VIRAL NETWORKS

Connecting Digital Humanities and Medical History

Edited by E. Thomas Ewing and Katherine Randall

This volume of original essays explores the power of network thinking and analysis for humanities research. Contributing authors are all scholars whose research focuses on a medical history topic—from the Black Death in fourteenth-century Provence to psychiatric hospitals in twentieth-century Alabama. The chapters take readers through a variety of situations in which scholars must determine if network analysis is right for their research; and, if the answer is yes, what the possibilities are for implementation. Along the way, readers will find practical tips on identifying an appropriate network to analyze, finding the best way to apply network analysis, and choosing the right tools for data visualization. All the chapters in this volume grew out of the 2018 Viral Networks workshop, hosted by the History of Medicine Division of the National Library of Medicine (NIH), funded by the Office of Digital Humanities of the National Endowment for the Humanities, and organized by Virginia Tech.

Viral Networks

Viral Networks

Connecting Digital Humanities and Medical History

Edited by
E. THOMAS EWING
and KATHERINE RANDALL

VIRGINIA TECH

Every effort has been made to contact and acknowledge copyright owners, but the editors would be pleased to have any errors or omissions brought to their attention, so that corrections may be published in future editions.

Book cover by Katherine Randall and E. Thomas Ewing in consultation with Robert Browder

Cover image courtesy of National Library of Medicine, National Institutes of Health

Contents

Note about Online and Print Editions

In an effort to achieve the widest possible distribution for this volume, both online and print editions are available. The online edition can be found, free of charge, on the VT Publishing website. Print editions (both color and black & white) can be purchased through online retailers.

PDF and EPUB versions of this book may be found at:
doi.org/10.21061/viral-networks

Many of the data visualizations in this volume rely on color to communicate information—information that is lost in the black & white edition. In some cases detail contained in a visualization may be lost due to limitations imposed by page size. For these reasons, readers are encouraged to consult the data visualizations and associated data sets, which have been made freely available for download on the web. The download contains 9 folders, one for each chapter that features a visualization. Each visualization found in the book can be found in the folder that corresponds to the name of the chapter author. In many cases, data sets and Cytoscape files are present. The download also includes interactive visualizations for figures 1.1 and 6.7. The READ ME provides further information on requirements to use each type of file.

Images, datasets, and interactive visualizations may be found at:
doi.org/10.7294/284t-bf10

Foreword

JEFFREY S. REZNICK

This book represents the culmination of a unique scholarly initiative located at the dynamic intersection of medical history and the digital humanities. It also represents an important outcome of the longstanding partnership between the National Endowment for the Humanities (NEH) and the National Library of Medicine (NLM) with Virginia Tech (VT) as a key collaborator.

The specific initiative which led to this book–*Viral Networks: An Advanced Workshop in Digital Humanities and Medical History*–was a landmark moment in the NEH/NLM partnership dating from 2012 when these agencies signed an agreement to "bring together scholars, scientists, librarians, archivists, curators, technical information specialists, healthcare professionals, cultural heritage professionals, and others in the humanities and biomedical communities in order to share expertise and develop new research agendas representing the commitment of the NLM to supporting scholarship in medical history and digital humanities."[1]

Since that initial agreement, the NEH/NLM partnership has achieved its goals–if not exceeded them every step of the way–thanks to unwavering mutual support and commitment to advance scholarship in medical history and digital humanities. Such commitment was evident at the public program associated with *Viral Networks*. Taking place on the centenary of the 1918 influenza pandemic, it featured Theresa MacPhail, PhD, Assistant Professor, Science and Technology Studies, Stevens Institute of Technology, speaking about her authorship of *The Viral Network: a Pathography of the H1N1 Pandemic* (Cornell University Press, 2014). The *NIH Record*–one of the agency's leading publications–covered her lecture as a feature story, and the global livestream of the occasion remains available for all to appreciate, archived permanently by NIH Videocasting.[2]

"Our responses to outbreaks are conditioned by what we know about past outbreaks," MacPhail observed, as quoted in the *NIH Record.* "They rely upon institutions and structures put in place as a result of prior outbreaks and are often as much about politics and economic constraints as they are about science." She continued:

> We have to think about outbreaks, epidemics and pandemics holistically. We have to look at everything—history, politics, economics, biology, culture—all at once in order to understand not only what happened, but also what is happening and what is likely to happen in the future.

Preceding the workshop, the leaders of NEH and NLM signed a memorandum of understanding reaffirming their inter-agency partnership, paving the way to additional collaboration on research, education, and career initiatives, and no less to help ensure that the trajectory of inquiry suggested by MacPhail continues. To these ends, during the introduction of MacPhail's presentation, NLM Director Patricia Flatley Brennan stated that partnerships like the one between the NLM and the NEH "are quite important to the NLM because they help to create and sustain an interdisciplinary and collaborative platform for discovery at the Library" and across the National Institutes of Health campus:

> Creating such a platform is a key goal of our new strategic plan and commitment to growing infrastructure and supporting data-driven scholarship and inquiry for the benefit of medical research as well as the disciplines that intersect with medical research, like the humanities and medical humanities.[3]

NEH Senior Deputy Chairman Jon Parrish Peede expressed a similar objective in his own welcoming remarks to the workshop participants:

> NEH is pleased to team up with the NLM to help support conferences and workshops aimed at training historians of

medicine on the latest research techniques and to bring together biomedical scientists and humanists to explore possibilities of a collaborative nature. We look forward to many more fruitful ventures between our two organizations as we push both the boundaries of the humanities and the biosciences together.[4]

And pushing these boundaries was the very hallmark of the NEH/NLM partnership leading up to *Viral Networks*, with a series of projects bringing the agencies together with key collaborators to engage an increasing number of scholars from across the disciplines in the process of defining and advancing common ground in twenty-first century research methods. In April 2016, the NLM hosted the workshop *Images and Texts in Medical History: An Introduction to Methods, Tools, and Data from the Digital Humanities*, bringing scholars together to explore emerging approaches to the analysis of texts and images in the field of medical history. The workshop was funded by the NEH through a grant to Virginia Tech and held in cooperation with Virginia Tech, The Wellcome Library and The Wellcome Trust.[5] In October 2013, Virginia Tech hosted at its Research Center in Arlington, VA, *An Epidemiology of Information: New Methods for Interpreting Disease and Data* to explore new methods for large-scale data analysis of epidemic disease.[6] In April 2013, through its own grant from the NEH, and with generous support from Research Councils UK, the Maryland Institute for Technology in the Humanities at the University of Maryland organized and hosted *Shared Horizons: Data, Biomedicine, and the Digital Humanities* to explore the intersection of digital humanities and biomedicine.[7] Coinciding with *Shared Horizons*–indeed in the spirit and practice of the collaboration and openness in research it represented–the NLM released the Extensible Markup Language (XML) for its IndexCat™ database, including more than 3.7 million bibliographic items spanning five centuries.[8] Such commitment to opening and sharing data of all kinds–and no less representing all formats of knowledge–remains a hallmark at NLM. The collaboration with the NEH and many more like-

minded partners inspires all of us to advance the open-research enterprise in new and exciting ways through tools of the digital humanities and knowledge of medical history.

About *Shared Horizons* itself, Erez Aiden and Jean-Baptiste Michel observed in their 2013 book, *Uncharted: Big Data as a Lens on Human Culture*, that the name of the conference was "dead on," and that the collaboration behind it pointed to "the most exciting terrain in our intellectual future" being at "the interface of all our work":

> No one knows quite what to call it. And no one knows quite where it's going. But one thing is certain: Science and the humanities are becoming, once again, kindred spirits. And just as Galileo transformed our understanding of the world in the seventeenth century, these two lenses, back to back, will do the same in the twenty first....[9]

This book—*Viral Networks*—fits in the dynamic trajectory described by Aiden and Michel, as it represents true collaboration and commitment among a group of dedicated scholars, two federal agencies and their strategic partners, and one of America's most important public, land-grant, research universities. And this book represents such collaboration and commitment even more because it is available from VT *Publishing* in an open-access format, for all to appreciate as the studies therein engage undiscovered or underappreciated primary sources, push methodological boundaries to define and articulate new arguments, and chart new research trajectories. Indeed, this book defines the scholarly times in which its organizers conceived and published it as much as these times define the book itself.

With its editors and contributors, I am thrilled to see this book appear, go viral—fulfilling the very promise of its name and its open-access format—and inspire further collaborative research and new platforms for discovery of the human condition located at the intersection of medical history and the digital humanities.

Endnotes

1. National Endowment for the Humanities and National Library of Medicine. "NEH and NLM Renew Partnership to Collaborate on Research, Education, and Career Initiatives." neh.gov. https://www.neh.gov/news/press-release/2018-02-15; https://www.nlm.nih.gov/news/NEH_and_NLM_Renew_Partnership_to_Collaborate_on_Research_Education_and_Career_Initiatives.html (accessed May 2018).

2. Carla Garnett, "Not A Fair Fight: 'Viruses Don't Play By Our Rules,' Says MacPhail." nihrecord.nih.gov. https://nihrecord.nih.gov/newsletters/2018/03_09_2018/story1.htm (accessed May 2018).

3. National Library of Medicine, History of Medicine Division. "The Evolution of Viral Networks: H1N1, Ebola, and Zika." videocast.nih.gov. https://videocast.nih.gov/launch.asp?23678, 6:04, (accessed May 30, 2018).

4. Ibid., 10:40.

5. National Library of Medicine. "NLM to Host *Images and Texts in Medical History: An Introduction to Methods, Tools, and Data from the Digital Humanities*." nlm.nih.gov. https://www.nlm.nih.gov/news/nlm_host_images_texts_med.html (accessed May 2018).

6. National Library of Medicine. "NLM to Participate with Partners in *An Epidemiology of Information: New Methods for Interpreting Disease and Data*." nlm.nih.gov. https://www.nlm.nih.gov/news/epidemiology_of_information.html (accessed May 2018).

7. National Library of Medicine. "NLM to Participate with Partners in *Shared Horizons: Data, Biomedicine, and the Digital Humanities* Symposium." nlm.nih.gov. https://www.nlm.nih.gov/news/hmd_shared_horizons.html (accessed May 30, 2018).

8. National Library of Medicine. "NLM Releases Extensible Markup Language (XML) for IndexCat™ Data. Data Includes More than 3.7 Million Bibliographic Items Spanning Five Centuries." nlm.nih.gov. https://www.nlm. nih.gov/news/indexcat_data_xml.html (accessed May 30, 2018).

9. Erez Aiden and Jean-Baptiste Michel, *Uncharted: Big Data as a Lens on Human Culture* (New York: Penguin, 2013): 207-208.

Acknowledgments

Introduction: Connecting Digital Humanities and Medical History through Viral Networks

E. THOMAS EWING AND KATHERINE RANDALL

Milestones in the development of a networked understanding of disease transmission are also milestones in the history of medicine. figure 1, adapted from a 1984 article in the *American Journal of Medicine,* demonstrates how the earliest research on the emerging AIDS epidemic used network analysis to identify relationships among patients who were spreading this disease.[1] This article relied on interviews with nineteen patients about their sexual partners, which generated the forty circles connected by lines indicating sexual exposure. One patient, marked as Patient 0, located at the center, was connected directly to eight patients and, through a second link, to another eight. Based on network analysis of clusters of infected patients, the article, written by a team of leading experts in the study of this new and frightening disease, endorsed the recommendation issued less than a year earlier by the Centers for Disease Control and Prevention: "Members of high-risk groups should be aware that multiple sexual partners increase the probability of getting AIDS."

At the time, and even more so in subsequent years, this single network visualization functioned on multiple levels: instrumental as a tool for epidemiology, limited as an analytical operation, powerful in its cultural impact, and tragic in its human costs. This chart

Figure 1: *Network Analysis of AIDS Patients*

resulted from a relatively small data sample, interviews with less than twenty individuals (or with close friends and family members, in the case of deceased subjects), in a year in which AIDS is estimated to have killed approximately four thousand people in the United States. The conclusions were expressed in guarded, clinical language, yet in practice may have reinforced hostility towards those engaged in what was then called risky behavior. Most

significantly, the man connected to all the other patients was originally called Patient O (the letter O was an abbreviation for Out of California), but this individual was identified as Patient 0 (in this diagram), and then popularized–and vilified–as Patient Zero, the alleged starting point for the spread of HIV/AIDS in the United States.[2] In other words, a network map used as an analytical tool to represent disease transmission between individuals was transformed into a symbol of a certain kind of behavior that fit into dominant narratives of the era in ways that continue to shape perceptions of disease in popular, scholarly, and even scientific contexts. The human beings whose behaviors were reduced to circles and lines, including Gaetan Dugas, the man later identified as Patient Zero, mattered as individuals, but also as nodes of a network connected not only to each other but to millions of AIDS victims around the world in the decades that have followed since this network was identified in the early 1980s.

This illustration serves as an effective way to introduce the subjects, partnerships, collaborations, and processes that produced the chapters in this volume. All of these chapters deal with topics, themes, and problems in medical history, yet their chronological, geographical, and thematic perspectives range widely and vary considerably. Just as the metaphor of the network illustrates connections while recognizing distinctiveness, these chapters share a common approach informed by network analysis; yet the types of data, the tools used, and the outcomes observed also varied considerably. Most important, whereas the AIDS network diagram simplified complex human relationships in ways that permitted and even encouraged distortions premised on stereotypes, each chapter in this volume engages critically, thoughtfully, and productively with the value of network analysis as an analytical tool. In other words, even as the AIDS network diagram inspired critical thinking about connectivity, it was consistently and creatively challenged, revised, and ultimately re-imagined as a way to think about both networks in medical history and networks among scholars.

The Viral Networks project thus approaches networks as an object of study, a tool for analysis, a framework for collaboration, and a means of scholarly communication. The scholars who participated in this project examined networks in medical history even as they became "nodes" in a network of scholars engaged in collaborative learning. The workshop, inspired by models of networked pedagogy, brought these scholars into a connected series of activities that began with reading proposals, included one face-to-face and two virtual conferences, and ended with final edits on revised chapters. This collaboration helped address many of the issues that came up for each author as they wrote for a wider audience, including questions about how much historical content to include or cut in order to focus the paper on methodology. In essence, the authors in this collection spent months not only on their own papers but on guiding and critiquing the papers of their co-collaborators. The chapters should therefore be understood and read as a fully networked project, not as chapters written individually and placed together.

The tools of network analysis made possible by the digital humanities were enhanced by more traditional humanities methods of close reading, contextual analysis, and layered interpretation. Each chapter author was a node in this network, connected to the other authors by the experience of reading, editing, and evaluating each other's work, yet also connected by the shared experience of using networks as a tool for historical analysis. Finally, each author studied the operation of networks in medical history as a relationship among ideas, people, institutions, or language. Much as the first visualization of relationships among AIDS patients represented a reality of social interactions even as it became a tool for understanding this disease, the Viral Networks workshop created a relationship among scholars working collaboratively toward a shared outcome of understanding the place and significance of networks in medical history by integrating approaches from the digital humanities and network analysis.

