arguments at the expense of crediting lesser-known figures” (p.93). But this is not a fair test. It’s not even a good test of the core claim of the constructivist position – that citation rates are influenced as much by the intentions and beliefs of the citing authors than the content of the cited paper.

Constructivists, however, have been hesitant to use citation data. Edge (1979) argued that quantitative analysis of citations commits one to an overly positivist stance that fails to interact with the construction of knowledge claims in the “the soft underbelly of science” (1979 p.117). For normative theorists, constructivists cannot explain apparent ‘universalistic citation patterns’ (White 2004), which apparently cannot be explained by citations acting as a tool for persuasion.

So what are these universal patterns of citation? To understand this, we need to understand the work of Derek de Solla Price.
2. Derek de Solla Price on Growth, Inequality and Citation Networks

On growth of publication

Price was one of the first to use bibliometric data to study the dynamics of science. In 1951, he demonstrated that the number of papers published since 1918 in *Physics Abstracts* had grown exponentially – doubling about every 10.4 years (though stalling during World War II). Later, Price ([1963]1986) presented findings suggesting that scientific output, as measured by either the number of publications or journals published, had grown exponentially for three centuries – doubling every 10 to 15 years depending on which field was focussed upon (discounting the dips during the two World Wars). Extending this to the number of professional scientists, Price estimated they, like the papers they produced, were doubling at a rate of ~10–15 years.

These findings have been replicated and refined many times. Bornmann and Mutz (2015) observed three phases of growth in the number of publications since 1650. In the period between the dips of World War I and II to 2012, these authors estimated that the size of the scientific literature doubled every nine years. Fanelli and Larivière (2016) found that, despite this growth, the rate of publication of individual scientists has remained unchanged for over a century after adjusting for co-authorship. Thus, the growth in scientific publications appears to be explained primarily by the growth in the number of scientists. Sinatra *et al.* (2015), after analysing over 2.4 million papers in 242 physics journals, reported that the number of authors grew at the same rate as the number of papers between 1900 and 2000.

We may conjecture from these findings that the growth of any particular field depends on the number of scientists in that field, and hence on the appeal of that field for scientists, and the funding available for them to work in that field.

This might be best demonstrated by a little-known study that Price appeared unaware of. In 1935, Wilson and Fred (1935) examined publications related to nitrogen fixation by plants. They had compiled a list of over 2,000
relevant publications for a review of the literature. While the literature on this subject could be traced back to the fifth century B.C., regular publications on this topic did not begin until after 1886. The period before 1886 they named the 'embryonic period', which, from 1850–1884, experienced a steady rise in publication output in scholarly journals. Over this period, 90 papers were published that were generally polemical in nature and contained many conflicting findings. However, the field then witnessed a period of sustained research interest between 1886–1914 that appeared to be provoked by Hellriegel and Wilfarth’s experiments in 1886 showing that legumes assimilated nitrogen from the air facilitated by tubercles in their roots. During this period, the number of publications per year averaged about 30 papers. From 1915, despite a dip in productivity during World War I, the field hit “maturity” with the topic attracting interest from scientist who saw its importance for agriculture. Following the end of WWI, a surge in publications occurred as researchers turned to translating findings regarding nitrogen fixation to help increase agricultural output (Fig 1). Much of this growth was contributed by researchers in the United States, which had become politically interested in the idea of agricultural self-sufficiency.

Fig 1: Number of publications on nitrogen fixation by plants published per year per contributing country, 1884–1931. Source: Wilson and Fred (1935 p.249)
They concluded “the subject of nitrogen fixation by leguminous plants is so intimately connected with the business of living that the research in this field is rather sensitive to upheavals in the political and economical world” (p.249).

In a later work, Price (1965) examined how research into “N-Rays”, a conjectured form of radiation proposed by Blondlot in 1903, witnessed a surge in publications in the years immediately following its reported discovery. Publication output, however, had collapsed by 1906 after problems with Blondlot’s experimental procedure became widely known in the scientific community.

Thus, while the global scientific literature conforms to exponential growth, specific fields have their own unique patterns of growth that can be influenced by funding decisions, research interest, popular discoveries, and by apparent refutations.

**Inequality in authorship**

Price ([1963]1986) also described a very unequal distribution in the authorship of the literature. Particular fields appeared dominated by prolific authors that produced far more publications than other scientists. This was not a novel observation. Lotka (1926) had previously shown that the number of authors producing \( n \) papers in *Chemical Abstracts* between 1907 and 1916 was roughly proportional to \( 1/n^2 \). Price interpreted this as implying that half of the literature in a specific field was contributed by the square root of the total number of authors ([1963]1986). If there are 1000 scientists in a given field, about 32 of them are responsible for 50% of all publications.

Nicholls (1988) demonstrated this underestimated the number of authors publishing only a single document, which, in several datasets, represented over half of publications. However, a skewed distribution of author productivity is the norm, and typically translates to about 60% of papers being authored by 20% of authors (Gingras 2014). Recently, Ruiz-Castillo and Costas (2014), in a study examining 17.2 million unique authors and their record of publication

---

6 Lotka counted *only* the number of authors with surnames beginning with A or B due to the time-consuming nature of manual counting.
between 2003 and 2011, demonstrated that authorship was highly skewed, and remained highly skewed in a similar fashion when examining authorship in the context of thirty scientific fields, despite these having field-specific levels of productivity. According to their results, 5.9% of authors account for 35% of all publications, 18% account for 50%, and 68% of all authors publish only a single publication.

With these findings, Price conjectured that science, at least in terms of productivity, was more similar to other areas of culture, such as an economy or other professions, than many had appreciated.

**Inequality in citation**

Price (1965) began reflecting on what implications such a system might have for scientific knowledge and turned his attention to citations. He measured the average reference list size of papers of all papers published in 1961 and indexed in ISI’s Science Citation Index. There were about 15 references per paper, but these were not distributed normally. About 10% of papers contained no references, “normal research type papers” contributed about 50% of all references, while 1% of all publications with an average reference list size of 170 (reviews) contributed 12% of all references. He conjectured that if this average held across time and the literature maintained its current exponential growth, then we “would find that, on average, every scientific paper ever published is cited about once a year” (p.511).

To illustrate, he provided a hypothetical example of a stage of scientific development in which only 100 papers existed and seven new papers were added to this. If these seven papers had an average of 15 references, then 105 references would be directed back to the previous 100. If all papers had an equal chance of being cited, then it ought to be the case that these citations would distribute normally across this body of literature.

Price now tested this prediction by examining the distribution of references across their target papers. Price then established that citations to papers from his sample were very unevenly spread:

[I]n any given year, about 35 percent of all the existing papers are not cited at all, and another 49 percent are cited only once (n=1).
This leaves about 16 percent of the papers to be cited an average of about 3.2 times each. About 9 percent are cited twice; 3 percent, three times; 2 percent, four times; 1 percent, five times; and a remaining 1 percent, six times or more. For large \( n \), the number of papers cited appears to decrease as \( n^{2.5} \) or \( n^3 \) (1965, p. 511)

With this, he proposed that a power-law distribution\(^7\) with an exponent of between -2.5 and -3 approximatively fitted his data. Not only was there an unequal distribution of production of papers, but there was also very unequal use of those papers by later papers.

Price’s discovery of a skewed distribution in citation data has been repeatedly demonstrated. While there is a debate in the literature over what distribution best fits citation data, a right skewed long-tailed distribution has been observed in every area of science so far analysed (Seglen 1992; Albarrán et al. 2011; Brzezinski et al. 2015; Golosovsky 2017). While 60% of publications are typically authored by about 20% of authors, about 80% of citations are typically directed at 20% of all publications (Gingras 2014). Price’s finding of a power-law distribution in citation data are now regarded as the first evidence of a scale-free network. Three decades after Price’s publication, Barabási and Albert (1999) proposed the term ‘scale-free’ to describe a network in which the fraction of vertices having \( k \) connections to other vertices follows a power-law.

For Price (1965), this inequality seemed to suggest that the flow of knowledge through the literature was very uneven. There seemed to be a strong selection pressure that determined how just a few papers would be disproportionately cited.

**Research fronts**

Price (1965) noted that references were disproportionately directed at papers published within the last five years. Previously, he had found that the half-life of a paper’s citation rate was about 5 years from an analysis of papers published in Physical Review Letters, and he assumed this was likely a general rule across the scientific literature (Price [1963]1986).

\(^7\) The proportion \( P(n) \) of papers with \( n \) citations is proportional to \( n^{-\gamma} \) for some fixed power \( \gamma \).
He conjectured this “immediacy effect” meant that new papers favoured referencing recent research, but he found a seemingly random distribution of references to papers published beyond this. According to Price, this was not likely to be because papers randomly cited old literature but because only some of this old, ‘classic’ literature was retained and because certain reviews extensively catalogued the older literature. For Price, this was evidence that, in particular fields:

Most papers, through citations, are knit together rather tightly. The total research front of science has never, however, been a single row of knitting. It is, instead, divided by dropped stitches into quite small segments and strips...Such strips represent objectively defined subjects whose description may vary materially from year to year but which remain otherwise an intellectual whole. If one would work out the nature of such strips, it might lead to a method for delineating the topography of current scientific literature. With such a topography established, one could perhaps indicate the overlap and relative importance of journals and, indeed, of countries, authors, or individual papers by the place they occupied within the map, and by their degree of strategic centralness within a given strip (1965 p.515).

These “dropped stiches of knitting” descended backwards in time and related to specific areas of research. Previously, Price ([1963]1986) proposed that science develops not primarily through large disciplines, but through work on specific research questions and topics. These were addressed by small communities of scientists that represented an ‘invisible college’ – a group not necessarily bound by a shared disciplinary tradition but by shared research interests who closely corresponded with one another. For Price, “We publish for the small group, forcing the pace as fast as it will go in a process that will force it harder yet. Only secondarily, with the inertia born of tradition, do we publish for the world at large” (p.80). It was this small community that reviewed one another’s papers, organised conferences, established specialised journals, and it was here that researchers gained their prestige.

For Price, “dropped stiches” reflected the development of a research area over time through their publications and referencing behaviour. Price ends the

8 See Crane (1969) for a test of Price’s ‘invisible college’ hypothesis, and Crane (1972) for a book-length elaboration and exploration of this hypothesis.
paper by envisioning the totality of scientific literature as a complex citation network clustered into expert communities that are collected around a core literature that expands over time – with highly-cited papers being pulled into the centre of segments of these “knitted strips”. Finally, he imagines these “knitted strips” to have a low level of cross citation and from this the “cloth of science is woven”.

Thus, Price suggested that the citations to papers primarily derived from authors working on similar research questions. Highly-cited papers were those that played some important role in advancing research for a specific community. As a research area develops, it relies primarily on the most recent publications, and only those of established importance are retained in the new wave of scholarship. The research front, by retaining a partial selection of older literature, meant that only some publications survived in current scholarship, at least as recorded by citations to them. However, he highlighted that a research front, in focusing only on certain questions and studies, might miss works of importance:

It is impossible to say how much of this loss is deserved and just, but a large body of jilted authors will feel that it is not. There are cautionary tales of rediscovered papers, like that of Mendel, to make us feel that the statistical loss of literature must be minimized ([1963]1986 p.73)

Today, the term ‘Sleeping Beauty’ is used to describe a paper that makes its biggest citation impact many years after its publication (van Raan 2004). While these are apparently rare, an explanation for their existence is that a research front, in holding particular assumptions of what is important, misses works that doesn’t fit with current understanding or interest.

Recent research has also reinforced the idea that citation rates are heavily influenced by a temporal dimension (Larivière et al. 2008), and techniques based on clustering documents together based on their citations or references relations, such as bibliographic coupling (Kessler 1963), co-citation analysis (Small 1973), or modularity maximisation (Newman 2006), have been shown to be an effective in capturing documents that share similar research topics (Boyack and Klavans 2010, 2017).
Cumulative advantage

In 1976, Price (1976) wanted to understand the rules that governed the inequality that he observed in citation data. This inequality had began to attract the attention of others.

Merton’s (1968) Matthew Effect proposed scientific success and failure can be summarised from a passage in the bible “For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath” (p.58). The Matthew Effect is a complex phenomenon initiated by some past achievement that provoked a cascade effect that reinforced and heightened a scientist’s authority in a particular field. While he examined the Matthew Effect in different spheres, from the career trajectory of scientists to the allocation of resources, his discussion of how this principle may play out in the scientific communication system touched on ideas relevant to the construction of knowledge.

Merton conjectured that the Matthew Effect “may serve to heighten the visibility of contributions to science by scientists of acknowledged standing and to reduce the visibility of contributors by authors who are less well known” (p.62). For Merton, it seemed plausible that scientists, in an attempt to cope with the mass of publications that confronted them, preferentially read works authored by established authorities and ignored those not recognisable. Merton framed an interesting problem:

[F]or science to be advanced, it is not enough that fruitful ideas be originated or new experiments developed or new problems formulated or new methods instituted. The innovations must be effectively communicated to others. That, after all, is what we mean by a contribution to science – something given to the common fund of knowledge. In the end, then, science is a socially shared and socially validated body of knowledge. For the development of science, only work that is effectively perceived and utilized by other scientists, then and there, matters…it is therefore important to consider the social mechanisms that curb or facilitate the incorporation of would-be contributions (pp.59–60)

For Merton, an established authority might have disproportionate influence on what is considered knowledge in a particular field. If what is held as knowledge depends more on the papers that are seen, than on the papers that are there
to be seen, then this has interesting consequences for how we understand the
dynamics of knowledge construction and evaluation.

Price (1976) translated Merton’s Matthew Effect into a model of citation
behaviour based on Simon’s (1955) adaptation of the Urn Model to understand
the stochastic processes involved in heavily skewed distributions:

In general, the model supposes that fate has in storage an urn
containing red and black balls; at regular intervals a ball is drawn at
random, a red ball signifying a “success” and a black ball a “failure.”
If the composition of the urn remained fixed the chances of success
and failure would not vary, but if at each drawing the composition is
changed by some rule, the chances will change as an aftereffect of
the previous history.

For Price, it made more sense that success attracted more success,
rather than the unsuccessful being punished. Therefore,

[A]fter each drawing the ball is replaced; if a red is drawn then c red
balls are added, but if a black is drawn no extra balls are put in the
urn. If we start with b black balls and r red balls, the conditional probability
of success after n previous successes will be \((r+nc)/(b+rtnc)\) and
the corresponding conditional probability of failure will be \(b/(b+r+nc)\)
(p.293)

By this account, new papers preferentially reference papers already well-cited.
Price demonstrated that models constructed under these assumptions fit well
to empirical data on citation distributions. He took this as evidence that
citations dynamics could be explained by two simple functions – cumulative
advantage (number of existing citations) and population growth (number of
papers published) – though he highlighted that that time ought to also play
some role in this equation in decaying the probability of being cited.

The ability to quickly see how many citations an article has received is a
recent development provoked by digital citation indexing. While author
cognition may explain this inequality, it is also feasible that highly cited
papers simply are seen by more people and, as a consequence, are used more
often. While the former suggest scientists base their literature selection, in part,
on finding papers written by scientists they recognise, the latter does not
necessarily imply any pre-existing knowledge of the author.
Three decades later, Barabási and Albert (1999), proposed the Barabási-Albert model (BA Model) to explain the stochastic processes that lead to the scale-free property. They developed a model similar to Price’s model of Cumulative Advantage; calling this process instead Preferential Attachment. Under this model, the degree \( k \) distribution obeys:

\[
P(k) \sim k^{-\gamma}
\]

where \( \gamma \) is an exponent typically in the region of between -2 to -3 in citation data (Albert and Barabási 2002). Price’s exponent sat between the values of -2.5 and -3 (Price 1965).

**Implications**

From Price, we might expect:

(i) A research field is composed of a group of scientists pursuing particular research questions.

(ii) A field’s publication output will be related to funding decisions, the number of professional scientists working on similar problems, and the beliefs of scientists regarding the promise of that field.

(iii) The published output will be dominated by a few highly productive scientists.

(iv) Scientists will mainly reference scientists from the same field.

(v) The citations to a field’s literature will reflect how many publications are published in that field – a large literature will produce more references than a smaller literature.

(vi) These fields will be pushed through time by a dynamic research front that preferentially references recent scholarship and applies selection pressures through their references.

(vii) This pressure will result in a heavily right-skewed citation distribution that can be approximated by a power-law. The citations will be highly

---

\( ^9 \) Neither referencing or mentioning Price, it is likely that they came to this model independently. Newman (2003) states “The work of Price himself, however, is largely unknown in the scientific community, and cumulative advantage did not achieve currency…until its rediscovery some decades later by Barabási and Albert (1999)” (p.215).
uneven – a few papers will gain a disproportionate share of the total citations.

(viii) The previous referencing of particular articles will influence, in some way, the material that future publications reference.

There is nothing in the above claims that can be used to dismiss the constructivist position – indeed, they hint at the importance of researchers sharing similar research interests and assumptions, and this may be developed through a process of persuasion. Nor are they evidence against the normative theory. The inequality in citation might be an outcome of the differing quality of publications, the rise in publication output might follow important discoveries and the process of their validation and expansion, and this might be explained by scientists obeying norms.

Price’s findings suggest, however, that, to understand how information spreads across the scientific literature, we need to understand how citation networks develop over time. We need to study systems.

But is there any evidence that the way information circulates influences the content of knowledge? Are constructivists right that different interpretations will arise and spread in particular receptive communities? Are normative theorists correct that the references are guided by Merton’s norms? Is error really detected and stripped from scientific thinking? Or is error introduced and amplified?

In theory, you could use citations to trace the evolution and acceptance of particular facts, methods, or theories from their first public articulation to any point in time.

…In 2009, Steven Greenberg did just this.
3. Webs of Belief: Claim-Specific Citation Network Analysis

In 2009, Greenberg, a neurologist at Brigham and Women’s Hospital, became interested in whether a claim that he had seen regularly in scientific papers was actually supported by the available evidence. It was widely ‘known’ that a particular protein, β-amyloid, was abnormally and specifically present in the muscle fibres of patients with a particular disease: inclusion body myositis (IBM). According to Greenberg (2009), he had seen this claim in at least 200 papers; papers that gave the impression that this was a ‘fact’.

Greenberg found 12 primary research papers that had investigated this claim directly; six supported it but six were unsupportive. By Greenberg’s assessment, when considered together, the collected evidence was unsupportive of the β-amyloid hypothesis: there were technical weaknesses in the supportive papers, “most notably a lack of quantitative data as to how many affected muscle fibres were seen and a lack of specificity of reagents” (2009, p.3).

Greenberg wanted to understand why this claim had been taken as a ‘fact’. He looked at how the first eleven primary studies, all published between 1992 and 1995, were cited in the years until 2007 – six unsupportive and five supportive studies. He found 242 papers that made specific claims about the presence of β-amyloid in patients with IBM. He recorded all 675 citations from one paper to another in this set to understand the interactions between them. He then classified every citation as being either: (i) supportive of the claim; (ii) neutral; or (iii) critical of it, in a claim-specific citation network analysis (CS-CNA).

From this, he constructed a citation network. It looked like a typical citation network with a power-law distribution of citation \(P(k)\sim k^{-2.4}\). A total of 214 citations were directed to the early primary studies. Of these, 94% were to four supportive studies, while only 6% were to the six unsupportive studies. These four primary studies were produced by the same laboratory, and all ranked highly in the top ten mostly highly cited publications in this network.
Other highly cited papers included five animal models of the disease and one review paper. These were all dependent on the findings from data derived from the four supportive primary studies. All of the top cited literature expressed the view that the β-amyloid claim was true.

Greenberg demonstrated that review papers played a major role in directing scientists to evidence; in this instance, knowledge about β-amyloid spread from four reviews written by the same group of scientists that had published the four supportive studies: 95% of all citations to the original primary data went through these papers. To understand this, Greenberg counted the number of citation-paths that traversed through particular papers. For example, if paper A cites paper B which cites paper C, then there is a path from A to C via B. Importantly, these reviews only cited the four supportive primary research papers – funnelling scientific attention to studies that supported the β-amyloid hypothesis.

Between 1996 and 2007 support for the claim grew exponentially, with the number of supportive citations and citation paths increasing sevenfold and 777-fold, to 636 citations and 220,553 citation paths. In contrast, the critical view grew to only 21 citations and 28 citation paths. No papers refuted or critiqued the critical data, but instead the data were just ignored (p.4).

Greenberg suggested the broader scientific community believed the hypothesis because they only saw supportive evidence. He went further. By examining how the evidence was cited by later papers, he found inconvenient findings were often misread and distorted, making it appear as though they supported the hypothesis.

One primary data paper reported no β amyloid precursor protein or β amyloid in three of five patients with inclusion body myositis and its presence in only a “few fibres” in the remaining two patients. Three papers cited these data reporting that they “confirmed” the claim. Whether such data confirm the claim is perhaps open to interpretation (pp.3–4).

Greenberg also found evidence that claims that began as conjectures were being converted into “facts” in the literature by a process of citation alone. One claim, repeated in 24 papers, was that the accumulation of β-amyloid in muscle fibers precedes other abnormalities. It had begun life as a hypothesis,
but then some papers claimed that it was probably true, and eventually papers were stating it as fact. But no data directly relevant to this had been published; papers were citing other papers that simply asserted it. Through chains of citations linking papers containing no relevant data this claim became a “fact” in the literature.

Thus, the beliefs of these scientists were guided by reviews that ignored unsupportive evidence. Belief in the claim was further reinforced by papers that erroneously interpreted the evidence. This demonstrates how a small group of scientists and papers can have a massive effect on the trajectory of scientific research and knowledge by establishing ‘cascades of information’ through the literature, where scientists mirror the evidence selected and interpretations articulated in previous papers. The ‘shape’ of the literature, as connected through citations, affects both what evidence is used and how it is interpreted.

For Greenberg, a form of “social citation” would be needed to explain his results, referencing practices that deviate from reflecting an objective assessment of the strengths and weaknesses of evidence. He conjectured three likely causes.

First, the four highly cited supportive trials were authored by the same laboratory, and the four reviews that directed many later citations to these trials were also written by this group. These reviews failed to cite unsupportive trials, but this was not because they were unaware of them. They were aware of at least two of the unsupportive trials because they had authored these, but had nevertheless failed to cite them. Possibly they were influenced by confirmation bias; the supportive trials conformed to their expectations and they dismissed findings that didn’t fit. When later scientists saw only supportive evidence cited by reviews and by other papers, they didn’t think to question whether unsupportive evidence existed.

Second, his results might reflect that scientists tend to present strongly supportive cases for particular ideas to win funds and attract interest. For Greenberg, his results could be explained if scientists had a general bias against results that undermined prevailing understanding. Scientists, in trying
to present their work as worth pursuing, tend to reference supportive findings more frequently:

Unlike “positive results” there is nothing exciting to be repeatedly written about how something was not found in an experiment...Within a single paper readers generally view new claims as false until proved true through convincing methods and results. Across a network of papers, however, the barrier to the propagation of negative results biases claims as being viewed as true until proved false. (p.7)

Finally, Greenberg suggests that scientists seem to have either been directed to papers by reading other papers or had copied their interpretations of previous studies without reading the original papers. This would explain how some errors spread and how a conjecture, backed by no experimental data, was converted into a fact through citations alone.

For those of a Popperian persuasion, this study represents a refutation of the normative theory of citation. Merton's norms cannot be invoked to explain these findings. The architecture of the literature appeared to influence what scientists cited, as too did biases and interests. However, to date only Greenberg has applied claim-specific methodology to trace the trajectory of a particular claim in the literature, but studies with subtly different methods have suggested that similar problems exist in other areas.

**Polarisation**

Trinquart *et al.* (2016) analysed the controversy over dietary sodium reduction – a debate over whether public health guidelines ought to continue to advise the public to consume no more than 2 grams of salt per day. For advocates, salt reduction guidelines play a vital role in lowering the prevalence of cardiovascular diseases (CVD) in populations via reducing blood pressure. Critics, however, argue that the relationship between salt intake and clinical outcomes conforms to a U or J shaped distribution. For critics, too little attention has been paid to the harms associated with low salt consumption. Puzzled by the persistence of strong disagreement, Trinquart *et al.* wanted to understand the dynamics of evidence selection in this controversy.
The authors identified a set of 269 documents that directly examined the link between dietary sodium and either all-cause mortality or CVD mortality. They then classified each document by whether the major conclusion of each study were supportive, inconclusive, or unsupportive toward the hypothesis that population salt reduction policies result in a reduction of either CVD or all-cause mortality. By their analysis, 54% of documents were supportive, 33% unsupportive, and 13% were inconclusive towards this hypothesis.

To test whether documents were more likely to reference other documents that shared the same position, the authors constructed a citation network and applied a multivariate exponential random graph model. While the distribution of citation to supportive, inconclusive, and unsupportive studies was heavily-skewed and was similar in both their range of citations and mean citations, documents were 1.51 (95% CI 1.38-1.65) times more likely to reference a document that shared the same position.

The authors then constructed a co-authorship network to understand how authors were collaborating in this area. First, they showed that authors tended not to change their position in regards to this hypothesis by counting the number of papers of particular classification associated with each author. Second, they showed that both the supportive and unsupportive camps were dominated by prolific authors – about 75% of supportive studies were authored by 25% of all authors contributing supportive papers, 78% of unsupportive studies were by 28% of contributing authors contributing unsupportive papers. For Trinquart et al. this meant that both the supportive and unsupportive perspectives were dominated by a few prolific voices. They then established that authors of supportive and unsupportive documents very rarely collaborated with authors who held the contrary position.

Thus, according to Trinquart et al. different camps were not only citing different studies, but scientists were also not collaborating with those who held different perspectives:

We found that the published literature bears little imprint of an ongoing controversy, but rather contains two almost distinct and disparate lines of scholarship, one supporting and one contradicting the hypothesis that salt reduction in populations will improve clinical outcomes. (Trinquart et al. 2016 p.7)
According to these authors, the salt controversy is comprised of two different groups of scientists arguing past each other because they draw from different sources of evidences: they exist in different, largely isolated, worlds.

Citation Distortions: Citation Bias, Amplification, and Invention

What is happening here? One feature of modern science is its vast literature. In many cases, all the literature relevant to a research question will be too large for a paper to reference adequately, and too time consuming for authors to evaluate. As a result, papers aim to provide a representative sample of the literature – an overview that establishes the major supportive and unsupportive evidence to a particular research hypothesis or theory when evaluating validity. However, currently, this sampling is unregulated by formal conventions except in the cases of ‘systematic reviews’ of the literature. In the two examples discussed, marked citation bias was apparent, where a paper disproportionately draws from literature that supports a certain position while systematically ignoring evidence that does not.

Citation bias

Citation bias is a well-documented phenomenon. Ravnskov (1992) demonstrated that trials reporting statistically significant findings that supported the use of cholesterol lowering-drugs or diet in the management of CHD were cited almost six times more than trials reporting non-significant, equivocal, or unsupportive findings. Jannot et al. (2013) examined 242 meta-analyses published in the Cochrane Database of Systematic Reviews between January and March 2010. Of these, they identified 89 research focusses from cardiovascular disease, infectious disease, and psychiatry amongst others. These 242 reviews referenced 470 unique trials – with statistically significant results accumulating “more than twice as many citations when the reported result was significant” (p. 298). In 2017, Duyx et al. (2017) performed a meta-analysis of fifty-two studies of citation bias – thirty-eight focussing on bias in the biomedical literature, seven in the social sciences, six in the natural
sciences, and one with multiple focusses. Of these, twenty-nine reported a “clear effect of outcome on citation”, while eleven reported no effect of study outcome on citation rate, and twelve had mixed findings. The authors reported:

Our meta-analyses show that positive articles are cited about two times more often than negative ones. Our results suggest that citations are mostly based on the conclusion that authors draw rather than the underlying data. (p.97)

What Greenberg (2009) succeeded in doing was to demonstrate that these biases can lead to long-term distortions in the literature – focussing the attention of later papers disproportionately on a segment of the literature supporting a particular position. The decisions to cite only certain papers have consequences that ripple through the literature.

In a citation network two types of papers have considerable influence; authorities and hubs. Authority in CNA is the “equivalent of social consensus or belief regarding a claim; the most authoritative papers in a network have been elected through citation ‘votes’ as stating the accepted belief” (2011 p. 390), and refers specifically to highly cited publications. According to Greenberg, if authorities in a network are papers that do not directly report experimental findings of primary studies, then amplification is likely to be present in that network. Amplification describes the propagation of a scientific claim in a network by citations to papers lacking primary data. Hubs are rather different, although a hub can also be an authority, and refer to papers that many citation-paths flow through – funnels that direct citation traffic to particular parts of the literature.

To detect citation bias, one examines the number of references to a set of trials. For example, if paper A references only two primary studies B and C (out of four - B, C, D, E – with B and C supportive and D and E unsupportive) then the citing paper is guilty of citation bias. Examining the role of citation-paths can illuminate the underlying dynamics. For example, if paper A references Y, and Y only references B and C, then we would say paper A could have been directed to B and C through Y.
As discussed previously, citation bias and amplification are sources of citation distortion; however, Greenberg identifies a third category of distortion – *Invention*, which is composed of sub-categories.

(i) Citation diversion refers to erroneous references that distort the meaning of a study’s findings – e.g., if paper A references B as supporting claim Z when in fact B’s conclusion contradicted it.

(ii) Citation transmutation refers to the conversion of a hypothesis into a fact through citation alone – e.g., if paper A references B’s assertion that Z is true but fails to recognise a lack of supportive empirical evidence.

(iii) *Back-door to knowledge* refers to referencing a conference or meeting abstract and misrepresenting these as full-length papers, which has the effect of introducing literature that has no published methods or data and that hasn’t been subject to peer-review.

(iv) *Dead-end citation* refers to a situation where a paper is referenced in support of a claim despite containing no relevant material to that claim.

While less frequently studied than citation bias, evidence is emerging that these problems are present in the scientific literature, sometimes with disastrous effects.

**Research underutilisation**

We might expect that when clinical trials are published then they ought to receive considerable scrutiny and attention because of their expense, ethical issues, and health implications. Robinson and Goodman (2011) looked at how clinical trials were referenced in later, similar trials. They identified 1,523 trials and tracked how these had cited others on the same topic. Only about a quarter of the relevant trials were cited, which also constituted about a quarter of the subjects enrolled in relevant trials. A median of two trials were cited regardless of how many had actually been conducted. The authors concluded, “Potential implications include ethically unjustifiable trials, wasted resources, incorrect conclusions, and unnecessary risks for trial participants” (p.50).
Citation diversion

In 1980, Porter and Jick (1980) published a five-sentence letter, ‘Addiction Rare in Patients Treated With Narcotics’, in the *New England Journal of Medicine*, after sifting through the records of 11,882 patients who had been prescribed at least one opiate. To their surprise, they found only four cases of addiction in these records:

> [A]ddiction was considered major in only one instance. The drugs implicated were meperidine in two patients, Percodan in one, and hydromorphone in one. We conclude that despite widespread use of narcotic drugs in hospitals, the development of addiction is rare in medical patients with no history of addiction (p.123).

A potentially important observation. As a letter, rather than a full peer-reviewed publication, it should have been regarded as an example of hypothesis-generating work, rather than hypothesis testing, motivating others to explore the link between narcotics and addiction.

Three decades later, in another letter published in the *New England Journal of Medicine*, Leung et al. (2017) found that Porter and Jick’s letter had been cited in 608 papers between 1981 and 2017, and they read each of these papers to see how this letter was cited. In total, 439 papers (72%) had used this letter as evidence that, in patients treated with opioids, addiction is rare. Importantly, 491 (81%) of the citing papers failed to report that the original letter had described the experience of patients that had been hospitalised during their prescription; patients in a well-controlled, safe setting under supervision. The authors conclude:

> A five-sentence letter published in the Journal in 1980 was heavily and uncritically cited as evidence that addiction was rare with long-term opioid therapy. We believe that this citation pattern contributed to the North American opioid crisis by helping to shape a narrative that allayed prescribers’ concerns about the risk of addiction associated with long-term opioid therapy (p.2195).

In 2010, Stang (2010) published a critique of the Newcastle–Ottawa scale (NOS) (Wells et al. 2000), a scale used in modern meta-analyses to assess the quality of observational studies. In this article, Stang came to an unequivocally critical conclusion:
Wells et al. provide a quality score that has unknown validity at best, or that includes quality items that are even invalid. The current version appears to be unacceptable for the quality ranking of both case-control studies and cohort studies in meta-analyses. The use of this score in evidence-based reviews and meta-analyses may produce highly arbitrary results (2010 p.6).

Yet, eight years after publication, Stang et al. (2018) published a second paper calling attention to its widespread inaccurate citation. In the intervening years, Stang (2010) had become the ‘go to’ reference for scientists seeking something to reference when they used NOS in their publications – accumulating, as of 2016, 1,250 citations according to WoS. Stang et al. (2018) noted that few papers cited the original publication accurately; most were referencing it as though it supported the use of NOS or even developed it.

To understand this, the authors read citing systematic reviews because, as systematic reviews, their evidence selection ought to be more rigorous than other forms of studies. However, in 94 of the 96 systematic reviews they identified, Stang (2010) was cited in a manner that suggested it was supportive of NOS – with only 18 of these providing an additional reference apparently supportive of its use. Stang et al. (2018) then demonstrated that many of the citing papers cited other papers that also cited Stang (2010), and the authors conjecture that, through chains of citation, this study took on a meaning opposite to its original content. Stang et al. (2018) concluded “…the vast majority of indirect quotations of the commentary have been misleading. It appears that authors of systematic reviews who quote the commentary most likely did not read it” (p.1030).

By May 2019, Stang (2010) had been cited 3,282 times according to WoS. In the most recent published paper citing this study, a meta-analysis conducted by Abdeahad et al. (2019), Stang was cited as follows:

The Newcastle-Ottawa Score quality assessment system adopted for case-control studies was utilized to examine the methodological quality of studies (Stang, 2010). Included studies were scored according to the three broad perspectives: selection, comparability, and exposure (p.10927)
Stang et al.’s (2018) attempt to combat the spread of this erroneous interpretation appears to have fallen on deaf ears – since 2018, Stang et al.’s analysis of this misquotation has been cited only once (as of May 2019).

Reference errors and dead-end citation

Jergas and Baethge (2015) systematically reviewed 27 studies reporting findings on the accuracy of references in the biomedical literature. The authors separated all results into two categories – (i) major errors that seriously misrepresented or had no resemblance to the referenced paper; (ii) minor errors that contained factual inaccuracies. The smallest reported rate of major error in these 27 studies was 2.2% and the largest 55%, with a median of 11.7%, while for minor errors the smallest was 1.4% to 83%, with a median of 22.5%. Overall, the rate of reported error ranged from 6.7% to 83%, with a median of 22%. The authors concluded that this meant that about one in four references are “wrong or problematic”, while one in eight or nine references are “seriously incorrect”.

We might assume that a paper has to exist to be referenced. At least this seemed like a sound assumption, until Kroonenberg and Harzing (2017) discovered a non-existent “phantom” paper that had been cited more than 400 times:


The phantom reference was part of a style guide used by Elsevier to illustrate how to reference in particular journals – it was a hypothetical example. The authors attributed the accumulation of large numbers of citations to this phantom to a simple error on the behalf of researchers using a styling template to produce their reference lists.

I decided to follow this up to see whether this phantom reference has spread to support any particular claims in the literature. By 2019, it had cited more 480 times, most citations coming from conference abstracts, but 79 derived from journal papers. Of these, 13 papers were connected together
through references, and in all 12 accessible papers\(^\text{10}\), this reference was being used to support the claim that a compound, rutin, could dilute the blood, reduce capillary permeability and lower blood pressure (see Chapter 2 Supplement for full data; Ch2.S1–3).

In Fig 2, after downloading all citing articles and their references and constructing a network from this, I positioned each citing paper (circles) vertically by year of publication and horizontally based on their citation relations, with arrows pointing from the citing to the cited paper. Only papers citing the phantom reference and their references to one another are retained. The most cited documented is by Sun et al. (2008), who likely made the style-sheet error described previously. However, the papers published after this repeated the claim Sun et al. had made and used the same reference.

---

\(^{10}\) One paper could not be accessed due to paywall restrictions.
It seems that the likely explanation for this is that papers copied references and claims from previous papers. Simkin and Roychowdhury (2003) analysed the number of repeated errors in the references to papers (e.g. author name, page number, publication date etc.). Errors in a reference happen from time to time, and this is only to be expected. However, the repetition of the same errors in papers by different authors, suggests that authors might be copying references from other papers. By examining their propagation, and fitting a model in which authors copied references from other papers at some uniform rate, Simkin and Roychowdhury estimated that “only about 20% of citers read the original” (p.1). In a later publication, these authors proposed a model, known as the random-citing scientists model, based on a simple rule – “When a scientist writes a manuscript, he picks several random recent papers, cites them, and also copies some of their references” (2007 p.1661). By this, they show that the model’s predicted citation distribution fit citation distributions well. Furthermore, their model could account for the decaying influence of time on a paper’s citation impact, which the cumulative advantage model failed to account for. As with any model, however, just because it can be fitted to empirical data does not necessarily mean that it is an accurate description of underlying generative processes.

4. Main Path Analysis

Do citations then tell us anything important about scientific knowledge beyond some quirky stories of selectivity in evidence and the spread of particular errors?

In 1964, Garfield et al. (1964) wanted to demonstrate that citation data could be used to write a history of the development of scientific ideas. They used historiographs – similar to the network in Fig 2 – to understand the history of the discovery of DNA. They conjectured that, if “new discovery is dependent upon earlier discoveries” (p.iii) then analysing the citations between papers on
a specific research topic should identify previous discoveries that contributed to a particular development.

From Asimov’s (1963) history of DNA they identified 65 scientific documents referred to, 40 of which were deemed major contributions. They aligned the documents temporally and according to the narrative developed by Asimov that linked these documents by 43 conjectured relationships. They then compared this to a historiograph analysis of these 40 publications that analysed the citations between these documents. Most cited at least one other paper in this set. Second, 65% of the “historical dependencies” identified by Asimov were confirmed by the presence of citations between documents. Third, there existed a further 31 citation connections between these that were not captured in Asimov’s analysis.

Unquestionably, bibliographic citation data, if presented in the form of network diagrams…reveal historical dependencies which can be easily overlooked by the historian. On the other hand, citation standards are not always sufficiently rigorous to eliminate the need for human memory and evaluation (p.iv)

A bold position; but one that did not attract a large amount of research interest – the paper was hard to follow, the diagrams presented hard to interpret. Brush (1965) raised the following concern:

[I]f the citation technique is to contribute anything new to the study of areas of science which have already been described by historians, it is just these articles, located by the computer in a completely mechanical way, that are of interest. The authors do not have much to say about this, nor have they provided a very exacting test of their method in this respect by choosing Asimov’s book for comparison (p.487)

In 1989, Hummon and Doreian (1989) saw something in Garfield’s attempt that others had missed – citation networks may contain a structure that directs papers to core publications by virtue of the citation-paths historically constructed. They re-analysed Garfield’s data by a depth-first search that sorts papers topologically so that each paper can be positioned temporally as following from another. Second, they noted that, while citations drew attention to direct connections, another important property of networks is the existence of paths – chains of citation that link together sequences of papers. They
demonstrated that the longest paths contained 10 citation-links; two of which emanated from the first document published. For Hummon and Doreian, these properties could be used to understand the development of the main line of scholarship in particular research areas.

To do this, they proposed Traversal Counting – a method that counts the number of times a citation-link would be crossed when traversing a network when following the references from one paper to another in a given network. By this technique, they demonstrated that the important texts identified by Asimov and Garfield, and their relationship to one another, could be discovered by establishing the papers residing on the path with the highest transversal count – a path that, structurally, a reader following bibliographies from any randomly chosen paper in this set would most likely be led on to, “Connectivity in this citation network converges to the main stream of this literature” (p.54)

While they had demonstrated that the structural properties of networks might help identify the key papers and routes through which research develops, their paper attracted little interest. They had also used Asimov’s references, failing to address the challenge of Brush (1965) who believed the only utility that such a method could have was to aid researchers in identifying literature. But here complexities lay. Bibliographic databases organised records not by the specific hypotheses or theories, but by broad disciplines and sub-disciplines.

To address this problem of selection, Hummon and Carley (1993) analysed the development of Social Network Analysis (SNA) by analysing all papers published in 12 issues published between 1968 and 1990 in Social Networks, the field’s flagship journal. They wanted to demonstrate that Main Path Analysis (MPA) did not depend on using previous histories of a discipline, but also wanted to examine whether SNA research conformed to Kuhn’s ideas about the structure of normal science – i.e. a group of scholars publishing scientific papers in journals that communicate primarily with one another and coalesce around a core journals and papers.
The 227 articles published over this period contained 5,573 references. The 227 articles were authored by 194 unique authors, but 142 authored or co-authored only a single article. Only 30 authors published more than two articles, and only 11 authors published five or more papers.

These papers directed their references to many journals, but five stood out as attracting the most citations. Of these, 646 (11.6%) were to other papers in *Soc. Netw.*, and 254 to *Am. J. Sociol.*, 149 to *Am. Sociol. Rev.*, 134 to *J. Math. Sociol.*, and 132 to *Psychometrika*. Thus, scientists depended on a core-set of journals.

The 5,573 references pointed to 3,580 unique articles. Of these, 2,758 were cited only once by papers from *Social Networks*, while the remaining 822 were cited by at least two papers from the journal – collectively drawing in 2,815 citations. Only a small proportion of the cited literature referenced by papers in the same journal is mutually shared, but these shared papers constituted ~50% of citations.

Fourth, they applied MPA and detected six paths that had high total transversal counts. There existed a major path which had the highest traversal count that focussed on the Role Analysis in SNA – an area focussing on how particular network positions come to exert forms of influence on other areas of a network. A further five appear to splinter off from this or converge onto it at a later point. The authors identified the documents in these five sub-areas as focussing on different, though related, techniques applied in different empirical areas.

Hummon interpreted the graphs as providing a good representation of his understanding of the major developments achieved in the papers published in *Social Networks*, and he believed the approach demonstrated the ability of this methodology to establish sub-topics that diverged from this main path.

Again, the broader scientific community appeared uninterested. By selecting a single journal to analyse, Hummon and Carley (1993) recognised that their approach had significant limitations, missing many relevant papers that were not published in *Social Networks*. This form of analysis appeared to be of dubious accuracy when applied a limited dataset.
However, Batagelj and Mrvar (1998) incorporated the algorithms required for MPA into the statistical network analysis software, *Pajek*, enabling researchers to apply MPA relatively easily. Since then, MPA has been used to study different areas – from the development of fullerene and nanotube technology (Lucio-Arias and Leydesdorff 2008) to the development of the concept of Social Capital (Moore *et al.* 2006), and the emergence of peer-review as a research topic (Batagelj *et al.* 2017).

While some modern analyses have been performed on papers within specific target journals, most recent studies have used specific key-words in the titles, abstracts, or key-words of papers to identify relevant literature on specific areas of the literature. Of note is Lucio-Arias and Leydesdorff (2008) who applied this methodology to 7,696 documents containing “fullrene” in the title and 9,672 documents containing “nanotube*”. A similar key-word searches were performed by Liang *et al.* (2016) for MPA of IT outsourcing research, by Barbieri *et al.* (2016) for MPA of environmental innovation research, and Mina *et al.* (2007) for MPA of the development of percutaneous transluminal coronary angioplasty. Such searches appear to be a better method of literature retrieval than the journal method or the mirroring of references used in historical research. And, in all of these analyses, researchers have concluded that their analyses produce meaningful results that track the major developments within a field.

However, in Liu and Lu’s (2012) MPA analysis of the development of the Hirsh-Index, they include a small section at the end of their analysis that may, in the context of the findings reported previously, be of considerable importance. Entitled “On the Voices Not Heard”, the authors note something odd:

> It has been reported that the “preferential attachment” or the “rich get richer” phenomenon does exist in citation networks...The Hirsh index has been a well-liked research subject since its inception. Inevitably, articles that supported the idea gathered a good amount of citation in the beginning and rode the phenomenon better than did those that were critical. **Accordingly, the articles on main path are mostly the mainstream articles that positively support the h-index.** Articles that are relatively conservative to the idea...
question its adverse effect on scientists’ behaviour…are certainly not favored by the method (p. 540)

If you examine any of the studies reporting results of MPA, you find stories of progress, where successful articles lead to other successful articles, and the absence of critical voices. In other words, this is a tool that reports the history of a field from the winner’s perspective. Similarly, Moore et al. (2006) interpreted the results of their MPA of the Social Capital literature as indicating the field had heavily relied on one perspective at the expense of another:

[T]he dominance of the communitarian approach to social capital has given disproportionate attention to normative and associational properties of places. Network approaches to social capital were lost in this translation (p.729)

This should raise a note of caution for those advocating the use of MPA in systematic literature reviews (Colicchia and Strozzi 2012).

5. Discussion

The findings reported in this chapter undermine the view that citation practices follow Merton’s norms, and suggest that how studies are referenced comes to shape how they will be referenced in future literature, and that this seems to have an impact on what is considered knowledge in particular areas.

I discussed findings that suggested long-term citation distortions undermine arguments that scientific knowledge always possesses a self-correcting mechanism. While scientists may try to cite in a normative fashion, choices made in particular papers over what literature to reference can have a disproportionate impact on the future shape of the scientific literature by influencing future citing behaviour. It may be the case the scientists don’t knowingly defy the norms of science; they might just not be aware they aren’t following them. Scientists may have an implicit faith that a description of a study in another study is accurate, and that the author probably knew the evidence-base well and didn’t exclude any important publications. This is a flawed and dangerous assumption that can lead to long-term distortions in the scientific literature.
But why were the results of these studies surprising? They jar with a common and comfortable view of how science functions. According to the normative theory, scientists evaluate all of the relevant literature and reference all of the work that is relevant to their claims. Scientists, free of biases, personal ambition, and undue respect for authority, evaluate studies by their methodological rigor, subject each claim to critical scrutiny, and come to objective conclusions. From time to time, scientists might disagree, but these disagreements are settled by new theoretical explanations or evidence. In situations where there is an absence of clear evidence, scientists acknowledge and embrace uncertainty.

This picture of science appears incapable of explaining the findings discussed in this chapter. While findings from studies, such as Greenberg’s (2009), are not generalisable beyond the areas of focus, and studies of biases or error are typically performed because authors suspect some problem to exist within a literature, such results imply than it is never appropriate to assume that the normative model describes citation practices. Accompanied by studies, such as Fanelli’s (2012, 2013), that indicate that science may have a ‘positive-outcome’ bias in general, how scientists contribute and use an evidence-base must now be scrutinised. This bias for ‘positive’ studies also jars with the view that scientists uphold organised skepticism and conflicts heavily with Popper’s view that science progressed through bold conjectures and refutations. If Popper’s view is correct, we would anticipate that the scientific literature is primarily populated by negative reports, but it isn’t; it is replete with apparent confirmations – ‘positive’ studies are not only more likely to be published, they are also more likely to be highly-cited.

So ought we now accept the social constructivist stance? According to this position, scientists reference to persuade others, but if so, why would choices made in the past by certain authors influence the citing behaviour of other scientists? Perhaps the social constructivists can argue that this is an example of successful persuasion, but this term doesn’t quite capture what occurs during the spread of errors. We use the term ‘persuasion’ to describe purposefully trying to change another’s opinion, but is this really what is
occurring? The architecture of the literature itself appears to influence what scientist’s reference – and this structure seems to be self-organising and complex, in which it is unlikely any specific author can anticipate how any of their particular papers or references will come to be read and used. Of course, the shape of the literature alone cannot explain the entirety of what scientists believe or do, but it must now nevertheless be taken into account as one of the likely contributing factors to wide-spread belief in science. None of this rejects constructivism, but it does suggest that the remit of constructivist scholarship, such as is practiced in SSK, requires expansion.

CS-CNA appears to be a powerful tool for detecting these distortions, and should add depth to our understanding of how scientists use and interpret published evidence in constructing knowledge-claims and their disagreements. The ability of CS-CNA to capture the dynamics of referencing and their content – the cumulative effects of the mundane work of scientists in communicating their ideas – opens up new lines of inquiry.

Second, I discussed the findings of a recent methodology thought to establish the major path of development through a field – MPA – but highlighted this appears to construct accounts of scientific history that give an impression of apparently linear progress. According to those who have conducted such studies, the results produced by this method cohere with their understanding of the landmark contributions in an area and their relationship to one another. This tells us that what is considered promising, important, or popular is picked up in diachronic citation dynamics. Yet, the results regarding citation bias and distortion raise a paradox – one group is arguing that citations do, indeed, identify the ‘best’ or most ‘important’ studies, while the other is claiming that citation behaviour is heavily biased and error prone. So what is happening?

Liu and Lu’s (2012) application of MPA provides an indication that this methodology might help explain why certain studies are cited or not. Critical papers of the Hirsch-Index did not sit on the main path. Taken together with Greenberg’s (2009) finding of citation cascades regarding the belief in the β-amyloid claim and the prevalence of citation biases and distortions in the
literature, there are good reasons to believe the structure of the literature, as constructed through networks of citation, strongly influences both citation practices and scientific opinion. Specifically, the structure of the literature over time becomes biased toward favoured hypotheses, entrenching them in a network that guides readers to supportive studies. If you were to apply MPA to Greenberg’s data, the main propagators of bias in that citation network should be identifiable. Indeed, Greenberg used a method of path counting to identify important papers that disseminated bias.

Studies of MPA and CS-CNA have not interacted yet in the literature. The findings of one set of studies is not cited by the other, and the two methodologies appear to be developing in separation. This appears to be because: (i) MPA developed in the Scientometrics literature, and this discipline is primarily interested in using it to understand emerging topics in particular research areas; (ii) Scientometricians, generally, subscribe to a normative theory of citation, which they use to interpret the meaning of the results from MPA; (iii) CS-CNA was developed by scientists interested in evaluating the reliability of knowledge claims in regards to specific hypotheses and theories; (iv) the results of CS-CNA appear to contradict the basic assumptions of the normative theory of citation.

This is the gap in the literature that this thesis seeks to fill. While Greenberg and Trinquart et al. began examining the how their findings might contribute to understanding how scientists form beliefs over the validity of theories and how disagreement may emerge, this has not yet been pursued in the sociological literature via these methods.

Compatibility with SSK

Network analysis, and the forms of analysis discussed, can be used by those following the methodological tenets of the Strong Programme; one is not tied to accept a normative account of citation behaviour. The Strong Programme aims to describe and explain belief in scientific knowledge by groups regardless of whether or not that knowledge is ‘true’. To do this, Bloor envisioned that the sociologist would prioritise the analysis of:
The distribution of belief and the various factors which influence it. For example: how is knowledge transmitted; how stable is it; what processes go into its creation and maintenance; how is it organised and categorised into different disciplines or spheres? (p.4).

The findings of the reviewed studies suggest that CNA can contribute to this. However, there has been skepticism about using quantitative methods in SSK. Since the 1990s, SSK has championed the use of ethnographic methods as the primary theoretical lens and methodology through which the construction of scientific knowledge is to be understood and studied. While Scientometrics has become almost exclusively quantitative after adopting the normative account of citation behaviour – a theory used to justify the results of evaluative metrics. CNA, accompanied with the reading and analysis of the content of documents, appears to represent a method that could incorporate ideas and findings from both disciplines – a bridge between two distant traditions.

**A statement of intent**

In what follows, I apply the variants described in this chapter, CS-CNA, citation mapping and community detection, and MPA analysis to understand aspects of the development of the diet–heart hypothesis and its controversies. First, by applying CS-CNA to the debate over the validity of the diet–heart hypothesis, I test whether citation bias can explain divergence of opinion in the scientific community over what the key lines of evidence are and their implications. Second, after capturing a large body of literature relevant to the diet–heart hypothesis, I construct a citation network representing the state of the literature prior to 1985. By modularity and temporal analysis, I analyse this network for regularities in citation behaviour and detect communities within it. Third, I perform MPA to understand what the scientific community championed over time. For now, however, the basics of network analysis require explanation.
Chapter 3: General Methods - Citation Network Analysis

Introduction

This chapter introduces the basics of citation network analysis (CNA). As network analysis, and CNA in particular, may not be familiar to the reader, I explain key terminology, measures, and data structures via several working examples. Each of the following chapters uses a specific variant of CNA to fit the particular questions asked in those chapters, and the materials and methods for each chapter are self-contained. Accordingly, this chapter is restricted to a basic introduction that ought to help the reader follow the more advanced methods introduced later in the thesis. For a detailed introduction to network science, I recommend Newman (2010).

1. Constructing networks

A network, also referred to as a graph in mathematics, is composed of vertices, which in a citation network refer to documents, and edges, which in this context refer to a reference between pairs of documents\(^\text{11}\). Thus, a network refers to a graph \(G\) composed of a set of vertices \(V\) and a set of edges \(E\):

\[
G = (V, E)
\]

By convention, the total number of vertices in a network is typically given by \(n\) and edges by \(m\). As references always point in a particular direction – from the citing to the cited – networks of citation data are examples of directed networks. Fig 3 shows a simple example of a small citation network; the blue circles represent vertices (papers), the edges (references) between these are

---

\(^{11}\) Vertices are synonymous with nodes and actors in social network analysis; Edges are synonymous with links and ties in social network analysis. Directed edges are often referred to as arcs, but I have retained the term edge because of popular use.
represented by arrows that also convey the direction of the reference (from the citing to the cited).

**Fig 3: Citation Network Example 1 (n=7; m=9)**

An edge between vertices \( i \) and \( j \) is denoted by \((i,j)\) and this order is reversed to establish the direction of an edge (\( j \rightarrow i \)). Formally, the above network can be described by \( G = (V, E) \) with \( V = \{1, 2, 3, 4, 5, 6, 7\} \) and \( E = \{(1,2), (1,3), (2,3), (2,4), (2,5), (2,6), (4,6), (5,6), (5,7)\} \). Typically, this information is recorded in an adjacency matrix:

\[
A_{ij} = \begin{cases} 
1 & \text{if there is an edge from vertex } j \text{ to vertex } i \\
0 & \text{otherwise}
\end{cases}
\]

Therefore,

\[
A = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\]
Networks produced with citation data resemble directed acyclic graphs (DAG), which means that there are no loops or self-loops of edges (e.g., (7,1), (2,2)) because most references refer to papers previously published. Given that the matrix is upper triangular and the diagonal elements are zero, the matrix is strictly upper triangular. This tells us, at a glance, that this is an example of a DAG. Citation networks are *asymmetrical* because two papers generally cannot both reference each other due to the temporal direction of referencing. In practice, citation networks can violate the conditions of a DAG. First, a paper may reference itself. Second, it is also possible that papers are published simultaneously and reference each other. Third, a paper may reference a paper awaiting publication and its predicted publication date. As a result, citation networks are often not strictly acyclic as there are occasionally a few short chain loops contained in a given network.