The Viral Networks project marks the convergence of three important trajectories: first, the fact that networks are an essential aspect of living the human experience; second, the development of more accessible and powerful network analysis tools; and third, the opportunity to make scholarship more collaborative and accessible through digital humanities tools. As illustrated in these chapters, networks were an essential aspect of the human experience in the form of communication between and among individuals, the operation of medical teams, the debate over the meaning of concepts, the use of tools for diagnosis and treatment, and personal appeals based on shared narratives of experience and established frameworks of order. Networks were central to the human experience; studying networks is thus an essential tool and step in the process of understanding the human experience. As humanities scholars, the participants in this workshop collectively and individually examined networks as an aspect of the experience of the people and processes central to human experiences. Some scholars were committed to network analysis from the inception of their studies; others used the opportunity to participate in this workshop as an inspiration to explore their subjects in a new way.

A recurring question during the Viral Networks project has been, "What can a network show you that another type of analysis can't?" The chapters in this volume demonstrate what a network analysis can reveal, but also how a network analysis can help a humanities scholar approach a problem in a different way, or understand what is *missing* in their sources or interpretations. A network methodology may not be the most appropriate to answer every research question and every project. But any humanities scholar can use network analysis when it *is* appropriate, and our intent with this collection was to demonstrate what that might look like. The scholars who contributed to this collection are all studying topics in the history of medicine—the common denominator for the Viral Networks project—but they vary in research area, familiarity with network processes, and level of comfort with network analysis software. As one of our outside readers for this collection

commented, each chapter "represents work in progress, opening a window onto the author's work at a particular moment in its development." The chapters are snapshots of a research process, meant in many cases to demonstrate methodology-in-process as scholars deliberately work through what network tools and techniques mean for their project, what they learned from their use, and how their work has changed because they have self-consciously applied this approach.

   In chapters one through three, the authors navigate the new terrain of network methodology as traditional historians, documenting research journeys that are valuable to other humanist scholars who are unfamiliar with network methods and tools. In chapter one, Runcie brings academic conversations regarding postcolonialism and the ethics of using colonial records in constructing historical narratives to network analysis. Networking healthcare teams in colonial Cameroon, Runcie demonstrates how varying data inputs in data visualizations can re-center the focus on Cameroonian medical auxiliaries and away from French colonist medical authorities. Smith's chapter two essay demonstrates how networks of psychiatrists, hospitals, and the government worked to maintain segregation in 1960s Alabama, while also tracing the process (and difficulty) of moving from analog to digital history work. Smith shows how historians can build upon hand-drawn mapping of people, places, and events to using digital tools with a more specific focus. In chapter three, Sorrels explores the intersections between allopathic and alternative medicine by networking citation data between practitioners, asking what can be learned about how these two seemingly disparate sects interact from where and how frequently their practitioners published. Sorrels also challenges new digital humanists to navigate the line between reducing the complexity of humanistic research and producing the specific questions and bounded data required for network analysis.

   A concern raised in these first three chapters is how to determine what data should be included in the analysis. The authors of

chapters four and five address this concern in more depth, walking readers through the process of preparing archival materials for network analysis. Engelmann develops in chapter four a genre of early epidemiology outbreak reports, arguing that pinpointing the concepts involved in data extraction for network analysis is in itself an epistemological exercise that opens up new ways of seeing for the historian. Though Engelmann does not use this data to create a network visualization in this paper, he theorizes multiple ways in which the data could be used in a revelatory network analysis. In chapter five, DiMeo and Ruis walk readers through an example of how to take a digitized data set—in this case, the mid-seventeenth century Hartlib papers—and determine how to ask the right research questions in order to glean the appropriate data to then feed into the epistemic network analysis. They challenge researchers to think about what makes network analysis appropriate for a project, how to determine which elements of the data should be included or excluded, and how a historical data set must be understood for a mixed-methods approach, among other considerations. They deliberately focus on the "work in progress" stage of a network analysis project in hopes of demystifying the process for historians new to digital methods.

In chapters six through nine, the authors offer reflections based on the results of their networks. Cottle's chapter six looks at the epistolary networks that emerge in the early-twentieth-century correspondence between two academic women, focusing both on what Cottle terms "macroscopic" and "microscopic" anatomy. While macroscopic anatomy is the level of analysis that comes from traditional historical research, Cottle argues that digital visualizations of connections and themes (microscopic anatomy) can help historians trace connections and networks among people, places, and ideas in written correspondence. While other contributors focus on specifying a research question for a network analysis project, in chapter seven Archambeau demonstrates how the unexpected results in a network analysis can change the trajectory of a research question and challenge assumptions a

researcher may have about data. Archambeau uses plague references made in witness testimonies during a canonization inquest in fourteenth-century Provence to look for characteristics and patterns in how people remember and engage with plague events. In chapter eight, Ruis maps the shifting conception of nutrition over the nineteenth and twentieth centuries, demonstrating how the computer modeling of epistemic network analysis can be used by historians as a tool of macrohistorical analysis to complement traditional close reading. Ruis argues that using this kind of mixed-methods approach can be a way to expand historical understandings of—and create new arguments about—the past. Finally, in chapter nine, Phillips uses network analysis as an exploratory tool, demonstrating through his study of how a core group of researchers at the National Institutes of Health brought statistics into medicine in the mid-twentieth century that historical researchers should not be afraid of thinking in networked terms, though there is no one precise way to apply network tools to archival research.

While the approaches to network methodology used by the authors in this volume vary widely, what is reassuring to network newcomers is that none of them is wrong. Network analysis, like the networks themselves, is often more flexible and open-ended than we might think. This flexibility in network methodology is both encouraging, in that it has room to accommodate humanist scholars, and daunting, in that it can take many shapes for different ends. As many of the authors demonstrate, using network methodology requires critical perspective and judgment in determining what data to include or exclude, and in finding the appropriate way to contextualize what the network shows (or doesn't show). Fortunately, humanist scholars are well-suited to these tasks, being intimately concerned with issues of how ideas spread, how people are connected, and who read/says what to/ by whom. The keynote speaker at the workshop, Teresa MacPhail, illustrated this approach to network analysis by connecting historical examples of epidemics to present and future strategies

by government agencies and non-governmental organizations for dealing with epidemic disease. Using her analytical methods as an anthropologist, MacPhail focused on the human beings within these medical establishments who gather information, evaluate evidence, make recommendations, and deal with the consequences. By focusing on the human element of networks, MacPhail's approach set the tone for the chapters to emulate this interdisciplinary perspective on digital humanities and medical history.

For the methodology—with which many of the Viral Network participants were previously unfamiliar—we benefited greatly from the assistance of data visualization and network scholars who were critical in demonstrating that networks have great potential as well as significant limitations as a tool for digital humanities projects. At the workshop's opening session, Amy Nelson of the Virginia Tech Department of History described how networked learning can enhance both the collaborative and individual contributions of students to research projects. The networked nature of learning is closely connected to the goals of public learning and open access, which provides further reinforcement to this project's emphasis on both the openness of the research process and the accessibility of the research outcomes. Ryan Cordell of the Department of English at Northeastern University described the Viral Texts project and how it explores networks of information constructed by American newspapers in the nineteenth century. By focusing on the changing nature of authorship in the interstices of these networks, this presentation provided a model for this workshop's emphasis on collective reviewing and editing of texts. Finally, Samarth Swarup of the Biocomplexity Institute at Virginia Tech discussed tools for network analysis used by computational analysts across fields, including epidemiology, for understanding and predicting large scale patterns of change. A common theme in all three presentations was the importance of recognizing the humans at the center of the networks, a theme that also connects all the chapters in this book. Finally, Nathaniel Porter, the Social Sciences Data Consultant at University Libraries at Virginia Tech, provided

guidance to the individual scholars, worked with colleagues to develop data visualizations in this volume, and contributed a chapter that discusses the advantages of integrating network analysis with humanities scholarship. Throughout the two days of the workshop, these scholars, as well as observers from the National Institutes of Health National Library of Medicine and National Center for Biomedical Information, contributed their critical perspectives on the chapters and made recommendations for expanding, refining, or reconfiguring tools in order to better understand source materials and analytical questions.

   As the volume editors, we can step back from the workshop and subsequent discussions of chapters to identify key themes that illustrate the scholarly contribution of this volume as a whole: there are connections that may not mean causation; the research questions in a network approach should be finely targeted; not all the complexities of the data can be shown in a single network; and there is bias in a network due to what is preserved, coded, and collected. In some cases, the authors and consulting scholars were able to find strategies to address and overcome these challenges. In other cases, the authors used these concerns to engage critically with the limits of using networks as an analytical tool. The Viral Networks workshop and the contributions to this volume demonstrate how digital network methodology expertise and humanities scholarship can work together to advance and provide new insights that benefit both fields.

   "We experience life as a narrative, not as a map and certainly not as a network," was the deliberately provocative claim made in 2016 by Mushon Zer-Aviv, in the equally provocatively entitled post, "If everything is a network, nothing is a network."[3] As co-editors of this volume, we also experienced this process as a narrative: the call for papers that allowed authors to propose topics; a first virtual meeting to review abstracts; two days of intensive discussion at the National Library of Medicine with contributing authors, consulting scholars, and observers; the substantial revision of chapters, which were then reviewed by other contributing authors; another virtual

conversation to discuss recommended edits; and the final stages of editing, proofing, and publishing this volume. In contrast to Zer-Aviv's claim, however, we also experienced this process as a network: the intellectual connections with scholars, the conversations in the conference room of the National Library of Medicine, and the shared editing space of folders, documents, and virtual discussions. Narratives and networks are not mutually contradictory; networks can be experienced as narratives and narratives can be experienced as networks.

Defining and mapping networks is central to several influential digital humanities projects, including *Viral Texts: Mapping Networks of Reprinting in 19th Century Newspapers and Magazines*, *Colored Conventions: Bringing Nineteenth Century Black Organizing to Digital Life*, *Six Degrees of Francis Bacon*, and *Mapping the Republic of Letters*.[4] All of these projects illustrate how network analysis, using easily accessible tools and digitally curated data, can become an insightful and accessible tool for humanities scholars. Network analysis is popular in digital humanities projects because scholars in fields such as literature, history, and anthropology have recognized connections among individuals to be powerful forces in shaping experiences, values, and relationships; yet these networks can also be transformed into data in ways that can be analyzed by computer scientists and others in data fields. The proliferation of visualizations in these projects illustrates the potential of network analysis to transform the textual evidence valued by humanities scholars into the charts, diagrams, and webs more familiar to scholars in computational fields. These projects directly address key questions for the humanities using new tools that provide fresh perspectives on available evidence: How do ideas spread among people and across communities? How can the diversity of participants be recognized while also exploring the commonality of ideas? How did networks allow ideas to be simultaneously debated at the more sophisticated levels while also penetrating every level of society in the form of published texts and spoken words? Yet the illustration of these connections has not always sufficiently engaged

with the core humanities challenge of understanding and interpreting meaning; or, to use language from the computational fields, the correlations among people, ideas, and places has not always been accompanied by sufficient attention to causation. The presence of network analysis in the digital humanities has been intellectually powerful in ways that have generated significant projects and inspired new research fields, yet the challenge is to move beyond these specific case studies to understand the value of network analysis as a research tool connecting disparate fields.

Viral Networks builds on these remarkable examples of successful implementation of network analysis in the digital humanities, but its larger goal has been to cultivate and support a broad community of contributing scholars, drawn from a range of institutions, thus building a model of collaborative and networked research and writing that can inspire more projects in the future. We encourage readers of this volume to take advantage of the flexibility of digital scholarly publication. The chapters, indeed the entire volume, can be read in a linear fashion, starting with the introduction and proceeding through each chapter, in either the digital form or a print edition. Yet readers may also choose to read across layers, moving from the text of the chapters to the networked diagrams to the data for each chapter, thus finding that the act of reading follows a networked structure similar to that experienced by workshop participants. These chapters should also be read as works-in-progress; in effect, as part of a networked conversation among the individual chapter authors, the workshop participants, and the readers of this volume. In this sense, the chapters are not a final definitive word, but rather an effort to engage both medical historians and digital humanities in continuing to think creatively and critically about the interpretive value of network analysis as a tool, a process, and a metaphor.

The cover image for this volume, a photograph of a training school for nurses in Illinois (figure 2), provides evidence that networking in medical history is neither a new phenomenon nor a product of visualization tools.[5] Professional associations of nurses and

physicians, conferences, and training programs have emerged over the centuries as ways to connect medical personnel, patients, and the general public.[6] The more formal gathering of nurses illustrated

SECTION OF ONE CLASS – SECTION OF ROOM
ILLINOIS POST GRADUATE AND TRAINING SCHOOL FOR NURSES

Figure 2: Illinois Post Graduate and Training School for Nurses

in this photograph became increasingly widespread in nineteenth and twentieth centuries, and serve in some ways as a model for the Viral Networks workshop hosted by the National Library of Medicine, funded by the National Endowment for the Humanities, and organized by Virginia Tech. Like the AIDS diagram in figure 1, this photograph captures a moment in time, with no indication of the specific steps that brought these individuals together, and certainly no way of predicting whether the connections made in this training program lasted in the months, years, and even decades ahead–or whether they ended as soon as the training school came to an end. Yet this photograph reminds scholars that even in a digital age tremendous value remains in the capacity to bring participants together in a single room, to discuss common research interests, to learn from experts and from each other, and to leave the session better educated and more committed to professional activities. We hope this collection is useful to medical historians

looking for new tools to understand research topics, to humanities scholars looking for ways to acquire and apply new analytical tools, or to students at any stage of learning who are interested in how networks might add new dimensions to their research.

Endnotes

1. D. M. Auerbach, et al, "Cluster of Cases of the Acquired Immune Deficiency Syndrome: Patients Linked by Sexual Contact," *American Journal of Medicine*, 76, (1984): 488.

2. The naming of Patient Zero reached the broadest public audience in Randy Shilts, *And the Band Played On. Politics, People, and the AIDS Epidemic* (New York: St. Martin's Press, 1987). For historical analysis of this mis-identification, see Richard A. McKay, "'Patient Zero': The Absence of a Patient's view of the Early North American AIDS Epidemic," *Bulletin of the History of Medicine* 88 (2014): 161-194; idem, *Patient Zero and the Making of the AIDS Epidemic* (Chicago: University of Chicago Press, 2017); Jon Cohen, "'Patient Zero' No More," *Science*, March 4, 2016.

3. Mushon Zer-Aviv, "If everything is a network, nothing is a network," https://visualisingadvocacy.org/blog/if-everything-network-nothing-network, accessed May 2018.

4. *The Viral Texts Project: Mapping Networks of Reprinting in 19th Century Newspapers and Magazines*, http://viraltexts.org/; *Colored Conventions: Bringing Nineteenth Century Black Organizing to Digital Life*, http://coloredconventions.org/; *Six Degrees of Francis Bacon*, http://www.sixdegreesoffrancisbacon.com/; *Mapping the Republic of Letters*, http://republicofletters.stanford.edu/.

5. Illinois Post Graduate and Training School for Nurses: section of one class – section of room (date of collection: 1914). Courtesy of the U.S. National Library of Medicine. https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101611030-img.

6. "American Nursing: An Introduction to the Past," *University of Pennsylvania School of Nursing*: https://www.nursing.upenn.edu/nhhc/american-nursing-an-introduction-to-the-past/; "Nursing History in Illinois," *University Library, University of Illinois, Urbana-Champaign*: https://researchguides.uic.edu/nursinghistory.

1. Networks of the Unnamed and Medical Interventions in Colonial Cameroon

SARAH RUNCIE

Historians focusing on periods of colonial rule and enslavement have long grappled with how to uncover the names, voices, and agency of the oppressed from an archival record often written by the powerful. As scholars have begun to explore the use of digital tools and visualizations of data, moreover, some have raised pertinent questions about how one might represent such "absences" in a visual form.[1]

As a complete novice of network analysis participating in the Viral Networks workshop, these questions quickly arose for me as well. Exploring network analysis offered an opportunity to work in new ways with my research on the history of mobile health teams in French colonial Cameroon. These teams, which were generally led by Europeans but staffed primarily by Cameroonian medical auxiliaries, traveled across the territory and became the primary basis of biomedical intervention in rural areas beginning in the 1920s. French colonial doctors also left detailed records about the work of the mobile teams. While I was at first intrigued by the sheer novelty of turning some of these records into datasets and visualizations, I quickly developed a healthy scholarly skepticism about producing visualizations based on colonial medical records. This experience led me to rethink my approach to the data drawing on my own training as a historian of Africa.

Scholars of colonialism have highlighted how we might best approach colonial records as representative of the logic, aspirations, and blind spots of the state.[2] In the case of the mobile health team service, the aspiration of reaching Cameroonians as patients and recording this encounter was not only a medical or clerical task,

but one of significant political importance.[3] Moreover, historians of disease and health in Africa have long pointed to the multiplicity of forms of healing in Africa during the colonial period, and they have questioned the hegemony of biomedicine in this context.[4] Critical methodological questions thus arise for the historian wishing to productively use colonial records in a new digital medium. How can we use data for network analysis while recognizing complexity? Does creating visualizations based on colonial records reify this information while obscuring other forms of information about medicine in colonial contexts that might be essential? Put another way, what kinds of questions about colonial medical records might network analysis be most helpful in exploring?

Through my participation in the Viral Networks workshop, I first aimed to draw on network analysis to explore the question of *how* colonial mobile health teams spread biomedical intervention in Cameroon. Examining networks presents a potential opportunity to move beyond analysis of the mobile health teams through description of an individual visit or in aggregate terms of how many people they examined, and instead move into closer analysis of how the individual visits of the teams were connected to one another. But what data do we have to connect these teams and visits to one another?

A significant pitfall of using colonial data to analyze the work of the teams arises in who is named and who is not named in these records. If we prioritize seeing the work of the teams as driven by specific, named, historical actors, for example, then we run the risk of focusing exclusively on the work of French doctors and thus reproducing a colonial narrative. This dilemma speaks to larger historiographical trends in the history of medicine that have turned away from a "diffusionist" model that explains medicine as something that traveled from the European metropole to African or Asian colonies. Scholars have rather highlighted how medical practices and forms of knowledge were actively, and messily, created on the ground in the colonies.[5] In my own research, one of my main focuses has been the work of Cameroonians as biomedical

workers in the colonial period and beyond. Yet, in the colonial records I worked with for this project, the Cameroonians constituting the clear majority of labor of the mobile health teams go unnamed. Do network analysis and data visualization have the potential to intervene in these historiographical questions by disrupting an idea of French colonial doctors as the "drivers" of medical intervention in the colonies and towards a focus on the work of known, but unnamed, African actors? This piece aims to provide a concrete example of how historians can bring these kinds of scholarly orientations to bear on choices in using data for network analysis.

Networks and Naming in Medical Work

Many accounts of the French mobile health teams in Africa, both scholarly and otherwise, focus on the work of one man in their creation and spread. French military doctor Eugène Jamot formulated the early mobile health team model to address an epidemic of sleeping sickness raging in Central Africa in the 1920s. The innovation of the teams was their mobility and the idea that medical personnel would travel directly to people within set geographic parameters, rather than only interacting with those who visited hospitals or dispensaries. Jamot continues to occupy an immense place in both scholarship and popular remembrances of French colonial medicine.[6] A bust of Jamot sits in Cameroon's capital city of Yaoundé to this day.

The work of the teams, however, extended long beyond Jamot's death in 1937. After World War II, an expanded mobile health service called the *Service d'Hygiène Mobile et de Prophylaxie* (SHMP) became the primary basis of rural health intervention in France's African colonies. In addition to continuing to screen and treat sleeping sickness, the teams expanded their mandate after the war to include focus on other endemic and epidemic diseases such as smallpox, leprosy, and malaria. Run by French military doctors and staffed by

African auxiliaries, these teams traveled on circuits through villages and gathered local people for examination, and sometimes treatment, in coordination with local authorities. The teams thus expanded biomedical interventions and diagnosis to new parts of Cameroon, and to new individuals, through their travel. Through acts such as physical examinations, injections, and vaccinations, the teams represented a key component of France's medical work in colonial Cameroon and elsewhere in Africa.

Cameroonian historian Wang Sonne broke new ground in moving analysis of the teams away from a singular focus on Jamot and other French military doctors to examine closely the role of Cameroonian medical auxiliaries.[7] Since Soone's early work, a broader historical scholarship has also grown focused on African "intermediaries" of the colonial state. This scholarship has highlighted how "Africans in the lower ranks of the colonial bureaucracy often held positions that bestowed little official authority, but in practice the occupants of these positions functioned, somewhat paradoxically, as the hidden linchpins of colonial rule."[8] Africans thus played key roles in the functioning of the colonial state in realms such as teaching, forestry and certainly medical services.[9] Other scholars of Cameroon and colonial medicine have followed suit, continuing to elaborate on the work of African auxiliaries and also examining the mobile teams as key sites for the unfolding of the agendas, contradictions, and disasters of French colonial medicine in Africa.[10] These works, including my own, have relied on qualitative assessments in their use of colonial records.

Network analysis offers a potential opportunity to use these same records in new ways to examine the fine-grain work of individual people, or groups of people, and how they connected in their work across Cameroon. In line with my broader scholarship, I am most interested in how visualizations might help to continue to challenge a portrait of the mobile health teams as an endeavor driven by a small number of French military doctors and re-enforce a focus on the Cameroonian medical auxiliaries performing the labor of the teams.

Significant challenges in visually representing this work arise due to hierarchies of authority over the teams and uneven naming of participants in the archival record. These hierarchies existed along lines of both "European" versus "African" medical personnel and in terms of the degree of medical training. The medical personnel leading each individual mobile team often show up most clearly as individuals in the archival record, although there are notable disparities in how the work of Africans leading the teams is described. The one to two Cameroonian medical personnel who led the teams in the late 1940s and early 1950s were a target of major critique. In a 1950 report, for example, the head of the mobile service for Cameroon complained that African medical personnel did not have the necessary "upper hand" with the population to ensure success.[11] In this regard, although the vast majority of the personnel of the mobile teams were African medical auxiliaries, colonial officials framed white Europeans as the drivers of the spread of biomedical intervention through the mobile health teams. The colonial record reproduces this interpretative slant in who it names and does not name. Some records from the late 1940s name the person heading each mobile team, both European and African, although these specifics slip out of many of the reports in the 1950s. No details other than professional rank, such as nurse, however, are provided for the African medical auxiliaries performing the work of the mobile health teams.

The dilemmas presented by this project reflect questions that humanities scholars have fruitfully explored in relation to digital humanities. Engaging the archive of slavery in the United States in her article, "The Images of Absence: Archival Silence, Data Visualization, and James Hemings," Lauren F. Klein offers a powerful exploration of how humanities scholars can think about maintaining the kinds of questions they ask and approaches to sources in delving into work in the digital humanities. Specifically, in examining the issue of "silences in the archive" of slavery, she proposes methods to try to move away from this focus on absence to bring forth pathways, connections and the "distributed impact of the labor"

of people seemingly lost to archival silences. Using digital tools in this way, she argues, "reframes the archive itself as a site of action, rather than as a record of fixity or loss." [12] The connection between Klein's article and my own project point to a shared challenge of historians wishing to explore data visualizations but working on subjects in which the voices, or even the names, of certain actors are rendered invisible by archival sources.

In this spirit, I offer three visualizations of the work of mobile health teams in colonial Cameroon that are identical apart from who they name or don't name. Through these visualizations I aim to highlight the kinds of small but meaningful choices that historians face in visually presenting data. Cameroonian medical auxiliaries played a key role in the work of the mobile health teams, but they remain a nameless mass in the colonial records I draw on here. To paraphrase Klein, I seek to explore here how data visualization can be utilized to move from a framework of namelessness of medical auxiliaries to one of networks of labor.

Data and Methodology

This network analysis draws on records produced directly by the mobile health service in Cameroon (SHMP) and published either in annual reports produced by the broader colonial Public Health Service in Cameroon, or in French governmental reports on Cameroon to the United Nations. The United Nations reports are available in the Columbia University library and I collected the annual reports of the Public Health Service through archival research in France. For this piece, I have used data on the SHMP only for the years 1947-1951.

Using these records, I compiled a database using Microsoft Access that lists each known visit of a mobile health team from 1947-1951. The database includes information on the location and date of the visit, the recorded population of that location, the number of people examined by the mobile team, and the number of people given

either a smallpox vaccination or a mixed smallpox/yellow fever vaccine. A connected table captures information about the individual mobile team performing the visit, such as their numerical designation within Cameroon's SHMP (i.e., Team 5), the name of the person leading that team, and the professional rank of this person. A limitation of my database is inconsistency of information. For some years, for example, the reports I used do not provide the precise dates of visits of the mobile teams, the names of the teams, or the names of the team heads.

Three visualizations were created through Cytoscape with this dataset. These three visualizations highlight the work of the teams in three forms. Figure 1.1 frames the network of mobile teams through the name of the team head, when available. Figure 1.2 highlights the professional rank of the team head. Figure 1.3 removes all information about the team head and highlights the administrative designation of the team, when available. All of these

Figure 1.1: Team head names shown, mobile health team Visits, 1947–1951 (portion of graph) The Blue ovals (nodes) represent the mobile health teams and the red represents locations of the visits. The red oval size is based on the recorded population for that location. The width of the connecting lines (edges) represents the number of patients seen at each visit.

graphs are based off an original Cytoscape query and visualization using my database and created by Nathaniel Porter of Virginia Tech.

In figure 1.1, the names of the team heads are highlighted and this visualization supports a framing of the mobile health teams through the work of named historical actors. We can see that certain team heads led multiple mobile health team visits and thus were contributors to the geographic reach of the teams and their encounters with the Cameroonian population. As highlighted by the square in the bottom portion of the graph, however, not all team head names are provided in the archival record. This blank spot highlights an important limitation of the data.

Figure 1.2, like the first, highlights the work of specific individuals as the drivers of the mobile health teams. Instead of the individual names, however, figure 1.2 shows instead the professional titles of those leading the mobile teams. Contractual Doctors, Captain

Figure 1.2: Team head titles shown (portion of graph)

Doctor, and Sanitary Assistants were all European medical personnel. The term "African Doctor" represents an official professional rank from the colonial period, designating medical training beyond that of a nurse or auxiliary but below that of a

French physician. In this case, the choice to shift to representing this data point highlights quite clearly how the prioritization of certain information over others re-enforces the "silences" of the archive. Here, we see in the bottom portion of figure 1.2 that an "African doctor" who goes unnamed was leading one of the teams. Historians naturally gravitate towards identifying clear historical actors, but in this case, prioritizing the ability to name individual medical workers obscures the work of African medical personnel in leading the teams.[13]

Figure 1.3 removes the names and titles of the individuals leading the teams and instead focuses on the administrative number of the teams, when available. When I first worked with this visualization in the context of the Viral Networks workshop, I was concerned about it uncritically reproducing the logic of the colonial state, in that it draws exclusively on French colonial records and their account of

Figure 1.3: Team numbers shown when available, with locations of visits (portion of graph)

the mobile health teams in driving medical interventions in rural Cameroon. However, especially when shown in comparison to the previous graphs, I suggest this visualization also presents an opportunity to move away from a presentation of the work of the teams as driven by the labor of French military doctors.

While the historiography of French colonial medicine focuses on the central role of these doctors in creating, growing and sustaining the idea of the mobile health service, the removal of their names and the focus on the "team" aspect of the mobile health teams can present a different view. Following the highlighted bottom portion of the graph through the figures, we move from labeling the same node through absence or namelessness, to highlighting the leadership of an African doctor, to highlighting the work of a team of people. These shifts in data presentation in turn correspond to the choice here to "give up" the names of the known European medical personnel. Only by "un-naming" them can we move towards other framings of the data on the mobile health teams.

This is an imperfect outcome still beset by limitations. The records I used to create this dataset and visualization neither name nor provide concrete numbers of the African medical auxiliaries working on each team. Yet, here we might see their organization into units or groups as the main driver of colonial medical intervention in French Cameroon. Moreover, we might see this graph as representing how Cameroonians living in specific locales became connected as well to a network of a new professional class within the colonial state: medical auxiliaries working for the teams.

This diagram also, however, further highlights additional challenges and limitations of my data beyond naming. For example, because I don't have an administrative designation of the mobile teams for many of the visits, I don't know (and can't show) if some teams that here are represented by blank blue nodes were really making visits to multiple locations. Second, this graph does not differentiate for change in the teams over time. The administrative organization of the teams changed over the years so "Team2" is not a static entity, which explains why there are multiple blue nodes with the same team name.

This diagram does, however, present helpful visualizations that can lead to more focused questions on the work of the mobile teams. The portion of the diagram displayed here, for example, shows the relative importance in terms of total population and

patients seen for locations such as Foumban and Bafia. But it also highlights how relatively smaller locales like Abong Mbang received multiple visits from the teams. This visualization thus invites a return to the records with questions such as why Abong Mbang was a place of importance for the teams over time.

Turning other Absences into Actions

Throughout the colonial period there existed a significant gap between ideals and reality in the practice of the mobile health teams. First, the teams were chronically under-resourced and staffed and thus fell far short of the aspiration of reaching the whole Cameroonian population. Second, the teams reproduced many oppressive colonial practices and met mixed reactions by Cameroonians. The teams relied on militaristic, often coercive, measures to ensure compliance with medical intervention and had localized histories of medical disaster.[14]

The reaction of Cameroonians to the teams thus grew from complex factors, but colonial officials framed non-compliance as an administrative issue to be overcome. In the 1940s and 50s, medical authorities focused heavily on the percentage of Cameroonians complying with their work as a measure of success. Reaching Cameroonians with biomedical intervention through the mobile health teams remained a central aspiration of the French colonial medical administration throughout the postwar and late colonial period.