2. Measures and partitions

**Degree**

The *in-degree* of a vertex represents the number of citations to a particular document. For example, in Fig 3, vertex ‘2’ has an in-degree of 4 and this can be read from the second row of the adjacency matrix. Thus, the in-degree ($k_i^{in}$) of a vertex is the number of edges from other vertices to that vertex, and, for a vertex $i$, is given by:

$$k_i^{in} = \sum_{j=1}^{n} A_{ij}$$

The *out-degree* of a vertex represents the number of references from a particular document. In Fig 3, vertex ‘2’ has as an out-degree of one and this can be read from the second column of the adjacency matrix. Thus, the out-degree ($k_j^{out}$) of a vertex is the number of edges from that vertex to other vertices, and, for a vertex $j$, is given by:

$$k_j^{out} = \sum_{i=1}^{n} A_{ij}$$
The degree \((k)\) of a vertex is given by the sum of \(k^{in}\) and \(k^{out}\), which provides the combined total of references and citations of a given document. While the average degree of a given network is given by the sum of edges over the sum of vertices.

The total number of citations to a document is generally considered to be a measure of its impact on the literature. However, following the findings regarding the spread or errors in the scientific literature that suggest that scientists may copy references and their descriptions from other papers (Simkin and Roychowdhury 2003), this assumption is questionable. Nevertheless, the use of a particular reference in-text appears to indicate that it plays some rhetorical function; at least some interpretation of the study being referenced plays some role in the citing article. As discussed in Chapter 4, measures of degree can be used to establish the presence of citation bias.

**Paths**

As Greenberg (2009) demonstrated, *paths* can be used to understand how particular claims and evidence selection behaviours may spread in the literature. A *path* refers to a sequence of edges that connect together a sequence of vertices in a given graph, while the *length* of a path refers to the number of edges between these vertices. In a citation network, a path refers to a chain of citations linking a series of documents together. In Fig 3, there are two paths of length three between vertices 6 and 1 (6–4–2–1) and 6–5–2–1), and a third path of length two (6–2–1). In Fig 3, if vertex ‘2’ misinterpreted vertex ‘1’ and vertex ‘3’ held the same interpretation as vertex ‘2’, then it is a possibility that vertex ‘3’ took this interpretation and reference from vertex ‘2’ without reading vertex ‘1’. For further discussion of paths and related analyses, see Chapter 9.

**Isolates and Components**

A network refers to an interconnected set of vertices, but in empirical data there can be different networks that do not connect to one another or vertices that do not connect to any other vertex. An *isolate* refers to a vertex with no incoming or outgoing edges – a document with no citations or references. A
weakly-connected component refers to a set of interconnected vertices that form a network in which there exists an undirected path between any given vertex and any other vertex, while the largest component describes the largest weakly-connected component – the network that holds the largest number of interconnected vertices in a given set of empirical data. In Fig 4, I’ve constructed an example that shows the largest weakly-connected component in blue, another weakly-connected component in red, and two isolates in green.

Fig 4: Citation Network Example 2. Vertices in blue are members of the largest weakly connected component \((n=7)\), vertices in red are members of a weakly-connected component \((n=4)\), while vertices in green are isolates \((n=2)\).

In a citation network, separate components would indicate that two sets of documents share none of the same literature with one another. If there is no interaction in terms of references and citations between two bodies of literature, it is likely that either (i) the two literatures focus on different topics; (ii) that they share a topic but are unaware of the others existence. Typically, isolates occur due to the presence of meeting abstracts, which are generally not cited or referenced. For further discussion, see Chapter 7.

Distance

The shortest path between any two vertices in a network is referred to as the geodesic path. The diameter of a network refers to the geodesic path between the two most distance vertices in a weakly-connected component – the longest geodesic path of all geodesic paths in a network. While the mean geodesic path refers to the average geodesic path length between any two vertices.
within a component. In Fig 3, the diameter is 3 – the longest geodesic path (length three) is between vertex ‘7’ and vertex ‘1’ – while the average geodesic path is 1.43 (nine paths of length one, four paths of length two, and one path of length three). These are distance measures and indicate how close particular documents are in terms of their reference and citation links to other documents in that network. Networks with small values for average geodesic path length compared to the total number of vertices are said to possess the small-world property, described informally by Milgram’s (1967) concept of Six Degrees of Separation. More recently, small-world networks have been formally defined by Watts and Strogatz (1998) as networks in which the average distance between two random vertices is proportional to the natural logarithm of the number of vertices in that network, and where the clustering (transitivity) coefficient is not small. For further discussion of distance measures, see Chapter 7.

Density

The density of a network is a measure used to establish how cohesive the vertices in a network are. To establish density, one first calculates the total number of edges present between all vertices in a given network and the number of possible edges that network could have if all vertices shared an edge. However, due to the temporal dimension of citation networks this can be difficult to approximate.

In an undirected network, density is given by:

\[ D = \frac{2m}{n(n-1)} \]

where \( m \) is the sum of edges and \( n \) the sum of vertices

If we regard the network in Fig 3 as undirected, then:

\[ \frac{2(9)}{7(7-1)} = 0.429 \]
A $D$ of 0.429 means that about 43% of possible edges are actually present. The number of edges is doubled to reflect that a particular edge both points to and from a vertex in an undirected graph.

For a directed graph, density is given by:

$$D = \frac{m}{n(n-1)}$$

For Fig 3,

$$\frac{9}{7(7-1)} = 0.214$$

Thus, by this calculation, there are about 21% of possible edges present. However, this measure is misleading in the context of citation networks, which typically contain documents published in different years. This is because of the temporal dimension of publication and referencing – not all documents have the chance to reference or be cited by all documents in a given citation network.

To illustrate this, imagine a citation network that grows by one document per year. In this scenario, a network of two vertices has at most one edge, three vertices have three possible edges, and four vertices have six possible edges. Therefore, the maximum number of possible edges grows in a triangular number sequence $(1, 3, 6, 10, 15…n)$. Thus, to calculate the density for this network, take the number of edges present in a network against the total number of edges established by summing the total number of vertices and finding the corresponding triangular number to this number. Therefore, to calculate density:

$$D = \frac{m}{n(n-1)/2}$$

If the network in Fig 3 had these properties, then:

$$\frac{9}{7(7-1)/2} = 0.429$$

This equation is equivalent to the density equation for an undirected graph. While this fails to take into account simultaneous publication of documents that cannot cite one another nor can it precisely establish the
number of possible edges in a network that grows faster than \( y = x + 1 \) (linear growth), it provides a closer approximation than the density measure for simple directed graphs by taking into account network growth.

A better approximation of density is possible by taking into consideration the actual growth of a given network. In a network with \( n \) vertices already, and there are \( z \) new vertices added in a given year, then there are \((n \times z)\) possible new edges added assuming no new vertices can reference other new vertices. If we factor in a time dimension to the network presented in Fig 1 and connect all vertices to all other vertices by this rule, then:

![Fig 5: Citation Network Example 3 (n=7; m=18)](image)

The vertices in the Fig 5 are labelled by their generation; \( t_0 \) represents the first year of publication, \( t_1 \) the second generation, \( t_2 \) the third generation and so on. The maximum possible edges in this graph following the above rule can be ascertained by \((t_0 \times t_1) + ((t_0 + t_1) \times t_2) + ((t_0 + t_1 + t_2) \times t_3) + ((t_0 + t_1 + t_2 + t_3) \times t_4) = 18\). Thus, as the network in Fig 3 has 9 edges, then the density by this calculation would be 0.5 – 50\% of possible edges are present.

While this can account for network growth, it still provides only an approximation of the number of possible edges in a genuine citation network because typically a small proportion of references in any given year are
directed at other documents published that year. To resolve this, we could add the sum of intra-year citations to the sum of possible edges. This remains an approximation, however, as a paper might have been written long before its publication and, therefore, it is not possible to know exactly the literature that was available to reference for every paper.

Density measures provide a way of comparing two networks or different cuts of a network. For citation networks, a network that has a high density compared to some other is more cohesive in the sense that the documents contained within it will be more interconnected through references and citations (see Chapter 4 and 7 for further discussion).

**Community detection**

In a large, complex network we are often interested in how vertices connect together to form particular structures within a given network. *Clusters* refer to sets of vertices that share a particular quality, typically based on edge distributions. *Density-based clustering* refers to the partitioning of a network into clusters that share more intra-cluster edges than inter-cluster edges. An example of this is Newman’s (2006) modularity maximisation, a technique used measure how well a network can be divided into ‘modules’ (clusters).

Calculating modularity in a directed citation network is difficult because it is not feasible to precisely calculate the expected number of edges as discussed above. However, Speidel *et al.* (2015), after constructing a modularity maximisation method for directed acyclic graphs, found this produced similar results to the modularity maximisation method for undirected graphs. According to the authors,

[S]imply applying modularity maximization methods for undirected networks to DAGs may be practically innocuous. We stress that we have reached this conclusion by actually developing a modularity measure for DAGs and testing it against previous methods using several data sets (p.9).

Importantly, two of the test datasets the authors used in arriving at this conclusion were citation networks – one containing all arXiv publications on high energy physics phenomenology and one containing all arXiv publications high energy physics theory. Thus, following Speidel *et al.* (2015), undirected
modularity analysis is suitable for citation networks. Modularity \( Q \) for an undirected graph can be ascertained by:

\[
Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{d_i d_j}{2m} \right) \delta (c_i, c_j)
\]

where \( m \) is the total number of existing edges, \( A_{ij} \) is the sum of edges from the adjacency matrix, \( \frac{d_i d_j}{2m} \) calculates the probability of an edge between two vertices proportional to their degree, the delta function, \( \delta \), equals 1 if \( j \) and \( i \) are in cluster \( c \) and 0 otherwise.

To perform a modularity analysis, computational methods are required to establish robust clusters of vertices. The most recent method for doing so is Traag et al.’s (2019) Leiden agglomerative algorithm for modularity-based community detection; an algorithm demonstrated by the authors to be superior to other algorithms (see Chapter 8 for discussion).

Within a given citation network, the partitioning of documents into clusters based on their reference and citation relationships is said to establish groups of documents that are members of some discipline, or a finer grained analysis can find documents that appear to share particular research topics (Klavans and Boyack 2017).

To help conceptualise modularity-based community detection, Fig 6 shows two communities of vertices detected by this method.

*Fig 6: Citation Network Example 4: Community detection (n=7; m=9)*

Here we see two communities of vertices identified by the Leiden algorithm by maximising modularity. The three vertices in green have a density by this method (undirected density) of 1, while the four vertices in pink have a density
of 0.83. There is no other way that we can increase the density of either of 
these two groups by moving vertices from one group to another. For example, 
we might move the pink vertex connected to the green vertex in to the green 
group, but this would lower the density of the green groups to 0.67 while 
increasing the density of the pink group to 1, which would be a sub-optimal 
partition because the total density of the two groups is smaller.

**Sub-graphs and Meta-vertices**

A sub-graph is a graph formed from a subset of the vertices and edges of a 
given graph. For example, if, after performing a modularity analysis, we find 
two communities of vertices, we might consider inducing two separate sub-
graphs containing only those vertices that are members of those communities 
and only the edges between these vertices.

A meta-vertex describes a vertex formed from the merging of groups of 
vertices together that share some attribute. A meta-vertex includes all outgoing 
and incoming edges of the vertices it represents.

**Claim-Specific CNA (CS-CNA)**

Greenberg (2011) articulated an overview of what his CS-CNA methodology 
is designed to detect. CS-CNA is an application of network analysis to content-
specific citation data. A researcher systematically identifies all papers related 
to a research claim, reads them, classifies each by type (e.g. primary study, 
review etc.), records the citations that link papers together, and categorises 
citations as supportive, neutral, or unsupportive to the claim via *quotation analysis*.

Quotation analysis refers to the reading of texts to establish (i) how a 
particular study is being cited in that text; (ii) the conclusion or focus of the 
citing article, and then capturing quotations representative quotations of these 
classifications from those texts (see Chapter 4–5 for further discussion). I 
expand quotation analysis to classify papers by research focus by classifying 
their titles in Chapter 5 and 8. As quotation analysis requires interpretation of 
text, I’ve included all quotation data for every analysis performed in the 
Supplement.
3. Storing Network Data

While adjacency matrices can help facilitate certain analyses, their format is impractical for recording big data. For example, in a network of 1,000 vertices, an adjacency matrix will have 1,000 rows, 1,000 columns, 1,000,000 cells. For analyses, I use edge-lists to store citation data, which converts citation data into two columns – ‘Source’ and ‘Target’. The source column contains the vertex identifier (an integer assigned to a document to identify it) from which an edge derives, while the target column contains the identifier of the vertex to which an edge is directed. Thus, the source column contains the bibliographies of papers, while the target column represents citations received by papers.

For Fig 3, converting the adjacency matrix to an edge-list format results in the below table:

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

In an edge-list, it is also possible to store edge-attribute information. In this project, the only edge-attributes I store are the date of publication of the source, the data of publication of the target, and a variable reflecting the difference in publication year between the source and the target.

An edge-list contains all information necessary to construct a network and examine its structure. However, to identify what vertices represent or to assign particular variables to them, a vertex-attribute-list is also required. By using the same vertex identifier as the edge-list, I store all relevant paper information in this format along with particular variables required for the different analyses. For all analyses, I routinely capture the full paper reference,
all authors, title of publication, journal of publication, year of publication, study type, global citations (total citations record to that paper from a particular bibliometric database), and accession number. For example:

**Table 2: Example vertex-attribute-list**

<table>
<thead>
<tr>
<th>ID</th>
<th>Reference</th>
<th>Year</th>
<th>Journal</th>
<th>Authors</th>
<th>Type</th>
<th>Global Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Burns M, 1989, J Fake, V1, P1</td>
<td>1989</td>
<td>J Fake</td>
<td>BURNS, SILEES, N</td>
<td>Article</td>
<td>92</td>
</tr>
<tr>
<td>7</td>
<td>Burns M, 1994, J Fake, V1, P1</td>
<td>1994</td>
<td>J Fake</td>
<td>BURNS, S</td>
<td>Article</td>
<td>100</td>
</tr>
</tbody>
</table>

All variables derived from network measurement, such as in-degree, are stored in alongside this information following computation. Further, all further variables based either on community detection or document analysis are further stored in these datasets – such variables are described in full in the relevant chapters.

**Software**

Both of these file formats can be read by the major network analysis software used throughout the project. As multiple different analyses are performed, this is important because certain analyses and visualisations are performed in different software packages in *Gephi* version 0.92 (Bastian et al. 2009), *R* statistical environment version 3.5.1 (R Development Core Team 2008), *Sci²* tool version 1.3 (Sci² Team 2009), *Pajek* version 5.06 (Batagelj and Mrvar, 1998, 2018) *MATLAB version R2016b and R2019b* (MATLAB 2016, 2019), *Microsoft Excel* (2016), *Notepad++* (2018).
Network visualisation

All network figures produced in this project were constructed in Gephi via the ForceAtlas2 layout algorithm unless otherwise specified. ForceAtlas2 works by positioning vertices closer to vertices with which they share an edge, and pushes vertices away from one another if they do not. Generally, it produces clear visualisations in which community structure is particularly apparent and is particularly effective for visualising clusters in directed graphs (Jacomy et al. 2014).

Data availability statement

All data used for the analyses performed in this thesis have been made available in the Supplementary Material. For each chapter that uses a novel dataset, I’ve included all relevant data in sub-folders that are titled by the chapter in which that analysis was performed – e.g., “Chapter 4 Supplement”. The content of datasets, their structure, and the software required for analysis are explained in the body of each chapter or in the ‘Data Note’ provided in the Supplement.
Chapter 4: Claim-Specific Citation Network Analysis, Citation Bias, and Sub-Graph Analysis

Introduction

This chapter reports the findings from my adaptation of Greenberg’s (2009) claim-specific citation network analysis (CS-CNA). I apply and adapt this methodology to understand how the first four randomised controlled trials (RCTs) evaluating the efficacy of cholesterol-lowering diets in the secondary prevention of coronary heart disease (CHD) were interpreted in reviews of the literature prior to the National Institutes of Health (NIH) consensus conference in 1984.

First, I identified the relevant RCTs and classified them according to whether their conclusions supported or opposed the use of dietary fat modification/restriction in the secondary prevention of coronary heart disease. I then identified all indexed publications in Scopus and Web of Science (WoS) that cited at least one of these trials between 1969 and 1984 and retrieved all review articles amongst these. In total, 62 reviews were published in this period that cited at least one of the RCTs. I classified each as supportive, neutral, or unsupportive to the use of dietary fat modification based on a quotation analysis of their evaluation of the findings of these RCTs. I then manually coded all references from these reviews to these four RCTs into a network dataset. Citation bias (the act of citing only trials of a particular direction) and research underutilisation (sub-optimal referencing of trials) were detected by applying a comparative density measure, in-degree, and out-degree in a series of sub-graph analyses.

Of the four RCTs, one supported dietary modification in the secondary prevention of CHD, while three were unsupportive. Of the 62 reviews, 28 suggested that data from existing trials supported dietary treatment, 17 were neutral in regards to whether the treatment was supported or not by the findings of trials, while 17 stated that the findings of trials did not support the
use of dietary treatment. On average, supportive reviews underutilised the available RCTs to a greater degree than other reviews. Amongst the supportive group, citation bias was common – 23 (82%) reviews cited only the one RCT that was supportive. Thus, most reviews that disseminated a supportive evaluation of the results of RCTs in the context of secondary prevention cited only data that supported this position.

Unlike the later chapters of this thesis, which are based on a different literature retrieval strategy designed to capture a population of papers relevant to the diet–heart paradigm, this study was designed to focus only on tracking reviews that cited these four RCTs. As this is one of my earliest studies, began early in 2017 and published in 2018 (Leng 2018), I lacked the larger datasets used in later chapters. The methodology I describe here is specifically designed to detect citation bias without the need to capture large volumes of literature. I see this methodology as having more practical utility for those exclusively interested in detecting citation bias to specific sets of studies and their influence on this bias in the literature. Specifically, I establish here a novel methodology that introduces sub-graph analysis to CS-CNA and a new measure of sub-optimal referencing behaviour, and believe these could be effectively applied to other areas of the scientific literature.

As this paper was published in *PloS ONE* (Leng 2018), I’ve reproduced its content as published hereafter. As this paper was referenced using in-text reference numbers and all figures depended on this numbering, I’ve included a separate bibliography for this paper at the end of this thesis.

1. **Background**

In 1961, the *American Heart Association* published guidelines recommending that people who have had one or more atherosclerotic heart attacks should reduce their intake of saturated fatty acids (SFA) and modestly increase that of polyunsaturated fatty acids (PUFA) [1]. The theoretical basis for this was the *diet–heart hypothesis*, that dietary intake of SFA, by increasing total serum cholesterol, contributes to coronary heart disease (CHD) via the development
of atherosclerosis, while PUFA decrease serum cholesterol and lower the risk of CHD [2]. In the 1950s, this theory was supported by: (i) epidemiological evidence of an association between fat intake, particularly SFA intake, and raised serum cholesterol levels; (ii) an association between serum cholesterol and coronary incidence and mortality; (iii) short-term metabolic ward trials that suggested that SFA increased serum cholesterol while PUFA appeared to modestly lower serum cholesterol [3].

However, some scientists warned of the dangers of mistaking correlation for causation, particularly in population studies [4] and in a multifactorial disease such as CHD [5,6]. By the time that the guidelines were issued, no randomised controlled trial (RCT) had tested the link between the consumption of different types of fat and coronary risk in secondary prevention (i.e., after a clinical manifestation of CHD), but, in the 1960s, five RCTs examined this. The Los Angeles Veterans Administration Diet Study [7] combined secondary prevention and primary prevention (i.e., before the development of CHD), but was generally regarded as relevant only to primary prevention due to the way the authors stratified their data and presented their findings. Four other trials specifically examined the effect of a modified diet in the secondary prevention of CHD in patients who had experienced a recent myocardial infarction (MI): the Research Committee Low-Fat Diet Study [8] examined total fat restriction, while the Rose Corn Oil Study [9], the Oslo Diet-Heart Study [10], and the Medical Research Council’s Soya-bean Oil Study [11] examined the impact of restricting SFA and increasing PUFA intake. These were the only published secondary prevention RCTs until the partial publication of the Sydney Diet–Heart Study [12] in 1978, which only published results concerning all-cause mortality.

The four RCTs published in 1965 [8, 9], 1966 [10], and 1968 [11] had similar methodologies, and were published in major journals, but produced discrepant findings (see Ch4.S1–S4). The Oslo study, published in 1966, found significantly lower levels of fatal reinfarction, non-fatal reinfarction, and angina pectoris in the intervention group, but found no significant relationship between diet and sudden death or all-cause mortality [10]. This study was a
doctoral study by Paul Leren and was published as a supplement in Acta Medica Scandinavica. The other three studies all returned equivocal or negative results regarding reinfarction, coronary mortality, cardiovascular events, and all-cause mortality. All three of these studies were directed by leading authorities of British cardiology on behalf of the Medical Research Council and published in either the Lancet or British Medical Journal. The author of the Oslo study regarded his results as supporting dietary fat modification in secondary prevention, while the authors of the other studies regarded their results as evidence against such treatments.

In December 1984, the National Institutes of Health (NIH) gathered a panel of experts to resolve whether the link between raised serum cholesterol and CHD was causal, and under what circumstances dietary or drug therapy should be initiated [14]. After reviewing the available evidence, the panel concluded that an elevated serum cholesterol level is a cause of CHD and that lowering it is protective. To lower serum levels of cholesterol, the panel recommended that all men, women, and children over 2 years of age should eat a diet consisting of no more than 30% fat, should reduce their intake of SFA to less than 10% of total calories, and should increase their intake of polyunsaturated fatty acids to 10% of total calories. Specifically, on secondary prevention, the panel concluded:

Studies are available that indicate a beneficial effect of treating high cholesterol levels in individuals with preexisting clinical disease [secondary intervention] (p.2084)

The conference was widely regarded as bringing a (temporary) end to the controversy over both the relationship of cholesterol with CHD and the relationship between dietary fats and CHD [15].

Yet, by 1984, the four RCTs discussed previously remained the only available unifactoral RCTs with data on the effect of dietary modification on coronary mortality and morbidity in secondary prevention. However, the results of these trials were conflicting – an evaluation supported both by the directors of each study and by recent meta-analyses [16–19]. According to Harcombe et al. [16], the NIH advice was not supported by the available RCTs, neither in primary or secondary prevention. As can be seen in the risk-ratio analysis in
the supplement (see Ch4.S4), the evidence from these RCTs available prior to 1984 leads to no clear conclusion as to whether dietary modification is an appropriate treatment in secondary prevention. Recent meta-analyses also agree that, when the findings of these four trials are considered together, no significant difference exists on the “hard” clinical end-points – the incidence of MI, fatal CHD, and all-cause mortality – between the intervention and control groups [16–19].

The present study examines how these four trials were used in the literature and their impact on published scientific opinion regarding the case for dietary fat modification/restriction in the secondary prevention of CHD, prior to the NIH consensus conference. By using a modified version of claim-specific citation network analysis, a methodology developed by Greenberg [20], it maps and evaluates the network of review papers that cited these RCTs over a defined period (1969–1984); the date by which all relevant RCT data would have been available to reviewers to the start of the NIH conference. The rationale for examining how these studies were cited is the claim that the diet-heart controversy stems from selective citation [21, 22]. In 1992, Ravnskov, in one of the first published demonstrations of citation bias, examined 22 cholesterol-lowering trials and found that supportive trials were cited almost six times as often as others [21]. In three highly-cited reviews associated with dietary policy [23–25], he went on to document citation bias and quotation bias, which describes the erroneous or selective interpretation of experimental findings [22]. Finally, in a response to an article in the BMJ in 2003, he claimed that belief in the diet-heart hypothesis was kept alive by selective citation [26].

This study tests these claims by examining how dietary RCT results were cited in reviews during the period leading up the publication of dietary guidelines and closure of controversy, following Greenberg’s [20] finding that reviews tend to be the major propagators of information regarding the findings of primary studies. It examines two outcomes of selective citation: citation bias – the selective citing of data supporting a particular outcome [20]; and research underutilisation – when authors fail to reference all available data relevant to a specific claim. These are outcomes of sub-optimal citation practices that distort
the evidence base by inflating the influence of some studies while under-representing others [20, 27]. To detect these factors, network analysis was used to map the citation links between RCTs and reviews that were themselves classified by quotation analysis.

2. Materials and Methods

Study selection

To be included, a trial had to be a unifactorial RCT that: (i) was published before 1984; (ii) examined the effect of dietary fat restriction (<30% of total calories derived from dietary fat) OR dietary fat modification (replacing SFA with PUFA) on coronary benefit/risk in the secondary prevention of CHD; (iii) published data on sample numbers AND serum cholesterol changes AND mortality and morbidity figures regarding incidence of MI, coronary mortality (including MI and sudden death), and all-cause mortality; (iv) included only patients with established CHD (MI, angina pectoris); (v) a minimum intervention and follow-up period of 12 months.

RCTs were identified by a literature search on Web of Science using the ‘all database’ function with the following Boolean string (((dietary fat OR fatty acids OR saturated fat OR low fat diet OR cholesterol lowering diet OR modified fat diet OR restricted fat diet OR corn oil OR soya bean oil OR unsaturated fat OR linoleic acid) AND (coronary heart disease OR ischaemic heart disease OR myocardial infarction OR cardiovascular disease OR atherosclerotic heart disease) AND (trial OR intervention)) limited to publications between 1950 and 1984. This returned 244 publications across four databases in March 2017 (MEDLINE, Web of Science Core Collection, BIOSIS Citation Index, FSTA® - the Food Science Resource). An additional examination of trials identified by recent systematic reviews and meta-analyses was conducted to ensure that all relevant studies were identified [16–19]. After screening publication titles and abstracts, 19 trials were identified as potentially relevant. After reading these, four RCTs were identified that met the inclusion criteria, all of which were published in the decade after the publication.
of the American Heart Association guidelines. The findings of these four RCTs were published in seven papers: the results of the Oslo Trial were published in four journals [10, 28–30]. These were merged, reducing the primary source studies to four. The results of these trials are included in the analysis in the supplement (see Ch4.S1–S4).

One trial, the Los Angeles Veterans Administration Diet Study [7], was omitted due to the inability to assess data on the effectiveness of dietary modification in secondary prevention due to the way the authors stratified their data and reported their findings. This was further justified by the prevailing opinion of the scientific community as expressed in reviews, which appeared to regard it relevant only to primary prevention, and its exclusion from meta-analyses covering secondary prevention specifically [18]. Another trial, the Sydney Diet–Heart Study [12], was omitted because only results on all-cause mortality were reported within the time frame and because it remained uncited in the literature by reviews until 1985 [30]. The Finnish Mental Hospital Study [31], a commonly-cited study in this period, was excluded on the basis that it was not an RCT; it was a “cross-over” trial in two mental hospitals and combined secondary and primary prevention. Further, it was not a standard cross-over trial: the patient groups were not the same in the diet arm and control arm, no blinding was mentioned, and investigators re-coded deaths post-hoc. Bierenbaum et al.’s [32] St Vincent’s Hospital Study was excluded because of the lack of appropriate randomisation due to the assigning of a control group post-hoc that had “similar characteristics” to the intervention. Five trials were excluded on the basis of being primary prevention trials [33–35] and/or multifactorial [36–37], and six early secondary prevention trials were excluded due to a lack of appropriate randomisation and/or control [38–43].

Review papers, published in academic journals in English between 1969 and 1984 and which cited one or more of the studies, were identified by searching Scopus and Web of Science by searching for references to the relevant RCTs in the bibliographies of indexed articles. To be included, a paper had to be a review or an extended editorial (with at least one citation) published in an academic journal within which the coronary mortality and morbidity
results of the identified RCTs were discussed. During this period, only narrative reviews covered this topic, and none described a systematic search strategy. Excluded from analysis were letters, notes, book chapters, other primary articles, any paper not published in an academic journal, and any review that cited the RCTs without discussing mortality and morbidity. The search in Scopus returned 110 citing articles, of which 30 met the inclusion criteria. The search on Web of Science returned 215 citing articles, from which an additional 32 reviews were recovered that met the inclusion criteria. In total, 62 citing reviews were included in the analysis [44–105].

Classification of papers

The RCTs were classified using ‘quotation analysis’ [20, 22, 27] according to whether the findings were evidence for or against the use of dietary fat modification/restriction in the secondary prevention of CHD. Quotations from the papers were collected to enable papers to be classified by the authors’ interpretation of their findings and these are reproduced in the supplement (S5 Table). For example, the authors of the Rose Corn Oil Study, an unsupportive trial, concluded that “under the circumstances of this trial corn oil cannot be recommended in the treatment of ischaemic heart disease” (p. 1533). By contrast, the author of the Oslo Study, the only supportive trial, concluded that the observed “reduction of the serum cholesterol level associated with a reduced CHD relapse rate strongly suggests a cause and effect relationship” (p. 79).

Reviews were similarly classified by quotation analysis by the authors’ evaluation of the RCT evidence. Quotations were identified either in the passage in which RCTs were cited or were stated explicitly in a conclusion, and these are reproduced in the supplement (see Ch4.S5). Reviews were then classified into one of three categories: supportive reviews contained statements implying that the available trial evidence supported dietary intervention; neutral reviews contained statements highlighting the conflicting nature of the trial results and contained no statement either supporting or opposing dietary intervention; unsupportive reviews contained statements implying that the available trial evidence did not support the use of dietary
intervention.

**Network data**

Each paper was given an identifier code ($Vn$). All authors, title of paper, year of publication, paper type (RCT or review), and classification were recorded to identify each vertex in a ‘vertex list’ format (see Ch4.S6). This attribute information was used to examine the relationship between classification variables and citations behaviours. The citations to (incoming edges) and from (outgoing edges) all papers were recorded. The bibliographies of each review were screened to ensure that all citations to RCTs were identified within this sample. Citation information on the relationship between papers was recorded in an ‘edge list’ format in an Excel spreadsheet (see Ch4.S7). This describes a network as a set of vertices (papers) connected by ‘edges’ (citation links). From these data, a graph was constructed and analysed using R’s ‘statnet’ package [106, 107] and visualised in Gephi [108]. Citation information was converted into an adjacency matrix that records edges (citations) between vertices (papers):

$$A_{ij} = \begin{cases} 
1 & \text{if there is an edge from vertex } j \text{ to vertex } i \\
0 & \text{otherwise} 
\end{cases}$$

A graph $G$ was constructed of the vertex set $V$ and edge set $E$:

$$G = (V, E)$$

The vertices were coloured by classification and sized by in-degree, labelled by number assigned in the bibliography of this paper, and visualised initially using Gephi’s ‘ForceAtlas2’ algorithm [108, 109] and later manually realigned in sub-graphs (see Ch4.S8).

Two outcome measures – **citation bias** and **research underutilisation** – were used. Following Greenberg [20], citation bias refers to the act of citing only evidence of a particular classification. For example, if paper $A$ references only supportive studies and no unsupportive studies then it displays citation bias. Citation bias only describes the most extreme cases of evidence selection, but any underutilisation may create distortions in the literature – i.e., if reviews fail to use all relevant evidence even if a clear preference for a particular outcome does not influence selection. Three measures are used to
measure citation bias and underutilisation – a novel modified density measure, in-degree centrality, and out-degree.

The network in this study is an example of a directed acyclic network [110] because papers can only cite previously published papers and can never cite future literature or cite themselves. To know the density, the proportion of potential edges present, a measure of network connectivity is needed. Generally, if $|P|$ is the sum of primary studies, $|S|$ the sum of reviews, and $E$ the number of edges from $S$ to $P$, then

$$D = \frac{E}{|S| \times |P|}$$

captures the number of times that a set of reviews cites a set of RCTs as a proportion of the maximum possible number of citations, allowing comparison between different classifications of review. Thus, this equation can be used to understand divergent utilisation of primary data.

**In-degree** reflects how often a study has been cited. The in-degree of a vertex is the number of edges from other vertices to that vertex, and, for a vertex $i$, is given by:

$$k_i^{in} = \sum_{j=1}^{n} A_{ij}$$

**Out-degree** reflects the number of references from a particular review. The out-degree of a vertex is the number of edges from that vertex to other vertices. For a vertex $j$, this is established by:

$$k_j^{out} = \sum_{i=1}^{n} A_{ij}$$

To establish citation bias, the network was divided into three sub-graphs of the RCTs with reviews of each classification (supportive, neutral, opposing) and only edges from reviews to RCTs. The vertices that cite only RCTs of a particular classification were then counted – providing an exact account of citation bias and research utilisation.

Finally, a Pearson’s chi square test was performed on the distribution of citations to these four RCTs between the three classifications of review to
establish whether the differences in citation behaviour is significant or simply the product of chance.

3. Results

Of the four RCTs, one was supportive of dietary intervention [10] in the context of secondary prevention, while three were not [8, 9, 11]. Between 1969 and 1984, these were cited in 62 reviews, 28 of which were supportive of dietary intervention [78–105], 17 were unsupportive [61–77], and 17 'neutral' reviews came to no clear conclusion on this issue [44–60]. Of these, 28 were published in the first 8 years (12 supportive; 12 neutral; four unsupportive), and 34 in the next 8 years (16 supportive; five neutral; 13 unsupportive).

Fig 7 shows the citation network. Of the four RCTs, the most cited was the Oslo study, with 57 citations. This was the only one of the four RCTs that was supportive of dietary intervention, but the other three (unsupportive) RCTs together attracted a similar number of citations (32 to MRC Soya-bean, 15 to Rose Corn, 17 to Research Committee Low-fat). Here, citation bias, particularly by supportive reviews, is signified by the many studies that cite only the Oslo study. Clearly, the distribution of citations to these studies is uneven. A high proportion of the reviews (27 of 62; 43%) cited only one of the RCTs, 20 cited two of them, six cited three and only nine cited all four. Nevertheless, taking the three unsupportive trials together, supportive trials are cited 57 times and unsupportive trials 64 times. Overall the graph has a density $D$ of 0.49 – indicating only 49% of the possible citations exist between reviews and these RCTs.
Fig 7: Citation network mapping the citations from reviews to RCTs testing dietary fat modification in the secondary prevention of CHD \((n=66; m=121)\). Vertices represent papers and the integer label corresponds to a paper in this thesis’ bibliography. Vertex colour corresponds to a study’s classification (blue—supportive, green—neutral, red—unsupportive). Vertex size is proportional to in-degree (total number of incoming citations). Each edge corresponds to a specific citation link and the direction of the edge is established by the direction of the arrow head. Visualised with Gephi’s ‘ForceAtlas2’ algorithm [109].

To create an exact account of the utilisation of the primary RCT data by reviews, three sub-graphs were constructed including only reviews of a particular classification (supportive, neutral, unsupportive) and only edges from reviews to RCTs.

Neutral reviews

Of the neutral reviews (Fig 8), one cites only supportive data, which represent the only case of citation bias in this group. Overall the density of this sub-graph \((D=0.68)\) is higher than that of the network as a whole, indicating that reviews which cited more of the available primary literature were more likely to come to no clear conclusion than the average. While the Oslo Study received most citations (17), its citation influence is nearly rivalled by the MRC Soya-bean Study, which received 15; the Rose Corn Oil Study and the Research Committee Low-fat Study were each cited seven times. Six of the reviews cited
all four RCTs and only one cited only one of them. Thus, compared to the whole network, neutral reviews were more likely to fully utilise the available RCT evidence and less likely to cite only one RCT.

Fig 8: Sub-graph mapping the citations from neutral reviews to RCTs testing cholesterol lowering diets in the secondary prevention of CHD ($n=21; m=46$). Vertices represent papers and the integer label corresponds to a paper in the bibliography. Vertex colour corresponds to a study's classification (blue—supportive, green—neutral, red—unsupportive). Vertex size is proportional to in-degree. Each edge corresponds to a specific citation link, and the colour of an edge indicates the classification of the cited article. This was visualised manually in Gephi [108] by lining up reviews in the centre, placing supportive the RCT below and the unsupportive RCTs on top.

**Unsupportive reviews**

The sub-graph of unsupportive reviews (Fig 9) had a density of 0.56. The *Oslo* study and the *MRC* study were each cited 12 times by unsupportive reviews, the *Rose* study six times and the *Low-fat* study eight times. Five of the 17 unsupportive reviews cited only evidence that supported their position. Three reviews cited only the supportive trial, implying a difference of interpretation between the reviewers and the authors of that study. Therefore, in total eight reviews (47%) exhibit citation bias for trials of a particular outcome. Three
reviews cited just one RCT and two cited all four; most (nine) cited two of the RCTs.

**Fig 9:** Sub-graph mapping the citations from unsupportive reviews to RCTs testing cholesterol lowering diets in the secondary prevention of CHD (n=21; m=38). Vertices represent papers and the integer label corresponds to a paper in the bibliography. The colour of a vertex corresponds to a study’s classification (blue—supportive; red—unsupportive). Vertex size is proportional to in-degree. Each edge corresponds to a specific citation link and the colour of an edge indicates the classification of the cited article.

**Supportive reviews**

Compared with the neutral and unsupportive reviews, the 28 supportive reviews (Fig 10) used much less of the available primary data ($D=0.33$ vs $0.56$ in unsupportive reviews and $0.68$ in neutral reviews). All 28 cited the single supportive RCT, the *Oslo* study, while the three other RCTs were cited only nine times. In all, 23 of the supportive reviews (82%) cited only the *Oslo* study, and only five of these reviews cited any of the other RCTs (two cited one other, two cited two others, and just one cited all four RCTs).
Fig 10: Sub-graph mapping the citations from supportive reviews to RCTs testing cholesterol lowering diets in the secondary prevention of CHD ($n=32; m=37$). Vertices represent papers and the integer label corresponds to a paper in the reference list. The colour of a vertex corresponds to a study’s classification (blue—supportive, red—unsupportive). Vertex size is proportional to in-degree. Each edge corresponds to a specific citation link and the colour of an edge indicates the classification of the cited article.

Table 3 gives the distribution of citations to RCTs from different classifications of review.

**Table 3: Contingency table of the distribution of citations from three different classifications of review to four secondary prevention RCTs.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neutral</strong></td>
<td>17</td>
<td>15</td>
<td>7</td>
<td>7</td>
<td>46</td>
</tr>
<tr>
<td><strong>Unsupportive</strong></td>
<td>12</td>
<td>12</td>
<td>6</td>
<td>8</td>
<td>38</td>
</tr>
<tr>
<td><strong>Supportive</strong></td>
<td>28</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>37</td>
</tr>
<tr>
<td><strong>Column Totals</strong></td>
<td>57</td>
<td>32</td>
<td>15</td>
<td>17</td>
<td>121</td>
</tr>
</tbody>
</table>

A Pearson’s chi square test was performed to test whether this distribution is significantly related to classification or the product of chance. The validity of a chi square test depends on the conditions that the expected count is greater than 5 for at least 80% of the cells, and that each cell has a count of at least 1: these conditions were both satisfied. The test result establishes that the difference observed in the citation behaviours between reviews is significant ($\chi^2(6, N=121)=18.19, P = 0.006$).
4. Discussion

This study examined the propagation of experimental evidence derived from RCTs in reviews of the literature during a period of scientific controversy. This was not a minor academic dispute; this was an issue of considerable public health importance at the time. It documented a major divide between reviews over what the available empirical evidence was, how it should be evaluated, and what the implications were for dietary intervention in the secondary prevention of CHD. Overwhelmingly, reviews that disseminated a supportive evaluation of the results of RCTs in the context of secondary prevention cited only data that supported this position. They did not critique the discrepant data, they simply ignored it, or were ignorant of its existence.

While the Oslo trial, a supportive RCT, was by far the most cited study, its influence was not uniform amongst different classifications of review. This was the only RCT to arrive at statistically significant findings. Several studies have reported evidence of a bias in the literature, in that studies with statistically significant results tend to be cited more often than studies that report non-significant or equivocal results [21, 111–114]. However, such a bias for significant outcomes fails to explain the discrepancies between the citing behaviours of different groups.

The degree of utilisation of primary data by reviews appeared to be mediated by a review’s position, with supportive reviews likely to use less of the available RCT evidence than others. Again, this may be explained by a preference for statistically significant findings, but, as discussed, there are problems with this explanation. An alternative hypothesis may be that the shape of the literature itself determines the selection of RCT evidence – where reviewers find evidence from other papers. In this case, if a reviewer believes another paper to have identified all relevant evidence, they may replicate selective citation. This has been described by Greenberg’s information cascade hypothesis – where chains of citations from one paper to another lead researcher’s to only a segment of the available evidence, if those authors simply follow bibliographies [20]. This hypothesis may be supported by Simkin and Roychowdhury’s [115] finding that on average only 20% of the cited
literature in bibliographies has been read by the authors. If scientists do tend to gain their information about primary studies by reading other papers rather than the originals, this is particularly worrying in this case due to the amount of bias in the selection of evidence by reviews and the degree of underutilisation.

Citation bias was considerably more prevalent in supportive reviews. This seems to be an example of homophily - the tendency for a vertex to share an edge with a vertex with which it shares an attribute, and is consistent with Heider’s [116] theory that individuals choose relationships that reinforce existing preferences and beliefs. However, this fails to explain the unsupportive reviews, which we would expect to only cite evidence that supports their perspective, but this was not observe this.

Few reviews cited all relevant data – with underutilisation common. In the period concerned, conventions governing systematic search strategies were in their infancy. The reviews were mostly of the narrative form, a type that has been criticised for selection bias and the promotion of partial perspectives of the evidence. Indeed, one finding of this paper is that scientists in this field were poor at considering all the available empirical evidence. Bushman and Wells [117] offer experimental evidence of the role of availability bias (a heuristic that describes the tendency of humans to judge the prevalence/probability of an event by the ease in which they are able to recall certain information), which the authors argue influence the propensity for selection bias in narrative reviews compared with systematic reviews and meta-analysis, which impose a rigorous structure of literature searching and study evaluation on the researcher. Importantly, narrative reviews are not the only papers in which citation bias has been documented, Robinson and Goodman [118] documented significant citation bias amongst RCTs addressing the same question and may be explained by the lack of systematic search strategies amongst these papers.

Finally, did the selection of evidence itself explain the divergence of opinion amongst reviews? That is, were these authors only aware of the data they cited? If so, it would seem those underutilising the literature arrive at positions that are based on partial perspectives. Or did pre-existing opinions
held by reviewers determine their selection of evidence? That is, were opinions formed first and evidence selected afterwards? If so, this would be a case of confirmation bias. Here, we see the complexities involved in attempting to determine which direction the arrow of causality points – is evidence formative in scientific opinion? Or do pre-existing opinions dictate the evidence we choose to examine?

These findings offer an alternative to the prevailing notion of what a scientific controversy is, i.e. that scientific disagreement stems from different theoretical positions, where researchers looking at the same evidence can justifiably come to different conclusions [117]. Disagreement in this instance appears to stem from the selection of different sources of evidence. Trinquart et al. [27] came to similar conclusions when examining the controversy over dietary salt restriction. They observed “two almost distinct and disparate lines of scholarship, one supporting and one contradicting the hypothesis” (p.57) – lines that relied on different sources of evidence and which differed in their citing behaviour.

**Implications**

Scientists, at present, are concerned at a wide range of factors that they understand to be distorting science. These include publication bias [119, 120], problems with replication [121], and citation biases [20, 21, 27, 111–114]. Their combined effect has undermined faith in the idea that our sciences are self-correcting, and increasingly commentators fear that our system of knowledge production is broken. Price [122] and later Ravetz [123] raised concerns that as science grew it may experience ‘senility’ – with the evidence-base and citing literature growing too large for effective quality control, and this is a view that is gaining increasing attention [124].

This study focussed on citation bias, a bias that leads to distortions of the evidence-base, whereby some positions are sustained by a biased selection of the evidence and where scientists do not interact with rival perspectives and unsupportive data [20]. This can lead to polarisation, where different scientists draw from different sources of evidence and arrive at opposing conclusions [27]. Unlike these two studies, which examined recent citation practices, the
present study focussed on a historical analysis of divergent citation practices amongst papers drawing different conclusions from the available empirical evidence. Citation bias and research underutilisation were exactly documented, as were the different citation behaviours of different classifications of review.

This case study represents a unique example where the relevant RCTs were published in quick succession (1965-1968), and these remained the only RCTs addressing this particular question prior to a major consensus building conference in 1984. This provided a 16-year citation window, from the time all results of the RCTs were available to researchers to the beginning of this conference, where the evidence-base, as derived from RCTs, remained constant (i.e. no additional secondary prevention RCTs were published that would have impacted the evidence-base and, possibly, citation behaviours). Further, as these RCTs were the first to test this particular treatment, the relevant evidence-base and citation behaviour could not have been influenced by prior studies, at least not evidence derived from RCTs, which may have distorted the results. This provided an opportunity to study how scientists use a small evidence-base over a long and undisturbed period. One surprising findings is that even with a small evidence-base, scientists did not generally cite these studies in a representative manner nor did their citation behaviours improve as this period unfolded, and this challenges our understanding that it may be the size of the literature responsible for this phenomenon.

The major findings of this study demonstrate that citation bias is not a new phenomenon – challenging narratives that suggest our science has become increasingly less reliable in recent years because science has grown too large – and that scientific disagreement may revolve around something simpler than divergent theoretical positions – scientists may just be looking at different sources of data.
5. Post-publication reflection

This study played an influential role in shaping the direction of my research. Prior to this, I had been writing historical case studies, but this study’s success in establishing interesting dynamics of evidence selection focussed my attention on this. I wanted to understand how network analysis could be used to understand how scientists use evidence and how this contributes to the endorsement or rejection of ideas. Further, this study produced several interesting lines of research: What mechanisms might explain why these unsupportive studies were cited less frequently? Could this be explained by the structure of the literature? Or is something else responsible? As network analysis required a large commitment to learn, I decided to devote my PhD to exploring its utility.

Further, as this was, and remains, an area that has received comparatively little scholarly attention, I saw this as an area that held promise for methodological innovation and for generating new insights into scientific behaviour. The published works that have focussed on this and used similar methods have been authored by scientists rather than sociologists or historians.

As an SSK researcher, I thought I could make a novel contribution because of my distinct disciplinary background. I am neither a dietary researcher nor a cardiologist and have no stake in what is ‘true’ in these respective fields, but what interests me is what these methods might tell us about science in general. I saw it as an exciting research area that I could contribute to, and, importantly, believed my discipline might benefit from the incorporation of this method to study selective citation and its consequences.

Following this, I focussed on trying to apply CS-CNA to other forms of trials that were important to the evaluation of the diet–heart hypothesis. However, as the next chapter demonstrates, this was not straight forward and the limitations of the methodology I describe here were exposed. Specifically, problems arose when I tried to apply this method to different kinds of studies, particularly prospective cohort studies. Prospective cohort studies, by virtue of the large number of results typically reported, can be cited in an enormously
varied manner. Further, the four RCTs examined here were published in close succession, but this was not the case with prospective cohort studies and this caused complexities. I realised that network analysis must be adapted to both the kinds of studies analysed, but that also only certain question can be asked in certain circumstances. However, because of the problems I encountered, I discovered the flexibility of the network approach to be moulded around specific questions that could be answered with these techniques. In the following chapter, I offer a modification of CS-CNA to understand how the results of a particular prospective cohort study were cited.
Chapter 5: Hidden in Plain Sight: A claim-specific network analysis of the selection of evidence derived from a prospective cohort study

Introduction

In a paper with many empirical findings, such as a cohort study examining the impact of a range of risk-factors on disease, what findings eventually become known in the scientific community? This question has implications for how we understand the impact of a study from its citation count. We tend to assume that a paper with many citations has had a large impact, but if a paper reports many different outcomes, which of its particular findings is it cited for? With a growing interest in dissemination biases, a study with multiple findings complicates analysis. For example, citation bias is typically detected by comparing how often supportive studies have been cited compared with unsupportive studies over a defined period (Jannot et al. 2013; Kivimäki et al. 2014). While this may be appropriate for analysing unifactorial trials, it may be misleading for multifactorial trials and prospective cohort studies that analyse many variables.

This chapter demonstrates the problem of using citation counts to assess bias via an exploratory study, and demonstrates how claim-specific citation network analysis (CS-CNA) (Greenberg 2009; Leng 2018) can overcome these problems by documenting what findings are actually used in the scientific literature. To do this, I examine how a prospective cohort study (Paul et al. 1963), examining the relationship between a plethora of factors, including dietary fat, and CHD, was cited over a seven-year period. This study followed a debate in the scientific literature over the diet–heart hypothesis – the theory that would form the justification for population-wide guidelines restricting dietary saturated fat in the late 1970s and early 1980s. Yet, Paul et al. (1963), the first prospective observation study to test this theory, returned unsupportive results: it found no association between intakes of dietary
saturated fatty acids (SFA), polyunsaturated fatty acids (PUFA), or total dietary fat intake and the development of CHD.

First, by classifying all citing papers by a quotation analysis of their titles, I establish the likely research focus of each paper and show how this classification is predictive of both what findings will be referenced, and how citing studies reference each other. Second, in a sample of 110 citing papers, I demonstrate how CS-CNA can detect the exact findings referenced in each citing paper. Third, by constructing a network, I show the relationship between title classification and the findings quoted by a study – particular findings are cited by groups focussed on particular research questions, and these groups cluster together by their references and citation links. From this, three major communities of papers were detected – one focussing on the relationship between diet and CHD (n=38), one on the relationship between caffeine and CHD (n=22), and one interested in the epidemiology of CHD, risk-factor research, or simply CHD, but which had no information on which factors were being focussed on beyond the two established risk factors of raised serum cholesterol and blood pressure (n=58).

1. Background

In the lead up to the 1984 NIH cholesterol consensus conference, ten prospective observational (prospective cohort) studies, covering 14 cohorts, had reported findings on the relationship between dietary fat and CHD mortality and morbidity within particular communities (Paul et al. 1963; Gordon and Kannel 1970; Medalie et al. 1973; Morris et al. 1977; Yano et al. 1978; Garcia-Palmieri et al. 1980; Gordon et al. 1981; Shekelle et al. 1981; McGee et al. 1984; Kromhout and Coulander 1984). Most of these studies were unsupportive of the diet–heart hypothesis. The evidence of these trials indicated that dietary fat intake had no significant association with CHD.

In three of these cohorts (Yano et al. 1978; Garcia-Palmieri et al. 1980; and McGee et al. 1984), a significant difference in SFA consumption was observed between those with CHD and those without CHD, linking higher SFA
intake (as % of total calories) with a higher risk of CHD. However, after controlling for other known risk-factors, this relationship was non-significant in two of them (Yano et al. 1978; Garcia-Palmieri et al. 1980). Thus, in 13 of these 14 cohorts no significant relationship was observed between SFA consumption and CHD after controlling for other known risk-factors.

Six of the cohorts showed a significant association between higher intake of PUFA (as % of total calories) and a lower risk of CHD (Yano et al. 1978; Garcia-Palmieri et al. 1980; Gordon et al. 1981 [Puerto Rico and Honolulu cohorts]; Shekelle et al. 1981; and McGee et al. 1984). However, only in three cohorts (Gordon et al. 1981 [Puerto Rico and Honolulu cohorts]; Shekelle et al. 1981) was this relationship significant after controlling for other known risk-factors. Thus, in 11 of these 14 cohorts, no significant relationship was observed between PUFA consumption and CHD after controlling for other known-risk factors.

Nine of these studies were published in peer-reviewed academic journals and one, Gordon and Kannel (1970), was published as a book. According to Web of Science (WoS), these nine papers were cited 836 times by the end of 1984 in 748 unique publications. As Gordon and Kannel (1970) was published as a book, this is not covered in the WoS Core Collection; however, by doing a ‘Cited Reference’ search for the authors and year of publication, a list of references from indexed files to this book were identified – a total of 126 citations by the end of 1984. This seemingly low number of citations is probably due to its publication as a book and because its significant findings (but not its dietary fat findings) were published in journal articles. Together, these studies were cited 962 times between 1963 and 1984 in 846 unique publications. In Table 4, I report, for each study, the number of citation received as of 1984 and January 2018.
Table 4: Citations to ten prospective cohort studies, 1984 and 2018, recorded by the Web of Science, as of January 2018.

<table>
<thead>
<tr>
<th>Study</th>
<th>#times cited as of 1984</th>
<th>#times cited as of 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul et al. (1963)</td>
<td>439</td>
<td>658</td>
</tr>
<tr>
<td>Gordon &amp; Kannel (1970)</td>
<td>126</td>
<td>193</td>
</tr>
<tr>
<td>Medalie et al. (1973)</td>
<td>118</td>
<td>195</td>
</tr>
<tr>
<td>Morris et al. (1977)</td>
<td>92</td>
<td>329</td>
</tr>
<tr>
<td>Yano et al. (1978)</td>
<td>36</td>
<td>73</td>
</tr>
<tr>
<td>Garcia-Palmieri et al. (1980)</td>
<td>13</td>
<td>117</td>
</tr>
<tr>
<td>Gordon et al. (1981)</td>
<td>28</td>
<td>208</td>
</tr>
<tr>
<td>Shekelle et al. (1981)</td>
<td>109</td>
<td>636</td>
</tr>
<tr>
<td>McGee et al. (1984)</td>
<td>1</td>
<td>206</td>
</tr>
<tr>
<td>Kromhout &amp; Coulander (1984)</td>
<td>0</td>
<td>109</td>
</tr>
</tbody>
</table>

But as these studies reported findings on a diverse array of other factors, were the results regarding dietary fat noticed? This is hard to study via CS-CNA for three reasons: (i) these studies were published over a reasonably long period, and we would expect their citation rates to vary with time and with additional studies added to the evidence-base; (ii) all papers published results on the relationship between other dietary and non-dietary variables and CHD; (iii) some studies also reported findings from between cohort comparisons and not only within cohort associations; (iv) a large volume of literature cited these studies, but it is not possible to know which findings were cited in these documents without reading them all. These issues meant that the claim-specific approach devised in the previous chapter would be very difficult to apply. For example, Medalie et al. (1973) report “over 100 single variables with the development of a first myocardial infarction” (p.329) in an Israeli population, while Gordon et al. (1981) report associations within three major cohorts in Framingham, Honolulu, and Puerto Rico, but also perform a between-cohort analysis.