Another way to approach this data, and one way that responds to some of the methodological challenges of colonial records, is to create a visualization of colonial *framings* of medical intervention. The SHMP consistently throughout the 1940s and 50s framed the success and failure of their work through the lens of "absenteeism" on the part of Cameroonians. The service also began to map where absenteeism happened most.[15] They measured recorded numbers of local populations against the number of people examined during

a mobile health team visit to calculate the percentage of the local population reached by each visit and, in aggregate, to measure the percentage reached of the total target population for that year.

Subsequently, colonial officials spoke about the relative success of the SHMP's work from year to year in terms of the rise and fall of this number. Officials lamented, for example, that the overall attendance rate to mobile health team visits in 1948 fell to 74.5%.[16] A governmental decree in June of 1948 rendered medical visits mandatory "for the detection of endemic and epidemic diseases and the treatment of recognized subjects suffering from these diseases."[17] Officials attributed a rise in the percentage of people reached by the SHMP the following year, from 74.5% to 77%, in part to this legislation.[18]

They also, however, recognized limitations to their own collection of data. The population counts for some locales were "fairly old," and they noted that in some places there had been significant emigration towards larger towns, thus suggesting that current populations were smaller than recorded, or as they put it, "justifying" some of the absences.[19]

Moreover, officials framed certain areas of the country through the lens of ethnicity and reported a particular recalcitrance towards the SHMP among these groups. In 1947, for example, the SHMP reported that attendance at visits had fallen overall to below 75% but in "Bamileké country" to 56%.[20] In 1950, officials complained that attendance in some areas had fallen into "ridicule" and again pointed to the Bamileké of Douala as being "particularly distinguished by their indifference."[21] As anti-colonial nationalist movements took root across southern Cameroon in the 1950s, this map of medical "indiscipline" became, moreover, imbued with ideas about ethnic groups and their ties to these political movements.[22]

Is there potential in creating a visualization of the concerns of colonial officials over compliance with medical interventions? Is there a way to do so such that the visualization offers additional insight beyond what colonial officials saw as a map of "indiscipline"? Addressing these questions fully is beyond the scope of this piece,

but I raise them to suggest how historians of colonialism and medicine might consider network analysis as a tool for new visual representations of subversions, adaptations, and negotiations around biomedicine in colonial contexts.

Conclusion

As humanities scholars turn to digital tools and data visualizations, we would do well to keep at the forefront of our minds the kinds of methodologies and approaches that guide our scholarship.[23] Data visualization can be alluring in its potential to simplify complex ideas, but my experience in the Viral Networks workshop led me to reflect most on how humanities scholars can offer framings of data that preserve complexity.

What I have presented here is a small example of the kinds of choices that humanities scholars must make in presenting visualizations of data. Visualizations that could be used to foreground how a small number of French doctors drove colonial medical interventions can also be reframed to explore how Cameroonians became connected, in both a conceptual and physical sense, through new kinds of relationships between bodies, disease, and medicine. In both recognizing these choices and communicating how they are informed by a much larger context of scholarship and methodological orientation, humanities scholars have an opportunity to continue to bridge disciplines while also insisting that data, and their representation, are never value-free.

Acknowledgments

I would like to thank the organizers and other contributing scholars of the Viral Networks workshop for their feedback and insights.

Endnotes

1. Lauren F. Klein, "The Image of Absence: Archival Silence, Data Visualization, and James Hemings," *American Literature* 85, no. 4 (December 2013): 661–88.

2. Ann L. Stoler, *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense* (Princeton, NJ: Princeton University Press, 2009).

3. After World War II, French Cameroon was a United Nations Trusteeship and the French were thus required to submit yearly reports to the UN on their governance of the territory, including statistics on medical work and health.

4. Steven Feierman, "Struggles for Control: The Social Roots of Health and Healing in Modern Africa," *African Studies Review* 28, no. 2/3 (1985): 73–147. Feierman points out that "much of the literature about healing in Africa assumes that biomedicine is based on objective knowledge of real phenomena whereas popular medicine is not, and that biomedicine works whereas popular medicine does not." He suggests that all systems of medicine be approached by scholars as forms of "ethnomedicine," that is, they are "products of history" and "embedded with the system of social relations," 105.

5. Helen Tilley, *Africa as a Living Laboratory: Empire, Development, and the Problem of Scientific Knowledge, 1870–1950* (Chicago: University of Chicago Press, 2011).

6. Jean-Paul Bado, *Eugène Jamot, 1879-1937: Le Médecin de la maladie du sommeil ou trypanosomiase* (Paris: Éditions Karthala, 2011). An example of popular commemoration in France comes from the website of the *Association Amicale Santé Navale et d'Outremer*: www.asnom.org, accessed April 19, 2018.

7. Wang Sonne, *Les Auxiliaires Autochtones dans l'Action Sanitaire Publique au Cameroun sous Administration Française, 1916-1945* (PhD diss., Université de Yaoundé, 1983).

8. Benjamin N. Lawrance, Emily Lynn Osborn, and Richard L. Roberts, "Introduction: African Intermediaries and the 'Bargain' of Collaboration," in *Intermediaries, Interpreters, and Clerks: African Employees in the Making of Colonial Africa*, ed. Lawrance et al. (Madison, WI: University of Wisconsin Press, 2006), 4.

9. Lawrance et al., 5. Mari Webel, "Medical Auxiliaries and the Negotiation of Public Health in Colonial North-Western Tanzania," *Journal of African History* 54, no. 3 (November 2013): 393–416.

10. Josiane Tantchou, *Épidémie et politique en Afrique : Maladie du sommeil et tuberculose au Cameroun* (Paris : L'Harmattan, 2007). Noémi Tousignant, "Trypanosomes, Toxicity and Resistance: The Politics of Mass Therapy in French Colonial Africa," *Social History of Medicine* 25, no. 3 (2012): 625–43. Guillaume Lachenal, *Le Médicament qui devait sauver l'Afrique : Un scandale pharmaceutique aux colonies* (Paris : La Découverte, 2014). Sarah Cook Runcie, "Mobile Health Teams, Decolonization, and the Eradication Era in Cameroon, 1945–1970" (PhD diss., Columbia University, 2017).

11. Archives, *Service historique de la Défense*, Toulon, France (SHD). 2013 ZK 005 110, *Rapport Annuel, Service de la Santé Publique, Cameroun Français, Année 1950*.

12. Klein, 665.

13. Note: one "African doctor" leading the teams is named in the records for 1947–1951. I have highlighted this particular example to show how a specific historical actor can be lost by focusing on the need for a name.

14. In the most infamous disaster in the 1920s, the teams administered an overdose of the drug tryparsamide and blinded an estimated 700–900 people in Bafia, Bado, 2011. In 1954, the teams administered lomidine injections using contaminated water in Yokadouma, Cameroon. These injections produced bacterial infections, leading to 300 cases of gangrene and 32 deaths. Lachenal (2014), 147–64.

15. Lachenal (2014), 67.

16. SHD 2013 ZK 005 110, *Rapport Annuel, Service de la Santé Publique, Cameroun Français, Année* 1948, 97.

17. Ibid, 97. Quoting Decree no. 2037, June 1, 1948: "Rendant obligatoire les visites médicales en vue du dépistage des maladies endémo-épidémiques et du traitement des sujets reconnus atteints de ces malade."

18. SHD 2013 ZK 005 110, *Rapport Annuel, Service de la Santé Publique, Cameroun Français, Année* 1949, 86.

19. SHD 2013 ZK 005 110, *Rapport Annuel, Service de la Santé Publique, Cameroun Français, Année* 1949, 86.

20. SHD 2013 ZK 005 110, *Rapport Annuel, Service de la Santé Publique, Cameroun Français, Année* 1947, 111.

21. SHD 2013 ZK 005 110, *Rapport Annuel, Service de la Santé Publique, Cameroun Français, Année* 1950.

22. Lachenal (2014), 67. Lachenal paraphrases Achille Mbembe in referring to record keeping by the SHMP as a mapping of the "terroirs de l'indiscipline," focused on Bamiléké areas.

23. Johanna Druker, "Humanities Approaches to Graphical Display," *Digital Humanities Quarterly* 5, no. 1 (2011). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html. Cited in Klein "The Image of Absence."

2. "A Rather Straightforward Problem": Unravelling Networks of Segregation in Alabama's Psychiatric Hospitals, 1966–1972

KYLIE SMITH

Racism in American psychiatry can be traced back to the intellectual justifications for slavery, and the early linkage of the black psyche with criminality.[1] The idea that the African American was inherently psychologically inferior, less complex, more childlike, or just inherently "bad," gave rise to centuries of neglect, abuse, and misdiagnosis of black people with mental illness, as well as justifying a system of separate and unequal treatment.[2] In Alabama, this system legally ended on February 11, 1969 when the Honorable Judge Frank M. Johnson, Chief Judge of the US District Court in the Middle District of Alabama, handed down his decision in what he called "a rather straightforward problem" in the case of *Marable v. Alabama Mental Health Board*. In this decision, Johnson laid out in plain detail the many ways in which the State of Alabama and the Alabama Mental Health Board were in breach of Title VI of the Civil Rights Act of 1964, and declared racial segregation in the state's mental hospitals unconstitutional. Judge Johnson gave the Alabama Mental Health Board 12 months to desegregate its inpatient facilities entirely, or it would continue to have its federal mental health funding withheld and would not be eligible for any further such funds.[3] In the context of the powerful Civil Rights Movement in Alabama, mental hospitals became sites of contested ideas about the nature of African American psychology and a challenge to the racist nature of American psychiatry itself.

This chapter is part of a much broader project called "Jim Crow in the Asylum: Psychiatry and Civil Rights in the American South,"

which is in its very early stages. The project will look at the impact of the Civil Rights Act on state psychiatric institutions in Georgia, Alabama, and Mississippi. In 2017 I began my research by focusing on archives physically located in Alabama. No single paper can tell this whole story; segregation was a complex process that took many years to achieve, and political positions, psychiatric practice, and community attitudes changed over time. Hence, this paper focuses on one particular series of events surrounding a government administrative hearing and two subsequent court cases in which the government of Alabama was both a plaintiff and defendant. These specific legal moments highlight the importance of psychiatric networks in maintaining segregation, but also demonstrate the importance and extent of the civil rights network, and the determination of the federal government and legal and judicial activists to challenge the medical racism that underpinned approaches to African American psychiatry.

At the same time, this chapter explores the methodological process of bringing network analysis to bear on a traditional historical project that uses non-digitized archival sources with inconsistent data. This is a complicated process in itself, but was made more so by a researcher inherently uncomfortable with a data science approach to a humanistic project. I am a historian working in a school of nursing, and much of my teaching life is devoted to asking critical questions about the effect of the biomedical and technoscientific hegemony on patient care. I ask my students to see beyond the data–to see the complicated forces and circumstances that make patients people. I am also one of those people who has been told her whole life that she is not good with math and should just stick with books. So why would I even venture into networks? Ironically, my interest in networks and the usefulness of network analysis comes from the sources themselves. My findings in the archives revealed a physical network of people who maintained segregation until they were challenged by an external network of civil rights activists and lawyers. I submitted my proposal to the call

for papers for the Viral Networks workshop because I wanted to learn how digital tools might help me make sense of this network and help with demonstrating its complexity to a wide audience.

A Traditional Historian

At the first meeting of the workshop, I described myself as a "traditional historian" without really thinking about what I meant by that. My focus is the history of ideas in psychiatry: how they are informed by political and social contexts, how they change over time and why. But these are not necessarily "traditional" approaches to history, nor are they unusual. By traditional, I suspect I actually meant "archival" and "analog" in that I tend to do things by hand with non-digitized sources. Really, I think I was just signalling my lack of digital skills. My natural method at archives is probably

Figure 2.1: Networks of psychiatric nursing in Alabama

similar to most historians working with non-digitized archives: I enter sources into Zotero and use the Notes function to add biographical detail about authors or main subjects of archival material. I also keep a running Word document open on my laptop where I make notes to keep track of people, places, and dates, as well as the relationship between people and events. I scan and print all the documents I can find, then I read them on paper and underline and highlight them. I have folders littered with colored sticky notes and piles of notebooks that I scribble thoughts in at the end of each day. I also draw maps, like figure 2.1.

I drew this map in May 2017 during my first week in the archives in Alabama. This research was conducted at the Reynolds-Finley History of Medicine Library and the University Archives at the University of Alabama Birmingham.[4] My goal with this map was to visualize the different institutions, people, and events that had any impact on the development of psychiatric nursing in Alabama. This map made it very clear to me that psychiatric nurses were led by a few key figures, were well connected across the South, and, interestingly, had strong connections between major nursing figures outside the state. Drawing this map also made me realize that I could not separate nurses from the broader context of changes in psychiatry in the state, nor from political events like the Civil Rights Act and its enforcement of desegregation. This map made me want to learn more about these broader connections, and then my research assistant came across a newspaper snippet about an executive order issued by the governor of Alabama overturning an attempt at integration. When I returned to Alabama I broadened my research to the Alabama Department of Archives and History (ADAH) in Montgomery and the papers of Governor George Wallace. At ADAH, governors' executive orders have all been digitized, and none of these orders mentions the mental hospitals at all. The archivists helped me sort through some of Wallace's other records, eventually delivering a box labelled "State Institutions." In the box was a folder named "Partlow" (the children's hospital).[5] Inside I found letters to and from the governor, telegrams between him and

his mental health administrators, and a newspaper article referring to "attempts at integration" by Superintendent of Asylums James Sidney Tarwater.

Figure 2.2: The Montgomery Advertiser, April 27, 1966, Alabama Department of Archives and History

This story explained that in March 1966, Superintendent Tarwater ordered that 30 black women from Searcy Hospital (the African American hospital in Mobile) be transferred to Bryce (the predominantly white hospital in Tuscaloosa) and, in exchange, 30 white women be moved to Searcy. As board member Dr. Robert Parker recalled, "[T]he consensus of the Board was that in order to get federal funds it was necessary to agree to comply with the Civil Rights Act of 1964." Parker added that "it was a bitter pill to take, but the decision was unanimous among the members present that the action should be taken."[6] Don Smith, assistant superintendent at Bryce Hospital, explained that the patients were carefully selected, and were fully consulted about the move: "We tried to take people in general who lived down that way...to get them closer to home. We

picked the type of patient who does not require intensive therapy."[7] The story reports that the media, probate judges, and the patients' family members were all informed on March 14 but that Governor Wallace was not informed. It was the actions of the Stokes family, who petitioned the State's US Senator to have their relative Pearl released from Searcy, that alerted Wallace to the patient transfer. The story reports that on April 26, 1966, Wallace demanded an emergency meeting with the Board and subsequently ordered that the patients be "returned to the hospitals from which they were transferred."