This study attempts to devise a method that overcomes these problems to investigate the prevalence of sub-optimal citation practices (citation bias, citation distortion, research underutilisation). To do this, I focussed on the earliest prospective study to gauge how it was cited to understand what would be required to evaluate how all of these prospective studies were cited.
However, after performing this small study, I deemed it too complex and time-consuming to cover all prospective cohort studies.

**The Western Electric Study**

This study examines how the *Western Electric Study* (Paul et al. 1963), was cited in the seven years between its between its publication date and that of the next prospective cohort study (Gordon and Kannel 1970). Paul et al. (1963) was a study of 1,989 men aged between 40 and 55 years and free of CHD from the Hawthorne Works of the Western Electric Company in Chicago. It reported results after 53 months of follow-up, by which time 88 men had developed CHD (47 angina pectoris; 28 myocardial infarction (MI); 13 deaths from CHD). On a univariate analysis, SFA, unsaturated fatty-acids, total fat, and dietary cholesterol were not associated with CHD. Those *without* CHD ate an average of 3,174 calories/day and 59 g/day SFA; 83 g/day unsaturated fatty acids; 757 mg/day cholesterol; 12.28 g/day linoleic acid. Those *with* CHD ate an average of 3,082 calories/day and 59 g/day SFA; 80 g/day unsaturated fatty acids; 721 mg/day cholesterol; 11.57 g/day linoleic acid.

The authors split the cohort into high fat and low fat sub-groups, each composed of 296 men who represent the top and bottom 15% of fat intake. There were 16 CHD cases in the low-fat group and 14 in the high fat group. Men in the high-fat group ate an average of 49.3% calories from fat compared to 36.1% in the low-fat group. There was a significant relationship between serum cholesterol level and CHD, but no significant association between dietary fat and serum cholesterol. The authors concluded that:

*It is true that our population does not include a low-fat group comparable to populations found in certain other parts of the world…The findings noted contrast with the step-wise association observed with certain other factors and would seem to inject a healthy note of caution at least toward attempts to alter the American diet within the modest range of fat intake we have described (p.30)*

While Paul *et al.* (1963) found no association between dietary fat and CHD, it did observe significant relationships between CHD and elevated serum cholesterol; elevated blood pressure; higher coffee intake; cigarette smoking;
family history (early death of father); history of peptic ulcer; history of chronic cough; history of 'non-cardiac' chest discomfort; increased skin fold thickness; arteriovenous nicking in the fundi; and abnormalities in the electrocardiogram. The authors also drew attention to the lack of association between CHD and body weight; all dietary variables (excluding coffee); mean blood sugar levels; lipoprotein lipase levels; job type; and physical activity associated with particular jobs. Finally, the authors also highlight their finding that serum cholesterol levels in this population appeared to fluctuate seasonally – peaking in the winter months, falling throughout the spring, reaching a low in the summer months, where it rose throughout autumn – but they could not explain this fluctuation by any of their collected variables.

Between 1963 and 1970, Paul et al. (1963) was cited 180 times (Web of Science, January 2018), but why was this study being cited?

2. Materials and Methods

Study selection

To be included, a study had to be: (i) a prospective cohort study – i.e. dietary information was collected on a defined cohort before the development of primary outcomes; (ii) published before 1985; (iii) examined the effect of dietary fat on CHD via dietary assessment in a sample OR the entire cohort population; (iv) published data on sample numbers AND dietary intake AND serum cholesterol AND mortality and morbidity figures regarding incidence of myocardial infarction (MI), CHD death (including MI and sudden coronary death), non-fatal CHD (angina pectoris, non-fatal MI), and all-cause mortality; (v) included patients either initially free of CHD (primary prevention) OR those with established CHD (secondary prevention); (vi) a minimum follow-up of 24 months.

Prospective observational studies were identified via examination of reviews, systematic reviews, and meta-analyses that have examined this area previously. From this, ten published studies were identified that met the inclusion criteria. From the ten studies, Paul et al. (1963) was identified as the
first prospective study to publish dietary findings, and it remained the only study to do so until the publication of the *Framingham* study’s dietary findings in 1970.

All papers published in academic journals in English between 1963 and 1970 and which cited Paul *et al.* (1963) were identified by searching WoS, Scopus, and Google Scholar using indexed citation links to this paper. In January 2018, a search in WoS returned 180 citing papers, Scopus returned only six (due to lack of coverage before 1970), and Google Scholar returned 146. Thus, as WoS was the most comprehensive citation set, this was selected for an initial analysis.

**Data acquisition**

Citation data for Paul *et al.* (1963) and 180 citing papers were downloaded via WoS “Full Record and Cited References Function” in a .txt file. Sci² Tool version 1.3 (2009) was then used to extract reference data from this file, and output files were exported for manual cleaning in Gephi version 0.92 (Bastian *et al.* 2009) to merge duplicate publication records and preserve citation links both between Paul *et al.* (1963) and its citing articles and between all citing articles. All network visualisations were produced via Gephi’s ForceAtlas2 algorithm (Jacomy *et al.* 2014).

**Title Classification: Research Focus**

All papers were classified by a quotation analysis of paper titles. For example, a paper was assigned “Diet” if a title referred to the relationship between diet and CHD or serum cholesterol levels. A paper was assigned “CHD” if a paper contained only terms indicative of CHD but had no information on which factors were being focussed on beyond the two established risk-factors – serum cholesterol and hypertension. A paper was later assigned the group “Other” if the focus of a paper was rare in the dataset (<3%) – e.g. two papers focussed on the role of lifestyle on the development of duodenal ulcer. On occasions, a paper’s title signified multiple focusses – e.g. diet and caffeine – and these papers were initially assigned to both groups. Thus, this exercise is designed to establish the likely research focus of a paper. As no key-words or abstracts
are indexed for the period examined here, this was the only feasible method for quick classification of paper topic.

From this, 38 papers were detected that focussed on the role of diet in CHD. These were further sub-classified by the specific dietary components focussed on. For example, a paper titled “Sugar and Ischaemic Heart Disease” is classified into ‘Carbohydrate’, a paper titled “Fat-controlled Diets” is classified into ‘Fat’, a paper with either multiple dietary elements in the title “Dietary fat + Dietary sugar” into ‘Carbohydrate/Fat’, and papers with only the word diet or nutrient rather than anything more specific are classified into the “Diet” group. This produced 19 papers in the ‘Diet’ group; 10 papers in the ‘Carbohydrate’ group; three papers in ‘Carbohydrate/Fat’; three papers in ‘Fat’; two papers in ‘Vitamin C’, one paper in ‘Dietary Cholesterol’.

**Quotation analysis**

All citing papers were read, and quotations copied into an Excel spreadsheet that captured how each citing paper was citing Paul et al. (1963) (see Ch5.S1). First, the exact findings being referenced were recoded. I coded into the dataset the general category of finding being referenced (e.g. dietary association, body weight, serum cholesterol). Second, any quotation regarding dietary fat was further classified as accurate or inaccurate based on whether the authors correctly reported these findings. Due to access issues\(^{12}\), an initial sample of 110 papers from this set of 180 citing papers was retrieved.

**Network construction**

Each paper was given an integer identifier code. All authors, title of paper, year of publication, journal of publication, paper type (article, review, editorial, letter, discussion, other), title classification, and the specific findings referenced from Paul et al. (1963) were recorded to identify each vertex in a ‘vertex list’ format (see Ch5.S2). This attribute information was used to examine the relationship between classification variables and citations behaviours. The citations to (incoming edges) and from (outgoing edges) all papers were recorded. The

---

\(^{12}\) The remaining articles were not accessible via the University library subscription. As a scoping study, I believed a sample would be sufficient for my purposes here.
bibliographies of each review were screened to ensure that all citations to other papers were identified within this sample. Citation information on the relationship between papers was recorded in an ‘edge-list’ format in an Excel spreadsheet (see Ch5.S3). From these data, a graph was constructed and analysed in Gephi. All associated data used in this chapter can be found in the Chapter 5 Supplement.

3. Results

Who cited these findings?

A total of 180 papers cited Paul et al. (1963) between 1963 and 1970. By quotation analysis, I assigned each paper to one or more research focus classifications. 38 papers were assigned to the Diet group, with 8 of these having an additional focus. Three were on the role of diet and exercise in CHD; two on the role of diet, stress and smoking on CHD; one on diet and obesity; one on diet and its relationship to hypertension; and one on diet and caffeine in CHD. 22 papers were assigned to the Caffeine group, with only one paper with a second research focus (diet, specifically focussing on sugar [carbohydrate group]). 58 papers were assigned to the group “CHD”. These included papers on the epidemiology of CHD, risk-factor research, or CHD, but had no information on which suspected causal factors were being focussed on. However, of these, five papers contained terms indicating an interest in factors influencing serum cholesterol levels, four papers contained terms also relevant to hypertension, and three papers contained terms indicating a secondary focus on diabetes.

Of the remaining 62 papers, 12 examined the effects of smoking on CHD and/or hypertension; 11 focussed on the role of socio-psychological variables and CHD; six on physical activity and CHD; four specifically on hypertension; five on screening and detection methods in epidemiological research; three on familial influences on CHD risk; two on stroke, two on obesity, two on respiratory disease, two on hyperglycaemia. Seven papers contained only German terms in the title and were coded simple into category German. A
further six papers with single focuses ranging from duodenal ulcers to gallbladder disease were classified into “Other”.

What findings were cited?

Out of a sample of 110 papers (out of a possible 168\textsuperscript{13}), 81 cited a single finding, 16 cited two, while only 13 cited three or more. In total, 13 papers cited this study for its findings regarding diet and CHD. Of these, eight explicitly report the dietary fat and CHD findings (Shillingford et al. 1964; Masironi 1970; Marquis et al. 1964; Papp et al. 1965; Epstein 1965), though only two referred specifically to the type of dietary fat (Yudkin and Roddy 1964; Yudkin and Morland 1967). One of these papers, Parkes et al. (1969), cited Paul et al. inaccurately by suggesting that it had found evidence dietary fat was correlated with CHD. The other five papers cited the lack of association of all dietary variables examined, but did not explicitly discuss dietary fat (Little and Shanoff 1965; Paul et al. 1968; Matsumoto 1970; Anonymous 1964); however, one of these (Weiner et al. 1965) reported inaccurately to suggest Paul et al.’s findings suggested dietary tracking may improve prognosis.

Dietary results were discussed in different contexts in another ten papers. Two papers cited this study for its finding of a lack of association between dietary fat and seasonal fluctuation in serum cholesterol levels (Stulb et al. 1965; McDonough et al. 1967), and another paper reported the lack of association between diet and serum cholesterol but not CHD (McGandy et al. 1967). Three studies quote the dietary intake findings to compare to intakes in different populations (Zukel 1965; Basset et al. 1966; Basset et al. 1969). One paper discusses the problem of underreporting in dietary questionnaires (Yudkin and Roddy 1966); one paper discusses the dietary findings in a vague and indirect manner – highlighting the lack of sharp differences in diet (and exercise) between those with and without CHD within particular cohorts, but does not report either the lack of association between diet and CHD nor its specific dietary fat findings (Keys et al. 1967). One study (Graber and Boltjes 1968) erroneously reports a significant relationship between sugar

\textsuperscript{13} Excluding seven German papers and a further five papers that were found on inspection to be in Spanish, French, and Russian.
consumption and CHD, which is likely a result of citing the 1963 publication rather than a follow-up study that examined sugar consumption specifically (Paul et al. 1968). One paper cites the lack of association between salt consumption and hypertension (Malhota 1970).

Other associations from Paul et al. (1963) were reported both more often and with greater consistency. The significant association between serum cholesterol and CHD was reported in 30 papers, and the significant association between blood pressure and CHD was cited 19 times. The significant association between coffee and CHD was cited 22 times, with only one paper inaccurately reporting this finding (reporting no association). The significant relationship between cigarette consumption and CHD was reported in 18 papers.

The body fat findings were cited nine times. In 12 papers, Paul et al. (1963) was cited as an example of a prospective cohort study. Four cited it for the detected electrocardiographic abnormalities; one for the lack of association with gallbladder disease; six papers for the methodology utilised, three citations to the lack of association between occupation and CHD, and a further three to the lack of association with physical activity.

Thus, the overwhelming majority of papers cited Paul et al. (1963) in a manner that suggests that its findings were used to propose or evaluate certain risk-factors of CHD. The most highly cited finding was the significant association with serum cholesterol and CHD (27% of sample). As the diet–heart hypothesis was the prevailing explanatory theory concerning the elevation of serum cholesterol and a leading hypothesis to explain the rise of CHD in the post-war years, we might expect papers that cited this serum cholesterol finding to also cite the lack of association with dietary fat. However, in 23 of the 30 (77%) papers that cited the association of raised serum cholesterol with CHD, there was no mention made of the lack of association between diet and either CHD or serum cholesterol.
Why were findings being cited?

Fig 11 shows the citation network of a set of 181 papers – Paul et al. (1963) and all citing papers between 1963 and 1970. All citations have been preserved between all papers and Paul et al. (1963), and between all citing papers to capture their interaction with one another. Paul et al. (1963) is the largest vertex with an in-degree of 180, which indicates the number of times it has been cited by other papers in this network.

Fig 11: Citation Network representing citations between papers referencing Paul et al. 1963 (n=181; m=514)

This visualisation reveals a clustering on the left and right hand sides of the graph. To improve on this, vertex attribute data was included that specified the title classification for each paper. In Fig 12, I colour only vertices in the ‘Diet’ and ‘Caffeine; title classification groups. Recall Paul et al. (1963) found no significant association between any of the dietary variables included in their study and CHD apart from between CHD and coffee consumption.
Fig 12: Citation network including diet and caffeine title classifications ($n=181$; $m=514$). Vertices coloured red refer to the diet group, blue to the caffeine group, and grey to all others.

Fig 12 shows that vertices in this network are clustered primarily around research focus, as expected. But how well do these communities, now established by their reference relationships and by their title classification, help us to understand which papers are citing what findings? Fig 13 shows which papers were citing the dietary findings ($n=11$), the caffeine findings ($n=20$), and both ($n=2$). To do this, quotation classification attribute data was included. Red vertices refer to papers quoting the dietary findings, pink vertices represent the two inaccurate dietary references; blue vertices refer to papers referencing caffeine findings, and the green vertices represent papers that cited both of these findings.
Fig 13: Citation network displaying papers citing diet and caffeine findings (n=181; m=514). Vertices coloured red refer for papers referencing dietary findings; pink—inaccurate dietary references; blue—papers referencing caffeine finding; green—both diet and caffeine; grey—all others.

Fig 13 demonstrates that the communities that appear to be clustered around research topics are also likely to cite the same finding; thus, what findings are cited from a study is reasonably well predicted from title classification. Ten of the citations to the caffeine finding come from papers in the caffeine title group, six were from the dietary title group, and one citation comes from a paper in the CHD title group. Regarding the dietary findings, six citations derive from the dietary title group, five from CHD title group, and two from papers on socio-psychological title group. Of the papers that cited both findings, one was from the dietary title group, while another was from the CHD title group.

Sub-graph Analysis

There appears to be structure to the interaction between papers in this network that appears to be non-random. To understand this structure, I removed Paul et al. (1963) from the network, as this vertex expresses no important relational information, all isolates (30 papers that do not cite any other citing papers) as these offer no relational information, and removed any foreign-language
documents. This produces a large weakly-connected component containing 145 papers. I re-ran the ForceAtlas2 algorithm to capture the shape of the network based only on the relationship between connected citing papers, re-sized vertices based on their relative in-degree, coloured vertices based on same quotation classification as in Fig 13, and labelled all vertices with more than ten citations within the network.

Fig 14: Weakly-connected component showing interaction between citing papers classified by quotation analysis of referenced findings ($n=145; m=308$). Vertices sized by in-degree, blue–caffeine; red–diet; green–diet+caffeine; pink–inaccurate dietary reference; grey—all others; labels attached to vertices with in-degree greater than ten.

Fig 14 demonstrates that documents that reference the same findings tend to cluster together via their reference and citation links, but also cluster around particular highly cited studies. Of the vertices in the caffeine and diet title groups, five studies have more than ten citations and these appear to be partly responsible for the clustering.

To understand this cluster further, I inserted title classification back into the graph, but this time included the CHD, smoking, socio-psychological, and physical activity groups alongside diet and caffeine (Fig 15).
Fig 15: Network displaying largest-component with vertices coloured by title classification \((n=145; m=308)\). Vertices sized by in-degree, blue—caffeine; red—diet; yellow—CHD; green—socio-psychological; physical exercise—pink; black—smoking; grey—all others; labels attached to vertices with an in-degree greater than ten.

Fig 15 demonstrates that papers with the same title classification tend to be clustered together, and these are also clustered around different studies than those in the diet and caffeine groups.

A close relationship between the diet group and caffeine group is indicated by the number of times these groups cite each other’s primary vertices. In Table 5, I report the number of citations from within this network to the five most highly-cited publications.
Table 5: Citations to highly-cited papers in the dietary and caffeine title groups and citations contributed by each group.

<table>
<thead>
<tr>
<th>Vertex</th>
<th>Title Class</th>
<th>Citations</th>
<th>Citing papers of the same class</th>
<th>Citing papers in the other class</th>
<th>Citations from Diet and Caffeine classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yudkin and Roddy (1964)</td>
<td>Diet</td>
<td>23</td>
<td>16 (70%)</td>
<td>3 (13%)</td>
<td>19 (83%)</td>
</tr>
<tr>
<td>Little et al. (1965)</td>
<td>Diet</td>
<td>17</td>
<td>9 (53%)</td>
<td>8 (47%)</td>
<td>17 (100%)</td>
</tr>
<tr>
<td>Leren (1966)</td>
<td>Diet</td>
<td>11</td>
<td>3 (27%)</td>
<td>0</td>
<td>3 (27%)</td>
</tr>
<tr>
<td>Little et al. (1966)</td>
<td>Caffeine</td>
<td>15</td>
<td>14 (93%)</td>
<td>1 (7%)</td>
<td>15 (100%)</td>
</tr>
<tr>
<td>Bellet et al. (1965)</td>
<td>Caffeine</td>
<td>15</td>
<td>14 (93%)</td>
<td>1 (7%)</td>
<td>15 (100%)</td>
</tr>
</tbody>
</table>

The caffeine group is tightly clustered around Bellet et al. (1965) and Little et al. (1966). These papers are primarily interested in understanding how caffeine consumption, particularly from coffee, could be implicated in CHD and whether caffeine affects serum cholesterol and blood glucose levels.

The diet group is more complex. While most dietary papers are clustered around Yudkin and Roddy (1964), and these tended to be the papers citing the dietary findings, a few documents with titles indicating a dietary focus are grouped between Leren (1966) and Epstein (1965) [from the CHD title group] and these appear different in both their research focus and citation behaviours. While Leren (1966) has a title suggesting a dietary focus, indeed this is the Oslo study discussed in the previous chapter, yet he does not cite the dietary findings from Paul et al. Leren (1966) appears, from its citation behaviour and the behaviour of the papers citing it, to have more in common with those in the CHD group than it does to other papers in the large diet and caffeine clusters. Why?

As discussed, the Diet group was also sub-classified into groups that reflected the specific nutrient or food group included in a paper’s title. This identified that papers with either sugar or carbohydrate constituted the largest group of papers (n=13; two of which also included terms related to dietary fat in the title, one also included caffeine). During this period, Yudkin, a British physiologist and long-time critic of the diet–heart hypothesis, proposed a rival
hypothesis linking sugar intake with an increased risk of CHD. According to Yudkin, the rise in sugar consumption in the post-war years better explained the increase in CHD experience in the Western, developed world than dietary fat. As we see from Fig 14 and 15, most studies examining the role of diet in CHD were collected around his 1964 publication – is this then a community examining the sugar hypothesis? To investigate this, the carbohydrate sub-classification (light blue) was added to the network.

Fig 16: Network displaying largest-component with dietary carbohydrate and other dietary title classifications (n=145; m=308). Vertices sized by in-degree, light blue–carbohydrate; red–other dietary papers; green–carbohydrate+fat; grey–all others.

Fig 16 shows a community of papers clustered around Yudkin and Roddy (1964) that appear to be interested in the sugar hypothesis. However, this group’s interest in Paul et al. (1963) could not have been regarding dietary sugar findings, as this study did not investigate this specific dietary variable. In 1968, Paul et al. would go on to investigate this in a follow-up study that re-analysed their data for sugar in-take. So what findings were the sugar community interested in? In the paper by Yudkin and Roddy (1964):
As regards diet, they [Paul et al. 1963] specifically mention the lack of association between any of the dietary constituents that they considered, and the development of I.H.D. [CHD] in 88 of their subjects. They considered calories, protein, total carbohydrate, fat, animal fat, total saturated and total unsaturated fatty acids, individual unsaturated fatty acids, and a number of mineral elements and vitamins...[However] they do not appear to have calculated the intake of sugar, or its relation to the development of disease (p.7).

Here, the authors correctly state many of the dietary findings of Paul et al. (1963) and highlight the absence of reported sugar intake in this study. These findings appear to play two roles in this paper: (i) an attempt to falsify or discredit the diet–heart hypothesis; (ii) a call to the scientific community to pay greater attention to the role of sugar in CHD. As Fig 14 demonstrated, the findings regarding the relationship between dietary fat or diet and CHD were primarily confined to a cluster surrounding this Yudkin and Roddy (1964), and many appear to mirror their position. Indeed, only three papers outside of this cluster cited the lack of association, and two of these did so inaccurately – while an additional three papers cited the dietary findings but were unconnected to any other citing paper (Fig 13).

However, one reasonably highly-cited paper was in the centre of the CHD cluster and this did correctly cite the lack of association between dietary fat and CHD – Epstein's (1965) review of CHD epidemiology. Epstein was a close ally of Keys, regarded by many as the founder of the dietary–heart hypothesis. Here, Epstein offers a subtly different interpretation to Yudkin:

A series of investigations have suggested that serum cholesterol and dietary fat intake are not correlated in Western countries. This is well illustrated in the study by Paul and his associates who found no difference in coronary disease incidence among men in the highest and lowest quintile of fat intake; however, even the men in the lowest quintile consumed 36 per cent of their calories as fat and their average serum cholesterol level was as high as 247 mg/100 ml so that, as the authors indicate, minor alterations in the so-called ‘average’ American diet are not likely to achieve the desired goal. At the same time, the experiences already quoted do not suggest that these alterations must be so drastic as to be unacceptable (p.762)
For Epstein, Paul *et al.* failed to find an association because the population studied were *all* eating too much fat. By this logic, the lack of association was not a demonstration that dietary fat had no influence on CHD risk, but, rather, the population under study concealed a true association because of its high average fat intake.

Yet, Epstein holds a different position in the network to Yudkin. In Fig 17, I colour vertices by whether they cited the dietary, serum cholesterol, or caffeine findings. As I’m interested in the dietary findings primarily, in cases where multiple findings are cited, the dietary finding takes precedence. In yellow, we see all the papers that cited the serum cholesterol finding, but not the dietary findings, red reflects paper citing the lack of association between diet and CHD, pink represents inaccurate reporting of dietary findings, while blue indicates caffeine finding. Epstein is in an area of the network that rarely cited the dietary fat finding, and this area of the network contains advocates of the diet–heart position – Keys, Stamler, and Leren.

**Fig 17:** Network displaying largest component with quotation classifications of references to serum cholesterol, diet, and caffeine findings (*n*=145; *m*=308). Vertices sized by in-degree, blue–caffeine; red–diet; pink–inaccurate diet; yellow–serum cholesterol; grey–other
In Keys et al. (1967), Paul et al. (1963) is cited in an indirect manner, but in a way that suggests that Epstein’s interpretation was being endorsed.

The incidence of coronary heart disease among middle-aged men is such that in many populations a few thousand subject-years follow-up will provide an adequate number of cases for statistical analysis. Such a study was started in Minnesota in 1947 […] and others began later at Framingham […], at Albany […], at Los Angeles […], at Chicago [Paul et al. 1963]…Important findings are emerging from these follow-up programs in the United States. It is gratifying to note the consistency in the results but this also points to several limitations…These American population samples are characterized by a relatively high degree of homogeneity in certain respects of mode of life so that they throw little light on the effect of differences in mode of life…Physical activity and diet, too, do not show great contrasts among the samples and even within the samples variability in these items is restricted…Universals in the disease will be clear only when populations of differing habits and cultures are included in broader studies (pp.13 – 14).

Yet, in an earlier publication, Keys (1966) references only the significant serum cholesterol association from Paul et al. (1963). Stamler (1968a, 1968b) references the serum cholesterol association and smoking associations, respectively. Leren (1966), despite reporting in his own dietary trial, references only the serum cholesterol finding. Thus, the dietary fat findings appear to have had a major influence on those interested in the link between sugar and CHD, but, oddly, there appears to be largely an absence of a community discussing these results from a group primarily interested in the relationship between dietary fat and CHD. This may be because this community accepted the theoretical interpretation offered by Epstein and Keys. Indeed, this was Keys’ rationale for launching his multi-country epidemiological study to compare diet and incidence of CHD across different cohorts in different countries because he believed it was only through this method that a true association would be uncovered.

But why was Yudkin citing the caffeine findings and why is the cluster that surrounded him wedged between the CHD and Caffeine clusters? In Yudkin and Roddy (1964), after discussing the dietary fat finding, the authors write:
Subjects who developed ischaemic heart-disease drank significantly more cups of coffee than the control subjects. Though they make no comment on this observation, it is likely that the amount of coffee drunk was an indication of a higher intake of sugar, just as we ourselves have found that our English patients with arterial disease drink more cups, mostly of tea. (p.8)

Here, Yudkin and Roddy (1964) are suggesting that the link between coffee and CHD observed by Paul et al. is a spurious association caused by the confounding influence of sugar intake. Thus, this suggests why the community clustered around the Yudkin publication is also in close proximity to the caffeine cluster – was there a debate between these camps over whether the coffee findings were the result of confounding?

Bellet et al. (1965) examined the effect of caffeine on serum free fatty acids. The authors cite Paul et al. (1963) to contextualise their own findings:

A recent epidemiological study has shown a positive association between coffee intake and coronary heart disease. The FFA [free-fatty acids] effect after caffeine observed in the present study may be related to this clinical observation (p. 752)

Similarly, Little et al. (1965) cited Paul et al. (1963) in a manner to help contextualise their own research:

We found no significant correlations between various nutrients and serum-cholesterol levels in healthy Canadians and patients with coronary heart-disease...The single exception was a positive correlation between daily cups of coffee and serum-cholesterol in the coronary group (r=0.31; P<0.01). Paul et al. (1963) reported a correlation between coffee and the later development of coronary heart-disease (P<0.025). We have extended our inquiry by correlating coffee with a number of serum-lipid and serum-lipoprotein fractions in patients with coronary heart-disease and in controls (p.732)

After Yudkin suggested that sugar may explain their own coffee finding, Little and Shanoff (1965) published a letter calling for caution in regards to Yudkin’s interpretation of the Paul et al. study:
Such a study by Paul et al. showed that, except for coffee, there was no association between diet and later incidence of C.H.D. Unfortunately, sucrose as such was not specifically mentioned. Possibly their data could be reanalysed for dietary sucrose, and we are suggesting this to Dr. Paul...[However] until results from prospective studies are available, it should not be accepted that C.H.D. patients ate more sucrose than normal subjects...The observation by Professor Yudkin that the consumption of sugar in our civilisation has greatly increased in recent time may be very significant [but]...replacing these speculative facts will require much more work (pp.184–185).

Thus, it seems that scientists were using Paul et al. (1963) in a manner that supported their pre-existing beliefs. Yudkin saw in these findings a *falsification* of the diet–heart hypothesis and potential support for the sugar hypothesis. Epstein saw quite the opposite; for him, Paul et al. represented *supportive* evidence for the current favoured approach to studying the relationship between diet and CHD and a sign that the whole of the American population would have to lower their fat intake to reduce the level of CHD. Bellet saw a *hypothesis-generating* finding, in the link between caffeine and CHD, while Little used this study to help *contextualise* his own findings regarding an association between the consumption of caffeine and CHD.

4. Discussion

This study demonstrated that the findings from a single prospective cohort study (Paul et al. 1963) were cited in many different ways that appeared to reflect the research interests of particular communities. By identifying all citing papers between 1963 and 1970 and classifying these into research topics by terms within their titles, I demonstrated that this classification helped to explain both community structure as judged by citation relationships, but was predictive of what findings were cited within these citing papers. The three largest groups of papers contained terms indicating a focus on CHD, diet, and caffeine.
In a sample of 110 papers, I demonstrated that most studies cited Paul et al. for a single finding, ~74% of sample. Most papers cited specific findings in a manner that suggested that these findings were used to evaluate risk-factors of CHD. The most frequently cited findings were the statistically significant associations of serum cholesterol, blood pressure, coffee, and smoking with CHD.

The lack of association between diet and CHD was discussed in just 13 papers. Of concern, this lack of association was not reported in 23 of the 30 papers that cited the association between serum cholesterol and CHD. As dietary fat was the leading hypothesis to explain why different populations had different levels of serum cholesterol and different rates of CHD, this finding should have been cited. As serum cholesterol was significantly associated with CHD in Paul et al. (1963) but dietary fats had no association with either serum cholesterol level or CHD rate, this finding required close scrutiny.

The dietary fat findings made the biggest impact not on those interested in the diet–heart hypothesis, but in a community advocating a rival hypothesis – the sugar–heart hypothesis. After analysing why particular findings were being cited by examining the discourse surrounding a citation, I demonstrated that the most highly-cited paper in this group (Yudkin and Roddy 1964) took the dietary fat finding as a likely falsification of the diet–heart hypothesis. This paper also interpreted the significant association with caffeine as a potentially supportive finding towards the sugar hypothesis because these authors believed it a confounded relationship that concealed the true culprit – not the coffee being drunk by the men in this population, but by the sugar that they likely added. Finally, this discussion of findings appears to help explain the structure position of Yudkin and Roddy (1964) in the networks analysed – wedged between the CHD and caffeine title classification groups.

Importantly, Yudkin and Roddy (1964) offered a very different interpretation of these findings than papers within the ‘caffeine’ and ‘CHD’ title classification groups. Bellet et al. (1965) and Little and Shanoff (1965) seemed to see the caffeine findings as hypothesis-generating finding, and used the finding to help contextualise their own studies examining the relationship
between caffeine and CHD. Epstein (1965) offered a very different interpretation of the lack of association between diet and CHD to Yudkin and Roddy. For Epstein, this finding was supportive of the current favoured approach to epidemiology and seemed to suggest it was further proof that Americans consume too much fat.

We might understand these findings by an appeal to Latour and Woolgar’s (1979) account of publication and citation behaviour from *Laboratory Life*. For these authors, scientists use references primarily to persuade readers of their favoured positions. Scientists devote both their time in the laboratory and in writing their papers to converting particular claims into statements of higher or lower credibility. Scientists, through citation, support and attack statements contained in other scientific publications that help them to alter the level of credibility of those scientific assertions.

By this account, the act of referencing is primarily a *rhetorical* tool; a tool used to convince and audience of the validity of an author’s particular position. Indeed, the results reported here seem to support this interpretation. How else can we understand the emphasis placed on reporting the lack of nearly all dietary variables by Yudkin and Roddy (1964) when other papers devoted less time and emphasis on these results or simply ignored them? How else can we understand the differing interpretations of the meaning of the lack of association of dietary fat between these authors and Epstein (1965) or the different interpretations of the meaning of the caffeine association? Why would the serum cholesterol findings be reported but not the lack of association between diet and *either* serum cholesterol and CHD? Why else would single empirical findings be cited rather than a more holistic evaluation of Paul *et al.* (1963)?

Sociologists of scientific knowledge have long been interested in how interpretative factors can lead to scientific disagreement. Previously, I reported results that suggested that scientific disagreement may be sustained by different evidence selection behaviours, but this study suggests that disagreement can also stem from divergent interpretations of the same empirical study. Accordingly, the findings of this study regarding the different
interpretations of the same finding can be read as an example of what Collin’s (1981) termed ‘Interpretative Flexibility’ – a term used to describe how different groups of scientists can interpret the meaning of particular findings in different, sometimes divergent, manners.

This study demonstrates that CS-CNA holds utility for those interested in how prevalent different interpretations of the same study are in the scientific literature, for identifying the communities of documents in which particular interpretations reside and circulate, and for understanding the selective citation of particular findings from a specific study.

**Afterword on Paul et al. (1963)**

In few of the 110 papers sampled here did authors devote their attention to considering the totality of evidence produced by Paul et al. nor is there a sustained critique; rather, most used single findings in a way that supported particular empirical arguments regarding the causes of CHD. As many of the findings reported by Paul et al. were generated from univariate analysis, many of the observed associations may, after controlling for other variables, disappear.

Furthermore, Paul et al. did not report all the findings from all the variables they examined; rather, they reported a selection of those that were either statistically significant or non-significant findings but were believed important because of existing theories of CHD. According to the authors:

> The data presented herein represent the initial compilation of data centring on this problem; a discriminate function analysis is to be undertaken next. The report which follows has been made possible through the efforts of a large group of physicians and other scientists who have volunteered their time in the project to permit the accumulation of a large body of data, a portion of which is presented below (p.20)

Thus, there is a risk that all of the observed associations were due to chance. For example, if Paul et al. had correlated a hundred variables with CHD and five of these produced statistically significant findings at P<0.05, then these results would be expected by chance alone (see Streiner 2015 for a clear discussion of the problem of multiplicity). Without knowing how many variables
were examined, of how the authors chose to select which findings to report, it is not possible to eliminate this possibility.

Today, this is one of the primary reasons why both prospective cohort studies and clinical trials are required to publish a protocol for analysis before collecting their data and, in cases where there are multiple comparisons, an appropriate statistical procedure established to deal with these. It is beyond the scope of this article to go into detail about the problems inherent in multiple comparison statistical testing, but Smith and Ebrahim (2002) and Young and Karr (2011) provide a clear introduction to such problems, while Pocock et al. (2004) survey the prevalence of these practices in modern epidemiology studies. Paul et al. (1963) can be regarded as an exploratory study – its findings are hypothesis-generating – but its statistically significant findings should not be considered examples of hypothesis-testing science nor can their inferential validity be ascertained. Indeed, this is the position of Paul et al. as set out in their introduction; their study was an initial exploration of these data prior to a later discriminate function analysis; one that these authors did not ultimately publish.

Further, Paul et al.’s findings regarding the relationship between variables and CHD are hard to interpret because they collapse myocardial infarction, angina pectoris, and coronary death into this category. Thus, the meaning of the reported associations or lack thereof is not clear. Because of the small numbers of the harder end-points – MI and CHD death – any association with these would have been unlikely to reach statistical significance.

Thus, what Paul et al.’s findings tell the reader is not immediately obvious, and this may have contributed to the manner in which it was cited. That is, the findings in this paper could be read in many different ways, in part, because of the way this study was designed.

**Implications**

The purpose of designing this study was to find a method capable of understanding whether citation bias existed in the evaluation of prospective cohort studies that reported findings regarding dietary fat and CHD. In most
studies reporting results on citation bias (Chapman, Ragg, McGeechan, 2009; Jannot et al. 2013; Kivimäki et al. 2014), the number of citations to papers of different classifications (supportive or unsupportive) are compared to evaluate whether there is a significant difference in their accumulated citations. However, it is rare that studies read and record how the citing literature is referencing previous literature beyond those utilising some variant of claim-specific analysis (Tatsioni et al. 2007; Greenberg 2009; Leng 2018).

The results of this study suggest that there may be problems of using citation counts to try and evaluate citation bias to studies that contain multiple findings. This stems from the fact that few studies evaluated the totality of evidence, many selected single findings to support particular empirical arguments, and studies offered different interpretations of the meanings of those findings. Paul et al.’s (1963) citation count alone tells us little about what impact it had on the literature, and comparing its citations to other prospective cohort studies listed at the start of this chapter would clearly be an inappropriate test of citation bias regarding dietary fat findings.

While my results are derived from focussing on a single study, and its results cannot be strictly generalised beyond this, they nevertheless raise an important note of skepticism regarding the utility of using citation counts alone to test for the presence of citation bias in the literature without additional claim-specific data. While it cannot be assumed that all multifactorial studies will be cited for various different findings, a researcher must be aware that this is a possibility and account for this in their assessment of bias.

While CS-CNA was successfully adapted to fit this particular case study and helped clarify what specifically Paul et al. (1963) was being cited for, it is clear that an extension of CS-CNA to all the relevant prospective cohort studies that examined the link between diet and CHD would be impractical for several reasons:

First, the large number of reported findings makes tracking how findings are cited in particular papers very time consuming and difficult to record, and this problem will be greatly amplified when trying to simultaneously track how findings from multiple prospective cohort studies were cited. This creates
problems not only for recording findings because of the numbers involved, but also creates problems for how to visualise and analyse these via network analysis. Second, because different findings not relevant to the dietary question will likely have been heavily cited, many papers will not be addressing the dietary hypothesis directly.

As demonstrated in the previous chapter (Chapter 4 Supplement, Ch4.S5), while RCTs can be interpreted inaccurately authors tend to discuss the same findings. This might be because of the unifactoral nature of the RCTs analysed, which, presumably, greatly restricts what these studies can be cited for. Indeed, the vast majority of work on citation bias has been conducted on citations to such trials. However, the results of this study raise questions of how far CS-CNA can be extended to other forms of studies without considerable complexity.

5. Post-study reflection

This, my second application of CS-CNA, proved difficult. While it constitutes a failure to develop a feasible method to detect citation bias in a set of prospective studies in a reasonable time frame, it played an important role in shaping how my research proceeded. I had encountered a level of complexity that was difficult to overcome in the case of cohort studies. Furthermore, no other set of RCTs existed that could be examined in the same manner as in Chapter 4, with only one primary prevention fully published in the literature prior to 1984.

To produce answers to my research questions regarding what evidence played an important role in the evaluation of the diet–heart hypothesis, I would need to alter my approach. The two claim-specific approaches that I have introduced so far in this thesis begin by identifying particular studies of interest and then identifying the literature that cited them. While this was unproblematic for the RCTs, this method ended up capturing far too many irrelevant citing studies when a prospective cohort study was the focus. Further, while citations networks around particular studies can help us to understand how particular
results are received in a community, it is difficult to appreciate how this affects the general opinion of the scientific community towards the validity of a hypothesis because this tends to depend on the evaluation of other studies.

However, this study gave me an idea of how to devise an approach that was quite different. First, I demonstrated that the terms contained within a title were reasonably accurate predictors of the findings that were ultimately cited in those papers. Second, the terms in a title both helped to identify research topics and helped to explain the structural position of papers in a citation network. I conjectured if I were to capture literature that used terms that clearly indicated a focus on the diet–heart question, then I could map the literature used by this group to understand the influences on that community over time. Indeed, according to Kuhn ([1962]1970), it is these small communities that are responsible for constructing and evaluating scientific knowledge, and my attention turned to capturing the community responsible for the diet–heart hypothesis. In the following chapters, I describe this approach.
Chapter 6: On Capturing a Paradigm, Part I: Complexity, Boolean Searches, and Systematic Literature Retrieval

Introduction

Here, I develop an approach to systematically capture literature relevant to the debate over the relationship between diet, atherosclerosis, serum cholesterol, and coronary heart disease (CHD), popularly referred to as the diet–heart debate, from 1900 to 1984. By a series of Boolean queries designed to target each of these elements, I capture a large set of publications from Web of Science (WoS). By layering these queries in sequence, I demonstrate how a core-set relevant to the debate can be established. By extracting the reference lists of all identified publications and constructing a citation network, I identify literature missed from the original queries. My ambition here is to capture literature representative of the ‘paradigm’ in which the debate unfolded. These data are later used for temporal analysis of key stages of scientific development, community detection, and Main Path Analysis (MPA).

While the following method is designed around a specific case study, elements of the approach may be transferable to other areas of the scientific literature. With a growing interest in dissemination biases in the scientific literature (Greenberg 2009; Leng 2018), the ability to capture literature on specific scientific debates is becoming increasingly important. To accurately measure the prevalence of these biases in particular debates and to understand their influence, literature searches must capture a large and demonstrably representative set of documents. Further, there is currently a revival of interest in understanding the evolution of scientific thought via network-analytic techniques, such as variants of MPA (Liu and Lu 2012; Batagelj et al. 2017), in quantitative historiography (Garfield et al. 2003, Garfield 2009; Marx and Bornmann 2014), in producing visualisations and
maps of the scientific literature (Börner et al. 2005; Chen 2013; Chen 2017), and in sociological studies of scientific controversy utilising digital methods (Venturini 2012; Shwed and Bearman 2010). If these approaches are to flourish, methods are required that capture literature relevant to particular scientific theories and debates.

In Section 1, I discuss the complexities involved in capturing relevant literature. To help guide my literature retrieval, I draw from Kuhn’s observations on the structure of scientific paradigms and convert these into network propositions. In Section 2, I introduce the fundamentals of literature searches, and discuss the merits of using WoS in this project by demonstrating the superiority of its coverage and the quality of its relational data compared to Scopus. In Section 3, I introduce queries designed to capture atherosclerosis (A), cholesterol metabolism (B), and CHD (C) records, respectively. A preliminary validation of the retrieved results is performed by comparing growth rates between these literatures, the global scientific output, and the growth of literature related to other biomedical areas. In Section 4, I layer dietary terms on top of queries A, B, C in order to detect documents likely directly involved in the diet–heart debate and fully retrieve the results. Finally, I retrieve all documents that were caught in the overlap between queries – e.g., B⊂A.

In the second part of this methodology (Chapter 7), I demonstrate how network analysis may be used to identify non-indexed literature and validate the retrieve results.

1. Complexities and Paradigms

Complexities beset attempts to capture the literature relevant to the diet–heart debate. These stem from the evolution of thinking about the relationship between diet, cholesterol metabolism, atherosclerosis, and CHD; a history that has produced variants of measures, concepts, and subtle adaptations of theory. Below are 13 interrelated hypotheses that populated the literature before 1985:
1. The *lipid infiltration theory of atherosclerosis* – an increase in serum cholesterol is implicated in the development of atherosclerosis.

2. The *combination theory of atherosclerosis* – serum cholesterol is a necessary, but not always sufficient, cause of atherosclerosis. Atherosclerosis is multi-causal phenomenon and other factors, such as hypertension, age, and infection are also implicated.

3. The *atherosclerosis–heart hypothesis* – atherosclerosis is a cause of CHD.

4. The *total cholesterol (TC) hypothesis* – an increase in TC is a cause of atherosclerosis and thereby a cause of CHD.

5. The *low-density lipoprotein cholesterol (LDL-C) hypothesis* – an increase in cholesterol carried by LDL is a cause of atherosclerosis and thereby CHD.

6. The *high-density lipoprotein cholesterol (HDL-C) hypothesis* – an increase in cholesterol carried by HDL is protective against atherosclerosis and thereby protective against CHD.

7. The *dietary cholesterol hypothesis* – that dietary cholesterol is a cause of raised serum cholesterol (TC or LDL-C), atherosclerosis, and CHD.

8. The *dietary lipids hypothesis* – that dietary fat and dietary cholesterol are causes of raised serum cholesterol (TC or LDL-C), atherosclerosis, and CHD.

9. The *total fat hypothesis* – dietary fat increases serum cholesterol (TC or LDL-C), and thereby causes atherosclerosis and CHD.

10. The *saturated fat hypothesis* – that dietary saturated fat increases serum cholesterol (TC or LDL-C) and causes atherosclerosis and CHD.

11. The *polyunsaturated fat hypothesis* – polyunsaturated fats lower serum cholesterol (TC or LDL-C) are protective against atherosclerosis and CHD.

12. The *individualised treatment approach* – that dietary intervention should be used to reduce serum cholesterol levels in those with pre-existing CHD or who have associated risk-factors, particularly hypercholesterolemia.
13. The *public health approach* – that dietary advice should be issued to whole populations in an attempt to reduce population average serum cholesterol levels.

These hypotheses co-existed, sometimes uneasily, during most of the history of the debate. None were abandoned quickly, and many remained for periods alongside other hypotheses. Although some of these hypotheses imply entirely different basic mechanisms and intervention strategies, they are tied together by the core claim that a disturbance in cholesterol metabolism is the primary cause of atherosclerosis; they are variants that bear a familial resemblance.

Although the case for the dietary guidelines presented an apparently simple causal chain between diet and CHD, diverse and intertwining research communities studied this chain. Several research areas were involved – from pathology and blood and food chemistry to epidemiology – and their relative importance (as estimated from citation analysis) varied with time, as did the popularity of certain measures and terms. In the first half of the 20th century, animal models dominated research in atherosclerosis. In the 1950s, human metabolic ward trials dominated. By the 1960s, there was a rise in intervention and large observation trials. Then, in the late 1970s, came basic studies of cholesterol metabolism. These research efforts used very different methods – from histology and animal experiments to vital statistics and randomised controlled trials (see Chapter 8).

As findings in regard to any of the above hypotheses or from any particular speciality might have altered opinion on other hypotheses, it is necessary to capture an inclusive set of literature. Is it possible to capture this ‘paradigm’?
Kuhn and paradigms

While many have pointed to particular scientific communities as the primary unit of analysis for understanding the development of scientific knowledge (Fleck 1935), it is probably Kuhn ([1962]1970) who most clearly highlights their importance, their structure, and the empirical problems inherent in their detection. Importantly, Kuhn also offers hints of how to find them systematically.

For Kuhn, scientific communities are groups of scientists that share in a common language, use the same technical literature and methods, and share a theoretical and empirical understanding. These communities are defined by a level of consensus regarding what questions are important, what methods are needed to solve them, and the importance of particular exemplars. These largely consensual and homogenous communities are self-regulating; scientists communicate and use the findings of other scientists in their communities and rarely venture outside of their ‘paradigms’. According to Kuhn,

Communities in this sense exist, of course, at numerous levels. The most global is the community of all natural scientists. At an only slightly lower level the main scientific professional groups are communities…For these major groupings, community membership is readily established except at the fringes. Subject of highest degree, membership in professional societies, and journals read are ordinarily more than sufficient…It is only at the next lower level that empirical problems emerge…For this purpose, one must have recourse to attendance at special conferences, to the distribution of draft manuscripts or galley proofs prior to publication, and above all to formal and informal communication networks including those discovered in correspondence and in the linkages among citations…(pp.177–8).

According to Kuhn, it is this latter kind of community that is responsible for the construction and evaluation of knowledge – “Communities of this sort are the units that this book has presented as the producers and validators of scientific knowledge” (p.178).

For Kuhn then, *community structure* is the key to understanding the operation of science. In a footnote beside the “linkages among citations”, Kuhn cites three works – Garfield *et al.* (1964) ‘The Use of Citation Data in Writing
the History of Science’, Kessler’s (1965) ‘Comparison of the Results of Bibliographic Coupling and Analytic Subject Indexing’, and Price’s (1965) ‘Networks of Scientific Documents’. These are famous Scientiometric documents that laid the foundation of that discipline. Kuhn saw potential in citation analysis for detecting communities, particularly citation network analysis. But why did Kuhn think this?

First, according to Kuhn, scientific communities should be structured around a core-set of findings and methods.

“Normal science" means research firmly based upon one or more past scientific achievements, achievements that some particular scientific community acknowledges for a time as supplying the foundation for its further practice. (p.10)

Scientific communities should coalesce around a core-set of documents that report important findings and ideas. If this is true, we should see the documents retrieved here clustered around a core-set of highly cited methods, findings, and theoretical ideas.

Second, the structure of communication between scientists ought to reveal the existence of shared paradigms:

Within such groups communication is relatively full and professional judgment relatively unanimous. Because the attention of different scientific communities is, on the other hand, focussed on different matters, professional communication across group lines is sometimes arduous, often results in misunderstanding, and may, if pursued, evoke significant and previously unsuspected disagreement (p.177).

If Kuhn is right, paradigms will not only be distinguished by documents clustered around a few highly cited documents, but by a dense network of communication between documents within a community. The density of citations within a specific community will be higher than the density of citations between communities.

Finally, throughout the Structure, Kuhn repeatedly draws attention to the shared understanding that is contained in these specific communities. The clearest sign that a community has reached “normal science” – defined explicitly by the sharing of a ‘paradigm’ – is in the types and quantities of publications produced:
Research communiqués will begin to change in ways whose evolution has been too little studied but whose modern end products are obvious to all and oppressive to many. No longer will his researches usually be embodied in books addressed, like Franklin’s Experiments…on Electricity or Darwin’s Origin of Species, to anyone who might be interested in the subject matter of the field. Instead they will usually appear as brief articles addressed only to professional colleagues, the men whose knowledge of a shared paradigm can be assumed and who prove to be the only ones able to read the documents addressed to them (p.20).

Thus, a ‘paradigm’ will be observable through a change in publication practices. This will involve an increase in journal publications, and these documents will be directed at a particular research community. Further, if scientists do share in a common language, and desire to communicate primarily with other scientists in their respective paradigms, then scientists need a strategy of ‘sign-posting’ their publications to their desired audience and possess the ability recognise publications relevant to them. In the period studied here, 1900–1984, the most obvious way of doing this is for scientists to use particular words in particular combinations in the titles of their documents.

Having identified the terms used by scientists to ‘sign-post’ their documents, I expect to find a major growth in their use that coheres with historical knowledge of the development of this debate. These documents should form a large interconnected community, clustered around a core set of highly cited documents, and whose references are disproportionately directed at other documents in this set.

This approach to capturing a ‘paradigm’ is drawn from Kuhn’s ideas, albeit reinterpreted in network terminology. A previous attempt at capturing paradigms was developed by Small and Griffith (1974) after their interpretation of Kuhn using co-citation analysis – an approach that clustered highly cited documents together if they shared many of the same citing papers. This was based on Kuhn’s conjecture that the exemplars of a paradigm would flock together via an analogy with gaggles of geese. According to Small (2003), however, Kuhn rejected co-citation as a method on the basis that it ignored other features of his concept by focussing only on highly cited documents and
authors, rather than capturing the community of researchers involved in expanding and evaluating a theory. Another method came with Callon et al.’s (1986) co-word analysis – an approach that identified documents on a similar topic via the conjoining of specific symbolic words. However, again, this method focussed on a single aspect of Kuhn’s concept – the sharing of a particular language. The approach I propose aims at capturing first the community of papers focussing on a specific topic via words used in the titles of papers, and constructing a network specific to these documents to understand how documents clustered together via references and citations (Chapters 7–8).

2. Boolean Searches, Bibliometric Databases, and Systematic Literature Retrieval

Boolean queries

To identify documents as members of a particular class, they must bear a common resemblance. Bibliometric searches identify documents that contain specific shared words in titles, abstracts, and associated key-words. Searches can be confined either to searching for terms in the title of a document or within the title, abstract, or the keywords associated with that document provided by the author or journal. Thus, the ability to classify documents as relevant to a particular research question depends on the authors of those documents sharing a common language and an understanding of which terms to use to ‘signpost’ their documents.

Alongside relevant words, a query will often include Boolean logical operators: (i) AND – to specify that two or more terms must both exist in the same record; (ii) OR – to specify that at least one term of a series must be in a record; (ii) NOT – to exclude records that contain certain words. Furthermore, both Scopus and WoS have functions that allow searches of strings of words, typically designated via quotation marks, along with other proximity functions.
Queries can be refined by searching for words popular in particular research areas. To trace the growth of research in, say, particle physics, we might start with a simply query - ‘particle AND physics’ – or search for “particle physics”. The former would be too inclusive, while the latter would be over-restrictive. After retrieving this literature, we would need to scan titles and abstracts to see whether the query caught the documents we expect, and whether it also caught irrelevant literature. If so, we would redesign the query to remove terms, like - particle AND physics NOT verb - which would exclude documents on linguistics that, for whatever reason, also had physics in the title or abstract. Alternatively, we can search for documents containing specific words only in either the associated research category, say ‘particle physics’, or in specialised journals, like *Progress in Particle and Nuclear Physics*. Both research categories and journal, are document attributes that can be searched for via the major bibliometric databases.

These methods are important when a key term is popular in entirely different research contexts. For example, to study research into tuberculosis from the late 19th century, a search is required for “tuberculosis” but also for “consumption”, still commonly used for tuberculosis in the first decades of the 20th century. However, consumption was also popularly used in studies of national consumption of food, water, and other goods, and by biologists to describe oxygen use in cells. As many documents on tuberculosis during this period contained only the word “consumption” in their title, and no abstracts or keywords are indexed for literature this far back, we cannot easily use other terms to retrieve only the documents we want. Therefore, it is necessary to understand both how a term is used in a specific research domain and how it is used in others. In this case, we may limit our search for consumption to the period before 1920, and then exclude words associated with irrelevant literature or particular research areas, like economics. When designing queries, an iterative process is needed to check how well a query is operating. This includes becoming acquainted with how terms are used, when they are used, and by whom.
Finally, meta-analyses and systematic reviews use searches that contain terms that define (i) a particular type of study (e.g. prospective cohort studies); (ii) measures (e.g. total serum cholesterol); (iii) end-points (e.g. myocardial infarction); and sometimes (iv) specific categories of patient (e.g. women). This refined specification of search terms allows those conducting systematic reviews and meta-analyses to efficiently and systematically evaluate the evidence-base behind particular hypothesis. This ability tends to come at a late stage of research – in the evaluative stage, once key terms, measures, and end-points have been established.

**Search strategies**

A literature search is directed by a particular question that determines the search strategy. For a meta-analyst, this involves identifying end-points, measures, trial design, etc. However, for historians and sociologists interested in how scientific research developed around a particular theory, a more inclusive strategy is required; one that can adapt to fashions in terminology, methods, and subtle variations of hypotheses.

Generally, the most complex Boolean queries come not from bibliometrics, but from systematic reviews and meta-analyses in the biomedical sciences. This stems from the different *kinds* of research questions that are addressed. The former has focussed primarily on the history of disciplines, such as economics (Claveau and Gingras 2016); the growth of scholarship within specific topics, such as apoptosis (cell death) (Garfield 1997); the growth of international collaboration (Wagner and Leydesdorff 2005); or the growth of the entire scientific enterprise (Bornmann and Mutz 2015). To do this, most studies used relatively simple search queries.