There was no information in this file about what happened next, and Wallace's papers were not forthcoming about any follow-up to this action. A quick discussion with the archivists at ADAH led to a search of newspapers.com using the words "segregation" and "Bryce." This search returned a February 1969 article that mentioned two court cases ruling that the hospitals must integrate.

Figure 2.3: *The Montgomery Advertiser, February 12, 1969*

I hoped that the court records would be available and that they might help fill in this three-year gap in proceedings. With the archivists' help, we tracked down the district court case records for the Southeast, which are located at NARA Atlanta. This led us to a case called *Marable v. Alabama Mental Health Board* (Civil Action

Case No. 2615-N). When the box containing the *Marable* files arrived, however, it became evident that this was a much bigger story than I had anticipated. Next to the *Marable* file was a large legal folder containing more than 2,000 pages of documents, all related to a Department of Health, Education, and Welfare (HEW) investigation and hearing into segregation in Alabama's mental hospitals. This bundle of papers was called Docket No. MCR44 and contained testimony, letters, and memos about the continuation of segregation in the psychiatric hospitals. As a result of the investigation and hearings, HEW found Alabama in breach of the Civil Rights Act, declaring that there was no medical justification for segregation. The US Surgeon General then ordered the immediate withdrawal of all of Alabama's mental health funds.[8]

Rather than comply with this finding and voluntarily integrating the hospitals, Governor Wallace took HEW to court, arguing that the federal government was overstepping its authority and that the State of Alabama was not in breach of Title VI (*State of Alabama v. Gardner*, 2610-N). This case was filed in October 1967. The *Marable* case (2615-N) was filed in November 1967 by Orzelle Billingsley and Demetrious Newton (both well-known civil rights lawyers from Birmingham) and Jack Greenberg, Michael Meltsner, and Conrad Harper from the NAACP Legal Defense Fund in New York City.[9] Both civil actions (2610 and 2615) were filed in the US District Court for the Middle District Court of Alabama, and an identical three-judge panel (Johnson, Goodbold, and Pittman) was convened for both cases, which were then consolidated to be heard together. There was no trial; instead, all parties (which now included the US Department of Justice and the US Attorney General) stipulated that the material from the HEW hearing contained in Docket No. MCR44 would be used by both sides to argue their respective cases. It was noted by Judge Johnson that by doing so, all parties "conceded that there are no genuine issues of material fact and that the only issues in dispute are issues of law."[10]

From Analog to Digital

In an attempt to piece this story together and to make sense of the connections, I drew more maps and diagrams of circles, trying to put on one page all the moving parts of this story. This complicated network of professionals, lawyers, government officials, and community and patient activists ran like a spider web across the state of Alabama, with threads extending to Atlanta, the District of Columbia, and New York City. This spider web was like a roadmap—the "viral network" through which racism had both traveled and been arrested.

Figure 2.4: Networks of segregation vs. integration

In figure 2.4 I tried to lay out in one visual every institution that had anything to do with either segregation or integration, linking these institutions to their various documents, roles, ideas, practices, and outcomes. My goal here was to lay out all the elements of the story and determine which ones I would focus on as I prepared for

the Viral Networks workshop. I clarified this document with a more linear narrative in order to pinpoint the main external forces that had acted on segregation.

Figure 2.5: The narrative of integration

In figure 2.5 I was trying to use colors to identify the *types* of groups acting in the narrative (i.e., legal, government, community) and how those interacted with each other in order to force or fight desegregation, as well as considering some of the aftereffects. The Civil Rights Act clearly became the defining moment in this narrative, as it provided the impetus for action and the mechanism for judicial enforcement. This diagram helped me narrow my thinking down to exploring the centrality of the Civil Rights Act and the networks that existed both before and after it.

By now I had done some preliminary reading about network analysis and was familiar with terms like "nodes" and "edges," but I hadn't quite made the leap to actual software. Before we convened at our workshop in DC, I made one more diagram that I hoped would lay out more clearly what my main question was and what sort of data I had to work with.

Figure 2.6: Networks of racism

What I really wanted to do with figure 2.6 was to think about how network analysis might help me track the movement of racist ideas in psychiatry through the network and what happens to those ideas once the Civil Rights Act technically makes racism (in its "discrimination in services" form) illegal. Drawing this diagram made me think seriously about what sort of data I had, and I realized that at this stage of the project I didn't have enough data to be able to tell this whole story. This is still my overarching goal for the bigger project, but it will have to wait for the book.

Unpacking Segregated Networks

The real challenge began when I presented these diagrams at the workshop. As I received feedback from the other participants and data scientists, and as I listened to other papers, it became obvious to me that network analysis was a whole other language that I did not speak. I hoped that I could still learn enough of it to make

something useful, and I focused on trying to refine my question and work with the data that I did have. With Nathaniel Porter's help, I set up an Excel spreadsheet to start logging my data in such a way that would help me 1) identify the main players in the networks identified in my maps, 2) show the connections among the players and relevant institutions, and 3) classify their role in the desegregation process. I focused on entering data about select significant people who had some executive role over treatment practices and decisions in the two adult hospitals, Bryce and Searcy, in the period immediately before the Civil Rights Act. Based on consistent values I wanted to highlight, I made columns titled Name, Location (the main geographic place in Alabama from which the person worked), Affiliation (hospital or government department or agency), Role (professional capacity in that affiliation), Context (categorized as either Treatment, Administration or HEW Hearing, or the two court cases designated by their Civil Action numbers 2610 or 2615), Action ("compliance" or "defiance"), and Side ("segregation" or "integration").

The process of compiling this spreadsheet was illuminating. I was limited immediately by the names listed in the annual reports or other documents and by the fact that some people had multiple roles and were defendants in one case or plaintiffs in another. The values of "Side" and "Action" were also complicated because they characterised only official positions taken in response to the Civil Rights Act, which were often utilitarian and not necessarily reflective of lived reality. That is, all managers, directors, superintendents, clinicians, and supervisors were asked to confirm their compliance with the Civil Rights Act, which they did in a formal sense, but this was due to the threat of withdrawal of funds and not because of any ideological or practical commitment. In fact, the written sources indicate that some clinicians retained a *de facto* segregation by claiming they had "no Negro patients suitable for this kind of therapy" or "no Negro staff were suitably qualified."[11] How could an either/or value in a spreadsheet account for this ambiguity? I was also struck by who was *not* in the spreadsheet.

Focusing on people by name meant that I could only include people who *were actually named* in the archives, and this meant omitting the hundreds of people who worked in the asylums and were not listed by name anywhere. It also meant there could be no mention of patients, which is further complicated by HIPAA legislation that has made archivists nervous and patient records elusive.

With these limitations in mind, I then took a crash course in Cytoscape using the online tutorials and created my first diagram (figure 2.7). For this diagram, I sorted the data to show everyone with a value of "segregation" and separated out the people with this value involved in "Treatment." These data created Edge and Node tables, which I then imported into Cytoscape. I then worked with Styles to label each "Role" a distinct color. Red indicates physician, pink is PhD-prepared psychologist, yellow is nurse, and green is social worker. The two blue nodes are the main hospitals, Bryce and Searcy.

Figure 2.7: Networks of segregation by professional role, 1964

Figure 2.7 demonstrates a number of things about the segregated networks. Firstly, far more people are employed in treatment and care capacities at Bryce, the predominantly white hospital. The

network is insular in that the four main executive positions (Director of Nursing, Superintendent, Director of Psychology, and Director of Social Services) were responsible for designing services and programs at both institutions. The implication here is that the four key people would have been well aware of the disparities in treatment between both institutions. There is no record of any of them finding these disparities problematic. All of these people are white.

I find it interesting to consider the role of Superintendent Tarwater, who appears in this diagram as just another dot the same size as the others around him. In fact, however, if I could have figured out how to weight his appearance in this diagram for influence, he would be represented more as a large circle linking both hospitals together. Tarwater oversaw the running of the whole system within Alabama from 1950 until 1970. He is not entirely to blame for its deficiencies. He worked in a severely underfunded system and was continually frustrated by the situation. In 1954 he had written a terse cover letter to the Annual Reports to the Governor in which he stated quite simply, "We need more money." He had maintained this frustration in every year since.[12] He was surrounded by a community and political system that cared little for its mentally ill and in which people could be committed with no medical advice at the petition of a family member to a single probate judge. This indifference was even more marked when it came to the situation of African Americans, who were yet to even be considered citizens by the voters of Alabama.[13] But I was curious to see how he would fare in other diagrams.

Negotiating the Civil Rights Act

The records in Docket No. MCR44 expanded significantly on the sketchy details of the story covered by *The Montgomery Advertiser* and revealed the extent of Tarwater's role in enforcing compliance with the Civil Rights Act. In 1965 the state Department of Health

in Alabama consolidated its mental health services with the establishment of the Alabama Mental Health Board. The Board appointed Tarwater as its first director, and in this capacity he was contacted by HEW to answer questions about Alabama's compliance with Title VI regarding mental health services. On February 2 of that year, Tarwater signed an official HEW compliance form, as did the state Departments of Agriculture and Education, which were receiving food surplus assistance from the Federal Department of Agriculture that they distributed to the state hospitals.[14] However, on July 30 Tarwater received a letter from Robert Brown, the Acting Regional Director of the Public Health Service in Atlanta, informing him that despite signing the forms, there was no actual evidence that the state psychiatric hospitals were in compliance. Brown asked for more detail on how compliance was being enforced and what measures Tarwater intended to take to bring about active desegregation for patients and staff.[15]

It was in response to this pressure that Tarwater had made his attempt at integration in March 1966. In the HEW hearing evidence, it was noted by members of the Alabama Mental Health Board that Wallace had threatened them, promising that if they did not move the patients back that "the highway patrol would do it for them."[16] As a result of Governor Wallace's reaction, on July 20, 1966, Tarwater was forced to tell the Regional Director of the Public Health Service that the Alabama Mental Health Board would not be taking any further steps to meet requirements for compliance with Title VI.[17] Not surprisingly, it was this disregard for federal authority that would ultimately bring the full force of federal law to bear against Wallace. In January 1967 the department commenced formal administrative compliance proceedings, with hearings held on April 11 and 12.

Attempting to represent or visualize this particular part of the network proved challenging. What exactly did I want to say about the network at this stage, and how did it translate into Cytoscape? I needed to determine which elements of the hearing I wanted to represent and what was significant about the people involved.

Figure 2.8 is a simplistic representation of the types of relationships within the Department of Health, Education and Welfare's administrative hearings, labeled as Enforcement, Testimony, and Certification. The "Enforcers" are people employed by the federal agencies (HEW in Washington, DC, and the Public Health Service regional office in Atlanta) who actively sought to enforce Title VI of the Civil Rights Act. The "Certifiers" are all heads of relevant mental health services within Alabama who were legally required to submit

Figure 2.8: Networks of evidence, HEW hearing, July 1966

letters of compliance, and the "Testifiers" all provided verbal evidence through interviews conducted by Marilyn Rose, Special

Counsel for the Department of Health, Education, and Welfare. The attributes of each of the nodes in the networks are extremely difficult to represent in diagrams like this because some people are many things at once, and I had to determine the most significant aspect of their work for this context. In figure 2.8 I have chosen to represent "affiliation" rather than the "professional roles" because, in this particular instance, people are acting as representatives of their institution or agency, and I am trying to show how many of these were internal and external to Alabama. The red circles signify evidence from within the Alabama state hospital and government system; yellow are state-based mental hygiene clinics (that operate with federal funding); orange are new, state-based mental health centers (operating with state funds since 1960); purple are state government administrators; pink are federal agency representatives; and the three dark blue dots are expert witnesses from outside of Alabama.

I could immediately see the problem with this diagram: it separates the Enforcement network entirely from the other two networks, when in fact it was the Enforcement network that both created and acted upon the other two. There should be a link through Tarwater to all of the networks, reflecting the fact that Enforcement processes acted almost entirely through him, but I had not set up the data in a sophisticated enough way for Cytoscape to build this connection. The process of creating this diagram made it clear to me that I needed more skill with the software. It also highlighted the importance in network analysis of knowing the kind of connections you might wish to analyze before actually starting to work with the data. I also wondered about the simplicity of the relationships in this diagram, as well as the profusion of colors, which then need explaining. I also questioned if my networks were too people-centric and if I would see more complex analysis if I used something other than "Name" as the key column. With these questions in mind, I turned to representing all those involved in integration or the enforcement of the Civil Rights Act process.

Networks of Integration

By the late '60s the NAACP's Legal Defense Fund (LDF) was a well-oiled machine in the prosecution of medical segregation cases. Michael Meltsner, LDF's first assistant counsel, was responsible for LDF's health docket. As the lead attorney in the landmark *Simkins v. Cone* (1963) case in North Carolina, Meltsner was well aware of the constitutional and civil rights precedents of which Alabama was in breach.[18] While the official record is not clear on the details, Meltsner suggests that the rapid launch of *Marable* (only three weeks after Alabama launched its own case against HEW) indicates that attorneys and activists in Alabama (with whom the LDF had close working relationships) had been watching the HEW investigation; and, when Wallace reacted with belligerence, they may have alerted LDF. Meltsner then sent a new LDF staff member, 26-year-old Conrad Harper, a Howard graduate and fresh out of Harvard Law School, to work with Billingsley and Newton on the case.[19] The case was brought as a class action by African American patients (and their family members): Loveman Marable, who had been a patient at Bryce for 12 years; Joe Brown, Jr., who was at Searcy Hospital; and Willie James Nichols, a minor from Selma, who was "confined to Searcy from 1966 until July 1967 [when] he was released on a trial basis but is subject to be recommitted in the discretion of defendants."[20] Once this case was launched, and then consolidated with Alabama's own case against HEW, the combined weight of Civil Rights Act enforcement and judicial activism was overpowering.

In the network visualization I tried to demonstrate this weight by logging all the people involved in each case and highlighting their roles on either side. In figure 2.9 the red circles denote anyone affiliated with the Alabama state government or the Alabama Mental Health Board, most of whom have been represented somewhere in either figure 2.7 or 2.8 (this is the first time the state governors appear as named people). In Case No. 2615 Alabama is the defendant; in Case No. 2610 it is the plaintiff. The federal government is again

represented in pink, this time consisting of the Department of Justice and the US Attorney General as well as the Counsel for Health, Education, and Welfare. The secretaries of HEW are the pink defendants in Case No. 2610 but are the prosecution in 2615. Newcomers to the network are patients (purple dots) and lawyers (green dots), with the three judges as dark blue dots forming the connection between the two cases.

Figure 2.9: Networks of enforcement, Civil Actions 2610 & 2615, 1967

The 2615 context is far more diverse and intense, with many more people from outside the state of Alabama now involved, whereas 2610 is almost entirely an argument between the state and the court. This is an interesting visualization in that it seems to convey the weight and power of the network as it related to enforcement of the Civil Rights Act, which swept through Alabama like a threshing machine through the 1960s.