**Journal-specific-search**

For example, in a study of the history of consumer research, Baumgartner (2010) searched for articles published in the *Journal of Marketing*, *Journal of Marketing Research*, and *Journal of Consumer Research* between 1934 and 2009. Similarly, Griffith *et al.* (2008) sought to understand emerging topics in the international business literature by retrieving articles from six top journals.
between 1996 and 2006, while Pilkington and Meredith (2008) retrieved articles from the three oldest journals of operation management to understand the intellectual structure of that sub-discipline. This journal-specific-search method appears to be a popular tool for capturing the state of disciplines in the business literature; but this method has tended to be restricted to disciplines with a small set of specialist journals.

Clearly, this method is unsuited for the present study. Major biomedical journals, such as the *Lancet* or the *Journal of New England Medicine* publish documents on a large variety of research topics. As most articles relevant to the diet–heart debate come from such broad journals, journal-specific-searches retrieve mostly articles unrelated to the debate.

**Lexical searches**

Another popular method has been to search for common phrases associated with a research area. Liu et al. (2015), in a study of innovations systems research, used - “Innovation System”, “Innovation Systems”, “System of Innovation”, and “Systems of Innovation” – to search the literature indexed by WoS. Again, this method is unsuited to capturing the diet–heart debate. A search for “Cardiology” and “Coronary Heart Disease” would be far too inclusive – capturing literature on other causes and treatments of CHD – whilst also being overly restrictive – missing studies that focus on the effect of particular foods on serum cholesterol, or the relationship between serum cholesterol and atherosclerosis.

However, similar key-term searches have been used to examine areas of the biomedical literature by focussing on specific phenomena. Garfield (1997), in a study of interest in cell death (apoptosis), used a simple query – 'Apopt(free ending)' AND/OR `(free) Cell (free ending) Death' – to retrieve articles with these terms in the title or keywords. Here, the (free) and (free ending) refer to a truncation function that allows WoS to search for variants of a term (e.g. Cell, Cells, Cellular). Similarly, Zheng et al. (2009), in a study of interest in the John Cunningham virus, used the following query to search WoS and PubMed – “JC virus OR John Cunningham virus OR JCV OR JC polyomavirus OR JC polyoma virus”. Robert et al. (2007), in a study of the
literature on sleep science, describe only that “a broad set of thesaurus terms from EMTREE were searched along with their equivalent ‘free-text’ terms” (p.233). Koester et al.’s (1982) examined 5,765 documents in high energy physics to understand the evolution of the weak-electromagnetic unification theory between 1950 and 1975, but do not state their search strategy.

As the diet–heart debate includes several interacting parts, I mimicked the strategies commonly found in systematic reviews and meta-analyses. A similar approach has been described by Trinquart et al. (2016) in their study of the salt controversy. They used variants of “salt”, several type of interventions, and several outcomes of interest in designing their query. Similarly, Nykiforuk et al. (2010) used a series of key-word searches, in various combinations, to capture the literature on smoke-free spaces in the policy literature. However, Greenberg’s (2009) study provides the major exemplar. Greenberg, in a study of the literature on the relationship between β amyloid and inclusion body myositis, utilised two queries. First, he queried PubMed for all records containing “Inclusion body myositis” in the title. He then used a second query to capture missed literature “(amyloid AND (muscle[ti] OR polymyositis[ti] OR dermatomyositis[ti] OR inflammatory myopathy[ti]) NOT smooth muscle[ti])”. Here, [ti] directs the query to the title of publications. He then examined the abstracts of all retrieved documents and removed any not relevant to his specific question, which reduced the sample from 766 to 293 documents. Finally, he examined the reference lists of all these articles to identify relevant literature not captured by his query, which added another nine documents (Greenberg 2009). That study demonstrated an impressive ability to capture literature relevant to a specific research question. It is the style of literature retrieval that I expand on in Section 3.
Databases

The selection of an appropriate database must be guided by its coverage; i.e., what proportion of the relevant literature it has indexed. As my research examines a period before 1985, I use WoS’ ‘Core Collection’. The Core Collection indexes about 73 million articles from over 20,000 journals from 1900 onwards, and contains over 1.4 billion cited references from bibliographies of all documents published in these journals. Furthermore, WoS (2018) has a transparent protocol regarding its selection of journals for indexing:

WoS bases its selectivity on Bradford’s Law of Scattering and Garfield’s Law of Concentration. Bradford argued that literature on any research topic will tend to cluster in a few top journals (about a third of all articles), another third will be collected in a slightly larger group of journals, while the last third will be scattered across a much larger set of journals. According to Bradford (1948), if journals are arranged in order of how many articles they contain on a given subject, “they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the number of periodicals in the nucleus and succeeding zones will be as 1:n:n^2…” (p.116). Garfield’s Law of Concentration explains this distribution by observing that journals contain articles on various research topics. For Garfield (1979), this meant “the tail of the literature of one discipline consists, in large part, of the cores of the literature of other disciplines” (p. 23). Garfield demonstrated that 80–90% of citations are to documents contained in 10–20% of all journals in a field, and these journals derive from the top third of journals as ranked by their density of publications on specific research topics. Accordingly, WoS only indexes the top journals that attract many citations.

The Core Collection has important differences with WoS’ ‘All Databases Function’ that includes all of the Core Collection, but also items indexed MEDLINE, and BIOSIS Citation Index, among others. These differences are relevant to both bibliometric investigation and network analysis.
In *all-databases*, WoS returns 135,651,187 indexed records, which include journal publications (articles, reviews, letters, editorials, abstracts, meeting proceedings, and notes) and other types of publications (books, patents, and conference proceedings). The Core Collection covers 71,064,086 of these; however, its primary focus is on high quality journals. Before 1990, the Core Collection holds no indexed records on conference proceedings, before 2005 no books, and it still has no patent data. These figures were retrieved on 20 May 2018 by a search for all documents from all research categories, markers included by WoS on all indexed studies, and first searching across “All databases” and then only the “Core Collection” via the advanced search function.

As my research focusses on the period before 1985, I refined my query to include a time-limit. In *all-databases*, WoS holds 25,292,743 records with the Core Collection covering 15,081,007 of these (~60%).

The *all-database* contains 3,258,500 patents, 355,011 books, and 331 records on legislation excluded by the Core Collection. Second, the *all-database* search captures more foreign-language literature than the Core Collection. In all databases, WoS holds 4,048,282 foreign-language records (16% of all records) with German, French, and Russian responsible for 63% of them. The Core Collection holds 2,079,845 records in languages other than English (14% of all records), with German, French, and Russian responsible for 86% of them. Thus, the Core Collection has a greater English-language bias, and in its foreign-language records favours German, French, and Russian over other languages.

I performed an equivalent search on the Scopus database via their tag “Subject Areas” in the period 1900–1984 producing 14,668,365 records in all subject areas. Fig 18 plots the number of indexed records per year by WoS versus Scopus between 1900 and 1984. As WoS has consistently better coverage, it is more suited to researching this particular period.
**Fig 18**: Number of indexed records from 1900 to 1984 in WoS and Scopus.

**Types of Data**

However, the *type* of data analysis desired is also important. For citation network analyses, *relational* data are required that includes the citation links between documents. WoS *only* holds comprehensive relational data for records in its Core Collection, and it only formally indexes citations between Core Collection documents and the documents they reference. To do this, it holds digital records of all references from all documents in its Core Collection regardless of whether those references link to other documents in the Core Collection. The number of citations displayed in WoS refers to citations from other documents *in this database*, not all citations. The same is true of Scopus. Thus, using citation scores to judge a document’s influence must be considered carefully – the number of citations reflects only the number of *indexed* documents that have cited it.

Scopus is unsuited for historical network analyses because it does not index full bibliographies of documents before the 1970s, and even after this its coverage is sporadic. To demonstrate this, I searched for all of Ancel Keys’ publications, one of the most famous scientists in the history of nutrition and cardiovascular research, published between 1948 and 1984. Of the 270 documents of which Keys was an author, only 19 (7%) were accompanied by full bibliographic data (i.e. reference lists). The first indexed bibliography was
associated with a 1971 publication. However, Scopus identified 49 documents between 1971 and 1984. Even after the period it began indexing bibliographies, Scopus only contained 39% of the bibliographic data of its own indexed articles. The same search in WoS’ Core Collection retrieved 329 documents of which Keys was an author along with their full reference lists. Therefore, Scopus is not suitable for the current study.

3. Capturing Atherosclerosis, Cholesterol, and CHD Research

Title searching

There are important differences between title and topic searches. I advocate title searches – as words in the title are generally reliable indicators of a document’s focus. Putting specific terms in a title is a decision taken carefully to sign-post it to a particular audience, accordingly the number of documents with shared terms in their titles is probably the best indicator that we possess of scientific activity/interest in a topic. Searches for terms anywhere in the title, abstract, or key-words, are generally too inclusive – particularly in biomedical research, where measures such as TC or LDL-C are used in studying other conditions, such as diabetes and hypothyroidism. In the period examined here (1900–1984), it is also the case that WoS does not index abstracts or key-terms. While title searches can be justified on substantive grounds, it is the case that here only title searches can be made systematically because of the absence of abstracts and key-terms, which are necessary for topic searching.

Capturing literature on atherosclerosis

Atherosclerosis, a sub-category of arteriosclerosis, is defined by the presence of atheroma in the tunica intima of large and medium sized arteries and is considered the underlying cause of most cardiovascular diseases, particularly CHD. It is this condition that is thought to connect increased serum cholesterol levels to CHD. However, one complexity of searching for related literature is the variability of terminology. From the beginning of the 20th century,
biomedical scientists were aware of two forms of arteriosclerosis – one distinguished by atheroma in the tunica intima of arteries, the other, described by Mönckeberg in 1903, involving the calcification of the tunica media. In 1904, Marchand coined the term ‘atherosclerosis’ to refer to the first form.

Both forms of arteriosclerosis were known before Mönckeberg, but many assumed they were one and the same condition. While Lobstein (1833) is credited with creating the term ‘arteriosclerosis’, knowledge of arteriosclerotic-like changes dates back further. According to Long (1933), knowledge of calcific arterial changes dates back to the time of Aristotle, but it was not until the 16th century that anatomists began systematically reporting on the ‘ossification’ of arteries discovered post-mortem. It was not until the 19th century, when quality microscopes became readily available, that atherosclerotic fatty changes in the intima began to attract interest (Gulliver 1843) and new theories of pathogenesis emerged, notably in the works of Rokitansky (1852), and Virchow (1863).

Unfortunately, referencing and publication practices make literature published before 1900 hard to detect. Important works regarding the pathogenesis of atherosclerosis were published as books (Rokitansky 1852; Virchow 1863), and many claims were not clearly referenced. WoS does not index any items before this period in its Core Collection, but records are held for some documents in its ‘All-Database’ function dating back further.

To capture literature only on atherosclerosis, I designed the following query after reading documents to detect common terminology in the titles of documents published before 1985:

TI=(Atherosclerosis OR Atherosklerose OR athérosclérose OR “atherosclerosis” OR “coronary sclerosis” OR atheroma OR atherogenic OR atherogenesis OR atherosclerosis OR atherosis OR Atherosclerotic OR atheromatous OR “cholesterol steatosis” OR “smooth muscle cell proliferation” OR ((lipid OR lipids OR cholesterol OR fat OR fats OR fatty) AND (accumulation OR infiltration OR infiltrative OR laden OR deposit OR deposits OR degeneration OR plaque OR plaques OR streak OR streaks OR lesion OR lesions) AND (intima OR artery OR arteries OR vessel OR vessels OR tunic OR tunica OR aorta OR carotid OR thoracic OR iliac OR pulmonary)))
This retrieved 18,875 items from all-databases and 9,226 from the Core Collection. It missed documents that use variants of ‘arteriosclerosis’. To correct this, I included the following query:

\[\text{TI}=(\text{Arteriosclerosis OR artériosclérose OR Arteriosklerose OR “arteriosclerosis” OR Arteriosclerotic})\]

Alone, this retrieved 8,369 records from all-databases, and 2,932 from the Core Collection. When added to the previous query, this retrieved 26,440 records from all-databases and 12,114 from the Core Collection. The inclusion of arteriosclerosis terms accounts for an additional 7,565 records from all-databases and 2,888 from the Core Collection.

To remove documents on Mönckeberg-type arteriosclerosis, I merged these Boolean strings, but excluded terms associated with Mönckeberg-type via the Boolean operator NOT.

\[\text{TI}=((\text{Monckebergs OR Monckeberg OR Monckebergs}) \text{ AND} (\text{arteriosclerosis OR media OR calcific OR calcification OR sclerosis OR calcium}))\]

Alone, this returned only 26 records from all databases, with 11 from the core collection. As some regarded this as a separate benign condition (Silbert et al. 1953), I excluded the small number of documents clearly related to Mönckeberg-type.

The final query to capture research into atherosclerosis and arteriosclerosis is as follows:

\[\text{TI}=(\text{Atherosclerosis OR Atherosklerose OR athérosclérose OR “atherosclerosis” OR “coronary sclerosis” OR atheroma OR atherogenic OR atherogenesis OR atherosis OR Atherosclerotic OR atheromatous OR “cholesterol steatosis” OR “smooth muscle cell proliferation” OR Arteriosclerosis OR artériosclérose OR Arteriosklerose OR “arteriosclerosis” OR Arteriosclerotic OR (lipid OR lipids OR cholesterol OR fat OR fats OR fatty) AND (accumulation OR infiltration OR infiltrative OR laden OR deposit OR deposits OR degeneration OR plaque OR plaques OR streak OR streaks OR lesion OR lesions) AND (intima OR artery OR arteries OR vessel OR vessels OR tunic OR tunica OR aorta OR carotid OR thoracic OR iliac OR pulmonary))) \text{ NOT} \text{ TI}=((\text{Monckebergs OR Monckeberg}) \text{ AND} (\text{arteriosclerosis OR media OR calcific OR calcification OR sclerosis OR calcium}))\]
This retrieved 26,426 records published before 1985 from *all databases*, and 12,108 from the Core Collection. The former contains 14,318 records from journals not indexed in the Core Collection, primarily derived from MEDLINE, BIOSIS Citation Index, and the Derwent Innovations Index.

The first record detected via *all databases* is from 1866, while the first record from the Core Collection is from 1900. Only six records are detected prior to 1900. To validate this query, I scanned titles to check whether this captured literature relevant to atherosclerosis. I examined the 50 most cited articles, the 50 least cited articles, and examined literature under seemingly unconnected research categories (i.e. documents designated to arts and humanities). I found no cases of error, and the documents assigned to unusual categories seem to be errors on the part of WoS.

While this query will miss documents that focus on basic biology of particular phagocytes involved in atherosclerosis, it was not possible to identify these without capturing literature related more to basic cell biology than to atherosclerosis. Further, this query captured documents focussing on aetiology and pathogenesis, but also publications on its relationship to sequelae. It captured documents on different conjectured causes, from hormonal influences to blood pressure, and literature on the two major competing theoretical explanations of pathogenesis – the lipid infiltration theory (Anitschkow 1933) and the thrombogenic theory (Duguid 1960) – but also lesser known positions, such as the monoclonal and the clonal-senescence hypotheses (Ross and Glomset 1976).

This query captures all records regardless of publication type. The Core Collection includes journal content only, while *all-databases* include books (1,226) and patents (539). The indexing of books is by BIOSIS, and the rules governing it are not transparent during this period. Therefore, all books were removed from the query because retrieving reference data from these is not possible due to their exclusion from the Core Collection. Patents derive from the Derwent Innovations Index, and these were also removed from the query. After removal, the *all-database* search identifies 24,661 publications from journals, with the Core Collection holding 49% of these records.
Fig 19 plots publications per year identified by the final Boolean query from 1900–1984. The top panel plots the output per year between 1900 and 1942, while the bottom panel plots the output per year between 1943 and 1984.

**Fig 19: Growth of indexed records related to atherosclerosis in WoS' All Database and Core Collection.** Top shows the distribution, 1900–1942; Bottom shows distribution, 1943–1984

Between 1900 and 1942, across *all-databases* 780 indexed items were published, with 640 of these from the Core Collection. According to this, atherosclerosis research appeared in the first decade of the 20th century, collapsed during World War I (WWI), increased throughout the 1920s and 1930s, and stagnated during World War II (WWII). Between 1943 and 1984, 23,875 indexed records were published, 11,468 derived from the Core
Collection. Using the all-database results, we see research into atherosclerosis stalled throughout WWII, but, even by 1945, research output had doubled (40 records) compared to the previous year (20). The field then underwent sustained growth until 1969 after which output remained steady.

However, the number of items from the all-database search and Core Collection search diverge significantly. This may be explained by the creation of MEDLINE’s database in 1950, which systematically indexes journals and articles with a biomedical research focus that meet particular standards. However, the different stories of growth after 1961 required further investigation.

I examined whether this divergence could be explained by the inclusion of more foreign-language journals in all-databases. The Core Collection search recovered 1,668 records in languages other than English (~14% of records), whereas the all-database search found 7,273 foreign-language records (30% of records). Across all-databases, 72% of articles are English, 7% in Russian, 5% in German, and 4% in French. The remainder (13%) are in 33 different languages. By comparison, the Core Collection set contains 86% of articles in English, 5% in German, 3% in French, and 3% in Russian. The remaining 2% are in 12 different languages. The Core Collection has an English-language bias in relation to this research topic.

In the post-war period, this bias is not problematic due to the dominance of English in the biomedical literature. However, in the early years of atherosclerosis research some major findings were published in German and French. After 1945, non-English documents will likely derive from either regional journals or constitute reprints of studies previously published in English.

To demonstrate this, I considered two periods: 1900–1942 and 1943–1984. If documents written in other languages had a major influence, this should be caught in the citations to those documents. That is, documents published in other languages in the early period ought to responsible for a greater proportion of citations to atherosclerosis research than those published later.
In the first period, 780 records were detected. As of 2018, these have attracted 9,857 citations, (8,187 <1985; 2,233 <1943). The 206 documents published in a language other than English (German, 75%; French, 25%) have been cited 1,670 times (1,329 <1985: 433 <1943). Thus, foreign language documents gathered 17% of citations, including 16% of all citations before 1985 and 19% before 1943.

In the second period, 23,875 documents were published that have attracted 248,292 citations (as of 2018). Of these, 7,067 were published in a language other than English, and these have been cited 5,227 times while the English-language documents have 243,065 citations. Foreign-language documents gathered only 2% of all citations. Foreign-language documents published in the first period perform much better than those published after – with an average citation rate of 8.1 compared to 1.3. This is compared to a modest rise in the influence of English-language documents from an average citation rate of 14.0 in the first period to 14.6 in the second.

On the basis of these results, I restricted my analysis to English-language documents from 1943 onwards, and included German and French documents published before 1943. This reduces the set to 17,586; a set that accounts for 98% of all citations to documents with terms associated with atherosclerosis in the title, and it reduces the set of documents in the Core Collection to 10,657. The Core Collection holds 61% of the all indexed literature after these exclusions. As Fig 20 demonstrates, there is still a discrepancy between the coverage of the Core Collection and the refined results.
To understand this, I re-examined the results manually. Most records not included in the Core Collection derive from foreign-language journals, primarily Russian and Polish, indexed by either MEDLINE or BIOSIS. These are indexed as being in English, although they are not held in digital databases. These strange results, and the difficulty in validating the retrieved items, means relying only on the Core Collection to preserve the validity of analysis. Without being able to examine these documents, I cannot be certain that they are unique publications or reprints of previous studies. The sudden rise and then abrupt fall in the revised atherosclerosis search may also indicate that these foreign-language journals stopped being indexed from around 1969. I also found a problem of duplication of records because of variations in the spelling or format or the title of documents between the Core Collection records and others. For example, Kannel et al.’s (1984) is indexed in BIOSIS Citation Index as ‘Optimal resources for primary prevention of athero sclerotic diseases athero sclerosis study group’, but in the Core Collection it is indexed as ‘optimal resources for primary prevention of atherosclerotic diseases’.

Therefore, it is necessary to select a single database for inspection, and because this project proceeds to network analysis, the Core Collection is the only option. I intended to use the all-database function to estimate how much

Fig 20: Growth of indexed records related to atherosclerosis across WoS All Database search function and the Core Collection, 1943–1984
of the literature I had captured via the Core Collection, but this wasn’t possible. In Chapter 7, by using the references from these collected papers, I capture literature not indexed in the Core Collection and assess whether the coverage of this database is sufficient for analysis.

Capturing literature on cholesterol

Cholesterol is produced in all animal cells and is vital for their functioning. The beginning of modern interest in this sterol probably begins with Chevreul’s description of ‘cholesterine’ (a now obsolete synonym for cholesterol) extracted from gallstones, although Olson (1998) notes that discovery should be attributed to Poulletier de la Salle in the late 18th century. Since the discovery of cholesterol in atheromatous plaques (Gulliver 1843; Windaus 1910), many have argued that serum cholesterol has a causal role in atherosclerosis via the infiltration of cholesterol through the lumen of arteries (Anitschkow 1933)

By the late 19th century, colour tests for detecting sterols in biological systems were invented by Salkowski (sulphuric acid-chloroform test) and Liebermann and Burchard (acetic anhydride-sulphuric acid test) (Truswell 2010). However, it was not until Bloor’s (1916) alcohol-ether method to extract cholesterol from the blood that serum cholesterol became of major interest to the scientific community, and it was not until the Abell-Kendall method (Abell et al. 1952) that a standardised blood cholesterol measure became readily available. From the 1950s onwards, this was widely used to measure total cholesterol (TC) in blood serum.

However, cholesterol is carried in the blood by lipoproteins. Macheboeuf (1929) is credited with the discovery of the first lipoprotein, and interest in the structure and variety of lipoproteins steadily increases from this period, though it was not until Gofman et al. (1949) began using the ultracentrifuge that a range of lipoproteins were isolated. In the 1970s, low-density lipoproteins (LDL) and high-density lipoproteins (HDL) attracted considerable interest. LDL carries cholesterol and other lipid (fat) molecules around the body, while HDL carries lipids to the liver to be excreted. By the late 1950s, alongside TC, researchers regularly used two other measures: (i) the total cholesterol carried
in LDL (LDL-C; also known as the $\beta$-lipoprotein fraction); (ii) cholesterol carried in HDL (HDL-C; also known as the $\alpha$-lipoprotein fraction).

To capture research related to research on cholesterol metabolism or serum cholesterol, I constructed the query:

$$\text{TI} = (((\text{Cholesterol OR cholesterin OR cholesterine OR lipid OR lipids}) \text{ AND (blood OR serum OR plasma OR metabolism)}) \text{ OR hypercholesterolemia OR hyperlipidaemia OR hyperlipidemia OR dyslipidaemia OR dyslipidemia OR lipoprotein OR lipoproteins OR LDL OR HDL OR LDL-C OR HDL-C})$$

This returned 49,714 items published between 1907 and 1984 across all databases; however, due to the problems exposed with the all-database function during the previous query, this set of data will not be used. First, there are reasons to suspect that the issues faced with the former will occur again. Second, by keeping the database constant, we can make more reliable comparisons between search results. The Core Collection holds 31,159 records published between 1907 and 1984. Fig 21 plots the number of publications per year, mirroring the split in distribution that was performed on atherosclerosis (1900–1942; 1943–1984).
From 1900 to 1942, 777 publications were detected. Publications on atherosclerosis preceded studies in blood cholesterol, but research into blood cholesterol consistently outperformed atherosclerosis research in terms of output after 1925. Between 1943 and 1984, 30,382 items were published. Cholesterol research declined during WWII, but recovered from the late 1940s.

Fig 21: Growth of indexed records related to cholesterol metabolism and atherosclerosis in WoS Core Collection, 1900–1984. Top shows the period 1900–1942; Bottom shows distribution, 1943–1984.
Capturing literature on CHD

CHD is a class of cardiovascular disease that occurs when atheromatous plaques build up inside the coronary arteries and restrict blood flow to the heart muscle. Over time, atherosclerosis can weaken the heart muscle leading to heart failure and arrhythmias by depriving the myocardium of oxygen rich blood via the progressive occlusion of arteries. In the event of total occlusion (obstruction) following excessive plaque build-up or rupture leading to coronary thrombosis, myocardial infarction (MI) and sudden cardiac death can follow.

Scientific interest in conditions now classed as CHD can be traced to Harvey’s description of the coronary circulation and Heberden’s identification of angina pectoris. However, it was not until the early 20th century that CHD attracted significant scientific interest. This interest stemmed from the work of Herrick on the relationship between atherosclerosis, thrombosis, and the occlusion of arteries (1912, 1919), Pardee’s use of the electrocardiogram to detect coronary artery obstruction (1920), and the rise in the incidence of heart disease in Western, developed countries from the 1920s. To capture the literature on CHD, I designed the following query:

\[ \text{TI} = (((\text{heart OR coronary OR cardiac}) \text{ AND (ischemic OR ischemia OR ischaemic OR ischaemia OR arteriosclerotic OR atherosclerotic OR infarct OR infarction OR occlusion OR occluded OR obstruction OR thrombosis OR thrombi OR atheroma OR embolism OR emboli)) OR CHD OR “sudden death” OR angina OR “coronary heart disease” OR “coronary artery disease” OR “atherosclerotic vascular disease” OR “myocardial infarction” OR “myocardial infarct” OR “heart attack” OR (coronary AND (event OR events OR death OR deaths OR morality or morbidity)) OR “heart disease” OR “coronary disease” OR “cardiovascular disease”) NOT \text{TI} = (rheumatic OR rheumatism OR Congenital OR Vincents Angina OR Vincent’s angina OR Vincent) \]

This returned 47,823 items published between 1900 and 1984 in the Core Collection. Fig 22 plots the number of publications per year, split as previously (1900–1942; 1943–1984).
Fig 22: Growth of indexed records related to CHD, atherosclerosis, and cholesterol metabolism in WoS Core Collection, 1900–1984. Top panel shows the period 1900–1942; bottom panel shows the period 1943–1984.

Between 1900 and 1942, 2,018 items were published. Fig 22 shows that the number of articles on CHD was greater than the output of atherosclerosis and blood cholesterol research from the beginning of this period. There are similar trends in growth – a period of sustained growth after WWI, but then a decline during WWII. Between 1943 and 1984, 45,805 items were published.
On growth

As major growth for each of these stands occurs in the second half of the period, we may compare their rates of growth. Atherosclerosis research is categorised by linear growth. In Fig 23, the linear trend equation is supplied along with an $R^2$ of approximately 0.96, approximately 96% of the variation in Y (atherosclerosis growth data) is explained by X (linear growth). While cholesterol research conforms to polynomial (base 2) and CHD research conforms to exponential growth.

$$y = 3E-79e^{0.0953x} \quad R^2 = 0.9788$$

$$y = 1.8256x^2 - 7115.2x + 7E+06 \quad R^2 = 0.9736$$

$$y = 13.455x - 26147 \quad R^2 = 0.9573$$

**Fig 23:** Growth of indexed records related to CHD, atherosclerosis, and cholesterol metabolism in WoS Core Collection, 1943–1984

These trends could be taken to suggest that there is a relationship between these research strands.

To understand this, I compared the growth in these areas to the whole of science during periods of interest, using all records from the all subject areas search in the Core Collection (see Fig 18). In all three areas examined, there was a decline in research during WWI and WWII. Similarly, Fig 24 shows a rise in all indexed records following WWI and then a decline during WWII.
Fig 24: Growth of all indexed records in WoS’ Core Collection, 1900–1945

In Fig 25, global scientific output, like CHD research, grew exponentially between 1943 and 1984, while atherosclerosis and cholesterol research grew more slowly.

Fig 25: Growth of all indexed records in WoS’ Core Collection, 1943–1984

To what extent then are these results simply an artefact of the database? To test this, I compared these three research strands to four other major biomedical research topics – tuberculosis, syphilis, cholera, and diabetes.

To identify documents on tuberculosis: TI=(Tuberculosis OR tuberculous OR tubercle bacillus OR Tuberkulse OR tuberculous OR Phthisis OR scrofula
OR “white Plague”). I also performed a search for “consumption” from 1900–1930 restricted to biomedical research categories. This returned 179 publications, of which 131 were related to tuberculosis detected and retrieved via manual addition. The last document to use this term to refer to this disease was a document from 1918. After combining these search results, this produced 34,832 records. To identify documents on syphilis: TI=(Syphilis OR syphilitic OR (Treponema AND pallidum)) – producing 10,588 records. To identify documents on cholera: TI=(cholera OR cholerae) – producing 7,276 records. To identify documents on diabetes: TI=(diabetes OR diabetic OR Diabète) – producing 44,167 records.

Between 1900 and 1942 scientific interest is primarily devoted to the infectious diseases, particularly in the first decade (Fig. 26; Top Panel). While cholera research collapses after this, both research into syphilis and particularly tuberculosis dominate. However, from around the 1920s we find a rise in interest in both diabetes and CHD.
Between 1943 and 1959, research into tuberculosis is dominant (Fig. 9; Bottom Panel). Around 1961, we observe the steep rise in research related to
both diabetes and CHD that remains growing exponentially to the end of the period. Note, the modest rise in cholera research in the 1960s that corresponds to a major cholera outbreak in 1961.

Thus, it is clear that the growth measured in CHD, cholesterol, and atherosclerosis is not an artefact of this database nor can their growth be explained simply by the global growth of scientific output. If this was an artefact of indexing practices, we would expect similar trends in growth across all biomedical conditions, particularly in the period 1943–1984, as this is the period that large scale indexing was initiated. These figures are consistent with existing understanding of the growth of these biomedical research topics. Thus, title searches of indexed databases seem to be a good method of estimating scientific interest around specific research topics; at least when confined to WoS’ Core Collection.

Of interest, cholesterol, atherosclerosis, and CHD research all witness sustained growth after WWII, particularly from 1948 onwards. In 1948, the United States passed the National Heart Act (1948), which created the National Institute of Health (NHI) to launch “a full-scale attack on the Nation's Number 1 destroyer of life – cardiovascular diseases” (US Public Health Service 1948, p.1059). Following the creation of NHI, the US Public Health Service directed substantial funds to research into CHD. According to Manton et al. (2009), the NHI funnelled around $5 million into heart disease research in 1950, in 1960 this had risen $60 million dollars, in 1970 around $65 million, and in 1980 received more than $75 million.

**Reflection**

Terms appearing in the title of a document will typically come after a period of use within the body of scientific texts. For a term to be used in a title, scientists need to possess an understanding of its meaning; the concept must have at least some stability and frequency of use before widespread use. We see this in atherosclerosis, cholesterol, and CHD research, as established by the short histories provided prior to the bibliometric methods.
4. On capturing the diet–heart paradigm

Overlap

The previous queries were designed to be inclusive in order to capture a large set of literature relevant to the conjectured process linking diet to CHD. By merging the results of the three queries, 86,679 records remain. This is fewer than the combined totals (91,090) because 3,100 documents are included in more than one query.

To understand how many documents were caught in this overlap, further queries were used to remove the words involved in other queries. For example, we can query WoS to first search for documents that include the terms relevant to atherosclerosis (A), then remove terms associated with cholesterol (B). By doing this, we can ascertain the proportion of (B⊂A)/A. In this example, 1,321 records from the cholesterol query are contained in the atherosclerosis records – 11% – while the proportion of (A⊂B)/B is 4%. If we consider the CHD query (C), then (C⊂A)/A totals 876 records (7% overlap), while the proportion of (A⊂C)/C is 2%. Finally, we need only consider the relationship between (B⊂C)/C, which identifies 991 records (3% overlap), with (C⊂B)/B has a 2% overlap. This overlap is expected as these research topics were not suspected to be completely independent.

The combined results, however, capture much literature only loosely related to diet–heart debate – the general research areas in which the debate unfolded. While the vast majority of documents relevant to the diet–heart debate will be contained within this set, we need to be more specific. To do this, we need to identify dietary terms within the titles of the identified documents.

Layering dietary terms

To detect literature directly related to diet–heart research, I merged the preceding queries via the Boolean operator OR with the following query:
A document that contained any of these words in the title AND terms related to atherosclerosis, cholesterol, or CHD research is likely to be directly relevant to diet–heart research. These additional terms mostly represent either dietary fats, either in the foods they are found in or their fatty acid classes. Note, however, the inclusion of the four broad terms at the end of the query – nutrition, nutritional, diet, or dietary – that were necessary due to the volume of literature that used only these terms in combination with terms related to serum cholesterol, atherosclerosis, and CHD. This is further justified by the fact that any discussion about lowering or increasing dietary content is inextricably bound up with a broader discussion about diet. Of necessity, changing the proportion of one nutrient simultaneously increases the proportion of others.

This retrieved 5,580 records. However, the query missed one important common phrase used in the diet–heart debate, with Many researchers often using the term “diet-heart” or “diet and heart”. I used the simply query to retrieve these.

\[
\text{TI=} (\text{"diet-heart" OR "diet and heart"})
\]

This produced 79 records; however, only 40 of these were missed by the layered query. These were added to the set. In total, this returned 5,620 records from the Core Collection (Fig 27).
Fig 27: Growth of indexed records in WoS Core Collection regarding dietary research related to atherosclerosis, CHD, or cholesterol metabolism, 1900–1984.

Before 1943, only 59 documents were identified, and these included discussion of dietary animal models of atherosclerosis, assessments of the role of diet on serum cholesterol, clinical studies on the relationship between diet, serum cholesterol, and diabetes in patients fed a high fat and cholesterol diet, studies linking atherosclerosis to milk consumption, and the relationship between serum cholesterol and heart disease. Thus, the link between diet and CHD or atherosclerosis attracted little scientific interest during this period. There is a small flurry of interest in the early 1930s after the increases in CHD and blood cholesterol research in the 1920s, and closely following the spike in atherosclerosis research in the early 1930s.

Between 1943 and 1984, 5,561 items were published. Diet–heart research output dramatically increased after WWII, particularly from 1948 to 1957, but from 1962 until 1971 there is a decline in the growth of the field. The major increase in records between 1948 and 1957 is likely explained, in part, by the massive increase of funds available to pursue heart disease related research.

Manton et al. (2009) reported that NHI funding increased dramatically per year from around 1948, began to decline from 1962, fluctuating between $50–60 million dollars per year until 1970, after which consistent increases raised funding to around $75 million dollars per year by the 1980s. According to the
authors, much of the NHI’s funding was devoted to epidemiology and other public health related research, and thus diet–heart research may have been affected by this. Note, the fluctuation in funding corresponds to the dip in diet–heart research output. While affiliation data is not available for all records retrieved from WoS, there is author country data for 3,309 records published between 1948 and 1984 (out of 5,505 records). Of these, 31% were published by researchers in the US. The second and third most productive countries, England and Canada, contributed only 3.4% and 2.9% of records, respectively.

Another possible contributing factor to the growth of publications throughout the 1950s is that several important breakthroughs were made in this decade. In 1950, Gofman et al. (1950) demonstrated that a raised level of low-density lipoproteins in the blood were associated with the development of atherosclerosis in a dietary cholesterol rabbit model, but also raised in humans who had recently suffered a myocardial infarction. In 1953, Keys (1953) proposed that a strong association existed between the consumption of fat in particular countries and CHD mortality. In 1957, Keys et al. (1957c) proposed the diet–heart hypothesis, that SFAs have a hypercholesterolaemic effect in humans, while PUFA has a modest hypocholesterolaemic effect. Furthermore, the US launched several major prospective cohort studies and RCTs over this period to test the relationship between diet, serum cholesterol, and CHD. The dip in research during the 1960s, while likely influenced by funding levels, may also reflect the fact that large studies launched to test the diet–heart hypothesis did not begin returning their findings until the late 1960s and 1970s.

**Overlap capture**

However, this set of 5,620 publications is probably too specific because it will miss studies that contributed to basic understanding of the mechanisms involved, notably those working on the lipid hypothesis of atherosclerosis, but also more general studies linking serum cholesterol, atherosclerosis, and CHD together. To rectify this, I merged the following queries with AND operators – atherosclerosis AND blood cholesterol terms \((A+B)\), atherosclerosis AND CHD terms \((A+C)\), and blood cholesterol AND CHD terms \((B+C)\). Any document with at least two of these elements in the title would suggest they
are related to the underlying assumptions of the diet–heart hypothesis (i.e., that a disturbance in cholesterol metabolism is the primary causal factor of atherosclerosis and that atherosclerosis is a cause of CHD). After searching for each of these combinations, I retrieved all associated publications and merged the retrieved lists. By doing this, 1,321 records were derived from A+B, 876 from A+C, and 991 from B+C (Fig 28).

**Fig 28:** Growth of indexed records in WoS Core Collection related to overlapping of atherosclerosis, CHD, cholesterol metabolism, 1900–1984.

The combined results of overlap queries retrieved 3,100 documents. To understand the overlap of results with the dietary query, I removed any document within this set that was identified in the dietary query. By doing this, the set was reduced to 2,718 documents – a 12% overlap. Fig 29 shows the growth of indexed records containing a mixture of terms related to atherosclerosis, serum cholesterol, or CHD research with dietary results removed.
Fig 29: Growth of indexed records in WoS Core Collection related to atherosclerosis, CHD, cholesterol metabolism – excluding dietary terms 1900–1984.

These were then combined with the dietary retrieval results to create a dataset for network analysis. By merging these results, 8,338 records with full bibliographies were retrieved (Fig 30).

Fig 30: Retrieved records from WoS Core Collection on diet, atherosclerosis, CHD, and cholesterol metabolism, 1900–1984

Of these, 4,844 were assigned by WoS as being ‘Articles’, which tend to be primary research studies, though other types of document are occasionally
misclassified by WoS. In all, 2,546 are meeting abstracts, 412 letters, 208 notes, 163 editorials, 100 reviews, 34 discussion pieces, 26 book reviews, four abstracts, and one biography (with a wonderful title, *Carl Friedrich Gauss: A Genius Who Apparently Died of Arteriosclerotic Heart Disease and Congestive Heart Failure* (Burch 1958)).

5. Discussion

In Chapter 7, this set of documents is used to construct a dataset of how relevant publications from different research specialities interacted with one another during the course of the diet–heart debate. By downloading all bibliographies of all of these documents, I demonstrate a method to detect literature both missed by my original query and literature not indexed by WoS but cited by any of these documents, including books. By using network analysis, I further demonstrate a method to validate the retrieved results.

If quantitative and network analytic methods are to be used to understand the development of particular research topics or debates, it is vital that methods are developed that are capable of systematically retrieving relevant literature. By removing individual selectivity in what is and is not included beyond the inclusion of specific words in queries the entire literature set can be retrieved. This allows not only for a transparent selection process, but always possible replication. All data underpinning these analyses are available in the Chapter 6 Supplement.

I have demonstrated a method that appears successful in capturing the published journal literature around a specific scientific debate. The trends in growth parallel historical accounts of the massive increase in research related to diet and CHD after WWII, and it conforms to my expectations from my understanding of the literature.

In Chapter 2, I highlighted that one of the major barriers to applying Main Path Analysis and for studying science via network analytic methods has historically been the inability (or perhaps lack of interest) of researchers to focus in on specific scientific theories and hypotheses, rather than disciplines
or broad fields. The only way to do this well, however, is to have a firm grasp of the literature you seek to find systematically before conducting a search. This requires spending time reading and learning the area of science that is of interest.
Chapter 7: Capturing a Paradigm: Part Two Dataset Construction, Validation, and Refinement

Introduction

This chapter details the conversion of the retrieved records described in the previous chapter into a dataset amendable to network analysis, further literature retrieval, and validation exercises. In Section 1, I describe the methodology used to construct the dataset. In Section 2, I discuss the cleaning of these data, focussing on the problem of duplication, and detail my approach to retrieving influential literature missed by the original search. This uses measures of in-degree to find documents highly cited by the retrieved documents, and retrieves any cited ≥20 times. In Section 3, to assess how comprehensively my literature retrieval strategy captured relevant literature, I retrieve the reference lists of 19 literature reviews, written by different authors and published in different years, and match these with records in the dataset. By applying inclusion and exclusion criteria to missed literature, I evaluate why relevant documents were missed and retrieve them. In Section 4, through network analyses, I show that the collected documents conform to the expectations of Kuhn in regards to community structure, but also appear to corroborate Price’s (1965) observations regarding the structure of citation networks (see Chapter 2).

Throughout this chapter, I introduce the reader to several datasets that reflect different stages of refinement. All datasets are available in Chapter 7 Supplement. For each, I have included an edge-list and a vertex-attribute list in .csv formats. These can be imported into various different software tools for network analysis. I have also included Gephi workstation files with key analyses performed within. One of the major values of network analysis, particular in Gephi, is its interactive nature. For each Gephi workstation, a full relational database is held in the ‘Data Laboratory’ window, which can be used to view either vertex attribute or edge data. Further, due to the text search
function, particular studies or attributes can be searched for manually. In the ‘Overview’ window, an interactive visualisation can be examined, as too can different statistics be applied and attributes filtered. All files are titled with the name of the dataset established in-text.

1. Dataset and network construction

From the initial set of 8,338 publications, I downloaded all associated bibliographic data contained in the WoS Core Collection via the “Full Record and Cited References” function in 17 .txt files and merging these in Notepad++ v.7.5.6 (2018). I detected and removed 16 duplicates, leaving a set of 8,322

Edge-list

I constructed an edge-list by parsing data into a ‘Source’ column (all retrieved publications) and a ‘Target’ column (all identified references) via the Sci2 Tool (2009, version 1.3) function ‘Directed Graph’. This describes a network as a set of vertices (documents) connected by directed ‘edges’ (citation links).

In the original .txt file, the source data derives from the “Cite me as” column, while target data derives from the “Cited References” column. In its original format, the .txt contains the full citation for each source document in capital letters in a text string, which includes the last name of author and initial, abbreviation of journal of publication, volume number, page number, and, occasionally, a DOI. For example,

GORDON T, 1977, AM J MED, V62, P707, DOI 10.1016/0002-9343(77)90874-9

The “Cited References” is typically in the same format, but some records are not capitalised or do not contain a DOI. Each reference from a source document is separated by a | and this can be used to parse the data into

14 This problem represented 0.002% of the collected data. Why this problem occurs is hard to assess. To examine this problem, one can search two accession numbers (WOS:A1981NF35200004|WOS:A1981NC61700004), which refer clearly to the same publication, but differ in the supplement number provided. However, two accession numbers (WOS:A1983QF90200001|WOS:A1983QD29100001) have no clear discrepancies in the reference or associated record.
individual records. For consistency, I reformatted both the ‘cite me as’ and ‘cited references’ columns to be in a standardised format in Excel. I then assigned a unique integer identifier to each unique text string.

**Vertex-attribute-list**

From the original .txt, I constructed a vertex-attribute-list. This records, for each integer identifier, the full formal citation (‘Cite me as’ column), year of publication, title, journal of publication, all authors, language of publication, document type, global citations (all citations received from WoS Core Collection), and WoS accession number. To do this, I assigned each unique citation its unique integer identifier, then used the VLOOKUP function in Excel to link all relevant data to this. By doing this, all retrieved documents have full attribute data.

Once parsed, I identified each record by the query used to retrieve it: “D” represents documents from the dietary query, and “H” represents documents from the combined cholesterol and atherosclerosis, cholesterol and CHD, and atherosclerosis and CHD queries. Another two categories of retrieval were included for the search methods described in Sections 2 and 3; “A” – documents cited ≥20-times; “R” – documents recovered during the validation.

However, vertices can also represent non-retrieved documents. These derive from the reference list of retrieved documents, but represent documents not retrieved from WoS. These do not have their reference lists, all authors, global citation, or WoS accession numbers recorded. For these, I include only their formal reference, year of publication, first author, and journal. To extract these data, I used the ‘text to column’ function in Excel to separate the formal references by their commas into these three attributes, and aligned them with their appropriate integer identifier along with their full reference via the VLOOKUP function.

**Network Construction**

In Gephi (v.0.92), I imported the vertex attribute list (in a .csv file) into the workstation, and the edge-list (in a .csv file). From this, a graph G was constructed of the vertex set V and edge set E:
By converting an adjacency matrix

\[ A_{ij} = \begin{cases} 
1 & \text{if there is an edge from vertex } j \text{ to vertex } i \\
0 & \text{otherwise} 
\end{cases} \]

If a document \( j \) references document \( i \), then the graph (network) contains a citation from \( j \) to \( i \). This produces a directed network with 65,879 vertices and 131,699 edges. This is many times larger than the original set of documents because it includes all of these and all of their references.

2. Cleaning and additional records

These data required extensive cleaning. As highlighted by Marx (2011), WoS data has a problem of duplication due to variations in the spelling of references. This a particular problem for records from before 1950, for records not formally indexed in the Core Collection, and for some documents that are regularly cited in various different manners. Variants stem from misspelt names, erroneous volume numbers or page numbers, inclusion or exclusion of initials, or references in different formats.

On the problem of variants

One can test this by using a ‘Cited-Reference Search’ in WoS, which pulls records from the base dataset of bibliographies. To demonstrate this, I take Anitschkow and Chalatow’s (1913b) review of the literature (not indexed in the Core Collection), Anitschkow’s (1933) chapter on atherosclerosis in Cowdry’s Arteriosclerosis (not indexed in the Core Collection), Leren’s (1966) secondary prevention trial (indexed in the Core Collection), and the Paul et al. (1963) cohort study (indexed in the Core Collection).

Anitschkow and Chalatow (1913b) published an influential review in the German Journal Zentralblatt für allgemeine Pathologie und pathologische Anatomie. I detected eight spelling variants of Anitschkow’s name in reference lists (Anitschkow; Anitchkow; Anichkow; Anitschow; Anitchkov; Anitschkov; Anichkov; Anitschkoff). I searched these variants without initials in the Cited
Reference search. The Core Collection does not hold a record for this, but by using the Cited-Reference search and merging the 30 variants, I found a total of 460 citing documents as of 2018.

In 1933, Anitschkow’s chapter in Cowdry’s *Arteriosclerosis* was the first exposition of his lipid-infiltration theory of atherosclerosis in the English-language. As a book chapter, the Core Collection holds no formal record for this. By searching for these variants of Anitschkow’s name confined to 1933, I found 65 variant records of this chapter, cited by 496 documents in the Core Collection.

This is not a problem confined to documents written by Anitschkow. I searched for Leren’s (1966), secondary prevention trial published as a supplement in *Acta Medica Scandinavica*. In the literature, I’ve seen both the author misspelled as Laren and an erroneous publication date of 1967 and these were variants were searched for. According to the Core Collection, this trial has been cited 12 times. By performing a Cited-Reference search and merging 66 variants, 416 citing documents were recovered.

However, for some documents, such as the Paul et al. (1963) document in *Circulation*, the problem of variants is less pronounced. In the Core Collection, this has a citation score of 659. Performing a Cited-Reference search, only an additional six variants were discovered that were cited by just eight documents.

As a general rule, variants typically pose a serious problem only for: (i) historical literature before 1950; (ii) literature from journals contained within special issues or supplements; (iii) well-cited literature that increases the risk of spelling mistakes in the reference; (iv) unindexed literature.

**Cleaning**

To clean these data, I considered using algorithmic methods to identify records that contain over 98% of the same letters and numbers via the Sci² Platform. However, after experimenting with this, I found this often merged unique documents, but also missed many duplicates (this probably accounts for why WoS has not been able to correct the problem of variants).

I cleaned these data manually. I used Gephi’s workstation to construct a
relational database composed of the ‘edge-list’ and ‘vertex-attribute-list’ as described above. I then used the ‘merge’ function, which merged two or more vertices together, preserving their edges but deleting variant vertices. By using the text search, I refined the results by year of publication and then ordered those years alphabetically in relation to the formal citation. This allowed me to merge records that clearly referred to the same document. In cases that this was not obvious but was suspected, I used PubMed to check the record.

Additionally, I performed a computational method that included using the formal reference, removing all spaces and punctuation, and retaining only certain information: (i) a text-string with year of publication, volume number, and page number; (ii) last name of author, year of publication, volume and page number; (iii) last name of author, year of publication, and journal of publication; and (iv) last name of author, journal of publication, and volume and page number. The first string captures errors in the name of author, the second errors in journal name, the third errors in volume and page numbers, and the last errors in year of publication. These four strings operate like the accessions numbers that I parsed into the vertex-attribute-list beside their full records via the VLOOKUP function. This allowed me to detect duplication of text-strings via highlighting duplication via Excel’s conditional formatting function and sorting the columns alphabetically. I then followed up these manually to examine whether the identified records were unique or not. This manual examination is essential as many highlighted records are false positives. In cases where duplication was clear, I merged these manually in Gephi.

This process reduced the network to 59,470 vertices and 130,407 edges – losing 6,409 vertices (10%) by resolving duplication. The reduction of 1,292 edges (1%) stems from removing multiple citations from one document to another that occurs when merging duplicates. This occurs when authors reference particular sections of a study, which tends to happen when authors reference books or reviews. For example, merging 57 variants of Fredrickson’s (1967) five-part review, ‘Fat Transport in Lipoproteins’, reduced the network by 476 edges. In the clean version, only 213 unique documents cited this, indicating that most citing documents cited multiple sections of this review.
Globally, this review had been cited 7,210 times by 2019; however, these derive from only 2,348 unique documents.

It is impractical to report all the changes to the dataset during cleaning. Gephi does not record previous merges; but, even if it did, these data would be hard to interpret. For example, during cleaning, a record is merged when it is a duplicate of another record. In merging, a new vertex is constructed that receives a new integer identifier in both the attribute and edge-list, deleting the previous vertices. However, more duplicates will often be discovered that will be merged with the previous merged record. Again, a new unique integer identifier is constructed. This makes tracking the changes overly hard.

Aware of the problems this poses for replication and validity, I compared the results of the fully cleaned dataset and the uncleaned dataset to test the robustness of my approach and have provided the full raw data (see Ch7.S1). Despite the extensive cleaning, the basic network properties remain relatively similar between the clean and unclean datasets. The major benefits of cleaning appear to be in correcting citations to non-indexed records, particularly books, and citations to duplicates associated with articles published as supplements.

**Additional documents**

Once cleaned, I ranked vertices by *in-degree*, which tells how often a document has been cited. From this, I detected another 386 documents that had been cited ≥20 times by this set of literature. I searched for each reference via the Core Collection. Of these, 306 were indexed and were retrieved following Batagelj *et al.* (2017) recommendation that this step is wise in constructing representative citation networks due to highly cited documents detected in this manner being likely of direct relevance.

However, 80 were not indexed in the Core Collection. Of these, 46 (57%) were to non-journal documents (primarily books, laboratory manuals, government guideline reports, and official statistics). The remaining 34 documents were journal publications. Of these, four were in supplements to *Acta Medica Scandinavica*, four from *Advances in Lipid Research*, three from *Journal of Chronic Diseases*, and two from *Scandinavian Journal of Clinical and Laboratory Investigation*. The remaining 21 were spread across 21
journals; primarily regional, such as *Minnesota Medicine*, *Journal of Mount Sinai Medicine*, and *Voeding*.

I recovered 33 of these 80 (23 journal publications; 10 books) by searching journal archives or by interlibrary loans. I prioritised records that appeared likely from their title/abstract to contain references lists directly relevant to the diet–heart debate. Due to the time-consuming nature of this, and limited returns, I did not pursue 47 documents that seemed of limited interest (primarily methodological documents and general methods textbooks; these documents can be found in the datasets under the code “Z” in the query column). For example, two statistics textbooks have been cited 58 and 166 times, respectively. These would have introduced noise by citing statistical methods documents not linked to the diet–heart debate.

This process recovered 339 unique documents that were retrieved, and their bibliographies and attributes were coded into the dataset. To understand why these records were missed by my systematic search method, I first classified each document into three broad categories by reading titles, abstracts, or full-texts: (i) *Methods* – documents introducing particular methods; (ii) *Human* – studies reporting the results of intervention or observation studies in humans or reviewing their results; (iii) *Animal* – animal studies, typically focussing on animal models of atherosclerosis.

In total, 89 methods; 198 human studies; and 52 animal studies (see Ch7.S2). On my reading, all document titles appear related, in some way, to serum cholesterol, atherosclerosis, and CHD, or dietary and other lifestyle factors related to these. The most indirectly related papers are those with a methodological focus. For example, Lowry *et al.* (1951), a paper describing a method for the measurement of proteins, is not itself a document addressing any substantive issue in diet–heart research; however, this method was commonly used in conjunction with serum cholesterol measurement to detect blood proteins. I decided to leave these methods paper in the dataset because, while they do not contribute empirical findings to this research area, their inclusion helps to identify groups of documents that share particular methods.
3. Reference validation

To assess the comprehensive of the dataset, I used references from different types of review published at different times. I used WoS and retrieved the references from these documents held in the Core Collection. This will miss records if the citations in those documents are not to indexed records. This can be because authors cite unindexed records or because of messy records – the duplicates discussed earlier. However, this approach allows for an efficient check via WoS accession number comparison between datasets via VLOOKUP matching. Thus, the number of references reported below is the number of indexed records retrieved from the documents concerned.

I examined all missed articles and detailed why these had been missed. This included either why relevant documents were missed (by examining the title terms) or why documents were missed due to irrelevancy. To judge whether a document was relevant, I established exclusion and inclusion criteria. This is necessary as reviews also cover other topics.

A document was deemed relevant IFF it sat within the framework of the diet–heart theory, i.e., focussing on the relationship of different types of dietary fatty acids and their effect on serum cholesterol levels, and via this mechanism either understands or critiques the development of atherosclerosis and CHD. A document had to have either a title that demonstrated this or an abstract which made this aim clear. Finally, any document evaluating the safety of dietary modification was included. All documents that met these criteria were included in the dataset with their full bibliographies. In most cases, this involved converting a ‘non-retrieved’ vertex into a ‘retrieved’.

A document was deemed irrelevant if it did not adhere to this framework. Studies on basic cellular mechanisms of atherosclerosis, on drug treatment to lower serum cholesterol levels, on the basic structure of lipoproteins, atherosclerosis or CHD prevalence studies that did not discuss diet, and on the basic chemistry of fatty-acids were excluded. While some would have influenced opinion about the validity of the diet–heart idea, their inclusion would pull in a lot of literature only loosely related. Any documents on these subjects that were frequently used during the diet–heart debate (as judged by
their citations) will still appear in the network, but will not contain their full bibliographies unless they were cited ≥20 times by documents in the network. Documents on platelet aggravation, blood coagulation and prostaglandins were similarly excluded.

All documents deemed relevant and not already contained in the assembled dataset were added in full. All included publications using this method are recognised by “R” in the query attribute column.

**Meta-analyses and systematic reviews**

I began with eight systematic reviews and meta-analyses that examined the effect of dietary fat on CHD/CVD risk (Sacks et al. 2017; Hamley 2017; Hooper et al. 2015; Ramsden et al. 2016; de Souza et al. 2015; Chowdhury et al. 2014; Ramsden et al. 2010; Jakobsen et al. 2009). I removed records published after 1984, reducing the set from 1,070 records to 125. In all, 80 of these references were amongst the previously retrieved documents (D – 62; H – five; A – 13).

The remaining 45 documents were examined to understand why these had not been included. As some of the meta-analyses included studies relevant to type-2 diabetes, I anticipated this might occur. I examined the 45 documents and included documents IFF their primary focus was on the relationship between diet and serum cholesterol, atherosclerosis, and CHD; 21 relevant documents were identified.

The 24 documents excluded comprise eight examining dietary influences on diabetes; three on methods in epidemiology or studies in which no dietary results were reported; two on platelet function in response to dietary fat; two on the pathology of atherosclerosis; and one each on: blood clotting times; diet and multiple sclerosis; diet in xanthomatosis; triglycerides; hypotension; dietary vitamin E; chromatography; linoleic acid in adipose tissue; and the effect of smoking and weight loss on HDL-C.

Due to their strict inclusion criteria, meta-analyses and systematic reviews cannot be used alone to retrieve literature, but are useful for validating literature searches, particularly for the presence of primary studies. Of the 101 relevant references, my search had retrieved 79%.
The 21 documents missed typically had brief titles, such as Dayton’s (1968) ‘Trial of unsaturated-fat diet’ – a letter responding to criticisms of a previous trial, Ahren’s (1976) critical review of the state of diet–heart research entitled ‘Management of Hyperlipidemia - Whether, Rather Than How’, and Heady’s (1974) article ‘Are PUFA Harmful’. Thus, my query method works when authors are being precise. However, commentary pieces that use simple titles are missed.

**Retrospective narrative reviews**

Another useful tool for validation is retrospective pieces that have analysed this debate. Unlike systematic reviews, these construct narratives; they tend to highlight the importance of particular reviews on shaping opinion and tend to capture differences of opinion as expressed by different scientists alongside important primary studies.

While there has not been a huge amount of work on this in the academic literature, a few pieces stand out. Most attention has been directed to understanding the acceptance of the lipid-hypothesis, the conjectured causal link between serum cholesterol and atherosclerosis. In an influential five-part thematic review, Steinberg (2006) surveyed the history of what he terms the ‘cholesterol controversy’. He was a participant in this debate and specialised in the basic science of lipid metabolism and atherosclerosis. Focussing on these aspects, he attempts to explain both why the lipid hypothesis is correct and why it took so long for the scientific community to accept it. While the debates around the validity of the lipid and dietary hypotheses are intertwined, it is not necessarily the case that if the lipid hypothesis is true then diet–heart hypothesis is correct. While Steinberg touches on relevant dietary aspects, this is not his primary focus.

I examined the correspondence between my dataset and the references supplied by Steinberg. Of the 259 references, 150 are to documents published before 1985. Of these, only 62 had been captured by my search (D – 23; H – 17; A – 21; R – one). Of the 88 documents missed, most are basic studies of the mechanisms involved in atherosclerosis – primarily on the role of macrophages or studies on endothelium permeability and fibroblast
development (24 documents); many others are studies of the impact of particular compounds on serum cholesterol or atherosclerotic processes (32 documents), primarily the role of HMG-CoA reductase inhibitors (statins), heparin, and clofibrate.

The remaining 32 are spread across various topics. Of these, 25 are not directly related to the dietary hypothesis: seven on the basic structure of lipoproteins; four on familial hypercholesterolaemia; five on xanthomatosis; five on methods [either statistical or experimental]; two studies on aggression linked to serum cholesterol levels; one on plant sterols, and one on basic cellular biology. Just seven of the missed documents appeared to be relevant. Of these, three were editorials or letters and four were reviews/commentaries.

Thus, the discrepancy between the present study and the history provided by Steinberg is explained by the different focusses. Steinberg focussed on the validity of the lipid-hypothesis, this study focusses on the debate over the validity of the diet–heart hypothesis. While these basic studies may have influenced thought, this will be picked up in my network IFF dietary relevant research cited these. Only a further seven documents were detected that were directly relevant; thus, the search captured 90% of relevant records cited by this history.

One of the most prominent modern critics of the diet–heart link is Ravnskov. In 1998, he published a narrative review attacking the evidence-base. Of the 102 references, 53 were to articles published before 1985. Of these, 30 had been captured by my search (D – 22; H – one; A – five).

However, 23 articles were missed, all of which were relevant. Most of these are retrospective observational studies in particular groups. The titles tend to include either atherosclerosis or CHD and the name of the group studied (e.g., Mann et al. (1972) ‘Atherosclerosis in Masai’ and Malhotra’s (1967) ‘Epidemiology of Ischaemic Heart Disease in India with Special Reference to Causation’). Thus, my search tended to miss such articles. Thus, only 57% of the relevant records in Ravnskov’s (1998) article were previously captured. The missed articles were also not included in the systematic reviews or Steinberg’s history.
Gordon (1988), a major figure in the diet–heart debate and lead investigator of the Framingham Study, published a short history. Of the 19 references, 18 were published before 1985. My search captured 17 of these (94%; D – one; H – four; A – 12). The missed study was, again, a retrospective study of a particular community.

**Narrative reviews**

I selected eight reviews that reflect different perspectives and specialities but which all had a focus on diet, serum cholesterol, atherosclerosis, and CHD. These were published between 1959 and 1983. There are few reviews before 1959 because the SFA link only clearly emerged out of the primary literature in 1957.

1. Stallone’s (1983) balanced review. Of 78 references, 57 (73%) had been captured by my search (D – 45; H – four; A – six; R – two). Of the 21 missed, 20 were relevant. The only exclusion is a previous document of Stallone’s (a self-citation) in which he reviewed approaches to epidemiology.

2. Coates’ (1983) sceptical review of dietary lipids and CHD. Of 85 references, 45 (53%) had been captured (D – 25; H – nine; A – 10; R – one). Of the 40 missed, six were relevant; 88% of relevant records captured.

3. Glueck et al. (1978) generally supportive review of diet and CHD. Of 86 references, 59 (69%) had been captured (D – 36; H – four; A – 17; R – two). Of the 27 missed, 19 were relevant; 74% of relevant records captured.

4. Shaper’s (1977) supportive review of dietary prevention of CHD in communities. Of the 31 references, 17 (55%) had been captured (D – 10; H – one; A – six). Of the 14 missed, six were relevant; 74% of relevant records captured.

5. Reiser’s (1973) critical review of the literature on the relationship between saturated fat and serum cholesterol. Of the 59 references, 41 (69%) had been captured (D – 32; H – 0; A – eight; R – one). Of the 18 missed studies, seven were relevant; 85% of relevant records captured.

6. Malmros (1969) supportive review focussed on dietary prevention of atherosclerosis. Of the 21 references, all had been captured (100%; D – 13; H – one; A – seven).
7. Stamler’s (1960) supportive review of the literature on dietary prevention of atherosclerotic coronary heart disease. Of 131 references, 61 (47%) had been captured by my search (D – 37; H – nine; A – 14; R – one). Of the 70 missed documents, 16 were relevant; 79% of relevant records captured.

8. Jolliffe’s (1959) supportive review on fat, cholesterol, and CHD. Of the 46 references, 26 had been captured (D – 21; H – three; A – two). Of the 20 missed documents, eight were relevant (three community studies, three on dietary fat and atherosclerosis or CHD, and two reviews); 76% of relevant records captured.

Summary

In total, 19 reviews published between 1959 and 2017 were examined. Of these, eight were the most recent systematic reviews or meta-analyses, two were retrospective pieces, one was a critical narrative review published after 1985, and eight were narrative reviews written between 1959 and 1983. In total, 656 references to unique indexed documents published before 1985 were retrieved from these. My search captured 318 (49%) of these (D – 206; H – 43; A – 69). However, of the 338 missed references, only 134 met the inclusion criteria; therefore, my search captured 70% of relevant literature.
4. Network analysis: A test of Kuhn and Price

The dataset contains a large proportion of relevant publications cited by reviews, but does it also contain irrelevant literature? There are two feasible methods for examining this. First, one could read the titles and abstracts of all documents to decide what is relevant. Second, network analytic methods can be used to establish whether the collected documents form a connected community. Here, I demonstrate the utility of the latter. As I designed the literature retrieval queries to capture relevant literature, and because I examined these during the literature retrieval phase, the former seemed unnecessary and impractical. For discussion of the content of documents, see Chapters 8–9.

If documents are part of a single community, then a network analysis should reveal a large weakly-connected-component composed of the vast majority of vertices surrounded by isolates (see Chapter 3). As citation networks are directed graphs that verge on acyclic, a ‘weak component’ is established by converting the asymmetric adjacency matrix into a symmetric adjacency matrix of an undirected graph. This replaces a directed edge from vertex \( j \) to \( i \) with an undirected edge between \( j \) and \( i \). A ‘weakly-connected-component’ is defined as a connected undirected graph in which there is a path between any given vertex and any other vertex in a connected sub-graph, while a ‘large component’ describes the largest connected sub-graph. An ‘isolate’ is a vertex with no incoming or outgoing edges. Isolates tend to be meeting abstracts, which tend not to contain citations and are generally rarely cited.

Following Price’s (1965) work, I expect that documents will be clustered by (i) time (documents are more likely to cite recently published documents); and by (ii) speciality (the type of research conducted) – both of these aspects are addressed in Chapter 8. No other large component should be in this dataset. Despite this clustering, it should display a high degree of connectivity: (iii) the diameter (the geodesic path between the two most distant vertices) and mean geodesic path (shortest path between any two randomly chosen vertices) should be a small fraction of the sum of all edges. If the largest
component has these two properties, it possesses small-world properties (Milgram 1967).

From Kuhn’s ([1962]1970) ideas of community structure, I anticipate that this network will be clustered around a core of highly cited literature. If this is a community sharing a paradigm, the literature should rely on shared sources, though, because of the time and speciality dimensions, I anticipate this will vary. I expect to find a heavily skewed distribution of citation approximated by a power distribution, with an exponent between -2 and -3, cohering with Price’s (1965) finding of power-law distributions in large citation networks.

As two major queries were used to extract the vast majority of publications (D+H), I expect to see that these sets are internally well-connected. When considered together, these should demonstrate signs of a clear relationship via similar referencing behaviours. D-documents should cite H-documents and vice-versa. Further, D-documents should share references (non-retrieved records) with other H-documents. While I treat these as separate entities here, I do so only to test whether they form different parts of a whole, rather than distinct and separate research topics.

**Dataset description**

These data were cleaned again to harmonise the addition of the new records by the same process described previously. This produced a dataset containing 8,795 unique documents and their full bibliographic data. In Table 6, these are matched to the search strategies used in retrieval, and into particular document type.

**Table 6: Dataset descriptors**

<table>
<thead>
<tr>
<th>Query</th>
<th>Articles</th>
<th>Meeting Abstracts</th>
<th>Letters</th>
<th>Notes</th>
<th>Reviews</th>
<th>Discussion</th>
<th>Editorials</th>
<th>Book Review</th>
<th>Biography</th>
<th>Book</th>
<th>Manual</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>3238</td>
<td>1784</td>
<td>423</td>
<td>159</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>2706</td>
<td>339</td>
<td>10</td>
<td>134</td>
<td>5616</td>
</tr>
<tr>
<td>H</td>
<td>1601</td>
<td>756</td>
<td>197</td>
<td>137</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>2706</td>
<td>339</td>
<td>10</td>
<td>134</td>
<td>2706</td>
</tr>
<tr>
<td>A</td>
<td>297</td>
<td>3</td>
<td>7</td>
<td>22</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>134</td>
<td>134</td>
<td>0</td>
<td>0</td>
<td>339</td>
</tr>
<tr>
<td>R</td>
<td>109</td>
<td>7</td>
<td>7</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>134</td>
<td>134</td>
<td>0</td>
<td>0</td>
<td>134</td>
</tr>
<tr>
<td>Total</td>
<td>5245</td>
<td>2550</td>
<td>634</td>
<td>329</td>
<td>27</td>
<td>10</td>
<td>10</td>
<td>8795</td>
<td>1001</td>
<td>10</td>
<td>10</td>
<td>8795</td>
</tr>
</tbody>
</table>
Network description

I converted these data into two datasets (Table 7): (i) a dataset composed of all retrieved documents and all of their references, including vertices representing non-retrieved documents (see Ch7.S3–4); (ii) a dataset composed of only the retrieved documents and only the references between these (Ch7.S5–6). These datasets will be referred to as ‘FULL’ and ‘BASE’ respectively and are available as Gephi workspaces in the Supplement (Ch7.S7–8).

To understand whether these formed networks, and their relative size, I performed a component analysis by converting the asymmetric adjacency matrix into a symmetric matrix and applying Tarjan’s algorithm (1972) to detect weakly connected components in Gephi.

**Table 7: Network descriptors**

<table>
<thead>
<tr>
<th>Network</th>
<th>Vertices</th>
<th>Edges</th>
<th>Weakly connected components</th>
<th>Vertices in largest component</th>
<th>Edges in largest component</th>
</tr>
</thead>
<tbody>
<tr>
<td>FULL</td>
<td>63970</td>
<td>147788</td>
<td>2509</td>
<td>60887</td>
<td>147208</td>
</tr>
<tr>
<td>BASE</td>
<td>8795</td>
<td>42458</td>
<td>2892</td>
<td>5870</td>
<td>42424</td>
</tr>
</tbody>
</table>

The BASE-dataset, despite containing <14% of the vertices contained in the FULL-dataset, constitutes ~29% of all edges. Thus, this tells us that the retrieved documents are substantially more likely to cite another retrieved document than a non-retrieved document.

The BASE-set represents 14% of the records of the FULL-set. If all of the BASE-set contained documents on completely separate topics, we would expect this value to be much lower. As we know the sum of references of the BASE-set (147,788), then this set has the potential to connect to 147,788 unique documents – creating a set of 156,583 vertices (147,788 + 8795). The true value, however, is 41% of this (63,970). Thus, many documents cite the same literature. To understand this in greater depth further analysis is required.

Components
The FULL-dataset contains 2,510 weakly-connected components. The largest has 60,887 vertices (95%) connected by 147,208 edges (99.6%). The other 3,083 are spread across 2,508 components, the largest of which has 29 vertices (a single document citing foreign language documents uncited by any other document). Of these, 2,376 are true isolates, while 707 form 132 weakly-connected components that stem from the reference lists of 140 documents. Thus, this set contains a large weakly-connected component surrounded by vertices that do not themselves form a substantial connected component.

The BASE-dataset contains 2,892 weakly-connected components. 5,870 (67%) of vertices sit within the largest component and are connected by 42,424 edges (99.9%). The remaining 2,925 vertices are spread across 2,891 weakly-connected components. Of these, 2,865 are isolates, while 60 are within 26 weakly-connected components, the largest of which is a string of five vertices focussing on ultrasound techniques in atherosclerotic lesion detection. The vast majority of documents not in the largest component are meeting abstracts; these tend not to contain references nor are they typically cited.

Even after excluding any reference to any document not within the BASE-set, we still observe a large weakly-connected component composed of the majority of documents and citations. While the BASE-dataset has a smaller proportion of total vertices in its largest component, this is explained almost entirely by the number of isolates. If we disregard these isolates in both sets (documents with no citations or references), about 1% of documents in both the BASE and FULL datasets are neither cited by nor reference documents from the largest component of these sets.

These results are consistent with Kuhn’s conception of a community, and provide validity to the results; the documents in these sets are related as judged by their citation relations.

**Connectivity**

How robustly connected is this network? Do the documents form a sparse (loosely) connected network, which would hardly provide strong evidence that these are a community. To understand this, I reverted to the original asymmetric matrix to establish a directed graph. After including component
membership as an attribute into the vertex list, I measured the diameter of both large components (the geodesic path between the most distant vertices) and the mean geodesic path length between any two randomly chosen vertices.

Of the largest-component in the FULL-dataset, the diameter is 14, while the average geodesic path is 4.58 – publications have, on average, 4.6 degrees of separation. Of the largest-component in the BASE-dataset, the diameter is 13, while the average geodesic path length is 4.1 – publications have, on average, 4.1 degrees of separation. Therefore, in the largest-component of both datasets, the diameter of the network is a small fraction of the total number of vertices, and the low average geodesic path is indicative of the small-world phenomenon (Milgram 1967).

**Density**

However, how dense is this network? Measuring density is complicated in citation networks, because citation networks are typically very close to being acyclic (a network that contains no loops) due to the temporal dimension of publication. That is, publications almost always reference backwards in time to documents already published; it is rare to find publications citing pre-prints or simultaneously citing one another; at least in the period studied here. Accordingly, the standard network density measurement is unsuitable because it assumes that all vertices can possess an edge with any other vertex (see Chapter 3).

There is no easy way around this problem. One could adapt the above equation to remove the possibility of documents referencing documents published in the same year and in the future, which would be an improvement, but not strictly accurate (see Chapter 3). One cannot establish conclusively exactly what the maximum number of possible edges is in a citation network. To understand how well-connected this graph is, we need to examine other measures of connectedness.
Edge distribution

Price (1965) proposed that distribution of citations to documents can be approximated by a power-law. I examined whether the documents collected meet this expectation in the FULL-dataset’s largest component. If Price is right, it is this dataset that ought to display a power-law distribution because it contains the full-reference lists of all retrieved documents.

In the FULL-largest-component, the average in-degree (citations) is 2.42, with a median of 1. This average (mean) simply takes the total in-degree of each vertex against the total number of vertices. The median in-degree is 1, which reflects a right-skewed long-tailed distribution; 2,308 documents are uncited, 39,578 are cited once, 8,037 twice, 3,311 three times, and 1,975 cited four times. Only 18% of publications received >3 citations. Only 2,149 documents receive ≥10 citations – about 3% of vertices – but these account for 51,176 citations (35% of all citations). This distribution fits a ‘70/30 rule’, 30% of documents receive 70% of citations. In Fig 31, I plot the cumulative distribution of citation with % of publications on the y-axis and % of citations on the x-axis.

![Fig 31: Cumulative distribution of citations to documents in the FULL-largest-component](image)

To examine whether this distribution conforms to a power-law, I exported the distribution data to MATLAB v.R2019b and applied Clauset et al. (2009) maximum likelihood estimate method for power-law fitting via Clauset’s MATLAB script (2012). However, the number of documents with zero citations in this set is simply an artefact of the collection strategy (only retrieved...
documents have a chance to have zero citations); therefore, these were removed. Fig 32 shows that the in-degree (k) distribution can be approximated by $P(k) \sim k^{-2.21}$.

Fig 32: Power-law fitted to log-log transformed distribution of in-degree (k) of the FULL-largest-component

Restricting the analysis to the BASE-largest-component, the mean citation is 7.2 and the median is 1; reflecting the right-skewed distribution of these citations. In all, 1,928 documents receive no citations, 1,098 receive one, 612 two, 368 three, and 274 four. About 38% of documents receive >2 citations, but only 1,073 documents receive >10 citations – 18% of vertices – but these account for 34,051 citations (80% of all citations).

This suggests that these documents are disproportionately collected around a small set of very highly cited documents. The probability of a randomly selected vertex having a degree $\leq 2$ is approximately 82% in the FULL-component and 62% in the BASE-component. From this, we know few documents ever receive many citations. Because of the skewed distribution, we also know that these networks are relatively sparse in that there exists only a small number of edges between vertices. In a maximally connected network, every vertex would share an edge with all other vertices. Even taking into account the time dimension of publication, it is clear that only a small proportion of documents that are available to reference are actually referenced by any individual document. While this reflects the conventions of reference list size
(and journal demands), this does not account for the fact that only a few documents receive the vast majority of citations.

**Set Analysis**

But are documents retrieved from the different methods related? To understand this, I compared the citations between documents captured in different retrieval sets (D, H, A, R) within the FULL-largest component. I included an attribute to all edges into the dataset identifying them as deriving from documents in these sets, then parsed this into 16 separate graphs that capture the interaction between these four sets.

Below (Table 8), we see the total number of documents cited, the number uncited, and the total citations to documents in particular sets. I have also included the proportion of documents in a set cited by a set. Here, 2,336 documents from the D-documents are cited by other D-documents, and 1,649 receive no citation from other D-documents. These 2,336 (59% of this set) receive 14,588 citations from other D-documents.

**Table 8: Total number of documents cited and uncited, and the total citations to documents from D, H, A and R literature sets**

<table>
<thead>
<tr>
<th>Dataset-Dataset</th>
<th>#documents cited</th>
<th>#documents uncited</th>
<th>#citations</th>
<th>Proportion of documents cited</th>
</tr>
</thead>
<tbody>
<tr>
<td>D–D</td>
<td>2336</td>
<td>1649</td>
<td>14588</td>
<td>0.59</td>
</tr>
<tr>
<td>D–H</td>
<td>546</td>
<td>1276</td>
<td>2419</td>
<td>0.30</td>
</tr>
<tr>
<td>D–A</td>
<td>337</td>
<td>2</td>
<td>10156</td>
<td>0.99</td>
</tr>
<tr>
<td>D–R</td>
<td>108</td>
<td>24</td>
<td>615</td>
<td>0.82</td>
</tr>
<tr>
<td>H–D</td>
<td>731</td>
<td>3254</td>
<td>1604</td>
<td>0.18</td>
</tr>
<tr>
<td>H–H</td>
<td>855</td>
<td>967</td>
<td>4394</td>
<td>0.32</td>
</tr>
<tr>
<td>H–A</td>
<td>307</td>
<td>32</td>
<td>4193</td>
<td>0.91</td>
</tr>
<tr>
<td>H–R</td>
<td>63</td>
<td>69</td>
<td>186</td>
<td>0.48</td>
</tr>
<tr>
<td>A–D</td>
<td>490</td>
<td>3495</td>
<td>1100</td>
<td>0.12</td>
</tr>
<tr>
<td>A–H</td>
<td>245</td>
<td>1577</td>
<td>652</td>
<td>0.13</td>
</tr>
<tr>
<td>A–A</td>
<td>286</td>
<td>53</td>
<td>1490</td>
<td>0.84</td>
</tr>
<tr>
<td>A–R</td>
<td>53</td>
<td>79</td>
<td>111</td>
<td>0.40</td>
</tr>
<tr>
<td>R–D</td>
<td>246</td>
<td>3739</td>
<td>391</td>
<td>0.06</td>
</tr>
<tr>
<td>R–H</td>
<td>97</td>
<td>1725</td>
<td>148</td>
<td>0.04</td>
</tr>
<tr>
<td>R–A</td>
<td>147</td>
<td>192</td>
<td>318</td>
<td>0.43</td>
</tr>
<tr>
<td>R–R</td>
<td>51</td>
<td>81</td>
<td>90</td>
<td>0.39</td>
</tr>
<tr>
<td>All (D,H,A,R)</td>
<td>3970</td>
<td>2308</td>
<td>42455</td>
<td>0.63</td>
</tr>
</tbody>
</table>

I counted the number of isolates (documents neither citing nor cited by other documents) in each set (Table 9), which provides more clarity on how
interconnected documents in particular sets are. Here, we see that 87% of D-documents are connected together by edges to other D-documents.

Table 9: Number of linked documents and isolates in D, H, A and R literature sets

<table>
<thead>
<tr>
<th>Dataset-Dataset</th>
<th>Cited and Citing Documents</th>
<th>Isolates</th>
<th>Proportion connected</th>
</tr>
</thead>
<tbody>
<tr>
<td>D–D</td>
<td>3474</td>
<td>511</td>
<td>0.87</td>
</tr>
<tr>
<td>H–H</td>
<td>1432</td>
<td>390</td>
<td>0.79</td>
</tr>
<tr>
<td>A–A</td>
<td>333</td>
<td>6</td>
<td>0.98</td>
</tr>
<tr>
<td>R–R</td>
<td>87</td>
<td>45</td>
<td>0.66</td>
</tr>
<tr>
<td>Combined (D,H,A,R)</td>
<td>5924</td>
<td>354</td>
<td>0.94</td>
</tr>
</tbody>
</table>

To understand the citations within and between sets further, I report the total number of citations from one set to another, the proportion of these citations in relation to the sum of all references from one set to all other retrieved sets, and the proportion of these citation to the sum of all reference to FULL-documents in the largest component (Table 10).

Table 10: Total number of citations from one literature set to another, the proportion of these citations in relation to the sum of all citations between retrieved documents, and the proportion of these citations to the sum of all references.

<table>
<thead>
<tr>
<th>Dataset-Dataset</th>
<th>#citations to D,H,A,R from a set</th>
<th>Proportion of Citations to D,H,A,R</th>
<th>Sum of all references</th>
<th>Proportion of full reference list</th>
</tr>
</thead>
<tbody>
<tr>
<td>D–D</td>
<td>14588</td>
<td>0.53</td>
<td>88433</td>
<td>0.16</td>
</tr>
<tr>
<td>D–H</td>
<td>2419</td>
<td>0.09</td>
<td>41394</td>
<td>0.04</td>
</tr>
<tr>
<td>D–A</td>
<td>10156</td>
<td>0.37</td>
<td>11412</td>
<td>0.08</td>
</tr>
<tr>
<td>D–R</td>
<td>615</td>
<td>0.02</td>
<td>947</td>
<td>0.00</td>
</tr>
<tr>
<td>H–D</td>
<td>1604</td>
<td>0.15</td>
<td>14112</td>
<td>0.08</td>
</tr>
<tr>
<td>H–H</td>
<td>4394</td>
<td>0.42</td>
<td>3269</td>
<td>0.12</td>
</tr>
<tr>
<td>H–A</td>
<td>4193</td>
<td>0.40</td>
<td>3269</td>
<td>0.10</td>
</tr>
<tr>
<td>H–R</td>
<td>186</td>
<td>0.02</td>
<td>947</td>
<td>0.01</td>
</tr>
<tr>
<td>A–D</td>
<td>1100</td>
<td>0.33</td>
<td>14112</td>
<td>0.08</td>
</tr>
<tr>
<td>A–H</td>
<td>652</td>
<td>0.19</td>
<td>14112</td>
<td>0.05</td>
</tr>
<tr>
<td>A–A</td>
<td>1490</td>
<td>0.44</td>
<td>3269</td>
<td>0.12</td>
</tr>
<tr>
<td>A–R</td>
<td>111</td>
<td>0.03</td>
<td>3269</td>
<td>0.12</td>
</tr>
<tr>
<td>R–D</td>
<td>391</td>
<td>0.41</td>
<td>14112</td>
<td>0.08</td>
</tr>
<tr>
<td>R–H</td>
<td>148</td>
<td>0.16</td>
<td>14112</td>
<td>0.05</td>
</tr>
<tr>
<td>R–A</td>
<td>318</td>
<td>0.34</td>
<td>3269</td>
<td>0.10</td>
</tr>
<tr>
<td>R–R</td>
<td>90</td>
<td>0.10</td>
<td>14112</td>
<td>0.03</td>
</tr>
<tr>
<td>All</td>
<td>42455</td>
<td>1.00</td>
<td>147208</td>
<td>0.29</td>
</tr>
</tbody>
</table>
These results support the ‘sign-posting’ concept described previously that formed the theoretical justification for the title-search retrieval method. First, a high proportion of D-documents are cited by other D-documents (59%), more are connected to other D-documents via citations and references (87%), and 53% of D-document references to other retrieved sets are directed at other D-documents. While citations to the D-set from other D-documents constitute only 16% of all references from D-documents, this is impacted primarily by low-cited literature. This is because all documents cited ≥20 times by this group were collected via the A-search. It is very rare to find two documents with identical reference lists, so the large amount of low-cited literature probably derives from idiosyncrasies in referencing of individual authors.

Of the H-documents, 32% are cited by other H-documents and 79% are connected together by citations and references. The H-documents are internally less well connected than the D-documents, which may reflect the overlap approach used to capture these documents. This set included documents that focused on the relation between serum cholesterol and atherosclerosis, serum cholesterol and CHD, and atherosclerosis and CHD.

The A-documents are, unsurprisingly, heavily cited by all other sets. However, this set is extraordinarily internally well-connected – 84% are directly cited by other A-documents, while 98% are connected by either references or citations. This means that the A-documents (documents highly cited by D and H but missed by those queries) also heavily rely on one another.

Interestingly, the internal connectivity of all of these sets is greater than within the R-documents; 39% are cited by other R-documents and 66% are connected together by citations and references. These were included according to strict inclusion and exclusion criteria based on dietary relevance.

All four sets only direct 25–35% of their references to other documents in these four sets. So what is happening beyond these sets? – i.e., how do these four sets cite the non-retrieved documents?
Meta-vertex analysis

I performed a meta-vertex analysis by merging members of a set into a single vertex by transforming the edge-list. I preserved all citation information by weighting edges to reflect how often members of a set cite either another set or the non-retrieved documents of the FULL-largest-component. For example, if eight D-documents cite the same non-retrieved document, then a single edge with a weight of 8 will connect the D-meta-vertex to this document. To do this, I included an attribute to identify edges as deriving from documents in these sets, and constructed a modified edge-list containing both meta-vertices and non-retrieved documents. All data used for these analyses can be found in the Supplment:(Ch7.S9–11).

In Fig 3, meta-vertices are set at an arbitrarily large size to highlight their presence, with the rest of the vertices (non-retrieved documents) set at a uniform size. The colours of the vertices represent: D–red; H–blue; A–yellow; R–green; non-retrieved documents–black. The edges are set at a uniform size and coloured by source vertex. In this first visualisation, I disregarded edge-weights and retained only edges from meta-vertices to non-retrieved documents. I then applied the ForceAtlas2 algorithm, which pulls vertices together that share an edge and pushes vertices away from one another if they do not. Thus, this graph simply shows the literature shared by different groups. One of the major advantages of using Gephi, particularly for historical work, is the interactive nature of the graphs it produces. In the below graph, each vertex can be clicked on to examine its attributes, including their formal reference, which can be used to retrieve the full publication from a bibliometric database.
In Fig 33, non-retrieved vertices, in black, are clustered into groups defined by the meta-vertices that cite them. All vertices in the centre are connected to at least two meta-vertices (sets D, H, A, or R). Collectively, these amount to 8,370 vertices. On the outside of the graph, a semi-circle of vertices (halo\(^\text{15}\)) surrounds each meta-vertex, these are non-retrieved documents only cited by that meta-vertex. There are 46,539 vertices in the four halos. Thus, only 18% of non-retrieved documents are shared by these different sets. I report the full results below (Table 11).

\(^{15}\) I found no technical term for this structure in the literature. Halo is simply used here because it is shorter (and slightly more intriguing) than semi-circle. As I show later, during weighted-degree analysis, these halos are further sub-divided into semi-circle or full circles when visualised.
Table 11: Meta-vertex analysis results

<table>
<thead>
<tr>
<th>Set</th>
<th>#documents cited by only one set</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>27708</td>
</tr>
<tr>
<td>H</td>
<td>14412</td>
</tr>
<tr>
<td>A</td>
<td>3540</td>
</tr>
<tr>
<td>R</td>
<td>879</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>46539</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multiple Sets</th>
<th>#documents shared by more than one set</th>
</tr>
</thead>
<tbody>
<tr>
<td>D,H</td>
<td>3417</td>
</tr>
<tr>
<td>D,A</td>
<td>1431</td>
</tr>
<tr>
<td>D,R</td>
<td>309</td>
</tr>
<tr>
<td>D,H,A</td>
<td>1492</td>
</tr>
<tr>
<td>D,H,R</td>
<td>197</td>
</tr>
<tr>
<td>D,A,R</td>
<td>124</td>
</tr>
<tr>
<td>D,H,A,R</td>
<td>317</td>
</tr>
<tr>
<td>H,A</td>
<td>839</td>
</tr>
<tr>
<td>H,R</td>
<td>131</td>
</tr>
<tr>
<td>H,A,R</td>
<td>55</td>
</tr>
<tr>
<td>A,R</td>
<td>58</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>8370</strong></td>
</tr>
</tbody>
</table>

Has the search method then missed a lot of relevant literature? Are these different sets really all that related?

However, the previous analysis disregarded edge weights (the number of times a document is cited), which gives a better idea of how different sets depend on certain documents. In Fig 34, due to the small numbers of documents in both the R and A sets, and because these documents were selected based on criteria that ensured their relevance, I compare only the relationship between D and H. This makes visualisation easier to interpret and allows us to examine the similarity between query-derived sets.

In Fig 34, vertices are set at a uniform size apart from the meta-vertices that are shown larger. Vertices are coloured by their types (D—red; H—blue; non-retrieved documents—black). ForceAtlas 2 was applied again, but with a condition to recognise the influence of edge-weight. Here, vertices are pulled together not only when they share an edge, but this is proportional to the weight of an edge. This means that target vertices are pulled towards the meta-vertices that cite them most. The visual size of the edges was held constant.
Fig 34: Weighted meta-vertex analysis of reference from the D and H literature sets \((n=50134; m=55555)\). Edges coloured red that derive from the D-set; blue for edges from H-set. Non-retrieved vertices in black; D-meta-vertex in red; H-meta-vertex in blue.

Here, we see the literature shared by D and H positioned between these two meta-vertices in the centre of the graph. These 5,423 vertices are clustered into groups based on the weighted in-degree of vertices, and are pulled towards or away from the meta-vertices based on the proportion of the weighted in-degree contributed by those meta-vertices. The clusters to the left of the centre are used more often by the H-set, while those to the right are used more heavily by the D-set. In the semi-rings (the halo is now subdivided into clear separate sections) surrounding each meta-vertex, vertices are clustered based on weighted in-degree. These halos hold 44,709 vertices. The furthest and largest semi-ring of each halo consists of non-retrieved documents cited only once by its meta-vertex, the next semi-ring consists of documents cited twice, the next with three citations etc. These semi-rings are readable down to about an edge-weight of seven, where they begin to lose their shape due to the small number of vertices with larger weighted in-degrees. Thus, the halo is structured by the citations to non-retrieved documents from meta-vertices. Here, we see that the vast majority of the halo is populated by documents with a single citation.
To clarify the relationship between vertex position and weighted in-degree, the following visualisation (Fig 35) colours vertices by their weighted in-degree.

**Fig 35: Weighted meta-vertex analysis of references from the D and H literature sets (n=50134; m=55555).** Vertices with a weighted in-degree of 1–black, 2–dark blue; 3–light blue; 4–green; 5–red; 6–pink; ≥7–yellow.

If we examine a single meta-vertex, we see pattern that reflects the power-law distribution of citation. Fig 36 shows the non-retrieved documents cited by the D-meta-vertex. Again, non-retrieved vertices are coloured black, while the D-meta-vertex is red. Vertices are sized by a constant, but the D-meta-vertex is sized larger to highlight it.
Like reading of the age of a tree by the rings of its stump, we can read this graph by its *rings of influence*. The largest ring is composed of non-retrieved documents cited only once, the next twice, the next three times etc. From this, we can see that most non-retrieved documents receive only a single citation, 22,897 documents populate this outer ring (70% of non-retrieved documents in this graph). A very similar pattern was found in the H-only neighbourhood, with 15,899 (77%) of non-retrieved documents cited once. This was observed also in the A-only neighbourhood, with 6,195 (79%) cited only once. The most striking distribution is in the R-only neighbourhood, with 1,886 (91%) cited only once. When combined, 38,454 non-retrieved documents are cited once by any of these sets – 70% of all non-retrieved documents.

Interestingly, of the 7,424 non-retrieved documents cited twice, 2,539 are shared by two of the sets (34%). Of the 2,942 documents cited three times, 1,519 (52%) are shared by two or more of the sets. Of the 1,701 cited four times, 1,061 (62%) are shared by two or more sets. Of the 1,041 cited five times, 734 (71%) are shared by multiple sets. Of the 3,047 non-retrieved documents cited six or more times, 2,517 (83%) are shared by multiple sets.
Thus, the likelihood of being cited (in-degree) by more than one set greatly increases as the sum of citations (weighted in-degree) rises.

5. Dataset refinement

Based on these results, I removed all non-retrieved documents from the dataset that have a weighted in-degree of two or less (45,878). This removes a considerable level of noise from the dataset. Further, their inclusion is of little empirical interest. By doing this, the set of non-retrieved documents is reduced to 8,731. Of these, 2,900 are set specific (cited only by D, H, A, or R), while 5,831 are shared by at least two sets. Thus, by removing the low-cited literature, we see that the different sets share a large proportion of the same literature.

A cut-off of an in-degree of >2 for non-retrieved literature was also used by Batagelj et al. (2017) in their citation network analysis of the history of peer-review research. Referred to as the ‘boundary problem’ in Batagelj et al. (2014), citation networks constructed from a set of documents from any specific search tend to contain many references to literature not retrieved, and that many of these non-retrieved documents are low-cited and not clearly related to a topic. One could formally retrieve all non-retrieved documents, adding their references into the network; however, this does not solve the problem. Upon adding these new documents into a citation network, a new population of low-cited, non-retrieved literature will appear. As the entire scientific literature is seemingly governed by the small-world phenomenon, if one were to continually repeat this then almost the entire cited scientific literature would eventually be retrieved. Thus, a boundary has to be erected to ensure that a citation network is sufficiently focussed on the intended research topic. Retaining all formally retrieved documents regardless of their number of citations is justified on the basis of having titles indicating relevance, while a cut-off for non-retrieved literature below a particular threshold removes non-retrieved documents of limited importance to the population of retrieved documents. Here, I have demonstrated that the cut-off of >2 is non-arbitrary in my dataset. It is derived from the finding that the likelihood of being cited (in-
degree) by more than one set (D, H, A, R) greatly increases as the sum of citations (weighted in-degree) rises.

To construct this refined dataset, I added an attribute to the FULL-dataset that classified documents as having met these conditions. This resulted in a dataset containing 17,526 vertices and 93,909 edges. However, I took this an opportunity to clean up another aspect of these data. The FULL-dataset contained 1,662 reference records that contained no date information, which were primarily references to unpublished literature and personal correspondence. I removed these records for two reasons: (i) most records contain only the last name and initials of author and variants of the phrases ‘unpublished’ or ‘personal correspondence’, which makes it very difficult, if not impossible, to verify what work is being referred to; (ii) if two or more documents cite, say, ‘Keys, A. unpublished’, then we cannot be certain that the two documents actually meant to refer to the same unpublished document or two different unpublished documents. While removing non-retrieved documents with a weighted-degree ≤2 removed 1,634 of these, 28 remained in the dataset. These 28 were collectively cited 95 times – thus, these made up <0.002% of the vertices, and 0.001% of edges.

These were removed, reducing the size of the dataset to 17,498 vertices and 93,814 edges. This has an average in-degree of 5.361, and 2,585 weakly-connected components. The largest component has 14,909 vertices and 93,809 edges, which has an average in-degree of 6.3, a diameter of 14, and an average geodesic path of 4.2. It is this largest component that will be primarily relied upon in following chapters. Thus, this network contains documents that either reference or are cited by other documents in this set, and together they form a connected whole.

From this, I extracted two network datasets to understand how this refinement impacted the data originally collected. First, ‘CUT-N1’ includes all retrieved documents and their references that met the conditions of the refinement described in this section (see Ch7.S12–14). Second, ‘CUT-N2’ includes a component with the exact same composition as the BASE-largest-component described earlier (see Ch7.S15–17). As CUT-N2 contains only full
bibliographic records, I report the composition of this in terms of document type in the below table:

**Table 12: Dataset descriptors CUT-N2**

<table>
<thead>
<tr>
<th>Query</th>
<th>Articles</th>
<th>Meeting Abstracts</th>
<th>Letters</th>
<th>Notes</th>
<th>Reviews</th>
<th>Discussion</th>
<th>Editorials</th>
<th>Book Review</th>
<th>Biography</th>
<th>Book Manual</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>2965</td>
<td>360</td>
<td>327</td>
<td>117</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3771</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>1301</td>
<td>120</td>
<td>118</td>
<td>90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1629</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>297</td>
<td>3</td>
<td>7</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>339</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>109</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>131</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4672</td>
<td>488</td>
<td>459</td>
<td>239</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>5870</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As anticipated, the major reduction compared to the BASE-dataset derives primarily from the removal of uncited meeting abstracts – 77% of the reduction comes from the removal of 2,062 meeting abstracts. However, some meeting abstracts do attract citations. And some, like Ahrens et al. (1955) - cited 63 times by my retrieved documents - attract an unexpected amount of interest.

The CUT-N1, however, contains a total of 6,206 retrieved documents, 337 more than the CUT-N2 because these are connected in the largest component via shared references to non-retrieved documents, rather than direct citation links with other retrieved documents. These include an additional 229 articles, 37 meeting abstracts, 48 letters and notes, 12 reviews, editorials and discussion pieces, and 11 book reviews. Thus, only 342 articles collected from the literature search described previously did not reference or were not cited by other documents in this set (about 6%), while 80% of meeting abstracts were not connected through either shared references or citations.

The first record in the CUT-N1 is to Vogel's (1847), *The pathological anatomy of the human body*, a publication in which he described the presence of a cholesterol in atherosclerotic plaque. A total of 13 records relate to documents published before 1900, which include Rokitansky (1852) and Virchow's (1856) major works on atherosclerosis, Liebermann (1885) and Buchard's (1890) work on cholesterol detection, and Osler's (1897) lecture linking angina pectoris to atherosclerosis.
The first record in CUT-N2 is Roger’s (1908) report on atheroma and dilation of the pulmonary arteries. The distribution of documents published per year in these datasets is as follows:

![Graph showing the number of documents published per year in CUT-N1 and CUT-N2, 1900–1984.](image)

**Fig 37: Number of documents published per year in CUT-N1 and CUT-N2, 1900–1984.**

Thus, despite a major reduction in the number of vertices, we see similar trends in publication as observed previously. The dip in publication output in CUT-N1 from 1979 is explained by the absence of documents with reference lists published after 1984. This provides a rough idea of how likely literature is to cite other literature in different time-windows (see Chapter 8).

Here, we may reasonable suppose that the history of the development of this area of research can be split into five major periods based on publication growth rates.

The first, a period before 1910, in which publication is infrequent – neither the WoS database detects much activity in these years, nor are there many references to records from the collected publications to documents published in this period. In CUT-N1, only 37 records derive from this period and CUT-N2 only three. Second, a period between 1910 to 1947, which is defined by modest, though interrupted, growth. In this period, CUT-N1 holds 871 records and CUT-N2 holds 112. Third, a period of dramatic growth between around
1948 to 1961 (CUT-N1 = 3,448; CUT-N2 = 1,113). Fourth, a period of stagnation and decline in the rate of publication from around 1962 to 1971 (CUT-N1 = 4,069; CUT-N2 = 1,346). Finally, a further period of sustained, intense growth lasting from around 1972 to 1984 (CUT-N1 = 6,484; CUT-N2 = 3,296).

Interested by the consistency in these trends in publication, I examined this further in the two CUT networks. While we know the number of publications per year, the number of citations attracted by documents published in a particular year provides a general idea of how influential particular years were. To understand this, I constructed two distributions recording the sum of citations to documents published in a given year for the two CUT networks. To do this, I took the in-degree of each vertex and aggregated these into individual years.

In Fig 38, the top panel shows the sum of citations to documents in a given year per year for CUT-N1 and the bottom panels shows this for CUT-N2.

**Fig 38: Total citations to all documents published in specific years in CUT-N1 and CUT-N2**

The similarity of these distributions is explained because CUT-N2 contains all the highly-cited documents in CUT-N1. The intermittent spiking in the number of citations to publications in particular years appear to cohere with
studies considered of fundamental importance to this particular research area. Specifically, the years 1913, 1925, 1933, 1950, and 1957 appear, from the aggregate of citations to all documents in those years, to be particularly important (see Chapter 9 for further analysis).

6. Discussion

In this chapter, after extensive cleaning and refinement of the original collected data, I quantitatively described and validated the datasets that will be used throughout the following chapters. By using insights from Kuhn and Price, I established three quantitative conditions of community structure to validate the retrieved results: (i) a large-component containing the vast majority of identified records; (ii) a power-law distribution of citation to documents in this component; (iii) and that the retrieved documents ought to cite other retrieved documents more often than non-retrieved documents.

First, I demonstrated that, despite being derived from different search strategies (D, H, A, R), these documents form a large-component composed of almost all documents that contain references. After excluding these isolates, I demonstrated that this component contained 99% of the literature retrieved and 99.9% of their references. Thus, these documents can be considered part of a single broad community – they form a single connected network.

Second, I demonstrated that in the largest component formed from all retrieved documents and their references a power-law distribution is present \( P(k) \sim k^{-2.21} \). Thus, the literature is collected around a small number of highly influential documents. This distribution has several consequences. From this, we know that this area of science is categorised by major inequality in citation. This inequality is one of the defining qualities of citation networks and widely believed to be generated by some form of the Matthew Effect or Cumulative Advantage, and this inequality has implications for how we understand the way in which evidence propagates across the scientific literature.

Third, I demonstrated that documents collected from different search strategies bear similarities in their reference behaviour; similarities that
suggest the sharing of similar topics. I established the sets from which particular documents derive and measured the proportion of references from these sets that were directed either within or between these sets. By doing this, I demonstrated that, not only are retrieved documents more likely to cite other retrieved documents than non-retrieved documents, but that retrieved documents also share a high proportion of non-retrieved literature. For the latter, I performed a novel application of meta-vertex analysis.

Thus, the retrieved documents met these conditions. Finally, as described previous the trends in publication appear to lend weight to the idea that the collected literature represents a ‘paradigm’, at least a large proportion of literature relevant to diet–heart research. Kuhn conjectured that the birth of a ‘paradigm’ will be observable by a dramatic change in publication behaviour. This was observed in the previous chapter. As measured by publication output per year, diet–heart research took off in the 1950s, hit a period of stagnation in the 1960s, and witnessed a second period of intense growth from 1972 onwards.

This trend in publication output was robust. It was present in the documents derived from D-query, from the H-query, from the combined D and H query. Further, this trend remained despite the addition of a further 533 additional documents from the A and R searches. Finally, this trend was also evident when considering the references of all the retrieved documents, and it remained present after refining the dataset to include only literature that sat within the largest-component and non-retrieved literature that had three or more citations.

Finally, these results offer support for the concept of ‘sign-posting’ described in the previous chapter. Literature that had particular words related to the logical structure of the conjectured causal mechanism linking diet to CHD in the title formed a cohesive network that had the characteristics of a ‘paradigm’ as described by Kuhn. Yet, this method missed literature that was relevant to the debate. To resolve this, I retrieved the top-cited literature from the references of the D and H queries. I then validated these results by checking whether my retrieved set of documents had been cited by 19 different
reviews of the literature, all written by different authors and published in different years between 1959 and 2017. From this, I demonstrated that my search method was able to capture ~70% of relevant literature cited by these reviews. Further, the 134 relevant documents missed were not well-cited in my data-set, the reasons why these had been missed by my original search strategy was established, and these too were included in my dataset.

While these results offer quantitative support that this literature is representative of the diet–heart paradigm, further analysis is still required. Quantitative analysis alone is not enough because it does not consider the content of the documents beyond the apparently representative words in titles of those documents and their reference behaviour. This, however, awaits the reader in the following chapters later in which I examine the substantive content of documents. In the following chapter, I examine whether my anticipations regarding the influence of time and speciality on citation behaviour are correct, and, if so, how these structure this citation network.

This represents the end of my search strategy for literature and the associated data collection. The advantages of this systematic approach to literature retrieval for the history and sociology of science should be clear. First, it reduces the danger of bias impacting literature selection. Second, quantitative analysis of this literature provides a way of justifying excluding particular documents from consideration (those that do not sit within the largest component and low-cited non-retrieved literature), and this will be increasingly important for researchers studying science due to the massive growth in publication since the 1940s. Finally, adopting this strategy allows for future replication.
Chapter 8: On Mapping a Paradigm: On time and community structure

Introduction

This chapter examines time and community structure in the CUT-N1 network. As discussed in previous chapters, I anticipate that both time and research type ought to play a role in structuring these networks.

Following Price (1965), we would expect diet–heart research to be primarily developed by a dynamic research front that directs the majority of its references to recently published literature (Chapter 2). If this is so, a citation network of a research area ought to resemble a “dropped stich of knitting”, with the majority of new papers at one end of the network and older papers descending backwards in time to the other side of the network roughly by year of publication. This structure ought to be visible by the citations links alone when visualised with a force directed layout (a layout that pulls vertices together that are connected by an edge and pushes vertices away from one another that do not).

First, I begin by visualising both the CUT-N1 and CUT-N2 networks to examine the structural similarity between these. Second, I examine the effect of time on network structure. There is an important temporal dimension to referencing in these networks; documents tend to cite documents in close temporal proximity, but highly cited documents appear more resistant to the decaying effects of time in terms of their citation impact. Third, I explore community structure in these networks via modularity analysis. I examine whether study type plays an important role in network clustering. However, I find that broad research questions (e.g., is diet related to serum cholesterol?) appear to better explain clustering than specific study type. Thus, the research focus of a study, defined as adherence to a general research question, appears to influence what a document references and how it is cited. Together,
temporality and research focus, alongside the power-law distribution of citation described previously, appear to be the major factors shaping network topology.

1. Network visualisation

In the previous chapter, I introduced two networks constructed from the literature retrieved from WoS. CUT-N1 is composed of the largest component derivable from *all* retrieved documents and any document not formally retrieved but cited three or more times by the retrieved documents, while CUT-N2 is composed of *only* the retrieved documents that constitute a largest component. As CUT-N2 is subset of CUT-N1, comparison between these allows for a visual examination of how including non-retrieved literature alters network structure.

In Fig 39, I visualise both networks with the structural position of vertices established via a Lin-Log adaptation of the ForceAtlas2 algorithm, which has been demonstrated to be effective for visualising clusters in directed graphs (Jacomy et al. 2014). I scaled vertex size relative to in-degree between the values of 20 (documents with no citations) and 548 (the highest cited document in both networks has 537 citations). To help compare these networks, I’ve set seven documents to different colours to aid the reader.
Fig 39: Citation network. Right panel is the CUT-N1 network \((n=14909; m=93809)\). Left panel is the CUT-N2 network \((n=5870; m=42424)\). Visualised in Gephi, Lin-log Force-Atlas 2. Vertices scaled by in-degree. Anitschkow (1933) is coloured dark blue (on the far left of each network), Ahrens et al. (1957) pink, Abell et al. (1952) in red, Fredrickson et al. (1967) orange, Keys (1970) in light blue, Folch et al. (1957), Miller and Miller (1975) in green (far right), and all other vertices are coloured grey.

There is little difference between these networks in the position of the seven highlighted vertices, and the networks appear, from a visual examination, to be structurally similar aside from the large difference in the number of vertices. Previously, I demonstrated that these two networks had similar mean in-degree (CUT-N1=6.3; CUT-N2=7.2), network diameter (CUT-N1=14; CUT-N2=13), and mean geodesic path (CUT-N1=4.2; CUT-N2=4.1). Thus, alongside the visualisation, the inclusion or exclusion of 9,038 vertices representing non-retrieved documents has little impact on these properties.

This suggests that retrieved documents contained in these networks are relevant to diet–heart research, and that they played some role in either introducing new information or in using the information produced by others.

Relevance is indicated by the absence of large, distant clusters of vertices, which would indicate papers on loosely-related topics. Further, CUT-N2 forms a connected component the encompasses nearly all of the retrieved documents from CUT-N1. This tells us that most of the retrieved documents
relate directly together even without the non-retrieved documents. It could have been the case that retrieved documents only formed a connected component with the addition of non-retrieved documents – e.g. vertex A has a path connecting to vertex C through non-retrieved vertex B. In CUT-N1, there are 337 retrieved documents that are only connected to other retrieved documents in this manner.

If the collected documents were examining very different topics, we would expect them to rely on different literatures, which would alter the structural position of vertices by clustering them into distinct literatures. While there are not many visualisations of phenomenon in the CNA literature, Kajikawa and Takeda (2009) detected this in the literature on organic light-emitting diodes (OLEDs). By retrieving papers that had specific key-words from WoS, the authors applied a recursive modularity analysis to detect community structure in the collected records. The literature on OLEDs could be partitioned into five large clusters that reflected research focus and year of publication (Fig 40). To categorise each cluster, the authors read a sample of the titles and abstracts of cited publications within each cluster.

![Fig 40: Citation network of organic light-emitting diodes (OLEDs) by Kajikawa and Takeda (2009)](image)

These clusters primarily reflect the material used to construct light-emitting diodes. In the above visualisation, we see that cluster #4 and #5 are not interacting frequently with the major clusters #1, #2, and #3. According to
these authors, this is because “Cluster #4 (polymerization) and Cluster #5 (LED light curing biomaterials) are not directly related to OLEDs” (p.1121), and they go on to explain why these were caught via their literature search methodology. Thus, this study demonstrates the ability of modularity analysis to detect sub-groups of papers within a large component, but also the benefit of visualisation in detecting noise a dataset.

In the following sections, I perform a temporal analysis before proceeding to modularity analysis of the CUT-N1 network. These analyses are designed to detect whether time is a factor in how particular publications are cited, and whether citations and references can be used to identify clusters of documents with a particular focus. That is, does time play a role in the decisions by scientists over what to cite? And does research type play a role in both how a document references and how it is cited?

2. Time

Aging references

As my networks cover a long period of publication – between 1847 and 1984 in CUT-N1 and between 1908 and 1984 in CUT-N2 – time should play an important role in structuring these networks (Price 1965). Burton and Kebler (1960) claimed that citation rates to scientific publications exhibit a regular half-life of decay. They proposed that the number of citations to most documents decreases exponentially after some peak time, and most documents in certain fields appeared to share a field-specific half-life. Recently, Parolo et al. (2015) demonstrated this decay trend in a larger sample of literature in Physics, Biology, Medicine, and Chemistry.