Working with Data Scientists

At this point in the process, and after receiving feedback from workshop participants, it was clear to me that my diagrams were not clearly demonstrating what was significant about these networks. They may have helped visualize certain characteristics

Figure 2.10: Networks of Segregation in Alabama, 1964

of it but they didn't address my central research question about insularity. The first network diagrams I made in Cytoscape were all people-centric; that is, they portrayed relationships that connected named individuals to their roles in the networks of segregation or integration. What struck me about my research conducted thus far was the way that clinicians and administrators in Alabama were (dis)connected to clinicians and administrators in other states, and the influence of this connection on segregation practices. I also wanted to do more with this information than make simple diagrams. I consulted again with Nathaniel Porter, and we talked about representing the institutions by geographical location instead. I then created a table of each of the institutions that had a role to play in desegregation and linked them to their precise geographic location. With this information in hand, Nathaniel and his team came up with two visualizations.[21] Figure 2.10 demonstrates the geographic spread, within Alabama as of 1964, of the network responsible for the maintenance of segregation.

This figure represents the segregated network in black lines that are weighted for influence. That is, the black lines indicate the multiple places where people from various institutions were located. They also signify the strength of connections between the white administrators, psychiatrists, physicians, nurses, and politicians working out of Tuscaloosa, Birmingham, and Montgomery in the northern half of the state. Some of those same people were responsible for the operation of Searcy Hospital in Mobile, which was also home to the Visiting Nurses Association for the southern half of the state. These facilities and administrative units were either actively segregated or administratively maintained segregation. The only integrated mental health units in the state of Alabama in 1964 were those operating with federal funds in Tuskegee, under the direction of the Department of Veterans Affairs (VA) or the Tuskegee Institute. These facilities, which were run by senior African American physicians and administrators, openly accepted white patients.

Data from figure 2.9 (the HEW hearing and subsequent court cases) was then also transposed over a map in order to demonstrate the long reach of the law from outside Alabama, and the impact of the Civil Rights Act within that state. Titles in red indicate those

Figure 2.11: Enforcing compliance with the Civil Rights Act, 1967

judicial or legal institutions responsible for enforcing compliance in Alabama's mental health institutions (HEW, the LDF, and the circuit

and district courts). Some previously segregated institutions from figure 2.10 are now represented in blue, signifying that they have indicated compliance with Title VI of the Civil Rights Act. New places on the map include mental hygiene clinics and mental health centers, which began opening in 1965 and needed to demonstrate compliance in order to receive funds. The only institutions that were not technically compliant in 1967 were the large state hospitals (Bryce and Searcy) along with the state government and its mental health board. This complicated internal network is more readily visible in figure 2.12, which is an inset of figure 2.11.

Figure 2.12: Inset – Networks of Compliance in Alabama, 1967

Much more could be done with these visualizations to enhance understanding. With more time and resources, they could be interactive maps that enabled the viewer to zoom in for clarity. It would also be possible to overlay maps on top of each other in a more dynamic demonstration of change over time. This process would then lend itself to analysis of a longer time period, with more data added from the complicated processes that continued throughout the 1970s and 1980s to bring the large hospitals more

fully into compliance, while they were simultaneously being downsized due to patients' rights and deinstitutionalization cases. The potential for these maps to more accurately demonstrate what I could not do in Cytoscape has given me food for thought for future expansions of this project.

Conclusion

Before the passing of civil rights legislation that was designed to overturn segregation, Alabama's mental health systems remained remarkably closed off from the rest of the country. This began to be challenged in the late 1950s as the National Institute of Mental Health tried to create Southern-focused programs and funding through regional collectives. Some of the professionals in the segregated networks, especially nurses, were a part of these efforts. The passing of the 1964 Civil Rights Act inflamed Alabama's more conservative politicians and voters through a "state's rights" rhetoric that fueled populist resentment about federal interference—especially interference that threatened segregated and racist practices. It was not until federal legislation was passed—and actively enforced through the courts—that any real change occurred. These network visualizations show the importance of a national network for bringing about this change. No diagram, however, can show the complex to-and-fro between and among judges, lawyers, and respective plaintiffs and defendants in the process of that change. From this distance, as Judge Johnson stated in his February 1969 decision, it seems a rather straightforward problem: segregation was illegal and unconstitutional, and it should be stopped by all means necessary. However, those who defended the old system and the "Southern way of life" did not view segregation in this way. It is perhaps not surprising, therefore, that the State of Alabama took another four years to be fully compliant with the orders handed down by Judge Johnson.

There are some limitations to this project that originate in my original data collection and in the use of network analysis. I started

the research in my usual fashion: taking photos or scans, entering items into Zotero, and making notes about people and places and events. I did not have network analysis in mind as a research methodology at the time, and none of my sources have been digitized. Similarly, the sources themselves, and the data contained therein, is haphazard and not consistently reported or formatted over the years in question. The images presented here tell only one very small part of the story and do so in a static visual form rather than using digital tools to actually analyze the data. In this sense, the visuals act as shortcuts to explaining complicated networks rather than testing for any cause or effect or statistical significance in these networks. Given more time and a longer lead-in period (not to mention some intense software training), I believe this project would be ideally suited to Dynamic Network Analysis,[22] which could more readily show the change over time that occurs in relation to the practice and attitudes of racism and segregation as a result of the Civil Rights Act. There are various other elements of the broader project about life for patients in these asylums that would also lend themselves to this kind of analysis. Figures 2.11 and 2.12, showing an overlay of the network with a geographic map, demonstrate some of the potential of digital tools for this kind of work.

In some ways, limitations in this project are also related to my own intellectual inclinations. Like many historians or humanists using network analysis for the first time, I am uncomfortable with simplifying or decontextualizing. I recognize no one project can tell a whole story, and we always make choices about what we can tell at any given moment. However, I could not shake the feeling that the need to provide data that could be analyzed by software necessarily required leaving out important complexities and grey areas that cannot be captured in this way. I would be interested to see if this holds true were I to pursue a more complicated Dynamic Network Analysis model, which would require a highly skilled team. The iterative process of this workshop and the writing of this chapter have helped me appreciate the importance of collaboration when a project is not "born digital." It is not the case that all

historical records of importance are digitized, ripe for text mining. Indeed, in some cases—especially in relation to sensitive issues like mental health or race—those records are deliberately hidden or buried. It takes a particular set of skills to find and make sense of them, and then a different set of skills entirely to translate them to a digital arena. It makes sense that rather than have one person, traditional historian or otherwise, be responsible for this entire process (or that traditional projects remain separate from digital analysis), teams of people with distinct skills and knowledge can more fruitfully combine to bring these projects to light.

At the same time, embarking on network analysis has given me new insight into the nature of historical data—along with some new ways of thinking about how I handle such data. I learned a great deal about the problems inherent in haphazard data collection techniques, and when I returned to the archives halfway through writing this paper, I used the spreadsheet that we had established as the data collection and recording tool. Using the spreadsheet really helped me think clearly about my categories of analysis and about the significance of each person to the broader history I am trying to recreate. I will continue to use this tool as I progress with the project and to explore avenues for further network analysis. At the same time, I am conscious of the need for vigilance when creating labeling categories. As I entered data into my spreadsheet, I found myself sometimes frustrated and sometimes concerned that I might be affixing artificial boundaries or forcing material into false categories that only serve to reify or privilege some people over others. By trying to label people as pro- or anti-segregation, for example, I ran the risk of making people look progressive when their motives may have been merely utilitarian. This is one grey area that standard social network analysis might not be able to account for.

The most important thing missing from this history is the voice of the people who suffered, and continue to suffer, at the hands of racism, indifference, neglect, and lack of funding in relation to mental health care in the United States. These people do not have a place in the records. They are not named. Their individual patient

records (where they exist) have become the property of a state that now hides behind HIPAA legislation. And how can I put an end date to a story that has no end? The same problems that beset Alabama's psychiatric institutions have now been replicated in prisons across the country, where millions of people are left to die for lack of diagnosis, care, or treatment. As we attempt to understand how digital and machine technologies can enhance our understanding of the human experience, we must not overlook the humanity at the heart of such a project. Good history is always analytical and contextual. As the papers in this volume demonstrate, counting and connecting alone should not be the end goal of this thing we call the digital humanities. While I am not sure that network analysis can capture the experience or the pain of those without a voice, I am sure that the need for the digital humanities to bring these histories into the public consciousness is more pressing than ever.

Endnotes

1. Alexander Thomas and Samuel Sillen, *Racism and Psychiatry* (Secaucus, NJ: The Citadel Press, 1972); Khalil Gibran Muhammad, *The Condemnation of Blackness: Race, Crime and the Making of Modern Urban America* (Cambridge: Harvard University Press, 2010); Anne C. Rose, *Psychology and Selfhood in the Segregated Self* (Chapel Hill: University of North Carolina Press, 2009); John Hoberman, *Black and Blue: The Origins and Consequences of Medical Racism* (Berkeley: University of California Press, 2016).

2. Sander Gilman, "On the Nexus of Blackness and Madness," in *Difference and Pathology: Stereotypes of Sexuality, Race and Madness* (Ithaca, NY: Cornell University Press, 1985), 131–49; Dennis Doyle, *Psychiatry and Racial Liberalism in Harlem 1936-1968* (Rochester, NY: University of Rochester Press, 2016); Jonathan Metzl, *The Protest Psychosis: How Schizophrenia Became a Black Disease* (Boston MA: Beacon Press, 2010); Gabriel N Mendes, *Under the Strain of Color: Harlem's Lafargue Clinic and the Promise of an Antiracist Psychiatry* (Ithaca, NY: Cornell University Press, 2015); Frantz Fanon, *Black Skin, White Masks* (New York: Grove Press, 1952); Thomas and Sillen, *Racism and Psychiatry*.

3. Johnson, District Judge, Decision in Summary Judgment Civil Action No. 2615-N & Civil Action No. 2610-N, February 11, 1969, p. 9. USDC Montgomery, AL Civil Case Files 74-C-0813 Box 149, National Archives and Records Administration (NARA) Atlanta.

4. This research was conducted while I was the Reynolds-Finley Fellow at the University of Alabama, Birmingham. I would like to take this opportunity to thank UAB for its support and acknowledge the unparalleled help I received from archivists Peggy Balch at the Reynolds-Finley Library for the History of Medicine and Tim Pennycuff at the University of Alabama Birmingham Archives, as well as Nancy Dupree and Scotty Kirkland at the Alabama Department of Archives and History and Maureen Hill at the NARA Southeast in Atlanta. I am also indebted to the personal knowledge and support from lawyers Michael Meltsner and Conrad Harper (formerly with the LDF), Ira Burnim, Legal Director, Bazelon Center for Mental Health Law and James Tucker, Director, Alabama Disabilities Advocacy Program.

5. "Partlow Segregation," in Alabama Governor Administrative Files: State Institutions SG21597, Alabama Department of Archives and History (ADAH), Montgomery, Alabama.

6. Tom Mackin, "Bryce, Searcy Inmates Swapped in Forced Integration Attempt: Governor Orders Inmates Returned," *The Montgomery Advertiser*, April 27, 1966.

7. Ibid.

8. Docket No. MCR44, USDC Montgomery, AL Civil Case Files 74-C-0813 Box 149, NARA Atlanta (hereafter referred to as MCR44).

9. State of Alabama v Gardner et al Civil Case 2610-N, Complaint, October 13, 1967 and Marable et al v Alabama Mental Health Board Civil Case 2615-N, Complaint, Filed November 17, 1967, USDC Montgomery, AL Civil Case Files 74-C-0813 Box 149, NARA Atlanta.

10. Johnson, Decision, February 11, 1969, ibid.

11. "Assurance of Compliance" forms, Documents GC1E and GC2A&B, MCR44.

12. "Report of the Trustees of the Alabama State Hospitals (for Mental and Nervous Disorders) to the Governor With Annual Report of the Superintendent," September 30, 1954.

13. Susan Youngblood Ashmore, *Carry It On: The War on Poverty and the Civil Rights Movement in Alabama 1964-1972* (Athens: University of Georgia Press, 2008); Wayne Flynt, *Alabama in the Twentieth Century* (Tuscaloosa: University of Alabama Press, 2004).

14. "Assurance of Compliance" forms, Documents GC1E and GC2A&B, MCR44.

15. Letter from Brown to Tarwater, July 30, 1965, Document GC3, MCR44.

16. Testimony to HEW Administrative Hearing, April 12, 1967, p. 68, MCR44.

17. Johnson, District Judge, Decision in Summary Judgment Civil Action No 2615-N & Civil Action No. 2610-N, Filed USDC for the Middle District of Alabama, Northern Division, February 11, 1969, p. 9. USDC Montgomery, AL Civil Case Files 74-C-0813 Box 149, NARA Atlanta.

18. Michael Meltsner, "Equality and Health," *University of Pennsylvania Law Review* 115, no. 1 (November 1966): 22–38; Michael Meltsner, *With Passion: An Activist Lawyer's Life* (Northport, NY: Twelve Tables Press, 2017).

19. Personal correspondence with Messrs. Meltsner and Harper, January 15, 2018.

20. Marable et al v Alabama Mental Health Board, Complaint, Filed November 17, 1967, p. 1, USDC Montgomery, AL Civil Case Files 74-C-0813 Box 149, NARA Atlanta.

21. Thanks to Nathaniel Porter and Angie Green in the dataviz studio at Virginia Tech for their work in creating these maps.

22. Kathleen Carley et al., "Toward an Interoperable Dynamic Network Analysis Toolkit," *Decision Support System* 43 (2007): 1324–47.

3. Can Network Analysis Capture Connections across Medical Sects? An Examination of Allopathic and Alternative Disability Research in Twentieth-Century Europe and the US

KATHERINE SORRELS

My research project concerns the international dissemination of a medical network rooted in 1920s Austria. My aim at the outset was to use social network analysis to do a bibliometric or citation analysis to determine the degree to which the network remained intact intellectually after its geographic dispersal. Publications can be a useful way to gauge connections within a network (or the very existence of one) because they are a record of communication between scholars. We can study the way ideas circulate among a group of authors by analyzing the platforms, in the form of journals and presses, that they used to communicate their work. My network is, however, somewhat unusual. I am working on alternative medicine and asking network analysis to do different tasks than those that traditional citation network analysis has accomplished. I hope network analysis can help me see the degree of isolation from the allopathic mainstream that alternative practitioners operated in. I ask whether my data show the boundaries between sects to be as clearly defined as we usually assume them to be, or whether networks of ideas and research trends transcended sectarian boundaries. In the process, I engage an ongoing discussion about the advantages and pitfalls we as researchers encounter when we reduce the complexity of humanistic research in order to produce

the unambiguous questions and clean data that network analysis requires. Finally, I reflect on whether the data we use in digital humanities research merely illustrates the divide between medical sects or in fact helps to create it.