To understand this in my data, I “aged” all edges via the inclusion of time attributes to the edge-lists. To do this, I used VLOOKUP in Excel to include year of publication for all ‘Source’ (r) vertices, year of publication of ‘Target’ (c) vertices, and measured the difference in years between these values (t); therefore, t=r-c (see Ch8.S1).
In the CUT-N1 network, there are 93,809 edges (references). The age of references ranges from 0 years (references to documents published in the same year) to 132 years. The mean age is 8.4 years, while the median is 6 years. About 48% of references are ≤5 years old, 73% ≤10 years, 85% ≤15 years, and 92% ≤20 years old.

In CUT-N2, there are 42,424 edges. The age of references ranges from 0 to 75 years, the mean is 8.6, and the median is 6. About 48% of references are ≤5 years old, 72% ≤10 years, 84% ≤15 years, and 90% ≤20 years old.

In the previous chapter, I excluded 45,878 non-retrieved documents that were cited less than three times. If these are counted, the average age of references in the full reference lists of the retrieved documents can be ascertained. These excluded documents attracted 51,620 citations\textsuperscript{16}. The range of age is from 0 to 292 years, mean age is 8.2 and the median age is six. About 50% of references are ≤5 years old, 74% ≤10 years, 86% ≤15 years, and 92% ≤20 years old.

Fig 41 reports the distribution of age of references in the CUT-N1, CUT-N2, and the largest component of the FULL-dataset.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig41.png}
\caption{Distribution of Edge (Reference) Age in the FULL, CUT-N1, and CUT-N2 networks}
\end{figure}

\textsuperscript{16} This excludes references that did not contain a year of publication, see Chapter 7.
Thus, reference age is remarkably similar whether we take the CUT-N2 references (42,424), CUT-N1 references (93,809), or all references from all retrieved papers in the largest component (145,429). In these sets, most references are to papers published in close temporal proximity. However, as the number of publications is skewed to more recent publications, further analysis is required to understand the average age of references in different publication years.

**Average age of reference lists**

I measured the mean and median age of the reference lists of papers published in different years across these three networks from 1948 to 1984 (Fig 42. Top panel, mean; Bottom panel, median)

![Graph showing mean reference age per year](image)

![Graph showing median reference age per year](image)

**Fig 42:** Mean (top panel) and median (bottom panel) reference list age per year in FULL, CUT-N1, and CUT-N2 networks
Note the mean age of the reference lists of documents published in any given year is older than 5-years. These results are similar to those reported by Larivière et al. (2008). In a study evaluating the age of references in the reference lists of all documents published in the medical sciences (MED) research category in WoS between 1900 and 2004, the authors reported a decline in the mean age of reference lists from around 10.8 years in 1948 to 8.6 years in 1984 for all MED indexed articles (after limiting the maximum reference age to 100 years; 100-year window). The median age of references is a better approximation, however, of the age of referenced literature in any given year because of the impact of very old references on the mean. Here, the 1950s were categorised by a particularly low median age of between two and four years after which the median age fluctuates between six and eight years. Importantly, what these statistics tell us is that, despite massively reducing the number of references across these different sets, the average age of referenced literature stays similar.

**Price Index**

Price ([1963]1986) conjectured that a field undergoing rapid development ought to direct about half of all references in a given year to works published in the previous five years. He came to this conclusion after studying different fields at different stages of development and noting that the average age of references decreased markedly as journal publication output increases (signifying development).

What then is the proportion of references of particular ages in the reference lists of papers published in different years? In Fig 43, I calculate the Price Index per year, a measure often used to assess whether a research field is progressing or not, and measures the proportion of references that are five years or younger in a given set of documents. I’ve also included the proportion of references 10 years or younger in Fig 43.
Fig 43: Proportion of references ≤5 years and ≤10 years old in the reference lists of retrieved publications per year

Again, these results are similar to those reported by Larivière et al. (2008). In their 100-year citation window analysis of all MED-articles, about 40% of references were ≤5 years old in 1948, this rose steeply to about 63% in 1955, then a steady decline to about 47% in 1984. In this analysis, 30% of references were ≤5 years in 1948, 69% in 1955, and 46% in 1984.

Larivière et al. (2008) interpreted the growing age of references overtime as indicating that the medical sciences had reached a ‘steady-state’ of development. Citing Kuhn ([1962]1970), they claim their results indicate that medical science had reached a period of ‘normal’ science categorised by “research firmly based upon one or more past scientific achievements, achievements that some particular scientific community acknowledges for a time as supplying the foundation for its further practice” (p.10).

My results could be taken as indicating that diet–heart research, after a period of intense growth in the 1950s, entered a period of ‘normal science’ by around the mid-60s, where the age of references remained relatively constant thereafter. However, as the reference behaviour of diet–heart related documents largely mirrors that of all medical sciences documents indexed in WoS over the same period, this seems questionable. Whether these trends tell us anything specific about diet–heart research is unclear, and these trends may well be explained by some overarching factor on the reference practices.
of the medical sciences unrelated to the stage of development of particular research areas.

These results are, however, consistent with Price’s (1965) claim that science progresses through a research front that primarily relies on recent literature, and this would explain why, in nearly every year, most documents direct at least 50% of their references to literature ≤5 years of age.

Yet, about a quarter of references are to documents over 10-years-old (27% in CUT-N1, 28% in CUT-N2, and 26% of all references contained in the retrieved literature). Larivière et al. (2008) interpreted the rising age of references as suggesting that ‘older’ literature has become more influential. However, they measured the average age of references, but not the average age of cited documents. The age of references can distort our understanding of influence due to the power-law distribution of citations. For example, Lowry et al. (1951), a paper on protein measurement has accumulated 337,657 citations as of 2019. It is the most cited document in scientific history, and, incidentally, the third most highly cited document in my own data (384 citations). According to WoS, this document has been most cited in the research area ‘biochemistry & molecular biology’. Thirty years after the publication of this paper, 30,161 documents were published under this research category in 1981, and Lowry et al. (1951) was cited by 11,018 of these – over a third of all documents. Thus, this single document will disproportionately affect the average age of references from these documents published in 1981. An analysis of reference age would make it appear that many older documents were being referenced, but 11,018 of them are directed at this single publication. The point here is that heavily cited older literature has a disproportionate impact on the average age of references.

Is the observed decay in influence true of all literature? Or does the likelihood of referencing a document of a particular age depend on how highly cited that document is? To understand this, I measured the distribution of references >20 years or >10 years of age from the largest component in the FULL-dataset to their specific target papers.
First, there are 11,374 references >20 years of age, but these are directed at only 5,379 unique documents, but 53% of these references are directed to only 691 of these that attracted at least 3 citations over the age of 20. The distribution of these citations is approximated by $P(k) \sim k^{-2.35}$. Second, 37,523 references are >10 years of age and these are directed to 16,717 unique publications, but 56% of these references are directed at only 2,606 unique publications that attracted at least 3 citations over the age of 10 years. The distribution of these citations is approximated by $P(k) \sim k^{-2.33}$.

As the distribution of citations >10 and >20 to particular documents follows a power-law, we know that the age of reference lists will generally be heavily influenced by a small number of highly cited publications.

From this analysis, we know that, on average, documents will tend to receive most of their citations in a relatively short space of time – half of all their citations occur in the five years following publication. For example, 27% of the retrieved documents are never cited again by other retrieved documents after five years. Fig 44 shows the % of citations to papers that are over five years of age for all retrieved documents from the 1950s.

![Fig 44: Percentage of total citations to documents published in the 1950s that are older than five years of age](image)

Interestingly, 7.5% of the retrieved documents published in the 1950s received their first citation from other retrieved documents more than five years after publication. While most of these publications received less than ten
citations, 53 documents collectively gathered 1,531 citations, but 77% of these are directed at three methods papers, one introducing a method for estimating the amount of lipid in animal tissue (Folch et al. 1957), Lowry et al. (1951) paper on protein estimation, and Havel et al. (1955) work on ultra-centrifugal analysis of lipoproteins.

Thus, most papers, if they are to become widely known and cited, have a small window of opportunity to do so.

**Influence of time on network structure**

But what impact does this have on network structure? In Fig 45, I colour vertices of the CUT-N1 network by membership of three temporal groups: (i) <1950 composed of 1,061 vertices; (ii) 1950–1967 composed of 5,696 vertices; (iii) 1968–1984 composed of 8,151 vertices. These groupings were established only to aid visualisation of the general impact of time on this network. While it is possible to assign vertices into groups that cover shorter periods, the resulting visualisation is overly hard to interpret.

![Fig 45: CUT-N1 citation network visualising the impact of year of publication on network structure (n=14909; m=93809). Vertices coloured green—<1950; red—1950–1967; blue—1968–1984](image_url)
We see that network structure is influenced by time even though the position of vertices is established purely by their connections to other vertices. On the furthest left of this network, most documents are published <1950, while on the furthest right is a large region dominated by documents published between 1968 and 1984. Note, that this is only generally true. Vertices published in these three different time periods do appear in regions of the network dominated by publications from a different period. This is because ForceAtlas2 (Jacomy et al. 2014) works by positioning vertices closer to vertices with which they share an edge, and pushes vertices away from one another if they do not. For example, say document A, published in 1980, cites documents B and C from the 1940s only, and documents B and C both cite other documents from the 1940s, then document A will appear in a region heavily populated by 1940s documents. If document A goes on to attract many citations in the 1980s, it will be pulled forwards into a region of the network populated heavily by 1980s documents. Thus, network structure is partly explained by the fact that most references are to documents in close temporal proximity.

**Price’s ‘Dropped Stich’**

To understand the extent to which diet–heart research progressed through a dynamic research front that cited literature published in the last five years, I constructed: (i) a network composed of only edges ≤5 years of age; (ii) only edges >5 years of age from the CUT-N1 network (see Ch8.S2). By doing this, it is possible to examine the extent to which the inclusion or exclusion of references of different ages alters the network structure.

Interestingly, of the 14,909 vertices of the largest component of the CUT-N1 network, 12,849 are connected together in a large component when including only the 44,644 references ≤5 of age – about 48% of all edges in CUT-N1. The average in-degree shrinks from 6.3 to 3.5, the diameter increases from 14 to 18, the mean geodesic path increases from 4.2 to 5.9. In Fig 46, I visualise the network with the position vertex established by their citation relationships *alone* via the ForceAtlas2 algorithm.
This visualisation reflects the development of the diet–heart research *if* authors chose only to reference literature published in the last five years. This resembles a progressively expanding ‘dropped stich’ – on the left of the network are documents from the 1920s and to the far right those of the 1980s. We see a continuous research stream connected tightly together through time. Few documents are lost when considering only reference links of this age – 1,929 become isolates (documents neither referenced nor cited), while 221 form 69 small components. However, one small component composed of 40 vertices connected by 63 included documents published between 1908 and 1919 and which were detached from the largest component because of the few documents with full reference lists in the 1920s in the CUT-N1 network. This component holds the beginnings of work on experimental atherosclerosis (Anitschkow and Chalatow [1913]1983) and cholesterol measurement (Bloor 1916).

If we now examine the network composed of only references older than five years, then we observe a rather different structure. In this network (Fig 47), 12,416 vertices are connected by a total of 48,952 edges. This has an average in-degree of 3.95. However, only documents published before <1979 have an opportunity to be cited, which increases the average citations to 4.6. The diameter of the network is 9, while the mean geodesic path length is 3.5.
While time plays some role in positioning the vertices, documents appear now to cluster primarily around core well-cited publications in the centre of the network. Price (1965) conjectured that an area of research ought to resemble a dropped stich of knitting, but he also claimed that highly cited documents would be pulled into the centre of these stiches. These results support such a notion of how scientific work proceeds over time. Importantly, according to Price, “such strips represent objectively defined subjects whose description may vary materially from year to year but which remain otherwise an intellectual whole” (p.515). So what else can this network tell us about how diet–heart research developed over time?
3. On clustering

In the previous section, I demonstrated that time has an important role in structuring the network. This effect appeared to stretch the network horizontally from left to right roughly following an arrow of time. However, what is shaping the rest of the network? I anticipated that study type (e.g. animal model, observational study, etc.) would cluster documents together, but is this assumption valid?

To understand this, I examined the CUT-N1 network for the presence of communities via modularity analysis. But why select this over the network composed of only the retrieved literature (CUT-N2)? This choice is based on the idea that the position of documents within a citation network will be a more accurate with the inclusion of shared non-retrieved documents than without. That is, a document might have more in common with another document by virtue of the number of shared non-retrieved documents referenced than they do through their references to retrieved documents. As community structure plays an important role in interpreting a network, the greater number of shared references in CUT-N1 is advantageous.

Modularity and the Leiden Algorithm

First proposed by Newman (2006), modularity maximisation is a technique to measure how well a network can be divided into ‘modules’ (clusters). To do this, one establishes clusters of vertices that maximise the difference between the number of edges in a cluster versus the expected number of edges if these were distributed randomly within the network.

Calculating modularity in a directed citation network is difficult because it is not feasible to precisely calculate the expected number of edges (see discussion of this in Chapter 3 and 7). However, Speidel et al. (2015), after constructing a modularity maximisation method for directed acyclic graphs, found this produced similar results to the modularity maximisation method for undirected graphs. According to the authors:
Simply applying modularity maximization methods for undirected networks to DAGs may be practically innocuous. We stress that we have reached this conclusion by actually developing a modularity measure for DAGs and testing it against previous methods using several data sets (p.9).

Modularity \((Q)\) for an undirected graph can be ascertained by:

\[
Q = \frac{1}{2m} \sum_{ij} \left( A_{ji} - \frac{d_j d_i}{2m} \right) \delta(c_j, c_i)
\]

where \(m\) is the total number of existing edges, \(A_{ij}\) is the sum of edges from the adjacency matrix, \(d_j d_i / 2m\) calculates the probability of an edge between two vertices proportional to their degree, the delta function, \(\delta\), equals 1 if \(j\) and \(i\) are in cluster \(c\) and 0 otherwise.

To perform a modularity analysis, computational methods are required to establish robust clusters of vertices. To this, I applied Traag et al. (2019) Leiden agglomerative algorithm for modularity-based community detection. The Leiden algorithm works in three phases. First, it constructs clusters of vertices by iteratively moving vertices from one cluster to another to establish clusters in which the fraction of edges between vertices within a cluster is higher than would be expected if edges were distributed randomly. Second, it aggregates clusters of vertices together into meta-vertices that contain self-loops that indicate how many edges are directed at itself and edges to other meta-vertices. It then repeats the first step, merging vertices together to maximise modularity. In this sense, it works like the Louvain algorithm; but with an important difference. Leiden’s algorithm allows initially clustered groups of vertices to divide up again and merge with other clusters if this increases the quality of the partition (modularity maximisation). By this, the Leiden algorithm ensures that clusters are well-connected, and that no movement of a particular vertex to a different cluster would improve the quality of the partition (see Chapter 3).
To apply the algorithm, I based the partition quality function on modularity. Following Traag et al. (2019), I set the resolution parameter to 1.0, which is the typical resolution parameter used by those who have examined modularity in networks of this size, and in citation networks in general. Lower values of the resolution parameter decrease the number of communities detected, while higher values increase the number detected. This decision is guided by the readability of the network – too many or too few communities lack utility for interpretation. There is a subjective element to community detection and I use this method here only as a heuristic tool to help identify structures for interpretation.

I allowed the algorithm 1,000 iterations – the number of times the algorithms steps are repeated. For each iteration, the results from the previous iteration are evaluated to judge whether the current iteration has performed better or worse (increased or decreased the value of $Q$). Thus, this takes the highest modularity value obtained on 1000th iteration as the basis for the partition. I also allowed the algorithm to restart 100 times, with each based on a different starting number produced by a random number generator. This random number determines the order of vertices visited during the algorithm’s identification of vertices that share an edge. To do this, the algorithm selects a vertex at random and then traverses the network via a random walk by following the edges it encounters, clustering vertices together that are densely connected. Accordingly, this process can be impacted by the order in which vertices are visited and can lead to subtly different community structures. By repeating the algorithm with different random numbers, the confidence in the resulting partition is increased, with the run with the highest modularity value being used as the basis for the final community partitions (see Waltman and van Eck 2013).

**Clustering in the CUT-N1 network**

I applied the Leiden algorithm to the CUT-N1 network in Gephi via the ‘Leiden Algorithm’ plugin developed by Traag et al. (2019). This produced an overall $Q$ of 0.553, which tells us that this network can be split relatively well into sub-communities. Values of $Q$ can range from -1 to 1, but, as a relational measure,
$Q$ does not tell us much by itself beyond providing a value for the algorithm to maximise. Higher relative values reflect better partitions of the network into well-defined sub-communities.

In total, 16 clusters were identified by this method (see Ch8.S3–4). In Table 13, I report the identifier for each community, the number of documents it contains, the percentage of the total documents that this represents, the median year of publication, and the average in-degree (citations) within each cluster from other documents in that cluster.

Table 13: Description of CUT-N1 following modularity maximisation clustering via the Leiden algorithm

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Documents</th>
<th>Percentage of Total Documents</th>
<th>Median Year of Publication</th>
<th>Average In-degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0</td>
<td>1801</td>
<td>12%</td>
<td>1958</td>
<td>5</td>
</tr>
<tr>
<td>C1</td>
<td>1686</td>
<td>11%</td>
<td>1976</td>
<td>4.6</td>
</tr>
<tr>
<td>C2</td>
<td>1650</td>
<td>11%</td>
<td>1978</td>
<td>4.9</td>
</tr>
<tr>
<td>C3</td>
<td>1613</td>
<td>11%</td>
<td>1967</td>
<td>3.6</td>
</tr>
<tr>
<td>C4</td>
<td>1582</td>
<td>11%</td>
<td>1969</td>
<td>3.5</td>
</tr>
<tr>
<td>C5</td>
<td>1558</td>
<td>10%</td>
<td>1971</td>
<td>5</td>
</tr>
<tr>
<td>C6</td>
<td>1349</td>
<td>9%</td>
<td>1950</td>
<td>4.5</td>
</tr>
<tr>
<td>C7</td>
<td>1325</td>
<td>9%</td>
<td>1975</td>
<td>3.9</td>
</tr>
<tr>
<td>C8</td>
<td>826</td>
<td>6%</td>
<td>1970</td>
<td>3.4</td>
</tr>
<tr>
<td>C9</td>
<td>616</td>
<td>4%</td>
<td>1966</td>
<td>2.5</td>
</tr>
<tr>
<td>C10</td>
<td>388</td>
<td>3%</td>
<td>1976</td>
<td>2.5</td>
</tr>
<tr>
<td>C11</td>
<td>199</td>
<td>1%</td>
<td>1968</td>
<td>2.2</td>
</tr>
<tr>
<td>C12</td>
<td>157</td>
<td>1%</td>
<td>1971</td>
<td>3.2</td>
</tr>
<tr>
<td>C13</td>
<td>82</td>
<td>1%</td>
<td>1972.5</td>
<td>2</td>
</tr>
<tr>
<td>C14</td>
<td>66</td>
<td>0%</td>
<td>1972.5</td>
<td>1.7</td>
</tr>
<tr>
<td>C15</td>
<td>11</td>
<td>0%</td>
<td>1965</td>
<td>0.9</td>
</tr>
<tr>
<td>All</td>
<td>14909</td>
<td>100%</td>
<td>1969</td>
<td>6.3</td>
</tr>
</tbody>
</table>

The average in-degree is lower in any specific cluster than for the whole network. This tells us that, although we can divide this network into clusters, doing so lowers the average number of times documents are cited. As 16 colours are overly confusing to visualise, I coloured the largest eight communities (C0-C7) that contain about 84% of all vertices in the below visualisation (Fig 48).
Fig 48: The CUT-N1 network clustered by modularity maximisation via the Leiden algorithm. Vertices sized by in-degree set between the bounds of 50 and 500 in Gephi. Vertices coloured by cluster: C0–dark red; C1–pink; C2–orange; C3–black; C4–dark blue; C5–light blue; C6–yellow; C7–green; C8-C15–grey.

This visualisation provides a snapshot of how the literature clustered together via references before 1985. Note that some clusters appear split, for instance cluster C3 (coloured black). This occurs because documents that make up a cluster can reference, or be cited by, documents in other clusters. As this a complex network, with over 14,909 vertices and 93,809 edges, the position of any single vertex is determined by their relationship to other vertices. The centre of the network is most affected by this because this position is populated by vertices that share an edge with vertices from across the network. For example, the large red circle in the middle of the graph is the Abell-Kendall Method (Abell et al. 1952), the second most cited document (497 citations), which was the standardised method to measure total serum cholesterol (TC) and was used by papers from many different areas of the
network. While it receives 132 citations from papers in its own cluster (C0), the largest contribution from any single cluster, it nevertheless gathers most of its citations from other clusters.

A co-citation analysis was also performed on this dataset, but a brief comparative analysis suggested that modularity-based techniques performed better on the full directed citation network in terms of community detection. This results coheres with work conducted previously. Klavans and Boyack (2017) performed a series of comparative tests using bibliographic coupling, co-citation analysis, and directed citation analysis. They concluded, “It is clear from this study that co-citation is inferior to direct citation if the goal is to create a knowledge taxonomy. Not only is the coverage lower, but the accuracy is much lower” (2017, p.996), and that modularity-based clustering of directed citation networks was the most accurate and inclusive computational method currently available.

**Categorising clusters**

To give an indication of the content of documents in each cluster, I constructed 16 sub-graphs that included only the vertices from each cluster and the edges between these. These sub-graphs were then analysed to establish the most highly cited documents from within these clusters (i.e. relative to references from documents only contained in that cluster).

From this, I read the title, abstract, or full-text of the top-ten most highly cited documents from within each grouping. I classified each by establishing the primary focus of each paper and the type of study performed. If methods papers were included in the top ten, I noted the aim of the methodology described.

In each community, I observed a right-skewed distribution of citation (apart from C15), similar in inequality to the citation distribution across the whole network. Thus, the top-ten studies in each are disproportionately utilised by other documents within these specific communities. Thus, from a small set of documents, a reasonable approximation can be made as to what a specific cluster is focussing on.
However, after analysing C0 and C6, coloured red and yellow in the above visualisation, it was clear that these clusters were heavily influenced by the temporal dimension described previously.

C0 (median publication date 1958) is composed of nearly all studies published over the 1950s debating the relationship between diet, serum cholesterol, and CHD. The top-ten cited documents are all controlled-feeding studies reporting results on the relationship between dietary fats and serum cholesterol levels (Keys et al. 1957; Bronte-stewart et al. 1956; Ahrens et al. 1954, 1957; Kinsell et al. 1952; Malmros and Wigand 1957; Beveridge et al. 1955, 1956; Groen et al. 1952; Sinclair 1956). These are joined by two methods of measuring serum cholesterol; Abell-Kendal Method (Abell et al. 1952) for estimating total serum cholesterol (TC) and Sperry and Webb’s (1950) adaptation of the Schoenheimer-Sperry method for cholesterol measurement.

C6 (median date: 1950) contains most of the works published before 1950 that focussed on developing animal models of atherosclerosis, identifying lipoproteins, and exploring the link between serum cholesterol and CHD. In the top-ten cited documents are two papers by Gofman et al. (1950a, 1950b) reporting their successful attempt at defining a range of lipoproteins and associating these with atherosclerosis in a replication of Anitschkow’s animal model of atherosclerosis. This is joined by Jones et al. (1951) study on lipoproteins and atherosclerosis. Following this is Gertler et al. (1950a) study of the interrelationship between serum cholesterol and phospholipids in those with CHD, which proposed the total cholesterol:phospholipids ratio as a measure for CHD risk, and Ahrens and Kunkel’s (1949) study of the role of phospholipids in atherosclerosis. There are two methodological papers; one serum cholesterol measure (Schoenheimer and Sperry 1934) and one on the detection of phosphorus (Fiske and Subbarow 1925). Five documents focus on animal models of atherosclerosis and basic mechanisms – Anitschkow’s (1933) *Experimental Atherosclerosis*, Hueck’s (1944/45) review of arteriosclerosis, Katz and Stamler’s (1953) *Experimental Atherosclerosis*, Duff’s (1935) review of experimental atherosclerosis, and Steiner and
Kendall’s (1947) report of atherosclerosis in dogs provoked by dietary cholesterol after thiouracil treatment (a drug that reduced thyroid hormone synthesis). Finally, one clinical study of serum cholesterol levels in those suffering coronary occlusion (Morrison et al. 1948).

As these periods were dominated by particular forms of studies, controlled-feeding studies and animal experimentation, it is not possible to ascertain whether it is study type that is clustering vertices together or simply the time period in question. C6 provides an indication that my assumption that study type, defined by methodological orientation, might be incorrect. While animal models dominate the top-ten, there are also clinical studies here. It could be that research focus, defined as a set of general research questions, plays a more important role at this level of analysis. For example, all documents in the top-ten in Cluster C0 are focussed on how diet impacts serum cholesterol in humans, while the documents in Cluster C6 focus on how serum cholesterol, or the particular lipoproteins that carry cholesterol within the body, contribute to atherosclerosis/CHD. However, other clusters by virtue of similar median publication years, should produce clearer results.

I began with the smallest cluster, C15 (median date: 1965), composed of only 11 papers. All of these were book reviews of the same book Kummerow’s (1965) *Metabolism of Lipids as Related to Atherosclerosis*. Neither the book or the book reviews received any citations from any other document in CUT-N1. These vertices were retained in the largest component by virtue of one of these documents referencing another document not in this cluster.

The closest cluster to this in terms of median publication date is C9, containing 616 documents. This group appears to be focussing on clinical descriptions of atherosclerosis, thrombosis, and CHD. The highest-cited paper is by Schlesinger (1938) describing a method of radiographing the coronary arteries to detect sites of occlusion. Three papers focus on the role of thrombosis in myocardial infarction (Roberts and Buja 1972; Blumgart *et al.* 1940; Levine 1929), one each on prevalence of atherosclerosis in men aged 18 to 39 (Yater *et al.* 1948), coronary aneurism (Daoud *et al.* 1963), the
pathology of coronary arteries following sudden death from myocardial infarction (Crawford et al. 1961), proposing a new class of angina (Prinzmetal et al. 1959), a report from the international atherosclerosis project (Guzman 1968), coronary arteriography (Proudfit et al. 1966), and hypertension (Bell and Clawson 1928). Finally, Herrick’s (1912) famous description of coronary occlusion and thrombosis.

C3 (median date: 1967) has 1,613 documents. The top-cited papers focus on three interrelated topics – measures of lipoprotein abnormalities, serum triglyceride levels, and dietary carbohydrate consumption and their effect on blood lipid profiles and CHD/atherosclerosis risk. The top-cited paper, by Fredrickson et al. (1967), is a major five-part review of research into fat transport via lipoproteins; a review that argued the case for greater use of direct measures of lipoproteins over total serum cholesterol measurements, as well as highlighting major classes of hyperlipidaemia that appeared to have a genetic basis. Below this is a method to establish the level of serum triglyceride in serum (Vanhandel and Zilversmit 1957), a constituent then thought to be raised following carbohydrate consumption. Another methodological paper focusses on a lipoprotein measurement (Lees and Hatch 1963). The remaining documents include seven on carbohydrate consumption and their effect on lipid profiles and risk of atherosclerosis/CHD (Dole 1956; Macdonald and Braithwaite 1964; Ahrens et al. 1961; Kuo and Basset 1965; Albrink et al. 1959, 1961; Carlson 1960; Antonis and Bersohn 1961). While not in the top ten, there are 19 papers in this group first authored by Yudkin, the originator of the sugar–heart hypothesis.

C11 (median publication date: 1968) has 199 documents focussing on fatty-acids in milk, their synthesis, and bovine lipoprotein profiles. The top-cited paper, Annison et al. (1967), examines milk-fat synthesis in the goat. This is followed by Storry’s (1970) review of milk-fat synthesis, Moore’s (1968) study of factors impacting the level of PUFA in the plasma lipids of sheep, Davis and Sachan’s (1966) study of factors effecting the fatty acids composition of bovine milk-fat, Raphael et al. (1973) study of bovine serum lipoproteins. Hartmann and Lascelles (1965) study of bovine lipoprotein profiles during lactation.
Barry’s (1963) study of fat transport in the goat. However, there is also a paper by Horwitt (1960) on the relationship between vitamin E and lipid metabolism in man, one paper describing a method for serum phospholipid analysis (Nelson 1959), and one on hydrogenation of PUFA (Scott et al. 1971).

Cluster C4 (median date: 1969) has 1,582 documents. These focus on the pathogenesis of atherosclerosis. The top-cited document is Constantinides (1965) book on *Experimental Atherosclerosis*, a large review of basic studies of atherosclerosis with an emphasis on animal models. In the top ten are six studies on animal models of atherosclerosis in the monkey, rabbit, and swine (Constantinides et al. 1960; Florentin et al. 1968; Armstrong et al. 1970, 1972; Newman and Zilversmit 1962; Kao and Wissler 1965). The remaining three papers include Ross and Glomset’s (1976) review of basic theories of atherogenesis, a study estimating cholesterol in serum (Leffler 1959), and an interesting study by Walton and Williamson (1968) advocating a role of ‘fibrin incrustations’ in atherosclerosis – an old theory, first proposed by Rokitansky (1852), and a historic rival to the lipid infiltration theory that forms the basis of the diet–heart hypothesis. Thus, this cluster appears to be populated by papers discussing basic theories of atherosclerosis and animal models.

C8 (median date: 1970) has 826 documents and is focussed on cholesterol synthesis, turnover, and excretion. All of the top-ten cited documents focus on this. The top two focus on measuring the amount of cholesterol excreted in faecal bile acids (Grundy et al. 1965; Miettinen et al. 1965), while the remaining eight focus on cholesterol metabolism (Grundy and Ahrens 1969, 1970; Grundy et al. 1968; Dietschy and Wilson 1964; Wilson 1964; Quintao et al. 1971; Connor et al. 1969; Moore et al. 1968).

C5 (median date: 1971) has 1,558 documents. Here, we find famous studies of the diet–heart hypothesis. The most cited document is Keys’ (1970) *Seven Countries* study. This is joined by Leren’s (1966) secondary prevention trial, the National Diet-heart Feasibility Study (NDHF 1968), Hegsted et al. (1965) modification of the Keys’ equation predicting the rise and fall of serum cholesterol in response to dietary fat and cholesterol, Dayton et al. (1969) LA veterans trial, Miettinen et al. (1972) Finnish Mental Hospital Trial, Keys et al.
(1965) modification of their 1957 Keys equation, Stamler’s report for the Inter-Society Commission for Heart Disease Resources (1970), a major review heavily backing the lipid and dietary hypotheses in primary prevention and recommending policy action, including specific population dietary advice. Finally, Paul et al. (1963) prospective cohort study and a report from the Framingham Study (Gordon and Kannel 1970). This cluster, then, appears to be debating the major evaluative studies of the diet–heart hypothesis in the 1960s and early 1970s.

C12 (median date: 1971) has 157 documents. All top cited documents are by Wexler et al. (1963a, 1963b, 1964a, 1964b, 1965, 1967, 1968, 1970a, 1970b), papers examining the role of spontaneous atherosclerosis, adrenaline, sex hormones in provoking atherosclerosis, and drug-induced myocardial necrosis in the absence of atherosclerosis in animals. Thus, Wexler was exploring the pathogenesis of atherosclerosis and CHD.

C14 (median date: 1972.5) has 66 documents primarily in Russian and German, which are rarely cited by other documents in the network. These appear to be on a range of topics. The top-cited papers are on lipid peroxides (Lankin et al. 1976; Lankin 1979), both published in Russian. This is joined by three on bile acids and lipid metabolism (Hunt et al. 1963; Leveille et al. 1963; Edwards 1961). Two on the toxicity of dietary lithocholic acid and its effect on liver size and lipid metabolism (Levelle 1962 [in Russian]; Eyssen and De Somer 1963). This cluster contains low-cited Russian papers beyond this.

C13 (median date: 1972.5) has 82 documents. Of the top-ten most cited papers, one is on an assay for serum cholesterol (Kim and Goldberg 1969), six on vitamin C and either lipid metabolism, atherosclerosis or CHD (Spittle 1971; Sokoloff et al. 1967; Elwood et al. 1970; Kothari and Jain 1977; Anderson et al. 1958; Roe and Kuether 1943) and three on the benefits of ‘garlic juice’ for the prevention of atherosclerosis (Bordia and Bansal 1973; Bordia et al. 1977a, 1977b).

C7 (median date: 1975) has 1,325 documents. Three of the top-ten papers focus on statistical methods (Snedcor and Cochran 1967; Steele and Torrie 1960; Duncan 1955), two methods for measuring serum cholesterol
(Pearson et al. 1953; Roschlau et al. 1974) and lipids in tissue (Folch et al. 1957), five documents focus on dietary protein on serum cholesterol, atherosclerosis, and CHD in both animal studies and metabolic ward studies (Carrol and Hamilton 1975; Hamilton and Carrol 1976; Sirtori et al. 1977; Kritchevsky et al. 1977; Huff et al. 1977), one on dietary fibre and CHD (Trowell 1972), and one focusses on the role of cholesterol-free, high-fat diets in inducing atherosclerosis in experimental animals (Lambert et al. 1958). Scanning down the list of papers, this appears to include many documents examining the role of different dietary factors, primarily protein, in inducing or retarding atherosclerosis.

C10 (median date: 1976) has 338 documents. The top-cited focussed on the role of eicosapentaenoic acid, an omega-3 PUFA typically found in oily fish, and its role in preventing thrombosis and atherosclerosis (Dyerberg et al. 1978). Following this are seven documents that focus also on establishing the role of fatty-acids derived from fish in atherosclerosis and CHD risk. These include four studies on Inuit populations who consumed large amounts of fish and were relatively free of CHD (Dyerberg et al. 1972, 1979; Bang et al. 1972, 1976), and three studies on the effect of the consumption of fish or fish oils on human lipid profiles (Lossoncz et al. 1978; Siess et al. 1980; Sanders et al. 1981) and one on mackerel oils effects on serum lipids in swine (Ruiter et al. 1978). Finally, a review by Goodnight et al. (1982) on the role of PUFA in hyperlipidaemia and thrombosis.

C1 (median date: 1976) is composed of 1,686 documents. The top-cited literature in this group is dominated by methods to measure lipoproteins. The top-cited is Lowry et al. (1951) on the folin phenol reagent for protein measurement. This is joined by four papers on methods for detecting lipoproteins in serum (Havel et al. 1955; Bartlett 1959; Noble 1968; Glomset 1968). Five empirical studies focus on elucidating lipoprotein metabolism. These include Shore et al. (1974) study of the role of Very-Low Density Lipoprotein (VLDL, which is converted in the blood into LDL) in provoking hypercholesterolemia in rabbit, and Bilheimer et al. (1972) study on the metabolism of VLDL. Goldstein and Brown’s (1977) study on the LDL-pathway
and its relationship to atherosclerosis reports their discovery of how the LDL-receptor regulates serum cholesterol levels, and how an inhibitor of HMG CoA reductase can induce the expression of LDL-receptors in the liver that, in turn, lowers the amount of cholesterol in the serum. This provided the basic understanding that underpinned statin treatment; an achievement that would result in their winning of a Nobel Prize in 1985. Finally, there are two documents by Mahley et al. (1974; 1978) on abnormalities in lipid and lipoprotein metabolism. This cluster thus appears to be focussing on basic studies of lipid metabolism.

Finally, C2 (median date: 1978) is composed of 1,650 documents. This group is composed of many documents on the protective role of HDL in preventing CHD. The top-cited paper is Miller and Miller’s (1975) article claiming that a low-level of HDL-C may be independently associated with the development of atherosclerosis and CHD, an observation not anticipated for most of the history of diet–heart research, with studies focussing on the dangers of serum cholesterol, either TC or LDL-C. This is followed by four large epidemiological studies all reporting an independent, inverse association between the level of HDL-C and the occurrence of CHD (Rhoads et al. 1976; Miller et al. 1977; Gordon et al. 1977; Castelli et al. 1977), and one paper focussing on the mechanism by which HDL may be protective against atherosclerosis (Carew et al. 1976). Accompanying these studies is the Lipid Research Clinics Program Manual of Laboratory Operations (1974) – a manual standardising particular measures for intervention and observational studies. Two studies focus on the role of “pre-beta” lipoproteins (VLDL) and CHD (Kannel et al. 1971) or its conversion to LDL (Wilson and Lees 1972). Finally, one study reports on an independent association between serum triglyceride and CHD (Carlson and Bottiger 1972). This cluster appears to be focussed on how HDL may be protective and what causes it rise and fall in the blood.

Study type or research focus?

Here, my anticipation that study type, defined by particular methods, would play an important role in clustering documents together is not supported. As modularity is a relative measure that considers the entire structure of a
network, changing the composition of a network (either by restricting it to a
time period or by focusing on a specific cluster) will result in different
modularity groupings. So it may yet be the case that study type does impact
clustering, it just wasn’t the major influence at this level.

What my analysis does show, however, is that when considering a large
set of documents published over a long period of time, study type does not
play the role that I expected. Rather, documents are clustered around specific
questions. For example, Cluster C1 is focused on the mechanism controlling
serum cholesterol levels, Cluster C2 on whether HDL is protective against
CHD and atherosclerosis, Cluster C0 on whether diet is implicated in the
raising of serum cholesterol and the risk of CHD, while Cluster C5 focuses on
the merits of the dietary fat link via observational and intervention trials.

However, it is important to highlight that these clusters are not silos that
rarely interact with one another.

**Interconnection between clusters**

I performed a meta-vertex analysis by aggregating vertices of one cluster
into a single vertex and retaining their edges via constructing directed-
weighted edges between meta-vertices. Thus, if five documents in cluster A
cite documents in cluster B eight times, then a weighted-directed edge from A
to B will have a value of eight. By doing this, I also allowed weighted-self loops
that establish how many times documents within a particular cluster cite one
another.

The CUT-N1 has a total number of edges (references) of 93,809 and the
proportion of these edges that are contained within and between clusters can
be calculated. A total of 61,820 (66%) are within clusters, while 31,989 (34%)
are between clusters. Thus, these clusters are interconnected.

Most clusters are connected to all other clusters via references, though
no cluster ever references C15 – the community formed around one book and
10 reviews of it. C0 through to C10 all share references and citations with all
other clusters, apart from C15. C11 does not reference [C13, C15], C12 does
not reference [C14, C15], C13 does not reference [C11, C14, C15], C14 does
not reference [C12, C13, C15], and C15 references only one community [C0].
As discussed, the clusters are not equal in the number of documents contained within them, but neither are they equal in terms of the number of citations accumulated. Fig 49 plots the number of documents within a cluster and the number of citations received by a cluster (including within and between cluster citations) with y-axis log transformed.

**Fig 49: Number of documents in, and number of citations to, each cluster in the CUT-N1 network**

To ascertain how related one cluster is to another, this must be taken into account. There are two feasible methods. First, to calculate the proportion of total references from documents in a specific cluster that are directed at documents in a different, specific cluster. Second, one can divide the incoming citations to documents in a cluster by the number of documents that make up that cluster, which establishes the average in-degree (citations) contributed to vertices from a specific cluster. After performing both, the strongest relationship by both measures is the dependence of C0 on C6. C0 directs 12% of its reference to C6, while C6 documents are cited on average 1.14 times by documents in C0. The second strongest relationship is between C5 and C0 – with C5 contributing 11.5% of its references to C0, while documents in C0 are referenced on average 0.82 times by C5. However, this is unusual – only 41 pairs of vertices (out of 199) direct over 5% of their references to documents in another single cluster. However, as Table 14 demonstrates, most clusters direct between 20–50% of their references to documents not within that cluster.
Table 14: Proportion of total references from within each cluster directed at other clusters.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>References to other clusters</th>
<th>Total references</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0</td>
<td>3679</td>
<td>12661</td>
<td>0.29</td>
</tr>
<tr>
<td>C1</td>
<td>3522</td>
<td>12861</td>
<td>0.27</td>
</tr>
<tr>
<td>C2</td>
<td>4786</td>
<td>12923</td>
<td>0.37</td>
</tr>
<tr>
<td>C3</td>
<td>3443</td>
<td>11219</td>
<td>0.31</td>
</tr>
<tr>
<td>C4</td>
<td>2885</td>
<td>7738</td>
<td>0.37</td>
</tr>
<tr>
<td>C5</td>
<td>5144</td>
<td>9266</td>
<td>0.56</td>
</tr>
<tr>
<td>C6</td>
<td>1690</td>
<td>8472</td>
<td>0.20</td>
</tr>
<tr>
<td>C7</td>
<td>2991</td>
<td>8218</td>
<td>0.36</td>
</tr>
<tr>
<td>C8</td>
<td>1874</td>
<td>4698</td>
<td>0.40</td>
</tr>
<tr>
<td>C9</td>
<td>680</td>
<td>2244</td>
<td>0.30</td>
</tr>
<tr>
<td>C10</td>
<td>723</td>
<td>1712</td>
<td>0.42</td>
</tr>
<tr>
<td>C11</td>
<td>251</td>
<td>689</td>
<td>0.36</td>
</tr>
<tr>
<td>C12</td>
<td>130</td>
<td>630</td>
<td>0.21</td>
</tr>
<tr>
<td>C13</td>
<td>144</td>
<td>307</td>
<td>0.47</td>
</tr>
<tr>
<td>C14</td>
<td>46</td>
<td>160</td>
<td>0.29</td>
</tr>
<tr>
<td>C15</td>
<td>1</td>
<td>11</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Thus, while we can cluster groups of documents together, it is clear that these clusters are heavily connected to other clusters. Cluster C6 is the most self-sufficient, but this is explained by virtue of its early median publication date. Most documents in this cluster do not have the opportunity to reference the documents published later in other clusters. While it might be tempting to refine the dataset to look only at particular clusters, this would be a distorting manoeuvre.

To help conceptualise the content of each cluster and its position in the network, I divided the clusters into five larger groups based on broader research focus.

First, clusters C6, C4, and C12 focus primarily on basic studies of atherosclerosis, with a particular focus on animal models. These three clusters make up 21% of all vertices and 13.8% of all edges.

Second, three clusters focus primarily on either developing or evaluating the dietary fat or serum cholesterol link to CHD via clinical studies (e.g., controlled-feeding studies, observational studies, and intervention studies) – C0, C5, and C2, which together hold 33.6% of all vertices and 31.1% of all
edges. C0 primarily develops the dietary fat hypothesis, C5 testing that hypothesis, and C2 focussed on an unexpected finding that derived from the studies launched to test the diet and lipid hypotheses.

Third, clusters C8, C11, and C1 appear to be primarily interested in developing basic understanding of cholesterol synthesis and metabolism – these make up 18.2% of vertices and 12.5% of vertices.

Fourth, clusters C3, C7, C10, C13 all appear to be focussed on examining whether other dietary factors (carbohydrate, protein, fish, and vitamin c) are implicated in lipid abnormalities and atherosclerosis/CHD. Together, these hold 22.9% of all vertices and 14.21% of edges.

Finally, cluster C9 appears to focus on understanding the pathology of CHD, which includes 4.1% of all vertices, and 1.67% of all edges. I’ve left out the two smallest clusters – C14 because it is dominated by Russian authors, and C15 because of its small numbers, which together make up only 0.5% of vertices and 0.13% of edges.

To simplify further, we might also make a distinction between: (i) basic science including studies interested in developing understanding of physiological mechanisms – C1, C6, C4, C8, C11, and C12; (ii) applied research, such as clinical and epidemiological studies – C0, C2, C3, C5, C9, C10. While C7 appears to be mixed – including both animal and clinical studies in the top-cited papers. Generally, basic studies appear to reside along the bottom half of the network, while applied research studies across the top half.

**Basic and clinical research**

To understand this further, I classified all retrieved documents in the CUT-N2 network (all vertices have full title data) by whether they included terms signifying animal experimentation or not (e.g. dog, rabbit, monkey, chick, “experimental”) in Excel (Ch8.S5). These were classified as “Animal” and titles were read to ensure human studies were not included. This process was repeated multiple times as I became aware of other animals used regularly (e.g. guinea-pig, hamster, quail, prawn cockerel, pig, hen, duck, lamb, calves, pigeon, baboon etc.). As I classified all 399 papers derived from the A search performed in Chapter 7, all methods papers detected here were classified as
“Method”. All other documents were presumed to focus on humans, and these classified as “Human”. If there is a split in this network between basic and applied research, it ought to be apparent by the number of animal studies contained in particular regions of the network. Fig 50 displays these results, and clearly indicates there is a divide in this network between basic research and clinical/applied research. Animal studies overwhelmingly reside in the bottom half of the network, while human studies dominate the top half.

Fig 50: Citation network of the CUT-N2 network with documents coloured by terms contained within their titles indicating a focus on animal (blue), methodological (yellow), or human (red) research (n=5870; m=42424).

As the position of documents in this network is established by their reference and citation links alone, this tells us that there is a divide between basic and applied research in regards to how documents relate to one another through their references. While this method of classification will not have correctly classified all papers, all papers appearing in blue contained terms that indicated that animals were used in the experiments reported in those
papers. About 37.5% of papers were classified as animal, only 1.5% as methods (these were pulled into the network by the A search), while 61% were classified as human studies, but these were only identified by the absence of terms indicating a focus on animal experimentation.

4. Discussion

In this chapter, I detected two factors that appear to contribute heavily to the structure of this citation network. From an analysis of time, I established the half-life of citation influence for documents. However, documents that received many citations appeared to be more resistant to the decaying effects of time. The tendency of documents to reference documents recently published stretched the network from left to right roughly following the arrow of time. Second, documents appeared to cluster into communities based around specific research questions, rather than specific study types. These results, taken together with the results reported in Chapter 5, which followed how the results of Paul et al. (1963) were used in the scientific literature in different ways by groups pursuing different research questions, suggest that, at least in this area, studies addressing particular questions reference one another more frequently than those addressing different, though related, questions.

This analysis of time and community clustering appears to highlight how diet–heart research focussed on different, though related questions over time. The controlled-feeding studies of the 1950s appeared to grow straightforwardly out of the work in the previous decades regarding the relationship between serum cholesterol and atherosclerosis. However, following these studies, communities began to diverge on the basis of a separation between clinical and basic scientific studies, pursuing different, though related, questions.

This distinction between clinical and basic science appeared to help explain structural features of this network. Clinical studies addressed questions such as: Is diet related to serum cholesterol levels? Is diet and serum cholesterol levels related to the risk of CHD? And what lipid measure is best in helping to predict the occurrence of CHD? The more basic studies
focussed on questions such as – what mechanisms explain the deposition of cholesterol in the walls of arteries? What carries cholesterol in the body? And how is serum cholesterol regulated? Generally, animal and human studies resided in different halves of the network when split horizontally.

When evaluating the diffusion of ideas or data, both this temporal dimension and community structure need to be taken into account. That is, ideas and data will spread relatively quickly and widely within a particular time-period and in a particular space (cluster).

Interestingly, no community clearly focussed on the refutation of any specific theory. Rather, in each cluster, the highest-cited documents focussed on attempting to establish and categorise something that is thought to ‘exist’, rather than proving that something does not. This picture of science sits starkly at odds with the ideal of scientific work that Popper proposed – a vision of science where bold idea and conjectures were first proposed and then scientists actively pursued rigorous attempts to refute them.

Importantly, at the outset of this thesis, I anticipated that network analysis when applied to this area would reveal a cluster of vertices dedicated to challenging the diet-heart hypothesis. Thus far, when considering a large body of literature, I have not found this; rather, one is left wondering exactly where the apparent long-running scientific controversy is?

**Reflection**

In this dataset, there are many stories to tell, and from my experience of developing it, I believe the history of science could benefit from adopting this method to retrieve literature and to explore it. By systematically collecting literature and visualising it in this manner, one is able to see how documents relate together through their citation relationships and cluster into distinct groups over time. However, it became clear to me that the history of this area is more complex than I had anticipated. I have observed a complex research ecology – with many different lines of research emerging over time that had varying levels of popularity. These stories need to be told, but I do not have time to do them justice here – this awaits a future project. For now, in the following chapter, I examine the CUT-N1 network via Main Path Analysis.
Chapter 9: The Paths Most Trodden

Introduction

Main Path Analysis (MPA), designed by Hummon and Doreian (1989), reveals paths that connect documents through chains of citations. To identify a main path, one first counts how many times particular edges (citations) are traversed to establish the path with the highest total traversal count (SPX). Citation networks are often composed of thousands of documents interlinked by a set of references many times larger. By identifying main paths, this complexity is reduced.

For most of the history of Scientometrics, citation scores have been the dominant method for measuring which documents are influential. By focusing on paths, rather than citation counts, MPA has been said to provide a better account of knowledge development. In recent years, this methodology has been refined (Batagelj 2003; Lucio-Arias and Leydesdorff 2008; Liu and Lu 2012; Batagelj et al. 2014; 2017). Advocates have promoted its use in systematic reviews (Colicchia and Strozzi 2012; Xiou et al. 2014), for algorithmic historiography (Lucio-Arias and Leydesdorff 2008), and for tracking technologies through networks of patents (Verspagen 2007).

I applied this methodology to the CUT-N1 network. To my knowledge, this is the first application of MPA on a large dataset assembled around a theory, rather than a research topic. Hummon and Doreian (1989) traced the history of DNA by applying MPA to a small sample of references from Asimov’s (1963) history (Chapter 1). The CUT-N1 network, by contrast, was generated from a systematic search, covering more than a century of publications. Furthermore, the inclusion of documents was not guided by preconceptions about how the field developed. While Batagelj et al. (2017) examined a research area (‘peer-review’) not a theory, they similarly aimed to “capture its emergence and evolution and identify the most influential publications” (p.525). There are similarities in network size and in how they constructed their
network. They dealt with the ‘boundary problem’ in the same manner as I have – removing non-retrieved vertices <3 citations and manually retrieving any document cited ≥20 times (Chapter 7).

2. Methods

The CUT-N1 network comprises a weakly-connected component of 14,909 vertices and 93,809 edges; 6,206 vertices represent retrieved documents, while 8,703 represent non-retrieved documents cited at least three times by retrieved documents (Chapter 6–7). The algorithms for traversal counting and MPA have been validated in Batagelj et al. (2014 pp.69-116), and are available via Pajek. I used Pajek Version 5.07 (Batagelj and Mrvar, 2018), to perform traversal counts and MPA. I exported files from Pajek to Excel for manual inspection and to edit the vertex-attribute-list and edge-list. Finally, I used Gephi version 0.92 (Bastian et al. 2009) for visualisations.

Traversal counting

Traversal counting refers to how often particular edges are traversed by paths. A path is a sequence of vertices connected by a chain of edges, and in a large network, very many paths may connect any two randomly-chosen vertices. The below figure displays a network with two paths between vertices a and z. The shortest path, a–b–z, has length 2; the longest, a–c–d–z, has length three.

Fig 51 Citation network path example
In traversal counting, however, the number of paths that traverse an edge are counted. For this, Batagelj (2003) proposed the Search-Path Count (SPC) method, based on a search through all possible paths between ‘Source’ (s) vertices (documents that are cited but contain no references) and ‘Sink’ (t) vertices (documents that reference but are not cited). In the above diagram, the traversal count for all edges is one, because only one path exists for each series of vertices.

Fig 52 Citation network traversal weight example

This example has one Source, z, and one Sink, a. Only one edge sits on more than one path, b–z. The edges are weighted by their traversal count and the thickness of the edge reflects this.

Other methods of traversal counting were proposed by Hummon and Dorain (1989), but Batagelj (2003) recommended that SPC is more efficient. While there is debate regarding which method best approximates knowledge diffusion (Liu 2019), the results are typically equivalent (Batagelj 2003). For SPC, a network must be acyclic or transformed to meet this condition by removing loops (Batagelj 2003). The CUT-N1 network contains just 26 mutual edges (vertices that cite one another). I deleted all edges from high-cited to low-cited literature for all mutual-degree relationships, removing 13 edges. I also performed the opposite to examine whether this altered the results, but found no important differences.

The CUT-N1 network has 2,240 Sources, 9,178 Sinks, and 3,491 intermediate vertices. Computational methods are required to calculate the
number of paths from all Sources to Sinks, and their traversal weights. I calculated edge weights without normalisation – the edge with the highest traversal count had a value of $5 \times 10^{22}$. To make these data manageable, I used flow normalisation. A flow normalisation can be read as the probability that a random s-t path passes through the edge $(j,i)$, and this does not impact the results of an MPA (Batagelj et al. 2014). Flow normalisation of SLP is given by:

$$w(j,i) = \frac{n(j,i)}{F}$$

where $n(j,i)$ is the number of paths from source ($s$) to sink ($t$) through the edge between vertices $j$ and $i$, while $F$ is the sum of all SPC values over a minimal edge cut-set (Batagelj et al. 2014).

I placed weights on all edges that reflect flow-normalised SPC values, and a value on each vertex that reflects the flow-normalised SPC weight. While the former calculates the probability that an edge sits along a random path, the latter reflects the probability that a vertex will sit along a random path – the total of the SPC weights of either its incoming or out-going edges (which are necessarily equivalent).

**Main Path Analysis**

There are several ways to identify the path with the highest SPX. Hummon and Dorien (1989) proposed a ‘priority first’ algorithm, by beginning at a randomly selected Source, selecting the citation to it with the highest traversal weight and following this to the next vertex, repeating until a Sink is hit. This is repeated for all paths from Sources to Sinks, and the path with the highest SPX is the main path. In Pajek, users can perform this by the Forward Search function ($F$). This approach can also be performed from Sinks to Sources by the Backward Search ($B$). A main path constructed by either technique will produce a path in which each vertex is directly connected to only one other vertex. Both approaches can miss the path with the highest SPX, and $F$ and $B$ searches can produce subtly different paths (Liu and Lu 2012). A Global Search ($G$) method calculates the path with the highest SPX. To do this, Pajek
is used to sum the traversal weights of all edges on all s-t paths. This finds the most traversed path regardless of how it developed.

**Key-Route Searching**

Both local and global search techniques can be overly restrictive. A complex network will consist of millions of paths, and, while most will have a low SPX, a single path might miss important documents and edges that have high traversal counts. Thus recent studies have applied all three searches.

Liu and Lu (2012) introduced the Key-route method, which combines elements of forward and backward local searches. It begins at a pre-defined number of influential edges (edges with the highest traversal weights). From this it searches backwards until a Source is hit, and then forward until it reaches a Sink, the resulting path is known as a key-route main path. The Global key-route establishes the path with the highest SPX in which a particular key-route exists. Increasing the number of initial key-routes will increase the number of high-SPX paths discovered, but typically, these paths either converge onto or split from the main path at some point.

Here, I apply these MPA analyses. From this, I generate an interpretative account of this area of research.

**3. Results**

**Forward, Backward, and Global MPA**

All data underpinning the following analyses and their results are provided in the Supplement (Ch9.S1-3). The F-analysis produced a path of 64 vertices from the ancestors (documents that are referenced) of Anitschkow (1913a) to Beynen et al. (1984), with a normalised SPX of 5.95. The B-analysis produced a path of 63 vertices from the ancestors of Anitschkow (1913a) to the descendants of (documents that cite) Oliver (1984), with a normalised SPX of 5.34. The G-analysis produced a path of 68 vertices from the ancestors of Anitschkow (1913a) to Rossouw et al. (1984), with a normalised SPX of 6.54.
As there was a large overlap in the vertices and edges detected, I visualise them together (Fig 53). Blue edges include edges included in all MPA searches, purple represents edges shared by F and G, dark green for edges shared by G and B. Red edges signify edges identified only via the F analysis, green edges for those identified only by B, and yellow edges for those identified via G, one edge is coloured pink that is shared in the F and B search but not present in G. The thickness of an edge is proportional to its normalised SPC, while the direction of the arrows reflects the citation relationship. Vertices are coloured grey, labels reflect the first author and year of publication. In Panel A, I show the full results, Panel B shows the network until 1961, while Panel C shows the network between 1962 and 1984.
Fig 53: Combined results of local forward (F), backward (B), and global (G) main-path analyses of the CUT-N1 network. Panel A shows full results, Panel B shows results until 1961, Panel C shows results from 1962–1984.
The documents in Fig 53 came from four search queries; 54 from the dietary terms query (D), 21 from the query used to captured terms related to the relationship between atherosclerosis, serum cholesterol, and CHD (H), 24 from the top-cited literature retrieval generated from the D and H networks (A), and two from the missed documents established from the validation exercise (R) (Chapters 6–7). Six documents were not formally retrieved — these the document referenced by Anitschkow (1913a). In this analysis, 18 documents on these main paths were published <1950, 31 from 1950–59, 18 from 1960–69, 19 from 1970–79, and 19 from 1980–84. The documents derive from seven of the clusters detected in the previous chapter.

*Table 15: Cluster membership of documents detected via MPA*

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0</td>
<td>29</td>
</tr>
<tr>
<td>C2</td>
<td>12</td>
</tr>
<tr>
<td>C3</td>
<td>1</td>
</tr>
<tr>
<td>C4</td>
<td>1</td>
</tr>
<tr>
<td>C5</td>
<td>28</td>
</tr>
<tr>
<td>C6</td>
<td>26</td>
</tr>
<tr>
<td>C7</td>
<td>8</td>
</tr>
</tbody>
</table>

The most represented clusters are C0 and C5, on the relationship between diet, serum cholesterol and CHD. Following this is C6, representing early research on the relationship between cholesterol and atherosclerosis. C2 is dominated by discussion of whether HDL-C is a predictor of CHD, while C7 examined the relationship between dietary protein and CHD. Thus, the main paths primarily run across the top half of the network depicted in the previous chapter. This MPA is largely blind to basic research studies, which tended to reside along the bottom half of the network.

**Key-route MPA**

For a key-route analysis, following Batagelj *et al.* (2017), I selected the 100 edges with the highest traversal weights and performed both local and global analyses (see Ch8.S4–6). While their network is roughly twice the size of CUT-
N1 and focusses on a broad research area over a fifty-year period, mimicking their choice allows comparisons to be drawn. The Local key-route analysis produced 128 documents, while the Global analysis returned 156, including all but 6 documents from the Local analysis. Below, I merge the results.

Fig 54: Local and Global Key-route MPA analyses with 100 initiating edges. Edges in blue reflect edges shared in both analyses, red=global only; green=local only. In the bar graph, the colours represent periods that overlap with the vertices in the above network, and blocks are proportional to the number of documents from those periods.

In both analyses, the first significant document is Anitschkow (1913a). The main paths split towards the end of the 1970s; one ends at Beynen et al. (1984); the other splits into two small paths. The global and local searches both end with Roussaw et al. (1984), while the local search splits leading to Oliver (1984). Thus, even after including the top 100 key-routes, the results do not dramatically differ.

These 162 documents (D –80; H–32; A–35; R–2; non-retrieved–6) derive from eight clusters:
<table>
<thead>
<tr>
<th>Cluster</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0</td>
<td>43</td>
</tr>
<tr>
<td>C1</td>
<td>1</td>
</tr>
<tr>
<td>C2</td>
<td>17</td>
</tr>
<tr>
<td>C3</td>
<td>3</td>
</tr>
<tr>
<td>C4</td>
<td>1</td>
</tr>
<tr>
<td>C5</td>
<td>39</td>
</tr>
<tr>
<td>C6</td>
<td>44</td>
</tr>
<tr>
<td>C7</td>
<td>14</td>
</tr>
</tbody>
</table>

Thus, again, the paths traverse primarily across the top of the CUT-N1 network.

**Cursory analysis**

The main path begins with Anitschkow (1913a) and ends shortly after the Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT 1984), generally taken as demonstrating a causal link between serum cholesterol reduction and a reduced risk of CHD. Although there are important absences, most notably Keys’ (1970) Seven Countries study, it provides a general outline of how the diet-heart hypothesis emerged, and how, eventually, it gained acceptance in the scientific literature.

For example, the MPA documents published before the 1950s primarily consist of reflections on experimental atherosclerosis in animals, particular Anitschkow’s works. The 1950s documents are overwhelmingly reports of feeding trials attempting to ascertain what dietary factors are related to serum cholesterol levels (Bronte-stewart *et al.* 1956; Keys *et al.* 1957c; Ahrens *et al.* 1957).

The 1960s are split between two periods. The first half contains continued work related to feeding trials, with Hegsted *et al.* (1965) and Keys *et al.* (1965) reporting predictive equations related to dietary fat/cholesterol intake and serum cholesterol levels. After 1965, we find intervention studies exploring the use of dietary fat restriction/modification in the prevention of CHD – Dayton *et al.* (1969), Leren (1966), Medical Research Council (1968), and the National Diet–heart Feasibility Study (1968).
The 1970s contains Armstrong et al. (1970) report of the production of atherosclerotic-like lesions in Rhesus monkeys following a high-fat, high-cholesterol diet, but also evidence that these lesions might be partially regressed by the cessation of this diet. This is joined by an influential review by the Inter-Society Commission for Heart Disease Resources recommending population dietary fat restriction (ISCHDR 1970). From 1975 to 1977, all documents report unexpected findings regarding a link between particular densities of lipoproteins and CHD, which caused potential problems of interpretation for the diet–heart hypothesis because studies examining this had relied on measures of total serum cholesterol (TC) and low-density lipoprotein cholesterol (LDL-C). Rhoads et al. (1976), Carew (1976), and Miller et al. (1977) reported findings suggesting a significant association between a low HDL-C level and a heighten risk of CHD.

These results appear to provoke a short-period of controversy, with critical reviews by Mann (1976) and Ahrens (1979) and defensive pieces by the American Heart Association (AHA 1978) and Glueck (1978), which were identified along the B-path. While on the F-path, these results appear to have motivated researchers to re-examine the link between other dietary factors, primarily protein, and lipid profiles in animals (Huff and Carrol 1980; West et al. 1983; Beynen et al. 1984). The G-path contains studies continuing to report findings regarding a significant inverse association between HDL-C and CHD (Enger et al. 1979), and studies presenting findings that suggested that HDL-C levels might be favourably altered by low-cholesterol, low-saturated fat diets (Hulley et al. 1977; Hjermann et al. 1979).

In the 1980s, the B and G paths join once more, and here we find the results of multiple intervention trials (Hjermann et al. 1981; MRFIT 1982; LRC-CPPT 1984), and a set of AHA guidelines (1982) advising the public to continue lowering fat consumption to prevent CHD. By most accounts of the diet–heart hypothesis, the Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT) findings represented proof that lowering serum cholesterol, specifically LDL-C, was causally related to lower CHD risk (see Chapter 1 for discussion of results).
The G-path splits from here, ending with two commentaries written in Afrikaans on serum cholesterol levels. There were identified purely on the basis that they cited many other documents identified by MPA (Serfontein et al. 1984; Rossouw et al. 1984). Neither of these end documents appear to have been cited often, and their inclusion in this analysis appears of limited significance.

The B-path ends with Oliver (1984) and his descendants (Ritchie 1984; Gotto 1984; Lenfant 1984). Oliver’s paper, entitled “Hypercholesterolemia and coronary heart-disease - an answer”, reflects on the implications of the LRC-CPPT (1984) results. Here, Oliver, a regular contributing skeptic to the diet–heart link, concedes that lowering serum cholesterol in those with established hypercholesterolemia is now established as an effective treatment for CHD.

Thus, MPA captures documents that reflect important research focusses in the diet–heart debate. Many of these documents tend to be cited in retrospective accounts by practitioners (Truswell 2010) and sociologists and historians analysing this case study (Garrety 1997, 1998, Olszewski 2015).

Plan for interpretation

However, those who have conducted MPA’s have claimed that such analyses reveal more than this in capturing how research develops through key documents and their relationships to one another. If MPA captures this, then reading each document in sequence ought to give a good understanding of how a field progressed.

To perform the following, I read each document detected in the sequence in which they were published, after reading each document I wrote a short summary of their content, and I generated an interpretive account. While this is my interpretation of the content of each document, I have tried to detach myself from these summaries as far as possible, summarising works by recourse to the major findings presented in those documents (often repeated in introductions and conclusions) and, as far as possible, I try to capture the arguments of authors in either their own words or short summaries of their positions.
As the results of the key-route analysis are similar to those from the G, F, and B analyses, I restrict this to the latter. I follow the MPA results only until 1961, when the AHA published its first dietary guidelines recommending a reduction in fat content and altering the kinds of fats eaten.

4. On the emergence of the diet–heart hypothesis

All paths identified share a common origin, a chain of 16 documents from Anitschkow’s (1913a) ancestors to Dock (1946). Anitschkow (1913a) fed rabbits cholesterol; after 4-8 weeks atherosclerotic-like lesions were discovered, with a similar morphology to human atherosclerosis. Before these studies, it was widely believed that atherosclerosis was the result of senescence. Anitschkow and Chalatow (1913b) summarised previous studies to explain why they decided to feed cholesterol to rabbits. Five years earlier, Ignatowski (1908) had fed rabbits animal protein and reported that they developed atherosclerotic-like lesions; he thought that this confirmed Metchnikoff’s (1908) hypothesis that high protein diets were toxic. However, Starokadomski (1908) had reported atherosclerotic changes in rabbits fed with egg yolk mixed with cow’s milk. Saltykow (1910) conjectured that the milk was the cause, but Stuckey (1910) found that atherosclerosis could be produced in rabbits using only egg yolk. Accordingly, Stuckey (1912) proposed that lipid, not protein was responsible, and tested a range of fats. The results were mostly negative, but one group of rabbits developed atherosclerosis after eating bovine brain. Chalatow (1912) reported that this provoked the ‘infiltration’ into the liver of a fatty substance that contained large quantities of cholesterol.

For Anitschkow and Chalatow ([1913b]1983), these results “…could be explained if the substances that produce the experimental ‘nutritional atheronecrosis’, occur in both egg yolk and brain” (p.179). As cholesterol is an essential constituent of the membranes of all animal cells, they conjectured that the cholesterol accumulating in the arterial walls was the result of excessive ingestion. This made sense in light of findings that atherosclerotic lesions contain large quantities of cholesterol (Windaus 1910), that cholesterol
in the blood of rabbits could be raised by feeding cholesterol (Pribram 1906), and that cholesterol crystals formed in tissues when large quantities were introduced into the blood (Kawamura 1911). Anitschkow and Chalatow (1913[1983]) declared:

[W]e have now reconciled all the investigations of the authors named above. It became completely clear why only such foods as egg yolks or brain, for example, evoked pronounced and characteristic changes in the organism. Since these same processes are also observed on feeding pure cholesterol, there is no longer any doubt that it is specifically this substance that is deposited in the organism in the form of liquid spherical crystals and exercises an extraordinarily harmful effect on different organs (p.181).

They proposed the *lipid hypothesis*; that hypercholesterolemia is a cause of atherosclerosis, and postulated that a high cholesterol diet might cause atherosclerosis: “[T]hese substances [cholesterol]...elicit the expression of pathological processes that are perhaps analogous to human pathology” (p.182).

Anitschkow’s (1933) chapter in Cowdry’s *Arteriosclerosis: A Survey of the Problem*, is his first work in English: his early experimental work was published in German and Russian, and many of these documents appear not to have been indexed. He began by critiquing other animal models:

Feeding with cholesterin is the only method which makes it possible for us to produce in certain species...changes that may be regarded as *equivalent* to those typical of human atherosclerosis (p.317).

This led him to a bold conjecture – “atherosclerosis *never* develops without cholesterin” (p.308). However:

[I]t would be entirely wrong if, on the basis of these conclusions, we were to describe cholesterin [cholesterol] or rather hypercholesterinaemia as “the cause” of atherosclerosis. But the cholesterin plays an important part of the process...has now been definitely established as a fact (p.308)

Anitschkow now introduced his *infiltration theory*: dietary cholesterol is absorbed into the blood and infiltrates the endothelium of arteries, injuring them; these injuries become sites where more cholesterol is deposited, leading to atherosclerosis. Anitschkow thought that serum cholesterol played a
necessary but not a sufficient role. He thought it likely that other causes had to be present to cause lipid infiltration, particularly high blood pressure; diet was probably unimportant in humans with a normal metabolism, which appeared able than herbivores to regulate serum cholesterol in response to dietary intake. Nevertheless, “experimental investigations based on [the rabbit model]...have opened the way to a systematic analysis of the pathogenesis and aetiology of the disease. Undoubtedly the results thus obtained will also provide valuable indications in respect to prophylactic and therapeutic measures” (p.317).

From here, documents on the main path develop Anitschkow’s idea, and propose alternatives. Leary (1934) replicated Anitschkow’s experiments and claimed that he had reproduced ‘human atherosclerosis’ in the rabbit. But he also drew attention to the importance of atherosclerosis in explaining the rising number of coronary deaths in the USA. Finally, he drew attention to the fact that humans are the only animal to consume milk and eggs throughout their lifetimes, and to the high prevalence of atherosclerosis in diabetics treated with high-fat, high cholesterol diets.

Duff (1935) highlighted differences between the atherosclerotic lesions in cholesterol-fed rabbits and in humans. For Duff, the rabbit probably was not a valid analogue; humans had consumed meat, milk, and eggs in high quantities before the rise in atherosclerotic heart disease. He pointed to the lack of evidence of a relationship between hypercholesterolemia and atherosclerosis: “the cholesterol feeding experiments provide no valid reason for believing that a disturbance in cholesterol or lipoid metabolism plays any parts...in human atherosclerosis”.

Following these criticisms, Leary (1941) argued that cholesterol was esterified in the liver and, from there, taken up by stellate macrophages whose role is to remove particulate matter from the blood. These were released from the liver and travelled to the lungs, from where they entered the circulation. Leary found evidence that these cells clung to the arterial walls, and invaded the endothelium. There, he conjectured, they spurred an inflammatory reaction that destroyed cells, leaving behind cholesterol. While this was a different
theory to Anitschkow’s (1933), Leary held that the consumption of cholesterol was the initiating factor. While Dauber and Katz (1942) showed that atherosclerosis-like lesions could be produced in chicks fed cholesterol. Previously, critics had quoted the failure to provoke atherosclerosis in omnivores as a reason to discount the cholesterol hypothesis.

Hirsch and Winehouse (1943) noted that dietary cholesterol provokes an extraordinary rise in serum cholesterol in the rabbit, but had less effect in in people. They pointed to evidence in rabbits that dietary fat might increase cholesterol absorption. However, in people, “there should be a closer relationship between the disease and the blood lipids than has been demonstrated… hyperlipemia may favor the development of atherosclerosis, factors in the tissues are concerned with the actual lipid deposition” (p.199). They proposed that degeneration associated with aging was responsible, but echoed Leary (1934) that dietary and blood cholesterol were important. They concluded, “atherosclerosis develops from disturbances in the metabolism of lipids infiltrated into the tissues of the intima” (p.200).

Hueper’s five-part review (Hueper 1944/45) surveyed more than 1,000 studies. He believed that evidence better supported the idea that hypertension played an important role by altering the permeability of arterial walls, but also proposed that atherosclerosis may be provoked by hypoxia (anoxemia theory of atherosclerosis).

All documents so far were written by pathologists. From here, we observe a transition into the clinical community. Dock (1946), a prominent American physician, highlighted the relationship between atherosclerosis and coronary deaths in the USA. He highlighted a report (French and Dock 1944) of atherosclerosis in 100 soldiers who had died from coronary occlusion. These were under forty years old, and the idea that senescence was the cause appeared untenable. He noted first, that men were far more likely to suffer a coronary event than women. Second, patients with congenital xanthomatosis, a condition in which serum cholesterol is abnormally high, had a much higher rate of CHD than the general population. Third, raised blood pressure is associated with atherosclerosis, and MI more common in hypertensives.
However, women, despite typically having higher values of both serum cholesterol and blood pressure, suffered less from atherosclerosis and CHD. Both atherosclerosis and CHD also had a strong familial component: those who had died from MI tended to have thicker intima, and family members appeared to share this trait. Finally, he considered diet:

[D]iets high in cholesterol, such as the American service men had while in this country, may hasten the process and lead to death decades earlier than if the individual had been on a diet poor in cholesterol but rich in whole grain, legumes and vegetable oils. In New York one is struck by the rarity of coronary disease in those on the southern Italian diet, except when very old, obese or diabetic. Even more striking are the histories of diets excessively rich in eggs, sour cream, butter and ice cream or of milk intake up to 1,500 cc. daily in cases of myocardial infarction occurring before 50 in men who are neither diabetic nor hypertensive (p.878).

Clinical Interest

From here, the paths split. On the G-path only are two documents discussing the results of animal experiments. In 1949, Gubner and Ungerleider (1949) reviewed theories of the pathogenesis of atherosclerosis:

[I]t is not possible to define with exactitude the relative importance and mechanisms of operation of the various factors which have been demonstrated to bear a relation to arteriosclerosis. Evident it must be, however, that there is no single cause...rather the evolution of the arteriosclerotic lesion appears to be due to multiple factors (p.61).

However, they proposed that enough was known to begin extensive work on two proposed causes – raised serum cholesterol and blood pressure. As advanced atherosclerosis could not be “cured”, the scientific community should regard the problem of atherosclerosis as “principally one of prevention” (p.78). They echoed Dock in highlighting that atherosclerosis is prevalent “in groups subsisting on a high fat, high cholesterol diet [that] contrasts eloquently with the rarity of arteriosclerosis in groups on a low fat, low cholesterol diet” (p.65), citing Achoff’s (1924) lecture in which he drew attention to the reduction of atherosclerosis following the semi-starvation diets in Germany following the end of World War I.
Gofman et al. (1950), sought to ascertain which lipoproteins were involved in atherosclerosis. Lipoproteins were distinguished by their Svedberg flotation (Sf) flotation rate – low-density lipoproteins (LDL), which can carry large quantities of cholesterol had a higher floatation rate than high-density lipoproteins (HDL). In rabbits fed cholesterol, Gofman et al. reported that a particular band of LDL was elevated, and concluded that these were “intimately involved in the production of atherosclerosis” (p.169). They then examined blood samples from those who had experienced a recent MI, revealing “an almost universal occurrence of cholesterol-bearing molecules of the Sf 10–20 class[LDL]” (p.186).

From here, the F-path and B-path appear to split, but the content of documents is similar, all concerning either controlled-feeding trials or serum cholesterol measurement in humans.

Boas et al. (1948) measured total serum cholesterol (TC) in patients with atherosclerosis and found that most had abnormally high levels. They suggested that these might be lowered by a transition to a “low-fat, low-cholesterol diet” (p.620).

Watkin et al. (1950) and Starke (1950) examined the effect of a diet consisting no cholesterol or animal fat in hypertensive patients. Kempner (1948) had designed this diet to lower blood pressure, but it also appeared to lower TC; Watkin (1950) reported that this diet lowered TC by about 40%. Starke (1950) examined the effect of this diet in hypertensive subjects: TC levels fell rapidly and remained at a lower level as long as the diet was adhered to. He concluded, “...the rice diet causes a decrease in total, free and esterified serum cholesterol in patients with hypertensive vascular disease” (p.498).

Gertler et al. (1950a, 1950b), proposed that TC ought to be considered in light of other lipids. They were seemingly unaware of Gofman’s studies, and here the main-paths split – Gofman’s studies on the G-path, Gertler’s on the B-path. Patients who had suffered a recent MI, had elevated levels of TC, but also an increase in serum phospholipids.

As the medical community began considering the virtues of a low-fat, low-cholesterol diet, Gertler et al. (1950c) examined cholesterol and fat intakes in
men who had recently suffered MI and controls. The groups ate similar quantities of cholesterol, but the diseased group ate more fat: “there is complete independency of the level of serum cholesterol and the amount of cholesterol ingested within the normal dietary variations” (p.703). They concluded, “there is no advantage to be gained from imposing a low cholesterol diet on patients with coronary artery disease” (p.703).

As skepticism over dietary cholesterol grew, Hildreth et al. (1951a) examined whether dietary fat alone might be the causative agent in raising serum cholesterol: “there is now ample evidence that in most individuals the serum cholesterol concentration will fall if the intake of fat is sufficiently restricted” – citing Kempner (1948), Watkin et al. (1950), and Keys et al. (1950).

Keys et al. (1950) was not identified by MPA, despite being cited 81 times in CUT-N1, because it did not reference any other paper identified by MPA. This study took 41 men and lowered dietary cholesterol by 50%. Their TC was unchanged. Keys et al., however, placed one patient with a genetic defect in cholesterol metabolism on a diet devoid of cholesterol and with a “negligible” amount of fat, and this markedly lowered TC. The authors remark this was a successful treatment for hypercholesterolemia, but too difficult to adhere to be of practical use. In this patient they introduced corn oil into the diet and found that this alone promoted a sharp rise in serum cholesterol. “Judging from the serum alone, this patient was accumulating cholesterol at a rate of 700mg daily on a diet devoid of cholesterol and animal fat, but containing moderate amounts of corn oil” (p.80)

Hildreth et al. (1951a) tested whether vegetable fat raised TC using themselves as subjects. They followed a severely fat restricted diet containing little cholesterol until their TC dropped. Then, high quantities of vegetable oils were added to their diets. In two of the three individuals, this provoked a sharp rise in TC, while for one this rise was gradual. Hildreth et al. (1951b) followed this up. They assert that “decreasing the amount of fat in the diet is an effective means of lowering the serum cholesterol concentration”, and that physicians were now prudently prescribing low-fat diets to their patients” (p. 903).
Moses (1952) reported that dietary cholesterol restriction had no impact on TC. Keys et al. (1952) surveying the evidence, stated:

[D]irect evidence on the effect of the diet on human atherosclerosis is very little and is likely to remain unsatisfactory for a long time. But such evidence as there is, plus valid inferences from indirect evidence, suggests that a substantial measure of control of the development of atherosclerosis in man may be achieved by control of the intake of calories and of all kinds of fats, with no special attention to the cholesterol intake (p.117).

They dismiss Gofman’s typology, claiming that it produced results no more meaningful than TC. “There is as yet no reason to suggest that the concentration of G substances [lipoproteins] in the serum is any more closely related to atherosclerosis than is the concentration of total cholesterol” (p.117).

**Escaping the laboratory**


Since the pathological condition cannot be precisely evaluated in life, and is all too seldom verified at death, it is proper to leave the pathologist in command of his own field and to stay within the limits of the facts on which this discussion actually depends: these are facts seen, recorded, and measured by the clinician, the biochemist, and the vital statistician. (Italicised for emphasis, p.118)

This marks a shift from a problem for pathologists to one of public health. Keys distances his research from rabbit experiments, “to extrapolate to man the findings from cholesterol experiments with rabbits…can lead to absurdities” (p.125). He argues that contemporary studies on atherosclerosis are not interested in this condition in itself, but in its consequences – CHD. Basic research had linked serum cholesterol and atherosclerosis, providing a promising possibility that CHD might be treatable by lowering TC.

Yet, for Keys, current methods were unsuitable for establishing if (a) diet was associated with TC; (b) whether raised serum cholesterol was related to CHD; and (c) whether diet was related to CHD. For him, these questions
demanded a “newer public health”, built on vital statistics, clinical science, and epidemiological population studies. In 1947, he received funding from the US Public Health Service for a prospective cohort study of Minnesota businessmen to examine whether serum cholesterol was associated with CHD. That year, the US Public Health Service also funded the Framingham Study to track serum cholesterol levels and dietary intake, among other variables, to establish causes of CHD. The following year, with estimates that heart disease was responsible for up to half of all deaths in the US, Congress passed the National Heart Act (1948) that established the National Heart Institute (NHI), which took over funding responsibilities.

Yet Keys believed evidence had already emerged that diet was likely the causative factor in CHD. Here, he presents a figure (Fig 55) demonstrating the difference in dietary total fat content and CHD mortality between six countries. He concluded that “dietary fat is somehow associated with cardiac disease mortality, at least in middle age” (p.134).

![Fig 55: Keys’ (1953 p.134) Six Country Graph](image)

He made two bold demands. First, “the facts and relationships indicated here are of such importance as to warrant a large extension of this type of epidemiological research”. Second, “It is difficult to escape the conclusion that public health programs must take cognizance of the information already at hand” (p.137). Thus, Keys backed the dietary-fat hypothesis, but modified it.
Before this, some clinicians advocated particular diets for patients with high TC or a family history of CHD, but, for Keys, population treatment was required.

Keys et al. (1954a) reports results from an excursion to Naples. The study surveyed healthy men and recorded their dietary intakes, TC, and body weight, and compared the findings to a sample from Minnesota, US. Total fat consumption in Naples was about half of the typical Minnesotan diet. Despite this, the groups had equal levels of obesity. However, Minnesotan men had a TC much higher than those in Naples. They interpreted this as supporting a link between dietary fat, raised TC, and CHD. This was followed by a similar study of men in Madrid (Keys et al. 1954b). The authors highlight differences in fat intake between working class and professional men, with professional men consuming more dietary fat, similar to the consumption in Minnesota. In professional men, the TC was similar to the average level in Minnesota, but in working class men, who subsisted on a diet in which about 27% of calories derived from fat, TC levels were similar to those in Naples.

Keys and Anderson (1954c), is a paper presented at the US National Research Council’s ‘Symposium on Atherosclerosis’ drawing on data from Naples, Madrid, London, and Minnesota. The authors argued that dietary fat appeared to be associated with TC, and appeared to explain the prevalence of CHD. The authors also dismiss the link between dietary cholesterol and atherosclerosis, “…there is no evidence that dietary cholesterol, other things being equal, has any influence on atherosclerosis in man” (p.193). Further, they highlight that their approach is progressing in the absence of clear mechanistic understanding:

At the start of this paper it was indicated that we would not attempt to discuss mechanisms here. This involves serious limitations, but these are not necessarily of first importance at the broad practical level. Historically, and until now, mankind’s most conspicuous successes in preventing or controlling major diseases have come about through measures applicable to the population, mainly developed from epidemiological researches, in advance of detailed understanding of pathological mechanisms. This is not to decry the value of fundamental studies on mechanisms, but it does point to the virtue of research on the relationships between factors in the mode of life and the development of disease, even in the absence of proper knowledge about the intervening processes (p.192).
Bronte-Stewart et al. (1955) reported their findings of South African men from three racial groups. Europeans were twice as likely to suffer CHD than Cape Coloured, and three times more likely than Bantu. They reported wide differences in serum cholesterol, measured both by TC and by a method that partially took into account Gofman’s ideas. This established the amount of cholesterol in two broad categories of lipoproteins: (i) α-lipoprotein, cholesterol carried in HDL (HDL-C); (ii) β-lipoprotein, cholesterol carried in LDL (LDL-C). The different profiles appeared to be associated with CHD. The authors argue that the differences stemmed from different dietary fat intakes.

They noted that a rise in the β-lipoprotein with increasing animal fat intake appeared to explain the rise in TC. Finally, they discuss the absence of correlation between vegetable fat and TC, which they argue is a cause for skepticism of reports from metabolic ward studies (Groen 1952) which suggested that vegetable fat might lower TC. “The results of this survey support the hypothesis that the dietary intake of fat influences the level of the serum cholesterol, particularly that in the β-lipoproteins, and in turn may be one of the major factors influencing the pathogenesis of coronary heart-disease” (p.1107).

The Good, the Bad, and the Essential Fats

Bronte-Stewart et al. (1956) looked for an association between dietary fat and TC in a metabolic ward study. “Animal fats and hydrogenated vegetable fat behaved differently from vegetable and marine fish and mammal oils. The most likely common difference between these fats and oils is the proportion of unsaturated and saturated fatty acids in the fats and oils concerned” (p.526).

From this point, the F and B paths split until 1965, merging only via a single shared document (Malmros and Wigand 1957). These capture the beginnings of a disagreement over which fat ought to be altered in the diet, at what level, and whether clinical or population dietary treatment was most appropriate.

Sinclair (1956) articulated a novel hypothesis linking deficiency in polyunsaturated essential fatty acids (EFA) with raised TC and CHD. He
believed that cholesterol was preferentially esterified in the body with unsaturated fatty-acids, but that people in the developed world were deficient in arachidonic acid – a polyunsaturated omega-6 fatty acid. He conjectured that this stemmed from a lack of linoleic acid, typically derived from vegetable oils, and that this was the cause of atherosclerosis and CHD. He proposed that, when deficient in this, cholesterol becomes esterified with SFA synthesised in the body following carbohydrate consumption. This created “abnormal esters” that were “less readily disposed of and so cause atheroma” (p.383). He believed diets high in SFA and cholesterol, by increasing the amount of cholesterol and SFA in the body, increased the amount of EFA needed in the diet to avoid deficiency. Accordingly, cholesterol and SFA ought to be restricted in the diet, and a large increase in the amount of linoleic acid was needed.

In 1957, a wave of small metabolic-ward trials reported on the relationship between dietary fats and serum cholesterol. Keys et al. (1957a) challenged two positions that had arisen. Several trials had reported that a diet high in unsaturated fat might be an effective treatment for hypercholesterolemia because this had a cholesterol depressing effect. Ahrens et al. (1955) argued that the degree of unsaturation, measured by the iodine number of fatty-acids, explained this. Sinclair (1956) argued that corn-oil, with its high concentration of linoleic acid, lowered TC by correcting an EFA deficiency. Keys et al. argues that these hypotheses had received undue enthusiasm. To test them, they compared diets high in corn-oil against diets high in sunflower-seed oil, sardine oil, and butter-fat.

If Sinclair’s EFA hypothesis was correct, sunflower-seed oil should lower TC more than other fats because it has the most linoleic acid. If Ahrens’s unsaturation hypothesis was right, sardine oil should provoke the largest effect. However, the corn oil diet achieved the greatest reduction in both TC and in β-lipoprotein cholesterol. Thus, “The attempt to discover whether the high content of essential fatty acids or the high degree of unsaturation is responsible for the cholesterol-depressant action of corn oil had the surprising
result of showing that corn oil is remarkable in some way not measured by either of these two characteristics" (p.953)

Nevertheless, Keys et al. (1957a) believed that dietary fat restriction ought to be the focus of research. They conclude by alluding to how the apparent virtues of unsaturated fats sit at odds with their previous population studies:

No natural human diets approach the situation of the present experiments where 50-60g. of linoleic acid was ingested daily...or where fish oils accounted for nearly 30% of the total calories...As populations increase the fat content of their diets, they do so by consuming more saturated and unsaturated fats, though the increase in the latter is smaller. It seems, then, that among most of the populations of the world the characteristic serum-cholesterol levels are directly related to the total consumption of fat rather than inversely related to the consumption of essential or highly unsaturated fatty acids. And, if we are to advocate dietary alterations on a population scale to control the serum-cholesterol level, it seems wiser to attempt to reverse the trend towards more and more luxurious high-fat diets than to advise consuming more vegetable and fish oils as a counter measure (p.68).

The year before, Keys had received the largest grant in the history of US medicine for his Seven Countries Study. This was the first cross-country longitudinal study designed to track lifestyle factors and CHD in different populations. A study whose conclusions, by the nature of the experimental design, would return results applicable only to population dietary treatment – yet it was not set to return results until the early 1970s.

Beveridge et al. (1957), in another small metabolic ward trial, found that corn-oil might lower TC for a different reason. They placed men on diets composed of large amount of corn-oil with either α-tocopherol or β-sitosterol removed from the oil. While the diet retaining α-tocopherol had no effect, the β-sitosterol diet lowered TC. They conjectured that the amount of dietary sterols may be important in explaining the effects of some unsaturated fats.

As opinion diverged over which dietary fat was important and the mechanism linking these to CHD, Ahrens et al. (1957) published a metabolic-ward study that added further complexity. They began by highlighting that the idea that TC is raised by the double-bond structure of particular fats: “none of
the experiments known to us has proven [this] conclusively”, yet “there has been widespread acceptance of this hypothesis by workers in human nutrition, and broad applications of the hypothesis have already been made...we think these extensions premature” (p.943). They demonstrate that a diet high in corn-oil was particularly effective in lowering TC. The authors also drew attention to a relationship between the degree of unsaturation of a fat, and its ability to lower TC. They proposed that it was the total amount of unsaturated fat that was important, similar to Sinclair’s recommendation, though an entirely different theory. However, the authors call for caution: “recommendations for radical changes in food habits, even by those populations most seriously threatened by atherosclerosis, should await a clearer definition of the specific food factors which control serum-lipid levels” (p. 953). For them, there were still unanswered questions:

If unsaturated acids cause lower lipid levels, is this due simply to the number of double bonds per unit weight of methylene groups, as our data suggest at present, or is the effect caused by certain unsaturated fatty acids and not by others? Do the monoenees have half the potency of the dienes? Are trienes more effective than dienes? Are conjugated double bonds as effective as unconjugated? Do saturated acids, per se, cause higher serum-lipid levels, and if so, do all chain lengths have the same effect? What is the effect of ingestion of the fatty acid isomers produced by hydrogenation processes? Assuming that the net effect of ingestion of a mixture of various fatty acids can be defined at one caloric level, what changes will be produced when the proportions of fat, carbohydrate, and protein are varied? (p.953)

They point out that other lipid measurements might also be important, but these were, typically, not measured. Further, they express concern that, while group averages of TC could be demonstrated to be lowered in particular groups, these groups tended to be examined because they had abnormally high TC, and that, in these groups, individuals appeared to vary in their response to changes in dietary fat.

This was followed by Keys et al. (1957b, 1957c) trial on schizophrenic patients from Hastings State Hospital. These had been placed on different diets that differed in either the proportion of calories derived from fat or the quality of dietary fat. SFA appeared to raise TC, PUFA lowered it, and
monounsaturated fats (MUFA) had no effect. The following equation “predicted” the change in TC in reference to dietary fat:

\[
\Delta TC = 2.74 \Delta SFA - 1.31 \Delta PUFA
\]

where the change (\(\Delta\)) in TC is expressed in mg/100ml, while changes in SFA and PUFA are expressed as changes in the % of total calories.

They concluded, “These experiments and their analysis offer no support for the suggestion that a deficiency of essential fatty acids produces the high serum-cholesterol levels characteristic of populations subsisting on luxurious American and Western European diets” (1957c p.966). As TC was lowered much more effectively by reducing SFA compared to the small benefit from PUFA, restricting SFA was a more practical treatment. This document is typically cited as the first instance of the ‘diet–heart hypothesis’.

Malmros and Wigand (1957) in a metabolic-ward study, emphasised that the terms ‘animal and vegetable fats’ were misleading because different fats had different effects on TC. For example, while milk-fat raised TC, whale fat lowered it. They highlighted that hydrogenation of unsaturated fats – a process that turned oils into solids at room temperature – removed their ability to lower TC. The authors argued that a diet high in unsaturated fatty-acids is an effective treatment for hypercholesterolemia, particularly oils high in linoleic acid.

From here, the main paths diverge, merging back in 1965. On the B-path, we find studies by Keys et al. that report findings on subjects at the Hastings State Hospital. These include a report that MUFA provoked a rise in TC in patients on a low-fat diet (Keys et al. 1958); results they use to criticise the unsaturation-hypothesis of Ahrens. Keys et al. (1959) refine their predictive equation by considering the initial TC, as “Men who are intrinsically hypercholeremic exhibit greater cholesterol response to dietary changes, while hypocholesteremic men are less responsive than reference men”. They assert that, generally, SFA “in ordinary food fats raise the serum cholesterol level, those of polyunsaturated fatty acids (mainly linoleic acid) lower it” (p.212). They claim that a rise in TC is always mirrored by a rise in the amount of cholesterol carried in \(\beta\)-lipoproteins, while “…no dietary manipulation is known
to affect the level of the alpha fraction [HDL]” (p.209). Keys et al. (1960) report that when the amount and type of fat in the diet is held constant, certain carbohydrates, specifically milk-sugars and sucrose, have a modest TC-raising effect compared to carbohydrates derived from leafy greens, fruits, and legumes. Anderson et al. (1961) report that hydrogenation of vegetable oils has a significant effect on lipid levels and advise against hydrogenation. Grande et al. (1961) suggest that the chain-length of dietary SFAs might have an important effect on serum cholesterol. Keys’ papers were beginning to touch on complexities – was SFA dangerous, or only fatty-acids of particular chain lengths? Was hydrogenation of vegetable oils safe, and if not, why not?

The alternative path begins with Kinsell et al. (1958), a paper defending the importance of linoleic acid. The authors survey evidence that dietary linoleic acid increases the unsaturated fatty-acid content cholesterol esters in an apparently favourable manner, and lowers TC in those on low-fat, low-cholesterol diets. They raise concerns that a low-fat diet is “biologically unsound” because “a low-fat diet inevitably is a high carbohydrate diet… ingested carbohydrate is probably transformed to fatty acids in the fat depots and eventually mobilised and metabolised through fatty-acid pathways” (p.339).

Lewis et al. (1958a) report that two patients fed and injected with unsaturated fats increased cholic acid secretion, which appeared to precede a fall in serum cholesterol, and suggest that unsaturated fats lower TC by accelerating the catabolism of cholesterol. Lewis et al. (1958b) discuss findings that the fatty-acids of cholesterol esters of those with atherosclerosis had more saturated-fatty acids and oleic acid (a MUFA) than healthy individuals. Further, cholesterol esters in atheromatous plaques were like those found in the blood of people with atherosclerosis. They interpret these findings as supporting Sinclair’s (1956) theory.

Bronte-Stewart (1958) reviewed the different theories:

[A]lthough the level of serum cholesterol and the nature of the dietary fat intake are closely related, the hypothesis that ischaemic heart disease results from a disturbance in fat metabolism is founded more on assumption than on fact. Undoubtedly other factors are important in its aetiology…Nevertheless, of all the
environmental variables constituting our present-day mode of life, no other at present known can be so closely related to pathological, metabolic and epidemiological features of the disease as the nature of the dietary fat intake. It should be emphasized again that the evidence for such a relationship is purely circumstantial and that the hypothesis is applied to a disease both difficult to measure in its prevalence and to predict in its mode of progression (p.251).

For Bronte-Stewart, while the dietary fat link was based more on assumption and ‘circumstantial’ evidence than on “fact”, it was still the best hope for a treatment for CHD because “dietary fat intake is the one most susceptible to measure and control.” (p.251).

Jolliffe and Archer (1959a) analysed vital statistics and food consumption statistics. For 14 countries, they recorded the number of deaths in men aged 55-59 from arteriosclerotic heart disease. All dietary variables were somewhat correlated with death rates, but the amount of dietary SFA could explain 83% of the variance in deaths.

Jolliffe et al. (1959b) reported results from the first year of dietary trial in 79 men aged between 50-59. Named the Anti-Coronary Club, this trial had been funded following an executive order of the Commissioner of Health in the US. While enrolling more than 600 men, they report on only a small sample to demonstrate that the diet was effective at lowering serum cholesterol. The ‘Prudent Diet’ was composed of 2000-2700 calories/day in which total fat did not exceed 33% of calories, and SFA was restricted to 8-9%. This was a diet to be eaten in the course of everyday lives; food sheets detailing the nutrient profiles of food were provided, as were specific meal plans. The men were split these into groups based on TC tertiles. The diet lowered TC regardless of the pre-treatment level, but was more effective in those with high TC.

The Return of Dietary Cholesterol

Stamler (1960) reviewed the evidence related to diet in the preventing heart disease. For Stamler, Anitschkow’s model had been replicated in many species, and hypercholesterolemia had been demonstrated to be associated with CHD in humans following reports that the Framingham study had found a six-fold increase in the rate of CHD in men with hypercholesterolemia. Finally,
diet had been “unequivocally” shown, both in metabolic-ward settings and in epidemiological studies, to be related to TC. Stamler thus asserts “diet is a key factor in hypercholesterolemia and atherosclerosis” (p.58). Stamler, however, attacked recent literature that had discounted dietary cholesterol as an important factor. For him, this was hard to reconcile with animal experiments, in which the inclusion of dietary cholesterol was a necessary element, and he criticises the Anti-Coronary Club for not bothering to report the dietary cholesterol supplied in their diet.

Stamler discussed his Coronary Prevention Evaluation Programme (CPEP) – a community-based prevention trial of men initiated in 1955 and funded by the NHI. This included a diet low in SFA and cholesterol, but also recommendations to reduce energy intake to 1,500-2,100/day and reduce sodium intake to lower the rates of obesity and hypertension. This diet lowered TC, blood pressure, and obesity. For Stamler, the evidence pointed to the need for a similar diet to be recommended to the US population; but systematic primary and secondary prevention trials had not taken place, though some had received funding. Regardless:

There is no reason to believe that a policy of ‘watchful waiting’ – pending the completion of research studies…over the next decade will in any way alter this grim picture. On the other hand, the measures proposed for long-term prevention [SFA & cholesterol restriction and hypertension/weight control] …are safe, moderate, sound, and free of danger…Their widespread prophylactic utilization by clinical medicine and public health would therefore seem to be the order of the day (pp.84-85)

This was followed by Berkson et al. (1960) and Stamler et al. (1961) reporting on a study of men at a Chicago utility company. Black employees were more likely to suffer from hypertension than white employees, but less likely to have hypercholesterolemia. Despite this, white and black men suffered CHD at the same rate. The authors suggest that both TC and blood pressure should be targeted in intervention programmes.

American Heart Association’s Dietary Guidelines

In 1961, the AHA published the first dietary guidelines, drafted by leading authorities of the field, aimed to combat the rise of CHD mortality in the US
While not on the main paths, it provides an interesting perspective on the literature. One might have followed the discussion with a sense that the MPA analysis had produced a seemingly random-walk through very different theories; theories only related in their shared belief that lowering TC was, probably, important. This is not an artefact of the MPA method, however, as the dietary guidelines demonstrate.

The AHA’s statement begins by asserting that heart disease, is “caused by the same condition…the problem of preventing or retarding these diseases is, then, one of preventing or retarding atherosclerosis”. It is “known that a number of factors influence or are related to the development of atherosclerosis…among these factors are a high content in the blood of a type of fat called cholesterol. elevation of blood pressure above normal, presence of diabetes, obesity, and a habit of excessive cigarette smoking. Age, sex and heredity are also important” (p.133). However, their articles title, ‘Dietary Fat and Its Relation to Heart Attacks and Strokes’, establishes its focus. The AHA provide a short history of the research:

Many years ago a scientist fed cholesterol and other types of fat to rabbits. The blood cholesterol content increased and the rabbits developed atherosclerosis; that is, cholesterol and other fatty substances were deposited in the walls of the arteries…These animal experiments indicate that diet may be an important cause of atherosclerosis…Global studies have shown that dietary habits of human populations differ. Evidence gathered from many countries suggests a relationship between the amount and type of fat consumed, the amount of cholesterol in the blood and the reported incidence of coronary artery disease. (p.133)

They highlight that the average US diet contains between 40-45% of calories from fat, much more than other countries, and this appears to be related to MI. They note that “other differences in these diets that may also be of importance, such as the amount and type of protein and carbohydrate”, but do not explore these further.

For the AHA, studies suggested that altering fat intake might lower TC and thereby lower CHD risk, but also altering fat intake promotes weight control that also lowers the risk of CHD. Despite the advice that follows, the AHA
caution that “there is as yet no final proof that heart attacks or strokes will be prevented by such measure” (p.133).

The AHA proposed dietary recommendations for men who have (i) a family history of CHD; (ii) an elevated TC; (iii) high blood pressure; (iv) overweight; (v) lead sedentary lives of relentless frustration”. For these, fat intake should be restricted to 25-35% of calories, and SFA should be ‘substantially replaced’ by PUFA.

After this, the statement highlights that “moderate amounts of fat, particularly those containing an appreciable quantity of the poly-unsaturated type, are necessary for good health” (p.134). Finally, it alludes to the idea that reducing dietary cholesterol is one way of lowering TC, but “if the amount of cholesterol in the diet is markedly decreased, but the caloric intake kept constant, the body may make more cholesterol from other substances, chiefly from other types of fat, sometimes nearly enough to make up for that which has been removed from the diet” (p.134). Thus, the best way of lowering TC is to lower dietary cholesterol, total calories, and SFA, while increasing that of PUFA.

The AHA appear to have taken every positive finding reported over the previous decade, without recognising that they often conflicted with each another. They relate Anitschkow’s findings to the exploratory epidemiological studies of Keys. But Anitschkow did not show that dietary fat led to atherosclerosis; his experiments ruled out any other substance than cholesterol. This is opposite to dietary fat theories of the 1950s that proposed that fat intake led to CHD and that cholesterol intake was unimportant.

The AHA reference Ahrens and Sinclair who held different theories, united only in belief that there may be some advantage in altering dietary fat. Advocates of both theories believed that a diet high in unsaturated fatty-acids in general or linoleic acid in particular, ought to guide treatment. However, Ahrens opposed guidelines because of the gaps in understanding of the mechanisms involved in the regulation of TC and the problem of individual variability. Sinclair and Kinsell believed a low-fat diet was 'biologically
unsound', as these high carbohydrate diets would increase synthesis of SFA in the body.

Before this, despite sitting on the AHA board, Keys had asserted that dietary cholesterol was not related to TC at any moderate intake. His 'predictive equation' did not include any parameter in regards to dietary cholesterol. For him, a sharp reduction in SFA and total fat was the most promising avenue. The AHA's dietary advice however, resembles the diets developed by Jolliffe and Stamler; diets constructed for population treatment, rather than to test any basic theory.

The AHA statement references only eight primary research articles reporting dietary relevant results. They reference Ahrens et al. (1957), Joliffee et al. (1959), Keys et al. (1957c), and Kinsell et al. (1959) who advocated different diets from different theoretical perspectives. The first three of these have been discussed already.

Kinsell et al. (1959) is another metabolic ward study of linoleic acids in support of the EFA hypothesis:

Linoleic acid appears to be the constituent of these fats which is responsible for such effects [lowering serum cholesterol]. Generally speaking, the greater the amount of saturated fat in the diet, the greater the amount of linoleic acid required to produce a desirable lowering of plasma lipids (p.185).

For them, given enough linoleic acid in the diet, the fat content was unimportant. The AHA also cite McCann et al. (1959), in which 1000 calories of peanut oil (high in MUFA but also in linoleic acid) was added to the diets of 20 subjects, lowering TC by 10-16%. They reference Pilkington et al. (1960) results of feeding a low-fat diet and a high unsaturated fat diet to out-patients following an MI, both reduced TC by ~26%. Brown and Page (1958) study is also referenced which reported lowering TC with a high unsaturated fat diet and a low-fat diet.

The AHA also reference a review of studies related to the relationship between diet and TC (Portman and Stare 1958). This review is damning of the state of nutrition research. The authors stress that, whereas dietary cholesterol
provoked increases in TC in all mammalian species studied, dietary fat has a particularly unpredictable effect.

It could be expected that since there is no general agreement about the importance of the various characteristics of natural fats that give them serum cholesterol-regulating activity, there would be no agreement about the mechanisms involved. When a particular fat is fed it is entirely possible that several cholesterol-potentiating and -depressing factors may be working synergistically or competitively (p.420).

The authors arrive at a conclusion directly contrary to the AHA’s advice:

*First*, it may become more profitable in the future to define effects in terms of the interaction of various dietary factors and the biological meaning of such interaction; and *second*, it is probably unwise to place too much emphasis on the effect of a single factor in the control of the serum cholesterol concentration unless that factor is evaluated under a wide range of conditions. It is still too early to predict whether the interaction of dietary factors in the regulation of serum cholesterol levels will prove to be of primary importance in humans. (p. 435).

The AHA’s advice was issued in the *absence* of any mechanistic theory of how diet affected TC, nor did there appear to be much clarity in regards to a consensus view on the pathogenesis of atherosclerosis.

Before the publication of this advice, no RCT or prospective cohort study had published any relevant findings about whether (i) diet was related to TC; or whether (ii) diet was related to CHD. Major trials had been launched to test these, but were still to return their findings. On the AHA committee sat Stamler and Keys who had launched two such studies – the Seven Countries Study and the Chicago study. Both appeared to believe that population-dietary treatment was required to combat CHD. By contrast, Ahrens believed in focussing on treating patients with hypercholesterolemia.

### 5. Discussion

This account, generated from documents detected via MPA, traces how an idea travelled through the literature, how it evolved, and its apparent acceptance.
The most influential citation relationship by SPC weight was that between Dock (1946) and Hueper (1944/45); a citation link involved in 28% of s-t paths in the CUT-N1 network. Dock (1946) represents the beginning of interest in the translational importance of Anitschkow’s theory, with interest moving from dietary cholesterol to the quantity and quality of dietary fat.

From here, we saw a progressive shift from interest in basic mechanisms of atherosclerosis to dietary treatment for CHD either in individuals with hypercholesterolemia or population treatment. The positions developed in the absence of a definitive account of the pathogenesis of either atherosclerosis or CHD or of the basic mechanisms of the regulation of TC. Rather, after 1950, interest centred on developing a treatment for CHD via the assumption that lowering TC was beneficial.

This shift was facilitated by reviews which argued that particular factors merited research by being the only practical avenues for a potential treatment. Gubner and Ungerleider (1949) highlighted that, for the pathology community, there were still questions over how cholesterol entered the arterial wall, whether endothelial damage was a necessary prerequisite, whether age was implicated, and whether blood pressure or raised cholesterol alone could be responsible. In humans, it remained an open question whether diet influenced TC in those with normal metabolism and whether TC was associated with CHD. But Gubner and Ungerleider (1949) concluded by recommending that scientists focus on TC and blood pressure as these were the only manageable factors that had arisen from basic research which might lead to a treatment for CHD/atherosclerosis. Keys (1953; 1954c) argued that research ought to progress in the absence of basic understanding, Bronte-Stewart (1958) recommended continued focus on dietary fat to control TC as this represented the best hope for a treatment, while Stamler (1960) suggested that both clinical and population dietary intervention ought to be initiated, and that this strategy should be preferred over a policy of ‘watchful waiting’ for definitive evidence.

From the 1950s, this led researchers to study the effect of different fats on TC. These used TC as an indicator of the harmful effects of foods, and appeared willing to advocate diets on the basis of their effect on TC in the
absence of evidence from trials initiated to test the links with CHD. Here, we saw sudden shifts in interest, from dietary cholesterol restriction being the most promising lead, to dietary fat, then different types of fat, and finally the return of dietary cholesterol as potentially harmful. By the end of the period, most scientists appeared to be advocating restricting at least one of these substances, but differed about which substance.

By the mid-1950s, a split between those advocating treatment strategies was apparent. Keys, Jolliffe and Stamler championed population treatment, while Ahrens and Kinsell supported treatment for those with established hypercholesterolemia. This translated into ‘parallel developments’ in the MPA – Keys along the B-path, Ahrens and Kinsell along the G/F pathways. Towards the end of the period, we see Stamler and Jolliffe on the G/F pathways and these works adopted wide-ranging dietary treatments in population studies designed for treatment, rather than to test hypotheses. By dominating the B-path over this period, Keys’ works would be more likely to be encountered when traversing the network backwards. Keys’ works became more popular over time, while works on the F-path attracted more initial interest.

Many of the studies identified by MPA are discussed in sociohistorical and retrospective pieces, such as Garrety’s (1997, 1998) analysis of the cholesterol controversy until 1984; Steinberg’s (2004) history of research into the pathogenesis of atherosclerosis; Truswell’s (2010) book describing the development of research on diet and CHD from 1900 to 2000. Each assigns Anitschkow’s work special significance, but, all highlight that it was not until the 1950s that his ideas were adopted by those interested in CHD. Garrety and Steinberg allude to a discontinuous jump to clinical interest, but Truswell suggests there was earlier interest in measuring TC in people with CHD, citing Davis et al. (1937) study of serum cholesterol levels in patients with angina. For Truswell, early studies had little impact because different cholesterol measurement techniques produced divergent results.

While not in my MPA, the CUT-N1 network contains a number of similar early studies, beginning with with Bacmeister and Henes (1913), that measured cholesterol levels in patients with atherosclerosis (Gorham and
Further, there are a number of studies examining how diet, specifically dietary fat, might play a role in serum cholesterol regulation in humans (Hunt 1929; Freyberg et al. 1936; Bloor 1932; Man and Peters 1932). Thus, it appears that the link between diet, serum cholesterol, and heart disease attracted clinical interest from the early 20th century. As Gorham and Meyer (1917) remark, “Judging from the extensive literature which has accumulated on cholesterol during the past ten years, this lipoid has been the subject of more varied and extended investigations than any other substance of physiologic importance” (p.599). It also appears, at least from Anitschkow’s (1933) chapter, that the cholesterol-fed animal model attracted a considerable interest in the pathology community; his chapter surveys more than 100 studies on the topic (mostly German, French, and Russian publications). Accordingly, it appears MPA analysis missed early interest in these topics, though this oversight is present in works not using MPA. However, these early studies tended not to be well-cited by the new generation of scientists enthused by diet–heart link in the post-war period, nor were they well-cited in their own day. While the CUT-N1 network contains 579 documents published before 1939, MPA detected the documents that were most cited before 1939 by documents this network, which also tended to remain reasonably highly-cited.