I begin with an overview of the larger book project to which my network analysis contributes. I then discuss my network analysis process, from the design of the research questions to the building of databases and the construction of network diagrams in Cytoscape. Finally, I conclude with some thoughts on the questions, observations, and next steps that came out of the project.

The Research Project

We are in the midst of a rapid transformation in our understanding of the autism spectrum and other intellectual and developmental disabilities (IDD). Activists with Down syndrome and autism have become powerful voices for a movement that challenges us to view IDD as difference, argues for inclusion, and champions self-determination. Along with this movement has come much scholarly and popular interest in the history of IDD, but the picture that has emerged misses a story crucial to understanding where we are today.

That story begins in April 1939, when nine-year-old Peter Bergel and his parents set out from Amsterdam for a small village in northern Scotland. Jewish refugees from Frankfurt, they had fled to Amsterdam in 1937 and applied for visas to the United States. Scotland was not their first choice. Although his parents were granted US entry, restrictions against "defectives" scuttled Peter's application.[1] He had contracted encephalitis as a three-year-old and was left with permanent brain damage. His Jewishness and his disability made him a double target in Nazi Germany. In 1933, eugenics legislation mandated forced sterilization of people with disabilities. Within five years, mass killing was sanctioned. The British Home Office granted Peter a visa because his parents found

a doctor in Scotland willing to care for him. In a small village outside Aberdeen, Dr. Karl König, himself a German Jewish refugee, had just secured permission to open Camphill Special School, a residential care village for children with IDD. Peter was to be his first patient.[2]

In an era when the response to disability was shame, blame, and institutionalization, Camphill was founded on the principle that children with IDD could enrich communities and that doctors should abandon the search for cures. König's radical position was rooted in his unusual approach to medicine. He was a follower of the Austrian occult philosopher, Rudolf Steiner, whose ideas spawned alternative medical, educational, and agricultural movements. Steiner began as a Goethe scholar, but soon discovered theosophy and became the leader of the German Theosophical Society. In 1912, he broke off from theosophy to establish his own occult movement. Called anthroposophy, Steiner defined the movement as a philosophy which held that higher, spiritual worlds could be accessed through what he called "spiritual science." Spiritual science was the inner work necessary to develop the tools to understand the spiritual world in a rational, scientific manner. These tools were not the kinds of gadgets that spiritualists used to detect ectoplasm. Rather, the anthroposophical tools for accessing higher worlds were the faculties of perceptive imagination, inspiration, and intuition. In addition to building on theosophy, anthroposophy drew on German idealism and mysticism, as well as Christian theology. Theosophical ideas about the origin of the world in Atlantis and the workings of karma and reincarnation blended with the belief that history is shaped by positive and negative *impulses*. Christ, for example, was understood as an impulse, as was a German or Middle European cultural mission to the world.[3]

König discovered Steiner as a medical student in Vienna. For a few years after his graduation in 1927, he tried to blend research and clinical care in allopathic medical institutions with his anthroposophical approach to medicine. This entailed bringing anthroposophical ideas about spiritual evolution into embryology, and incorporating homeopathy and a spiritual approach to

diagnosis into medical care. Within a year or two, he abandoned the attempt to blend traditions and moved to anthroposophical headquarters in Switzerland. There, he worked at the Clinical Therapeutic Institute in Arleshiem, near Dornach, under Dr. Ita Wegman, the Dutch physician who had co-founded anthroposophic medicine with Rudolf Steiner. König got involved in a growing network of doctors and teachers around Wegman who were interested in *Heilpädagogik* (curative education) for children with disabilities. By 1930, he had married a member of this network and settled in Pilgramshain, lower Silesia, where he established a successful anthroposophic pediatric practice. In 1936, König and his family fled Nazi Germany for Vienna, from which they fled again in 1938 for Scotland, where they established Camphill Special School.[4]

In spite of his unusual credentials, König was able to secure state support and a loan from the Scottish Council for Refugees. This allowed Camphill Special School to grow and establish a network of sister villages. By the 1950s, the network had spread from Scotland to England and Ireland. As the movement grew, it inspired and made connections to sister movements, extending a transnational network of intentional communities caring for people with disabilities.[5] In the 1960s and 1970s, hippies, activists, and conscientious objectors flocked to the villages and started new ones in the UK, North America, Southern Africa, and Central Europe. Camphill became a center of the counterculture. König guided this expansion, serving as the intellectual and spiritual leader of the movement and continuing to publish on IDD and a wide range of other topics until his death in 1966. Today, Camphill includes over 130 communities extending to Eastern Europe, the Middle East, and South and East Asia, and it continues to attract support from prominent artists and public intellectuals.[6] Its story lies at the intersection of some of the defining events and cultural currents of the last century, including mass migrations, the emergence of the counterculture, the rise of alternative medicine, and the growth of the disability rights movement.

Camphill has grown into a global movement, but its story is rooted in the history of medicine in Central Europe. Karl König was part of a generation of Viennese physicians and psychoanalysts working toward new understandings of child development. This group included Hans Asperger, of the eponymous diagnosis; Leo Kanner, who introduced the autism diagnosis; and Bruno Bettelheim, the psychoanalyst who popularized the "frigid mother" theory of autism. The network dispersed in the interwar period, but its members continued to transform the field. Leo Kanner (b. 1894, Klekotiv, Austria-Hungary) emigrated to the US in 1924. After four years at the state hospital in Yankton, South Dakota, he moved to Maryland and spent the rest of his career at Johns Hopkins. During the Second World War, Kanner, who was Jewish, helped get hundreds of Jewish physicians out of Nazi Europe. He retired from Johns Hopkins in the early 1970s, but remained active in the field until his death in 1981. Bruno Bettelheim (b. 1903, Vienna, Austria-Hungary) emigrated to the US in 1939 after imprisonment for just under a year in Dachau and Buchenwald. He was also instrumental in getting other Jewish refugee physicians out of Nazi Europe and into positions in the United States. He spent his career as a professor of psychology at the University of Chicago. There is much controversy around Bettelhiem, the PhD in Art History which he misrepresented in various ways, his falsification of evidence and plagiarism, and his abuse of students and patients. Much of this controversy came to light after his death in 1990. Hans Asperger remained in Austria, served as a medical officer in Croatia during the Second World War, and resumed his work on autism in Austria after the war until his death in 1980. Under Nazi rule, he modified his analysis of disability to accommodate Nazi ideology and collaborated with the euthanasia program.[7]

The literature on the history of IDD in the US acknowledges, but assigns no particular significance to, the Central European origins of its protagonists.[8] Yet IDD research in interwar Vienna and in Central Europe more broadly drew on an interdisciplinary cultural and intellectual milieu that produced strikingly original and creative

work in science and medicine.[9] To give just one example, all three figures had a serious interest in poetry as students, which they maintained and even published on later in their careers. König also shared this interest.

Against this background, I would like to use network analysis to determine to what degree, if at all, this dispersed group of doctors continued to constitute a medical/intellectual network. This seems like a straightforward undertaking, but it has broad implications. If a network persisted and encompassed both alternative and allopathic practitioners, it would reveal continuities across medical traditions. In line with recent literature that explores and contextualizes what were seen as eccentric, heretical, or simply embarrassing works by great scientists and writers (e.g. Newton's alchemy or Goethe's science), an account of the pioneers of IDD research that includes both allopathic and alternative traditions might not only include new figures, but also previously ignored work.[10] Kanner's first book, for example was *Folklore of the Teeth.*[11]

Methodology (or, Trial and Many Errors)

First Attempt

My first step was to get a sense of the kinds of questions that network analysis is well suited to answer, as well as a basic command of the field's vocabulary.[12] Then, to tackle my question about the degree to which dispersed Austrian IDD doctors continued to constitute a medical/intellectual network, I decided to start with an analysis of each figure's publications. A problem presented itself immediately: the bibliographies turned out to be vastly different in length and character.

After a handful of articles on embryology in allopathic journals early in his career, König worked exclusively with anthroposophic publishers. And once he made this shift, he became tremendously prolific, publishing over 520 articles and books on a wide variety

of topics including disability, curative education, folklore, animals, history of medicine, and spirituality. Even after I culled publications in newsletters and material printed for use within the Camphill movement, König's 496 entries dwarfed Bettelheim's 204, Kanner's 133, and Asperger's 27.

These numbers reveal the difficulty of running comparisons across sects. Kanner and Bettelheim published under similar conditions and in the same professional context, broadly speaking, so a comparison of their works pulled from American library databases rendered a fairly reliable basis for comparison. Adding König made the comparison lopsided. It is safe to hazard that König's vastly longer bibliography reflects the fact that he became the leader of a spiritual movement. His followers have gone to great lengths to publish everything he wrote, however short or informal. There may also be duplicates in the list, as texts were sometimes edited and reprinted under new titles when older versions went out of print. Moreover, Asperger's contrasting short bibliography may also be misleading. I generated it based on data from the Austrian and German National Libraries but will have to follow footnotes in the literature to determine whether this is complete. My impression is that it is not.

Setting these concerns aside for the moment, I built a database of publications for each figure to use as the basis for a bimodal edge list consisting of titles and publishers, with an additional column of tags for each publication identifying its primary field and/or topic, and color coding to indicate the years in which texts were published. I hoped to use these edge lists to create visualizations that would show where and on which topics each figure was publishing and reveal change over time through color coding. I had a vague notion of producing something sort of like a citation index visualization.

Figure 3.1: Example of a citation index visualization[13]

I started building a database of König's works. I then used this as a basis to create a bimodal edge list that showed König's publications and presses/journals. I then tried to create a node list that would allow me to visualize the publications by field, but I soon realized that this was impossible. As noted, König published on a wide variety of topics. The problem is that, with the exception of the *British Journal of Homeopathy*, the journals in which he published were not field specific. Articles on medicine could appear in the same issue with works on art, spirituality, pedagogy, etc. His books often had vague or esoteric titles and were similarly difficult to classify. Disciplinary keywords would have been easy to collect from metadata for Kanner and Bettelheim's publications; for König, I not only needed to create the data myself, but realized I couldn't. This brought into focus the ways in which the classifications I took for granted were generated specifically for allopathic medicine. While they are extremely useful in making comparisons *within* allopathy, they impede comparisons across sects.

Instead of a topical node list, I created one for publication year so that I could color code by decade. I thought this would at least allow me to visualize change over time in König's publishing. I uploaded to

Year	Title	Journal/ Publisher	Citation Details	
1932	The Being of Man and the Festivals	Anthroposophy	vol. 7	no. 4
1933	On the Illness of our Time Encephalitis and Angina pectoris	Anthroposophy	vol. 8	nos. 3/4
1932	Der Mensch und die Jahresfeste	Arbeiten aus dem Heil- und Erziehungsinstitut Schloss Pilgramshain	Paper 1	August 1932
1966	Music Therapy in Curative Education	Aspects of Curative Education		
1964	Denken – Schauen – Sinnen. Ein Hinweis auf die letzthin erschienenen	Bände der Schriftenreihe		
1954	Versuch einer geisteswissenschaftlichen Theorie der im Electro-Encephalo-gramm erscheinenden Phänomene	Beiträge	vol. 7	no. 1
1950	Der dreifache Eisenprozess im Menschen	Beiträge	vol. 2	nos. 7/8
1951	Die Bedeutung des Kosmischen Eisens im Menschen	Beiträge	vol. 3	nos. 9/10
1952	Buchbesprechung: M.M. Moncrieff The Clairvoyant Theory of Perception	Beiträge	vol. 4	nos. 7
1952	Zum Problem der kindlichen Taubheit	Beiträge	vol. 4	nos. 9/10
1953	Eugen Kolisko Im Gedenken an den Freund	Beiträge	vol. 6	nos. 11/12
1955	Die Nerventätigkeit kann nur durch eine Methode der Ausschliessung erfasst werden	Beiträge	vol. 8	nos. 3/4
1955	Samuel Hahnemann und seine Zeit	Beiträge	vol. 8	nos. 1

Title	Journal/ Publisher
Superintendent's Report, 31st January 1952–31st January 1955	
Über schwere Kontaktstörungen im Kindesalter und deren Behandlung mit der Substanz Thalamos	Der Merkurstab
Die menschenkundlichen Grundlagen des Rechnens	????????
The Human Soul	?????
The Foundation Stone	??????
An Inner Journey through the Year: Soul Images and the Calendar of the Soul	Floris Books
The Calendar of the Soul	Floris Books
Becoming Human: A Social Task	Floris Books
Communities for Tomorrow	Floris Books
At the Threshold of the Modern Age: Biographies Around the Year 1861	Floris Books
Brothers and Sisters: The Order of Birth in the Family	Floris Books
Kasper Hasuer and Karl König	Floris Books
Animals: An Imaginative Zoology	Floris Books

Cytoscape and the result was a huge visualization. It is essentially a series of balls of different sizes, which is helpful in as much as it is clear at a glance which journals and presses published the bulk of König's work. And if one zooms in and looks at titles, one can begin to get a sense of the topics on which he published with each journal or press. The color coding was largely unsuccessful; I picked one color per decade but, because the publications covered 10 decades, the differences between shades of color had to be too slight to distinguish easily. Also, each node was too small to see without zooming in so far that only a few points could be seen together.

Table 3.3: Node list of König's works

Title	Year
Superintendent's Report, 31st January 1952-31st January 1955	1955
Über schwere Kontaktstörungen im Kindesalter und deren Behandlung mit der Substanz Thalamos	2007
Die menschenkundlichen Grundlagen des Rechnens	2002
The Human Soul	2006
The Foundation Stone	2002
An Inner Journey through the Year: Soul Images and the Calendar of the Soul	2010
The Calendar of the Soul	2010
Becoming Human: A Social Task.	2011
Communities for Tomorrow	2011
At the Threshold of the Modern Age: Biographies Around the Year 1861	2011
Brothers and Sisters: The Order of Birth in the Family	2012
Kaspar Hauser and Karl König	2012
Animals: An Imaginative Zoology	2013

Figure 3.2: Bimodal visualization of König's works

Absurditäten.

Wie Karma wirkt. Reisebriefe aus Süd–Afrika.
Buche über die gegenwärtige... Brief.
der 'Tragik' der Bewegungsgestalten und
heilpädagogischen langgebilde
Diagnostik und Therapeutische Schicksale im Schatten des
erlebt, aber nicht erkannt
Von der Würde des... Darwinismus. 2
Menschen... Keyserling, Sa...and Zw Brief.
Vor hundert Jahren... Aus Justinus Kerners
Gedichten und Gedanken.
Die Elefanten.
Das Ringen um... Charles Darwin und das
Menschen zu... Schicksal des Darwinismus.
Jahrhundertwende.
Schicksale im Schatten des... Kaspar Hauser...Vom Leben der Pinguine.
Darwinismus... Mitteleuropas im
Die Wanderungen der Aale... Jahrhundert.
und Lachse. Göttingen. Die Taube als heiliger Vogel.
Bruder Pferd, part 1.