Truswell, Steinberg, and Garrety’s accounts highlight the works of Gofman, Ahrens, Bronte-Stewart, Keys, Stamler, and Jolliffe as influential in the 1950s. However, Steinberg and Garrety give the impression that their findings were mutually supportive. Truswell, however, discusses the debate that arose from the metabolic-ward trials. His account mirrors the MPA in tracing a path through interest in dietary cholesterol, dietary fat, and then particular types of fats over the 1950s. He notes the tension between the competing hypotheses of Keys, Ahrens, and Sinclair, and highlights the complexity in regards to other dietary factors.

While MPA performed poorly in identifying early research interest, the account developed from its results has some strengths. Specifically, it traces how certain assumptions became taken as facts over the course of the 1950s
within highly-cited documents. Stamler (2002), in a letter to Blackburn, provides an interesting perspective on how the diet-heart hypothesis arose.

We [Stamler and Keys] were not the originators of the hypothesis. This, I am convinced, reflects itself in such landmark articles as Ancel’s Mt. Sinai Hospital lecture [Keys 1953], where he displayed relationships between percent calories from fat for different countries as related to their CHD mortality rates. There may not be an explicit formulation of the “hypothesis” there, or in other of his or my multiple publications, the reason being – I am quite sure – that it was superfluous, since we were already standing on the shoulders of those whose initial work made the hypothesis self-evident.

Here, Stamler suggests that the diet–heart hypothesis emerged in such a way as to make it appear ‘self-evident’, and this is what my account appears to reflect. That TC is a cause of CHD and that lowering TC via diet is beneficial began to be accepted as ‘fact’, as evidenced by the willingness of scientists to advise the public to alter their diets. This advice was proposed despite the absence of evidence from the large trials designed to test these conjectures; trials launched by the strongest advocates of dietary treatment – Keys, Stamler, and Jolliffe.

On the Voices Unheard

Some have claimed that MPA could be used in systematic reviews (Colicchia and Strozzi 2012; Xiou et al. 2014). Systematic reviews aim to evaluate all the relevant evidence, but my results demonstrate MPA cannot do this. Not only was early research interest misused, but there is largely an absence of unsupportive studies of the diet–heart hypothesis – only two studies report unsupportive findings – the Medical Research Councils (1968) secondary prevention trial and the MRFIT trial (1982). These are joined by two critical discussions by Mann (1977) and Ahrens (1979).

Why were unsupportive views underrepresented? The supplement of this thesis includes the distribution of citations to all studies, and shows that unsupportive studies generally received few citations. Even discounting methods papers, the unsupportive trials of Rose et al. (1965) and the Research Committee’s (1965) low-fat trial appear in position #369 (26 citations) and #328 (28 citations) respectively. The MRC trial is at #131 (48 citations) – but still
attracts less than half of the single supportive secondary prevention RCT by Leren (1966), at #19 (107 citations).

MPA excluded documents that were more critical, as the account of MPA documents until 1961 demonstrates. While Ahrens, Sinclair, and Keys disagreed on which dietary fat was important and how dietary treatment should proceed, they all held that TC probably played a causal role in CHD and that altering the fat content of the diet represented the best avenue for treatment. However, other studies disputed this, and these were excluded:

Morris (1951) analysed autopsy records from the London Hospital; his findings contradicted the narrative that an increase in atherosclerosis could explain the rise in CHD:

The evidence here does not suggest that there has been any increase of atheroma – any progressive deterioration in the coronary arteries of the population during the past forty years. On the contrary there seems to be less of advanced coronary atheroma now than formerly (p.7).

Yerushalmy and Hilleboe (1957) attacked Keys’ ‘newer public health’ paradigm. For them, the lack of mechanistic understanding rendered his conclusions unsound, and the relationship between dietary fat consumption and CHD in six countries was based on a biased selection of data. Data on another 16 countries was available to Keys, and when included, the association dramatically weakened. A stronger association with CHD could be found for animal protein rather than fat, and a negative association was found if fat or protein was analysed against mortality:

[A]n association does not by itself constitute “proof” of cause-effect relationship. It merely serves either as a guide for further research or as supporting evidence for a hypothesis which has been formulated on the basis of other theoretical or empiric evidence in the sense it “fits” with the formulated etiologic pattern (p. 2351).

Yudkin (1957) reflected on the weakness of the association between fat and CHD, He suggested “there is a better relationship with intake of sugar than with any other nutrient we have examined” (p.157), but the strongest association of any lifestyle factor and CHD in the UK was with the number of television licenses issued per year.
Mann (1957) claimed the ‘large increase’ in CHD deaths in the US was an artefact of death certificate practices and increased awareness of CHD. He challenged the association between dietary fat and CHD, arguing that physical activity was more important, and argued that this might explain why Keys found different rates of CHD in rural and urban communities. Mann (1959) launched a vitriolic attack on scientists who promoted a link between dietary fat and CHD. Mann was a prominent experimental pathologist involved in the Framingham Study. This was not a minor controversy; this was warfare. To the laboratory scientist: “learn how to measure the attributes you use. Learn to live within your intellectual income” (p.102). He criticised Keys for ignoring variability in the cholesterol response to dietary fats, for his selective choice of statistics, for advocating in the popular press, for publishing a diet book that had become a best-seller (Keys and Keys 1959), and for pressuring health organisations to offer similar advice to the public. To scientists: “avoid the sawdust trail in defence of a shaky hypothesis. Problems are solved by experiment—not by debate” (p. 102). To clinicians: “Beware…the academic faddists who, under the guise of science, advise your patients in the press and sometimes advise you on how to prevent and cure CHD. They are selling short” (p.102).

Perhaps the most important omissions are Gofman’s works (Gofman et al. 1956, 1958). These risked undermining most research related to diet, serum cholesterol, and CHD. He argued that the prevailing measures of serum cholesterol “can be a dangerously misleading guide in evaluation of the effect of diet” (1958 p.282). Gofman et al. (1956) showed that several different lipoproteins were associated with MI. For example, lipoprotein S₀ 0–20 increased after eating fat, lipoprotein S₁ 20–400 after eating carbohydrate. S₀ 0–20 was the predominant carrier of cholesterol, so would explain a rise in serum TC, while a high S₁ 20–400 appeared to lower TC. Yet, if either of these were raised there was a greater risk of CHD. Further, Gofman highlighted that lipoproteins appeared to be raised in unique manners from individual to individual. Accordingly, he rejected the population-based approach “Rational management of patients with coronary heart disease or of individuals
attempting to avoid coronary disease depends upon knowledge of the lipoprotein distribution in the individual patient” (1958, p.283).

None of these studies appear on the main paths, but all are in the CUT-N1 network, and all challenged the works identified by MPA. Do the results of the key-route analysis avoid this problem? The network constructed with the addition of 100 initiating routes includes another 32 documents. Yet, the results still contain none of these critical studies.

This seems to reflect how difficult it was for alternative views to be taken seriously. Gofman, reflecting on this period, claimed that his “views were ignored”, but attributes this to the fact that “he was a physicist and not part of the orthodox medical scientific community” (quoted in Reynolds and Tansey, 2006 p.5). Taken together with the findings in Chapter 4 of citation bias favouring the single supportive secondary prevention trial, these results indicate that this area of research may have developed, in part, by failing to reference inconvenient data and critical perspectives.

A complex research ecology

My MPA account indicates that there may be a bias in the literature for ‘positive’ results, but to understand this, one needs to understand the results of particular studies and trace how they were used in the literature. For this, CS-CNA is better, both for evaluating bias, and for tracking how studies depend on previous evidence.

Analysing the full citation network as a directed graph may also hold more promise for historical research. Citation counts can provide an account of how influential particular documents were at particular times and places. Sub-graphs can be constructed that include only documents of interest and the documents that cite them – and graphs that can be analysed for citation bias via the claim-specific approach. Co-authorship networks can be constructed to examine the social structure that underpins these citation networks. Documents can be clustered together by shared methods: as methodological choices were important in the development of this area, this is an important avenue for future research. Finally, understanding why critical documents were neglected requires interaction with the larger network.
However, MPA also had limitations that probably extend to citation analysis in general. It was difficult to explain why research proceeded as it did without recourse to external factors. The most obvious is the massive increase in US funding for CHD research following the 1948 National Heart Act. Fig 3 shows the growth of publications <1962 in CUT-N1.

Fig 56: Distribution of documents per year related to diet–heart research contained in the CUT-N1, 1900–1961.

As MPA had important limitations, and as I came to suspect funding decisions were important in explaining the diet–heart debate, I brought my study to a close.
Chapter 10: On Endings and Beginnings

The Frontiers of a book are never clear-cut: beyond the title, the first lines, the last full stop...it is caught up in a system of references to other books, other texts, other sentences: it is a node within a network...[it] is not simply the object that one holds in one’s hands; and it cannot remain within the little parallelepiped that contains it

And so I come to the end of this journey; a journey that has weaved its course through significant episodes of the history of diet–heart research. Through a progressively widening lens of citation network analysis, I have explored the utility of variants of this methodology for studying how scientists use the available literature to develop and evaluate knowledge claims. From revealing the biased evidence selection of particular reviewers and tracing the various ways findings from a single study can be used, to the mapping of a scientific ‘paradigm’ and the evolution of a hypothesis within it, the techniques developed and refined here open new avenues for researchers to explore.

1. Claim-Specific Citation Network Analysis, Bias, and Diversity of Use and Interpretation

In Chapter 4, by claim-specific citation network analysis (CS-CNA), I produced findings that suggest that our system of scientific communication does not always have the ‘self-correcting’ mechanism often ascribed to it, and that scientists, far from being ‘objective seekers of truth’, are sometimes biased partisans of their favoured hypotheses. In the citations from reviews of the literature to four RCTs examining the efficacy of fat-controlled diets in secondary prevention of coronary heart disease, I documented a bias that overwhelmingly favoured the single supportive trial; a bias that persisted over time despite the small number of relevant trials to cite.

By introducing sub-graph analysis to CS-CNA, I demonstrated that citation bias can be precisely documented in a body of literature, as too can
research utilisation, and this simple method ought to be transferable to other areas of science. However, a number of factors must be taken into account when extending this technique. First, an analyst must either select a case study in which the available trial evidence was published in a narrow time window to control for the decaying effect of time on the citations to particular studies, as my study did, or control for time in some other way. This latter option is difficult; citation rates appear to express a regular half-life, the number of citations to papers will be associated with the size of the literature published over the following years, and trials published many years apart will likely have important differences in trial design and key measures.

Second, to identify citing literature, especially literature published before 1985, an analyst must ensure that the underlying citation data is accurate. As demonstrated throughout this thesis, the citation scores reported by both Web of Science (WoS) and Scopus are not accurate enough to be used without downloading all reference lists and manually cleaning these. To identify the citing literature of a particular article, one should use the ‘Cited Reference’ function in these databases and search for several variants of the spelling of author names and common errors in year of publication, and then manually examine the bibliographies of all papers to ensure citation data is accurate. Failure to do the following risks leading to false conclusions. For example, Leren’s (1966) supportive trial is recorded by WoS as having 12 citations as of 2018, but, as discussed in Chapter 6, when performing a cited-reference search, the total number of citing documents was 416 by 2018. In my study of citation bias, I used the cited-reference search and manually recorded all citation data from identified citing papers, but I will have failed to identify citing papers that only cited Leren’s paper but introduced some error into the author’s name or year of publication. Thus, while the citation data for all identified publications is accurate, I probably undercounted the number of papers citing only Leren’s paper. Indeed, the bias towards this paper was more pronounced in my large cleaned network of diet–heart relevant publications (Chapters 6–9) – with Leren’s (1966) trial attracting more than double the citations of the highest cited unsupportive trial (Medical Research Council 1968), though this
count includes all citations from papers regardless of paper type. This discovery of error in reported citation scores might have consequences for some large studies of citation bias that have used WoS citation counts without cleaning bibliographic data or searching for variants or references (Jannot et al. 2013), although this may be a problem restricted to the period and area examined in this thesis.

**Causes of citation bias**

Whether the bias reported here was produced by scientists knowingly selecting only evidence that supported a preferred conclusion, or was produced through a series of accidents, confirmation bias, the mimicking of the evidence selected in other papers, or some other unknown factor was beyond this study’s scope. Understanding why citation bias and the underutilisation of available trial evidence emerges and persists in the literature will require additional methods. While Greenberg (2009) demonstrated that tracing the papers through which many citation paths flow might establish the papers that spread awareness of particular findings and amplify biases, it is likely that a range of factors will need to be examined beyond this. Scientists are humans, and, as humans, are shaped by their social circumstances, their actions guided by pressures and expectations from within and beyond their own fields, their views influenced by the views of others, by their own preconceptions and the preconceptions embedded in the literature, and their reasoning and memories are subject to an array of seemingly innate flaws and quirks. From individual to individual, these factors will manifest in different behaviours and views. Accordingly, citation bias is likely a complex and multi-causal phenomenon that will require analysis by sociologists and psychologists amongst others.

**Beyond bias**

CS-CNA has a broader utility beyond studying citation bias. In Chapter 5, by tracing how findings from a single cohort study were used in the citing literature, I demonstrated that CS-CNA can track both what specific findings are cited from a paper and also the presence of different interpretations of those findings. I found that predominately only single findings were picked up
by citing authors; few reflected on or evaluated the totality of findings reported or the strengths and weaknesses of study design. Findings appeared to be cited on the basis that they held apparent utility for supporting, contextualising, or attacking particular hypotheses.

Specifically, the first prospective study to report findings regarding the relationship between dietary fat and CHD was rarely cited for this finding in the period studied. The absence in this study of an association between dietary fat and CHD was largely ignored by advocates of the diet–heart hypothesis, but it was cited regularly by its critics. When advocates did cite the dietary finding they interpreted it in such a way as to support the hypothesis, while critics saw an apparent refutation. This suggests that papers may be cited insofar as they are useful to the one who cites them, not for their intrinsic merit nor according to the intent of the cited authors – so when there is flexibility of interpretation or emphasis, this might be exploited by the citer. The dietary fat findings, however, were not the most used findings among citing documents, rather the statistically significant associations between CHD and serum cholesterol, blood pressure, and coffee consumption were cited most frequently. These results suggest that scientists were more likely to reference statistically significant findings than non-significant findings.

CS-CNA could be used by those interested in studying the different interpretations of evidence that scientists can come to, and where those interpretations reside and spread in the literature. There is a rich body of scholarship in the sociology of scientific knowledge (SSK), particularly the works of Collins (1981), which have seen importance in understanding differences in interpretation during times of scientific controversy. Further, work on the use of rhetoric and persuasion in scientific writing might be extended by the use of CS-CNA (Gilbert 1977; Latour and Woolgar 1979). If scientific knowledge comes to be widely endorsed and its controversies settled, in part, because certain studies and interpretations become widely known while others fade into obscurity, then this ought to be revealed in a CS-CNA analysis that tracks the use and interpretations of particular studies over time.
CS-CNA, SSK, and Scientometrics

My findings, taken against the backdrop of findings from other empirical studies regarding the prevalence and implications of citation bias and other citation distortions (Greenberg 2009; Fanelli 2013; Leung et al. 2017; Druyx et al. 2017), indicate that the study of how scientists use and interpret published evidence is an important area for continued research. CS-CNA enables a researcher to do this systematically, and the variants explored and refined here ought to be capable of extension to new cases. Of course, CS-CNA can play only a part in understanding how and why a particular study, or a specific interpretation of a study, becomes widely used and endorsed, but it is an important tool for analysing these processes in the literature.

Moving forward, this area would benefit from the involvement of sociologists of science, whose insights should help to contextualise such findings and explore their implications. While Scientometrics has produced an impressive body of methods for analysing citation data and developed understanding of publication and citation dynamics, and while Meta-science continues to find new ways of detecting and evaluating apparent problems in the scientific literature, SSK and the broader field of STS currently appear uninterested in this area. This is unfortunate, particularly because perspectives on citation behaviour by sociologists (e.g. Gilbert 1977; Latour and Woolgar 1979; MacRoberts and MacRoberts 2018) appear to be well corroborated by recent findings reported in the literature. As the integrity of the published evidence is currently of considerable concern to scientists, SSK’s perspective and contribution to this area could be vital.

These results appear to have serious implications for the use of citation metrics in research evaluation. First, there are good reasons to believe that citations tend to be directed at studies reporting ‘positive outcomes’, as was the case in this study. If citations tend to be directed at ‘positive’ studies, then the use of evaluative metrics based on citation data becomes suspect, as too its underpinning normative theory of citation. Are citation counts perhaps better indicators of the hopes and expectations of a scientific community than the ‘objective’ worth of papers (e.g. methodological rigour, accuracy in critique,
reporting fundamental novelties)? If scientists tend to favour referencing positive outcomes, what consequences does the use of citation data in research evaluation have on the behaviour of scientists (e.g. to publish negative results and critical pieces)? Here, I have reported results indicating that a tendency to reference supportive studies can be found in the literature as far back as the 1960s, long before citation metrics became a feature of scientific life, but this finding, obviously, shouldn’t be used to defend the use of such metrics.

Second, if papers are cited not only for their inherent worth, then the meaning of citation counts becomes problematic (Edge 1979; MacRoberts and MacRoberts 1989). Are highly-cited papers highly-cited because they make some objectively important contribution? Or because of the ways others have added values to them, by interpreting them in particular ways, by selecting particular findings from them, or by embellishing or distorting the intended meaning of the original author? Without reading cited papers and understanding how they are used in later citing papers, citations counts tell us too little about what role a paper is playing in the literature. As we saw in the Chapter 5, a single study can be cited in many different ways, with scientists picking out specific findings and interpreting these in different manners. What’s more, the findings reported in Chapter 2 regarding the spread of a ‘phantom’ reference in the literature demonstrated that a paper doesn’t even need to exist to be highly-cited. In thirteen papers, a non-existent paper came to be used to support the claim that a compound, rutin, had particular health benefits. This claim appears to have spread through authors copying references and interpretations from previous studies. We know this copying happens in the literature because of the repetition of error, but we do not know how prevalent this practice is. If authors copy references and interpretations from other studies without reading the original papers but make no obvious errors, how would anyone know that they had done this? We cannot assume that just because a paper is highly-cited that it has actually been read by those who cite it.
However, such findings have consequences beyond undermining the already flimsy case for the use of citation metrics. They have implications for how we understand how scientific knowledge develops through the literature. If the meaning of a study isn’t established only within the confines of its own pages, but in the works that extend, embellish, distort, or critique it, how do we study the impact of papers? How do we study the history of particular knowledge-claims when not only do we have to read those papers, but also how those papers are cited and reinterpreted across a population of citing papers?

As demonstrated in this thesis, if you trace the development of knowledge surrounding a seemingly simple scientific proposition, such as the claim that dietary fat is causally related to heart disease, you find a research effort that spans decades that consists of thousands of authors and papers. These publications link to other publications via references, and the network of interaction between these papers is many times larger than the set of relevant papers; interactions that modify, re-interpret, praise, and criticise previous work. If each individual paper can alter the course of scientific knowledge, not only by introducing new experimental or observation evidence, but by reinterpreting past studies, then understanding how scientific knowledge develops over time, achieves stability, or how particular debates are settled becomes particularly complex.

Science, by common accounts of its progress, seems like a linear system with a simple causal chain: paper A leads to paper B which leads to paper C. But when paper C can lead to a re-interpretation of paper A, and paper D can introduce an error to the interpretation of A, while paper E can mimic the interpretation of A from D, then we have to study science as a complex system – where the meaning of a study as upheld in the literature is not necessarily the same as the intended meaning of the original author nor predictable from its findings and arguments.

In one of the earliest works of the sociology of science, Fleck (1935) noted that scientific knowledge might only be understood as an empirical phenomenon by studying the dynamics of complex communication systems:
Thoughts pass from one individual to another, each time a little transformed, for each individual can attach to them somewhat different associations. Strictly speaking, the receiver never understands the thought exactly in the way that the transmitter intended it to be understood. After a series of such encounters, practically nothing is left of the original content. Whose thought is it that continues to circulate? It is one that obviously belongs not to any single individual but to the collective. Whether an individual construes it as truth or error, understands it correctly or not, a set of findings meanders through the community, becoming polished, transformed, enforced, or attenuated, while influencing other findings, concept formation, opinions, and habits of thought. (pp.42-3)

CS-CNA is one method by which this can be studied, and doing so might be of interest to those interested in developing Fleck’s idea, or perhaps extending Barnes’ (1983) theory of Bootstrapped Induction to understand how the meaning of studies is constructed through a complex social process in the literature. However, as Collins (1999, 2000) highlights, publication and citation data alone will not provide a full account of why scientists reference particular studies or interpret them in particular ways. For this, traditional methods of STS remain vital – from ethnography to historical case studies – to understand the social and historical context of action.

A note of caution

Importantly, as provocative as my findings are regarding the biased nature of evidence selection, I urge caution in drawing too bold conclusions about the reliability of this area of research. While my results could be taken as supporting claims that the diet–heart hypothesis has been preserved by a biased selection of the available evidence (Atrens 1994; Ravnskov 1992, 1998), my study was exploratory in nature and did not intend to contribute to the ongoing controversy regarding the validity of the diet–heart hypothesis. While evidence selection and evaluation may have diverged from an ideal, we currently understand little of how this impacts the course of scientific work. This is a new research front, and one that needs time to develop before making evaluative pronouncements.
Indeed, some scholars have proposed that a bias in how scientists use evidence plays a necessary role in the development of scientific knowledge. Kuhn ([1962]1970) argued that scientific work ‘progressed’ only when a particular scientific theory or ‘paradigm’ was adopted, and researchers devoted their efforts to developing and refining that theory. For Kuhn ([1962]1970), once a particular paradigm has arisen and attracted adherents, science progresses in a rather dogmatic fashion. He saw that scientists often ignored fundamental novelties and difficult findings. In cases where inconvenient findings were confronted, scientists appeared to “devise numerous articulations and ad hoc modifications of their theory in order to eliminate any apparent conflict…though they may begin to lose faith and then to consider alternatives, they do not renounce the paradigm that has led them into crisis” (p.78).

Kuhn’s view of science appears to fit well with the results reported in this thesis, accommodating both the findings regarding citation bias but also the interpretive strategies of diet–heart advocates when confronted by unsupportive data. Importantly, this behaviour was a rational response; science progressed in circumstances where a particular theoretical perspective was communally accepted and which, for time, did not yield to apparently unsupportive evidence. This was because of the considerable time and resources invested in developing theory – of scientists’ time and career prospects, of investment in equipment and skills – which he proposed explained a general hesitancy in science to embrace difficult findings.

2. Complexity, Paradigms, and Citation Networks

Following my CS-CNA studies, I became increasingly conscious of the complexity of citation behaviour and its relationship to knowledge making and evaluation. Accordingly, I sought a way of studying this complexity. In Chapter 6, I drew from Kuhn’s ideas as to what a ‘paradigm’ was and how its literature may be captured. I proposed a concept of ‘sign-posting’ – where scientists, wanting to speak to a particular audience of interested researchers, use
particular indicative terms in the titles of their publications. I captured these terms via Boolean queries designed to mirror the cognitive structure of the diet–heart hypothesis; queries designed to capture each element of the postulated causal mechanism linking diet to CHD by the terms used by scientists over this period. By layering these Boolean queries, I isolated a large body of literature likely relevant to diet–heart research. To validate this, I demonstrated that, as Kuhn predicted, journal publications with these terms in the title demonstrated remarkable growth that cohered with scientists own retrospective accounts of how this theory developed, and that the literature retrieved held a large proportion of the documents cited by reviews and historical pieces that have examined this area.

From this literature, I constructed a large citation network capturing the interactions between these documents over time (Chapter 7). Using network analysis, I validated the results by recourse to ideas from both Kuhn and Price (1965). First, I demonstrated that most of the documents collected formed a large interconnected network. Second, I demonstrated that this network conformed to Price’s power-law distribution – a dramatic inequality of citation was present – and related this to Kuhn’s ideas that a ‘paradigm’ would cluster around a set of core achievements and methods. Third, I demonstrated that the documents retrieved by my search strategy were more likely to reference other documents detected from this search than other documents – a sign of shared research focus between the retrieved records – and related this to Kuhn’s ideas about the preference scientists had for communicating with those working on the same research topic. Finally, I developed a method to address the ‘boundary problem’ (Batagelj et al. 2014) – of what non-retrieved literature ought to be retained in a citation network – by demonstrating that the likelihood of a non-retrieved document being cited by more than one of my four retrieved literature sets was a function of the total number of citations received by a non-retrieved document. Accordingly, the network constructed cohered with both Kuhn’s ideas about community structure in scientific paradigms and Price’s findings regarding the structure of citation networks. Importantly, while some have claimed that this area of science is corrupted and unusual (Ravnskov
1998), the network constructed from it resembles science as a whole, at least in terms of its overarching citation dynamics expressing a power-law of a similar exponent to science as whole.

**Community structure and the research front**

Following this, in Chapter 8, I mapped the network and demonstrated major structural features beyond the power-law distribution. First, time played an important role in structuring the network – documents tended to cite other documents in close temporal proximity. Second, documents tended to cluster together into groups focussing on particular research questions that were popular in different time periods. In different clusters, the top-cited literature was related to understanding the relationships between diet and cholesterol metabolism, atherosclerosis, or CHD. I found a complex research ecology – one that was structured around particular methods and findings, that appeared to evolve over time, and which was divided on lines of basic and clinical research. These results appeared to lend support to Price’s (1965) notion that science progresses through a densely interacting research front of scientific publications that preferentially references recently published articles. As the research front moved through time, it split along the lines of basic and applied research, but also began to investigate new, though related, questions regarding the link between diet and CHD.

**On the absence of a critical community**

I anticipated at the outset of this thesis that if I captured a large body of literature related to the diet–heart hypothesis, then I would find a “community” of critical papers. I did not observe this. I failed to detect such a clustering of documents both by computational methods, such as modularity analysis, but also by my own manual examination of the network. Critical papers are contained within the network, but they don’t seem to cluster together via references and citations, at least not to the degree to which a clear cluster emerged. Over the years of reading these papers, I have developed a working knowledge of this research area, and, to my knowledge, my network contains all of the major critical studies of the diet–heart idea.
The absence of a community of critical papers might be because of the level of analysis; however, I have explored temporal cuts of the network and have failed to find a critical cluster after exploring 10-year cuts of the network and have examined each individual modularity cluster. Accordingly, either no such critical grouping exists in this network or it is here, but I haven’t been able to find it. However, I am of the opinion that no such grouping of papers exists here – critics did not appear to have formed a cohesive group of publications that knit together by references and citations. This seems likely. The major critics proposed various different criticisms of the diet–heart hypothesis, and came from different areas. Ahrens (1957, 1976), a lipidologist, was concerned over the strength of evidence regarding the relationship between dietary fat, serum cholesterol, and CHD because he was aware of the weaknesses of feeding trials and problems of their generalisability. Mann (1959, 1977), an experimental pathologist turned epidemiologist, began as an advocate of the lipid hypothesis of atherosclerosis in the early 1950s, but came to believe that exercise was the major cause of CHD. His criticisms, some of the most strongly worded in the history of the diet–heart, were often followed by his advocacy of the link between exercise and CHD. Yudkin (1957, 1964), a British clinical scientist, attacked the diet–heart hypothesis, but simultaneously argued that the consumption of sugar was the chief cause of CHD. Gofman (1956, 1958) attacked the use of serum cholesterol measurements used by diet–heart advocates, while simultaneously arguing that his typology of lipoproteins and methods of analysis ought to supersede it. Reiser (1973), a chemist, critiqued the diet–heart idea for its vague use of terminology, such as “saturated fat” and “animal fat”, which, for him, hid a complex reality of different fatty-acids that had different effects on serum cholesterol, regardless of saturation, and which differed from individual to individual. McMichael (1979), an authority of British cardiology, dismissed the link between serum cholesterol and CHD, which he believed unfounded, and this seemed to motivate his attacks on the link between diet and CHD. Yerushalmy and Hilleboe (1957), statisticians, critiqued claims of a causal relationship drawn from statistical associations between diet and CHD because of an absence of a plausible mechanistic
theory. Thus, the critics attacked the hypothesis for different reasons and levelled different criticisms at it. Indeed, most of these critics appear to have little in common with one another beyond their scepticism of the dietary fat link. These disjointed criticisms will require further analysis, but it does raise the question as to whether the failure by critics to band together might have played a role in the failure of their criticisms to undermine the diet–heart hypothesis in the scientific community.

According to Kuhn ([1962]1970), “a scientific theory is declared invalid only if an alternate candidate is available to take its place” (p.77), and this alternative must be able to “attract an enduring group of adherents away from competing modes of scientific activity”, while also being “sufficiently open-ended to leave all sorts of problems for the redefined group of practitioners to resolve”, and capable of sufficiently explaining anomalies (p.10). Accordingly, it may be that critical evidence is only taken seriously when a rival theory is capable of explaining that data and solving research problems currently of interest. And, perhaps, only when this rival theory attracts a community of advocates that are willing to attack the prevailing theory, will certain findings become widely known and quoted. Thus, it might be the case that to understand why the diet–heart hypothesis survived we have to examine why alternative theories failed to attract a large group of adherents, or whether they did but didn’t feel the need to attack the link between diet and CHD.

In a recent study (Leng et al. 2019), I applied a novel form of CS-CNA to a case of controversy in a different area of science to examine this. To understand the dynamics of scientific debate in the literature, we examined the rise and fall of the vasopressin-memory hypothesis, a hypothesis linking the peripheral administration of vasopressin to improved memory. Here, we saw that, following its first articulation in the literature by de Wied (1971), the hypothesis quickly attracted a number of supportive publications in the literature. By 1980, the hypothesis had attracted scientific attention with a steep rise in publications devoted to exploring this hypothesis. Most publications (86%) were supportive, either contributing supportive empirical results, extending its arguments, or citing the theory in a supportive manner.
However, the first critical result emerged in 1975, and by the end of the decade ten papers were critiquing the hypothesis (13% of papers as of 1980). In the 1980s, however, de Wied’s hypothesis came under sustained attack. Between 1981 and 1991, 57 papers critiqued the hypothesis compared to 166 supportive papers. By analysing these papers by CS-CNA, we observed the critical papers, at first unconnected by reference relationships, began to form a tight critical group that drew the diverse lines of criticism together to make a case for refutation. Critical papers began to rival supportive papers in their citations counts. By the late-1980s, after launching their attack, the critics largely left the field, seemingly uninterested in continuing the debate. However, their work had been done. By 1990, the hypothesis had been largely abandoned by all authors apart from de Wied and his collaborators. Thus, in this study, we observed that a promising hypothesis attracted an initial spike in support, came under sustained attack, and this attack appeared to provoke the abandonment of this hypothesis by independent scientists. Following the period of intense criticism, the citations to de Wied and his collaborators body of relevant work decreased dramatically. The citations to the critical studies, high during the peak period of controversy in the 1980s, also collapsed. Thus, it appeared that critical studies had an impact beyond just the number of citations they accumulated, their major impact was in provoking a fall in both the citations to particular papers and causing a collapse in publication output. Interestingly, critical studies formed a clear community within the network, and this community appeared to be effective because it could draw together the findings from several independent teams and several lines of criticism. A similar, tightly knit community of critical papers was not observed in my analysis of the diet–heart hypothesis.

**Utility**

My hope is that I have demonstrated an approach here that can be used by others who desire to capture and analyse large bodies of literature relevant to particular theories and debates. This might be of interest to those seeking to establish archives of the scientific literature, but also for those interested in understanding the dynamics of publication and citation in specific research
areas. The methods described here, particular CS-CNA, are a valuable addition to current work aiming to map and analyse the dynamics of scientific and public controversies and efforts to incorporate qualitative and quantitative analysis in STS (Venturini 2010, 2012; Venturini et al. 2017). Currently, there are only two major citation network datasets that focus similarly on particular areas of research. One contains all arXiv publications on High Energy Physics Phenomenology and the other containing all arXiv publications High Energy Physics Theory, which are held in the Stanford Large Network Dataset Collection (Leskovec, and Krevl 2014a, 2014b). All data generated for this project will be made available in an open data repository, and I hope that this can be used by others, either to explore the dynamics of diet–heart research or to compare the dynamics of another body of literature against it.

3. Main Path Analysis: On the paths most trodden and the voices unheard

Finally, following reports in the literature that a version of Main Path Analysis (MPA) could capture the major research stream responsible for the development of a scientific area, I applied this to the constructed network. MPA analysis identified a collection of documents that are frequently discussed in scientific reviews and retrospective accounts of the development of the diet–heart hypothesis; however, few critical studies were identified by this method. By this point, this was an unsurprising result – the top-cited literature was populated primarily by supportive studies of the diet–heart idea or studies reporting statistically significant findings relevant to this research area. However, following reports that MPA analysis tends to exclude critical studies, this remains an important finding. As the diet–heart hypothesis involved one of the most famous, and likely the longest-running, medical controversies in scientific history, this indicates that MPA analysis is an unsuitable methodology for both systematic reviews, which necessarily aim to evaluate all relevant existing empirical evidence, and for understanding the history of an area of
research, unless one wants to pick out only those studies that were particularly popular and how these connected to other popular papers.

What then did MPA reveal? To understand this, I read all identified documents in the sequence by year of publication, summarised each, and wrote a narrative review using only these studies (Chapter 9). My account traced the emergence of the diet–heart idea, spanning from Anitschkow’s original animal model of atherosclerosis to Keys’ diet–heart hypothesis.

My account suggests that scientists, in the face of a growing understanding of the complexities involved in the pathogenesis of atherosclerosis and CHD, serum cholesterol regulation, and the relationship diet had to these, championed a theory that appeared promising despite the lack of clear evidence because it appeared to be a promising avenue for the treatment for CHD. Interestingly, in the period examined, there was no clear agreement as to which dietary fat, nor whether other dietary factors, such as dietary cholesterol, were the most important to focus on. While it captured a level of a disagreement between scientists championing different perspectives of which type of fat was important and at what level it should be in the diet, these scientists appeared to share the assumption that lowering total serum cholesterol via dietary means was a promising avenue to treat CHD by retarding atherosclerosis. By the early 1950s, the assumption that raised serum cholesterol levels were a causal factor in the development of CHD had been largely accepted by the documents that sat along main paths of development, and dietary research appeared to progress on the basis that lowering serum cholesterol would be beneficial. The most critical studies of this assumption, however, were excluded by the MPA analysis.

Accordingly, while this method is inappropriate for use in systematic reviews or historical research, it might be useful for those interested in how a position rises to dominance in the literature over time, and how certain assumptions, at first conjectural, appear to became taken for granted over time. The slow emergence of the diet–heart hypothesis in this analysis, and the absence of a clear single theoretical paper that first introduced it, was certainly an interesting finding. I had expected that, at some point, the diet–
heart hypothesis would be extensively outlined in a scientific publication. I neither found this in the MPA analysis, nor can I recall reading such a paper. This jars with my expectation that science is driven by explicit theory. Indeed, this might be how most areas of science operate, but it is nevertheless interesting that the diet–heart hypothesis does not seem to conform to this.

Yet, I also observed that this area of science is more complex than the picture constructed from this method alone, and on this basis I brought my study to a close. Rather, a complex ecology of research existed with different views being championed by different scientists, some of whom held different beliefs over what factors were important to measure in the blood, over which dietary fat was important and why, and over whether population or clinical treatment strategies ought to guide research and policy-makers. These beliefs appeared to structure how scientists pursued their research in both their selection of methods and their interpretation of data, and these factors require a great deal more historiographical analysis than was possible in this project. The network constructed provides a valuable resource for this in providing both a list of relevant publications and their citation relationships, and, in a follow-up project, I will explore how to analyse directed citation networks in the writing of history.

4. Limitations and Future Research

The findings of this thesis suggest that scientists, in this area, tended to use evidence selectively in support of their favoured ideas and bold conclusions. If, as others have suggested, scientists are in the business of trying to persuade others of the merits of their ideas, and that scientific writing is one of the tools by which scientists persuade one another (Gilbert 1977; Latour and Woolgar 1979), then perhaps this finding is unsurprising. Recently, however, scientists have voiced their concern over such practices and have drawn attention to the need to recognise degrees of advocacy in scientific work. Some forms of advocacy might be expected and reasonable, while other forms
should draw concern from the scientific community if such practices risk distorting published evidence.

**Regulatory Science and Policy Advocacy**

As diet–heart research informed both clinical and public health guidelines, and because some scientists came to advocate for particular policy choices, this area might be considered to be an example of ‘regulatory science’ (Jasanoff 1990, 1995, 2003; Irwin et al. 1997; Wynne 1992; Shackley and Wynne 1995a, 1995b; Wagner et al. 2018). Regulatory science is a term used to refer to any science that exists in dialogue with a regulatory or policy-making community, and which is oriented towards providing scientific evidence and advice to address particular regulatory or policy issues.

Sociologists have been interested in regulatory science because a range of novel pressures, motivations, and practices are thought to arise when scientific, political, and legal communities are drawn together to address particular issues. For Jasanoff (1990), regulatory sciences aim for ‘serviceable truths’, knowledge that can be used by policy-makers to devise workable policies or to ensure compliance with existing regulations. This is contrasted with the typical ‘curiosity driven’ research that scientists normally pursue. As policy-relevance may have altered the publication and citation behaviour of scientists in this area, it is possible that the dynamics of evidence use and selection described here will not be similar to other areas of science.

But why might this have influenced publication and citation dynamics of diet–heart research over the period studied here? Merton (1968, 1968b, 1988) proposed that publication and citation patterns were intimately linked to the scientific reward system. For Merton, it was through publication that scientists claimed ‘priority’ for particular discoveries and developments, it was through consistent contributions that they built their reputations, and their record of publication played a vital role in their career progression. Citations were also important for academic progression and status. How often a paper was discussed in other works was used as a signifier of esteem because scientists were thought to use their references to pay homage to those that they felt indebted to, and this was essential for the recognition of priority. Latour and
Woolgar (1979) proposed that all scientific work was aimed at persuading scientific peers and research funders of the credibility of particular claims, and this included how scientists wrote their papers and referenced previous works. In both of these examples, scientists aim their work at specific scientific communities. However, the intended audience expands in cases where policy communities or regulatory bodies are also involved. This might alter how reward is distributed in such circumstances, altering the motivation for conducting research (Jasanoff 1990). Expectations of policy-makers regarding the required strength and certainty of evidence can place unrealistic demands on scientists, either forcing them into position whereby they are expected to take bolder position than they otherwise would have or exacerbate disagreement between scientists (Jasanoff 1990, 1995; Wynne 1992; Shackley and Wynne 1995b). Finally, scientific research that contributes to decisions that impact government, industry, or other interested parties, then that research becomes important for a broader array of actors than may be typical in science; actors that may influence how that research progresses (Shackley and Wynne 1995b; Irwin et al. 1997; Wagner et al. 2018).

This thesis focussed on documenting selective citation in this area via citation network analysis and developing methodological capacity for this. Accordingly, many factors that might have contributed to particular evidence selection behaviours and the formation of scientific opinion were not investigated. Future studies to contextualise or expand upon the findings of this thesis by exploring the motivations of, and pressures on, diet–heart scientists should do so though other methods more suited to those aims. Indeed, understanding what political or economic pressure scientists faced in this area might explain why certain studies were performed and published, why certain scientific positions rose to popularity in the literature, and why some studies were ignored. Previous works by Garrety (1997; 1998), Teicholtz (2015), Schleifer (2012), and Olszewski (2015) provide a valuable starting point, but research exploring how political and industry involvement might have specifically influenced publication and citation behaviour will need to be conducted.
One option for this would be to accompany an analysis of selective citation via CS-CNA with existing methods and theories regularly used to understand how actors influence, and are influenced by, policy. Advocacy Coalition Framework (Sabatier and Jenkins-Smith 1988), Discourse Coalition Theory (Hajer 1993), Discursive Institutionalism (Schmidt 2008), or actor-network theory (ANT) (Callon et al. 1986; Latour 1987) would be suitable for this. Indeed, Callon et al. (1986), in *Mapping the Dynamics of Science and Technology*, suggested that ANT could be used to understand this:

…as actor-networks send out texts and hope to grow by translating the forces cited in their texts they inevitably come up against other actor-networks. The dynamics of the production of knowledge arise from the confrontations, competitions, conflicts, and their eventual (more or less temporary) resolution. The problem then becomes: how can we distinguish between powerful texts which allow their actor-networks to grow and those that are weak and vanish into oblivion? For victories, defeats, growth and disappearance all arise from interaction between the texts of actor-networks. Part of the problem, then, is to study texts at a higher level of aggregation. (p.107).

The methods discussed in this thesis should help to facilitate such analyses when accompanied by additional methods and data that aim to understand the interests of, and social relations between, actors. Researchers may wish to explore whether papers written by particular actor-network groupings express the same tendency for citation bias, and whether this can be shown to be related to their particular interests or goals.

A related avenue for future research is to understand whether diet-heart scientists were regularly advocating for particular policy choices in their peer-reviewed publications, and whether this correlates with a propensity for selective citation. Recent evidence suggests that policy advocacy is common in peer-reviewed publications. Scott et al. (2007) examined the extent to which peer-reviewed articles in conservation science were advocating for particular interventions. For them, a sign that a science was becoming a form of policy ‘advocacy’ would be the increase in use of value-laden language – moving from statements about what is the case to what ought to be the case. Selecting six flagship journals for analysis and covering 270 unique publications, the
authors searched for normative statements made in-text. Their results suggest that almost all papers in these journals contained normative statements somewhere in the paper (94% of their sample), while more than half (55%) included explicit statements supporting particular policies. In the *Journal of Conservation Biology*, the authors found that 58% of paragraphs contained normative language, which was significantly higher than other journals in this area.

However, to understand the prevalence of advocacy behaviours in the context of the diet–heart and whether this explains the findings reported here requires future research. The data collected for this project should help to facilitate such an analysis. As a small literature is emerging concerning the dangers of what has been termed ‘white hat’ bias in scientific research – where research is distorted by strong beliefs of scientist regarding the apparent benefits of their research for society (Cope and Allison 2010; Brown *et al.* 2013; Brown *et al.* 2014) – this appears a promising avenue for future research.

While policy-relevance might have contributed to the tendency to selective citation of confirmatory work in this case study, other factors might be responsible for this that are more common to science in general.

Throughout this thesis, I have drawn heavily on Kuhn’s theory about the structure and development of scientific paradigms. Kuhn, however, illustrated his ideas in reference to famous episodes from the history of science that would be considered classic examples of fundamental research, particularly from the physical sciences. Yet, using his ideas about the structure of paradigms and the development of scientific thinking, I showed that these proved useful both for capturing diet–heart related literature but also how the field progressed (Chapters 6 and 7), but also interpreted the bias for ‘positive’ findings as consistent with Kuhn’s view of the functioning of science.

My work also drew heavily from Price’s work on the development of science and the dynamics of publication and citation. Unlike Kuhn, who conducted historical case studies, Price drew his conclusions from studying large volumes of scientific publications and their references largely blind to
their substantive content. I demonstrated that the development diet–heart research through the scientific literature displayed similar dynamics to those observed by Price of the whole of the peer-reviewed literature. In Chapter 7, I demonstrated that the citation of papers from other diet–heart related papers followed a power-law distribution, a striking inequality that is present when considering the entire scientific literature (as Price did) in specific disciplines, or in a specific research area (as demonstrated here). In Chapter 8, I also presented findings the corroborated Price’s ideas of science advancing through a dynamic research front, but also on how the literature on a topic would structure into clusters within a citation network of a particular research area.

Diet–heart research did not necessarily need to fit with these two pictures of science, but my results suggest that it does so comfortably in the aspects studied here. There will be important substantive differences between diet–heart research and other areas, as there will be between any two different research areas, but nevertheless it is interesting that diet–heart research appears typical, at least by certain quantitative metrics of the structure of its literature, despite its policy-relevance.

Advocacy and the scientific reward system

Gitzen (2007) has drawn attention to what he believes is a problematic style of advocacy within peer-reviewed journal articles – IMRAD (Introduction, Methods, Results, and Discussion) advocacy. IMRAD advocacy refers to situations where:

Once bold claims about a poorly tested method or weak result are published, their sins are forgiven and they can be worked into future introductions and discussions at will. IMRAD advocates often stretch available data, gloss over uncertainties in their evidence, and ignore contrary results.

Gitzen proposed these practices could be found even in the most prestigious journals, which appear to favour papers that draw sweeping and seemingly definitive conclusions. He drew attention to this practice because he feared scientists were gaining personally from this style, and he believed this was diminishing the quality of the scientific literature and rigour of peer review. The
evidence that this may occur across many areas of science can be found in the studies of citation bias, publication bias, and of the prevalence of statistical weaknesses in the literature summarised in the introduction and Chapter 2 of this thesis. This alone would explain the findings reported in this thesis – a tendency for scientists to make bold claims and selective use evidence in support of favoured ideas.

Today, the prevalence citation and publication bias, statistical weaknesses, and a lack of replication studies are provoking growing concerns over the integrity of the scientific literature. Many suspect that the rise in reported problems is linked to the rise in use of publication and citation metrics and the behaviours they may have inadvertently incentivised (Abbott et al. 2010; Rawat and Meena 2014; Smaldino and McElreath 2016; Grimes et al. 2018; Biagioli and Lippman 2020). As metrics have become increasingly relied upon by academic institutions and funding bodies to inform decisions on resource allocation and employment, scientists increasingly feel the pressure to maintain high publication and citations rates. This is suspected to have incentivised scientists to produce more publications, especially publications that can attract many citations, and this might have lowered academic standards and influenced the questions scientists decide to pursue.

Simultaneously, academic journals appear to have prioritised increasing their ‘Impact Factor’ as a symbol of high quality, attracting with it increased profits, and to do this some believe that such considerations will have impacted how journals decide what to publish (Leng and Leng 2020). That is, journals publish papers they believe will accumulate citations quickly, with the scientific quality of those papers a secondary consideration. If, as recent evidence suggests, studies reporting statistically significant ‘positive’ results are both more likely to be published and cited (Fanelli 2012, 2013), then a system in which resources and recognition are distributed on the basis of publication and citation metrics might reinforce ‘positive outcome bias’ and provoke the range of ‘Questionable Research Practices’ that appear to driven by the desire to arrive at significant positive results (Ioannidis 2005b; Young et al. 2008; Ioannidis et al. 2015).
However, the impact of evaluative metrics on scientific work does not explain the dynamics of publication and citation in regards to the history studied here due to the time period studied. Accordingly, the findings of this thesis are especially relevant to current discussions over what might be contributing to the prevalence of dissemination biases because it provides a case study of such bias in a period before citation metrics made their impact on scientific work, and this means that other factors can be responsible for producing such distortions. Indeed, whether dissemination biases have actually increased in recent years or whether the apparent increase is simply due to increased academic interest is currently an open question (Fanelli 2018).

The use of evaluative metrics is now widespread, yet the pressures they are often blamed for exacerbating are not new. As discussed above and in Chapter 2, previous work suggests that science may have had a culture that risked amplifying dissemination biases before evaluative metrics made their impact, and such a culture might explain the results of this thesis. Merton’s (1968, 1968b, 1988) work on the reward system of science suggests that a culture incentivising the pursuit of high publication and citation rates is a long established feature of scientific life. But Merton (1968) also began to raise concerns over some of the unintended consequences of having the scientific communication tied to the scientific reward system. The Matthew Effect proposed that in this system scientists that were able to gain recognition and prestige found it easier to secure research funding, to publish in top journals, and to have their works read and used by others, and that this might obscure the contributions of lesser known scientists. Merton believed scientists were committed to a set of norms that ensured the rigour of scientific work, but if scientists don’t follow these norms, or do so imperfectly as Merton noted, then this system might enshrine favoured ideas regardless of the scientific rigour of work and entrench problems such as citation bias.

Gilbert (1977) described the complex reality of how scientists selected articles to reference, which included scientists reading other scientific papers to find relevant literature that was well-liked in particular communities. Gilbert
proposed scientists did this to ingratiate their work to a particular audience, and he described the desire of scientists to situate their work alongside the influential and respected works of a field as one of the main mechanisms involved in the recognition of exemplars in particular communities and the construction of consensus. Latour and Woolgar (1979) proposed that all scientists were advocates of their ideas, and that the scientific community was organised around a self-reinforcing ‘cycle of credit’ – where scientists that were effective at persuading their target audiences were more likely gain research funding, attract quality graduate students, publish regularly, publish in top journals, and attract more citations. This increased their authority and opportunity in a field thereby increasing their credibility and future ability to persuade:

In this respect, scientists' behaviour is remarkably similar to that of an investor of capital. An accumulation of credibility is prerequisite to investment. The greater this stockpile, the more able the investor to reap substantial returns and thus add further to his growing capital (p. 197).

If scientists do behave in this way when searching for literature, and if this is how the scientific reward system operates, then the risk of problems like citation bias are made more likely because scientists are likely to repeat the evidence selection in previous papers and because they have an incentive for selecting and interpreting evidence to support their particular arguments.

Indeed, despite the now widespread use evaluative metrics, evidence suggests these have had little impact on individual publication rates over the last century (Sinatra et al. 2015; Fanelli and Larivière 2016). Furthermore, a marked inequality in the citations received by published works was described by Price (1965), and this inequality appears to have persisted despite the introduction of citation metrics for research evaluation (Seglen 1992; Albarrán et al. 2011). The seemingly universal skew in citations across science strongly suggests a common underpinning generative mechanism, such as Cumulative Advantage (Price 1976), is in operation. This mechanism is likely related to the scientific reward system and/or the product of scientists using past papers to find relevant literature (Merton 1968; Latour and Woolgar 1979; Gilbert 1977;
Simkin and Roychowdhury 2007). Accordingly, it may be the case that problems like citation bias are an inherent danger because of this extreme inequality in citation. If this system is now drawing concern from scientists, then measures to address such problems will need to go beyond addressing the problem of research metrics (Hicks et al. 2015) and look to the underlying reward system in science.

Of course, this is not to say that citation metrics have not reinforced such dynamics in the scientific community, but that the use of citation metrics did not introduce practices of selective publication and citation to scientific work. Indeed, Biagioli and Lippman (2020) have recently documented a range of novel questionable practices and cases of fraud that appear to be aimed specifically at gaming research metrics, and this indicates that the use of metrics has introduced a range of perverse incentives and problematic behaviours into scientific practice. But to understand whether the introduction of evaluative metrics have actually increased the prevalence of dissemination biases, it is important the researchers examine pre-metrics scholarship. While my study of citation bias covers one of the earliest periods analysed for this problem (Leng 2018), no study has looked for citation bias in the scientific literature <1960\(^\text{17}\). Furthermore, the majority of research that has looked at citation bias has done so in the biomedical and social sciences (Duyx et al. 2017). Accordingly, to understand if citation bias is in fact worsening and if it is inherent in scientific practice or just certain areas, future studies should be initiated that focus on literature published <1960 and studies in other disciplines, such as physics or chemistry.

However, whether the findings of this thesis are explained by the reward system of science, policy involvement, some other, unknown factor, or by some combination of these requires additional research. I highlight these here only as potential hypotheses that may be explored in future research.

\(^{17}\) Sackett (1979) was the first to propose citation bias may have distorted understanding of the inheritance of hypertension in a review of the literature exploring this topic published in 1963. Gøtzsche (1987) was the first systematic study of citation bias and explored citations to trials of non-steroidal anti-inflammatory drugs from papers published between 1966 and 1985 – a similar time period to my study and also in area of biomedical sciences.
5. An End

While we currently understand only weakly how the selectivity of scientists in the evidence they choose to reference influences scientific knowledge and opinion, this is a promising avenue for future research. I have discussed, developed, and demonstrated a set of methods that should aid researchers interested in this, and I have tied those methods back to existing theories on the structure of science in the works of Kuhn, Price, and Latour. By this, I hope to have made the case that there is value in understanding how knowledge is shaped across the literature, and that such an understanding is possible.
Bibliography


Bell, E., Clawson, B. (1928). Primary (essential) hypertension - a study of four hundred and twenty cases. *Arch Pathol.* **5**(6):939–1002


Brush, S. (1965). The Use of Citation Data in Writing the History of Science by Eugene Garfield; Irving H. Sher; Richard J. Torpie. *Isis.* 56:487


Cope, MB., Allison, DB. (2010). White hat bias: examples of its presence in obesity research and a call for renewed commitment to faithfulness in research reporting. *Int J Obes (Lond).* 34(1):84-8


Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics.* **90**(3):891–904


Groen, JJ., Tjong, BK., Kamminga, CE., et al. (1952). The influence of nutrition, individuality and some other factors including various forms of stress, on the serum cholesterol: an experiment of nine months duration in 60 normal human volunteers. *Influence of nutrition, individuality and some other factors, including various forms of stress on serum cholesterol. Voeding (the Netherlands).* 13:556–87


Jannot, AS., Agoritsas, T., Gayet-Ageron, A., et al. (2013). Citation bias favoring statistically significant studies was present in medical research. J Clin Epidemiol. 66(3):296–301


Kawamura, R. (1911). Die Cholesterinesterverfettung (Cholesterinsteatose) [Fatty degeneration of the cholesterol ester]. *Jena*


Kessler, MM. (1965), Comparison of the results of bibliographic coupling and analytic subject indexing. *American Documentation.* 16(3):223–33


361


Leng, RI. (2019). ‘An introduction to citation network analysis’. In. ‘Bibliometrics and social network analysis’, a workshop run on behalf of the Economic and Social Research Council’s Advanced Methods Training Scheme.


Mann, GV. (1959). Diet and Coronary Heart Disease. Arch Intern Med. 104:921–9


Marquis, RM., Yudkin, J., Oliver, MF. (1964). Work after a coronary. BMJ. 2(541):874


Moed, HF. (2005). *Citation Analysis in Research Evaluation.* Dordrecht (Netherlands): Springer


Rogers, L. (1908). Extensive atheroma and dilatation of the pulmonary arteries, without marked valvular lesions, as a not very rare cause of fatal cardiac disease in Bengal. *QJM.* **2**(1):1–18


Saltykow, S. (1910). Further investigations on the staphylococcal-atherosclerosis of rabbits. Verhandlungen der Deutschen Pathologischen Gesellschaft, Fourteenth Convention, 1910 (in German) [Cited in Anitschkow and Chatalow (1913)]


Simkin, MV., Roychowdhury, VP. (2003). Read before you cite! *Complex Syst*. **14**:269–74


Chapter 4 Reference List