Figure 3.3: Bimodal visualization of König's works

I had done some reading on the potential hazards of bimodal
networks and how attempting to measure centrality in them can be
misleading.[14] I thus followed Miriam Posner's tutorial on converting
bimodal edge lists into unimodal ones.[15] This involved downloading
R and RStudio, following the tutorial, doing some troubleshooting,
and making some mistakes (like unnecessarily converting an Excel
spreadsheet into a CSV file, which threw off the whole process). I
uploaded the finished unimodal edge list to Cytoscape and ended
up with a visualization that, frankly, didn't tell me anything new.

It was fun learning a little bit about R and getting a sense of the
possibilities for more advanced network analysis, but I was unsure
what to do next. Simply repeating the process for my other three
key figures was not going to get me very far in understanding

the relationships among them. The Viral Networks meeting at the National Library of Medicine helped me realize I needed to return to basics and refine my questions in order to build edge lists and create visualizations that would help to advance my project.

Figure 3.4: *Bimodal visualizations of König's works*

Figure 3.5: *Unimodal visualization of König's work*

Figure 3.6: Unimodal visualization of König's work

Figure 3.7: Unimodal visualizations of König's work

Second Attempt

In my first attempt I had aimed for a visualization that was too complicated and was supposed to illustrate too many different things: year, publisher, topic, and change over time. The literature on network analysis also makes clear that this is a typical mistake. Most humanists, when they first begin working with network

analysis, try to make visualizations that show too much. We tend to be reluctant to let go of complexity and we resist the necessity to break questions down into very basic, component parts.[16] In fact, my experience was not so much that I was worried about obscuring complexity, but that I was not used to breaking questions down into components well suited to network analysis or to using sources as data. It is simply conceptually foreign for me to take apart a bibliographic reference and to discard parts of it irrelevant to an edge list. As a cultural and intellectual historian, I am not used to using my sources as data points.

Ultimately, I realized that I could do a series of discrete analyses using bibliographic data in order to answer the various questions about the strength and character of the network I am studying. But for now, a first step toward illustrating whether my four figures were part of a professional network or not involved simply illustrating the overlap (or lack thereof) in publishers among the four authors. If they shared publishers, I could infer that they were writing for some of the same audiences and they were recognized as authorities on a shared set of fields by the editors and peer reviewers who accepted their work. This required one edge list that included all four authors and the presses and journals they published with. Titles and years were irrelevant to this one, discrete visualization.[17]

Before creating this edge list, I had to finish building databases for all four key figures. I made one each for works by Leo Kanner, Bruno Bettelheim, and Hans Asperger, covering their early work in Vienna through their careers in the UK and the US. I then cleaned up the databases, creating consistent entries for data pulled from various libraries with different referencing conventions and in different languages. As noted above, I was left with a lopsided dataset. I was working with comprehensive lists of Karl König's publications, which included privately published manuscripts, pamphlets, and lectures printed for circulation in the Camphill movement. Thus my database of over 520 items for König was more than twice the size of the others combined. To combat this problem, I eliminated all works

by König that were privately published as well as articles published by individual Camphill communities, keeping only articles published in books and journals.

Figure 3.8: Combined visualization of all authors' works

Figure 3.9: Close-up of combined visualization of all authors' works

Table Panel	
Shared Name:	Name:
American Journal of Orthopsychiatry	American Journal of Orthopsychiatry
Archives of Neurology and Psychiatry	Archives of Neurology and Psychiatry
C.C. Thomas	C.C. Thomas

As the visualizations illustrate, all four figures were relatively isolated from one another as measured by publishers. Kanner and Bettelheim were connected by two journals and one press: *American Journal of Orthopsychiatry*, *Archives of Neurology*, and C. C. Thomas (each represented by a yellow node in figure 3.9). Asperger and König had no publication links to anyone else. This tells me that the four figures were not part of a publication network in Vienna before three of them emigrated. Bettelheim and Kanner both published in major journals, but Bettelheim tended to publish in more social scientific venues, such as the *American Journal of Sociology* and *The Elementary School Journal*, whereas Kanner tended to stay more strictly within medicine, publishing in the *Journal of Pediatrics* and the *American Journal of Psychiatry*. This reflects their training: Bettelheim's was in Philosophy and Art History while Kanner's was in medicine. And the fact that König is not the only one isolated in this visualization suggests that, in my focus on the question of divisions between medical sects, I had been overlooking the importance of geography. Kanner and Bettelheim worked with a few of the same publishers, not only because they had similar research interests and operated in the same medical sect but also because both worked in the American academy. Even if there had been more fluidity between medical sects, it is unlikely that König would have shared publishers with Kanner and Bettelheim; most of his work came out in British, Swiss, and German journals and books. And Asperger published exclusively in German. Nevertheless, I remain surprised by the complete lack in overlap at the beginning of their careers, when they were all in Vienna. This suggests that I should pay more attention to divisions at the University of Vienna, which was famously fractured in the interwar period.

The diagram clearly illustrates divisions more than connections, and I wondered whether narrowing the dataset to show only the journals and presses in which each author published most would reinforce or weaken that finding. I narrowed the edge list to include only those journals and presses with which authors published five

or more works (see below). Two things stood out. First, the shared publishers (represented by the yellow nodes) remained in the leaner diagram, which shows that the professional network linking Bettelheim and Kanner was perhaps tighter than the previous diagram seems to suggest. Second, the disparity in the number of publications between König and the other three figures is more accurate and apparent. His network diagram dwarfs those of the other three. Again, this is misleading, because the diagram cannot represent the vastly different professional culture and publication conventions within which the four figures worked.

Figure 3.10: Overview of visualizations of dataset reduced to presses and journals with which authors published five or more texts

Figure 3.11: Close-up visualizations of dataset reduced to presses and journals with which authors published five or more texts

Conclusions

I have learned that a) designing a network analysis project involves working backward from complex questions to simple, discrete ones; and b) the only way to learn to do this is through trial and error. For

example, I had begun with what I thought was a suitable question: do chosen publication topics and publication venues illustrate the existence of a network among four figures, and does the strength of the network change before and after emigration? I first broke this down to ask: what is the degree of overlap in four figures' publication venues before and after emigration? At this stage, I left publication titles in, failing to recognize that they were, essentially, clutter. Finally, I ended up with the question: to what extent, if any, did four figures publish in the same venues?

This has been a very helpful process. It has forced me to think of each component part of large and complex questions, questions that I had assumed were simple and discrete. It has helped me realize that I often give short shrift to pieces of evidence that I see as obvious. My visualizations do not reveal something that I couldn't have correctly guessed by sorting and reading through bibliographic databases I created for each figure. But I can now demonstrate the professional isolation between my four figures concretely, rather than simply anecdotally. In the process, I have had to slow down and think more about how that isolation came about and what it means, which in turn has added more depth to my research.

Finally and most importantly, this first network analysis project has raised new questions. For example, would a citation analysis or even a full-text analysis of all four figures' work reinforce König's and Asperger's isolation, or might it reveal a shared set of concerns among some or all of the figures, which they explored in different professional contexts? Such projects would undoubtedly advance my project, but I can already anticipate more problems posed by the attempt to transcend medical sects. I also anticipate new concerns about what my visualizations miss or obscure.[18] Finally, the practical obstacles remain. I cannot use existing databases to search for citations. I know from traditional, close-reading and archival research that König cited Asperger and that Asperger referenced König in a conference talk. The only way to build a citation edge list would be to search full texts.

In conclusion, network analysis offers a basis from which to discuss my key figures' relationships and the ways in which they are situated in a broader context, but the methods traditionally used to visualize professional and intellectual networks are not well equipped to work across disciplinary and national boundaries. In order to move forward with this project, I will need to rethink the questions I ask with the complexity and unevenness of my source base in mind.

Acknowledgments

Endnotes

1. Douglas Baynton, *Defectives in the Land: Disability and Immigration in the Age of Eugenics* (Chicago: University of Chicago Press, 2016).

2. On Peter Bergel's escape from Nazi Germany and his life in Camphill, see Alan Potter, "Intentional Community as a Continuing Response to the Holocaust: The Life of Peter Bergel and the Camphill Communities," published by Camphill Community Botton Village (n.d.).

3. Peter Staudenmaier offers a summary of key anthroposophical ideas and beliefs in his study of anthroposophy under Nazism. See Peter Staudenmaier, *Between Occultism and Nazism: Anthroposophy and the Politics of Race in the Fascist Era* (Leiden: Brill, 2014).

4. Karl König and Peter Selg, *My Task: Autobiography and Biographies* (Edinburgh: Floris Books, 2008), Hans Müller-Wiedermann, *Karl König: A Central European Biography of the Twentieth Century* (Botton Village, Malton, UK: Camphill Books, 1996).

5. In 1964, Jan Vanier, a Canadian Catholic, visited a Camphill community in North Yorkshire, England and went on to found L'Arche, a movement that operates on very similar principles to Camphill. Strong ties continue to exist between two movements. In 2015, Vanier won the Templeton Prize, which honors exceptional work in spiritual matters. He was nominated by John Swinton, a professor in the Divinity School at the University of Aberdeen who has also done work on (and commissioned by) Camphill. For Camphill's characterization of the influence on Vanier, see *Camphill Pages* (newsletter of the Association of Camphill Communities UK and Ireland), Spring, 2015. For Vanier's Templeton Prize announcement, including commentary by Swinton, see http://www.templetonprize.org/previouswinners/vanier.html (accessed May 23, 2018). For Swinton's work on Camphill, see John Swinton, Aileen Falconer, and Stephanie Brock, "Sensing the Extraordinary within the Ordinary: Understanding the Spiritual Lives of People Living and Working within Camphill Communities," a work commissioned by the Camphill Village Trust and funded by them with contributions from the Anthroposophical Medical Trust and Camphill Medical Practice Ltd (n.d.).

6. Friedwart Bock, ed. *Builders of Camphill: Lives and Destinies of the Founders* (Edinburgh: Floris Books, 2004).

7. Herwig Czech, "Hans Asperger, National Socialism, and 'Race Hygiene' in Nazi-era Vienna," *Molecular Autism* 9, no. 29 (2018), doi: https://doi.org/10.1186/s13229-018-0208-6; Edith Sheffer, *Asperger's Children: The Origins of Autism in Nazi Vienna* (New York: W. W. Norton, 2018). On all three figures, see Steve Silberman, *NeuroTribes: The Legacy of Autism and the Future of Neurodiversity* (New York: Avery, 2015), John Donvan and Caren Zucker, *In a Different Key: The Story of Autism* (New York: Broadway Books, 2016), Adam Feinstein, *A History of Autism: Conversations with the Pioneers* (West Sussex: Blackwell, 2010).

8. Their research and perspectives were connected from the start through shared influences. See John E Robison, "Kanner, Asperger, and Frankl: A Third Man at the Genesis of the Autism Diagnosis," *Autism* 21, no. 7 (2016): 1–10.

9. See Karl Schorske, *Fin-de-siècle Vienna: Politics and Culture* (New York: Knopf, 1979); Deborah Coen, *Vienna in the Age of Uncertainty: Science, Liberalism, and Private Life* (Chicago: University of Chicago Press, 2007); Eric Kandel, *The Age of Insight: The Quest to Understand the Unconscious in Art, Mind, and Brain, from Vienna 1900 to the Present* (New York: Random House, 2012).

10. See, for example, Sarah Dry, *The Newton Papers: The Strange and True Odyssey of Isaac Newton's Manuscripts* (Oxford: Oxford University Press, 2014).

11. Leo Kanner, *Folklore of the Teeth* (Detroit: Singing Tree Press, 1928).

12. My discussion below gets into detail, so familiarity with those basics will be helpful. For a brief and very accessible overview, see Scott Weingart, "Demistifying Networks," *Scottbot* (blog), accessed May 23, 2018, http://www.scottbot.net/HIAL/index.html@p=6279.html.

13. Stephen Carley, et al., "Visualization of Disciplinary Profiles: Enhanced Science Overlay Maps," *Journal of Data and Information Science* 2, no. 3 (2017), doi: https://doi.org/10.1515/jdis-2017-0015.

14. See Scott Weingart's discussion here: "Networks Demystified 9: Bimodal Networks," Scottbot (blog), January 21, 2015, accessed July 16, 2018, http://www.scottbot.net/HIAL/index.html@p=41158.html.

15. See "Get a Unimodal Network from a Bimodal Network," *Github*, June 30, 2016, accessed July 16, 2018, https://github.com/miriamposner/cytoscape_tutorials/blob/master/get-a-unimodal-network.md.

16. Miriam Posner, "Digital Humanities 101: Network Analysis," *Digital Humanities 101* (blog), accessed May 23, 2018, http://miriamposner.com/classes/dh101f16/tutorials-guides/data-visualization/network-analysis/.

17. I'd like to thank Nathaniel Porter for helping get me to this point.

18. Some of the texts are on sensitive topics connected to disability and the Holocaust. For a discussion of the ethical implications of doing computational analysis of such work, see Todd Presner, "The Ethics of the Algorithm: Close and Distant Readings of the Shoah Foundation's Visual History Archive," *History Unlimited: Probing the Ethics of Holocaust Culture*, eds. Claudio Fogu, Wulf Kansteiner, and Todd Presner (Cambridge: Harvard University Press, 2015), 175–202.

4. Mapping Early Epidemiology: Concepts of Causality in Reports of the Third Plague Pandemic, 1894–1950

LUKAS ENGELMANN

The science of epidemiology has always had an intricate relationship to the history of diseases. The design of models of the dynamics that govern diseases in their relation to population is ultimately based on information and data gathered from past outbreaks. Epidemiology belongs to what Lorraine Daston has recently called "Sciences of the Archive."[1] Like astronomy, zoology, demography, or meteorology, the study of epidemics operates with objects of superhuman scale. The discipline deals with plagues that exceed historiographical periods and geographical regions; and, thus, it always requires elaborated practices of collecting, accounting, and archiving to establish its status as a discipline. Daston reminds us that despite this reliance of some "hard" sciences on the historical record, their conduct of history often differs from the perspective of humanists on the same historical event. Where exegesis, commentary, and interpretation of contexts and niches might characterize a history of diseases and epidemics, the epidemiological grasp on the historical record seeks to collect quantifiable data.

But epidemiology wasn't always a science of mathematical analysis, concerned with the production of formal expressions and the elaborate design of stochastic models. The epidemiology of the late nineteenth and early twentieth centuries is best described as a broad interdisciplinary project, suspended between isolated academics in medical schools and a growing group of governmental medical officers applying a mixture of methods, integrating

historical, anthropological, sociological, statistical, and medical approaches to understand diseases in relation to populations and environments.[2]

Nineteenth-century epidemic outbreaks of cholera, smallpox, or bubonic plague were not captured in statistical data alone, but were regularly packaged into narratives. These narratives were built around detailed observations to discuss and propose arguments about causes, the significance of local conditions, and the efficiency of mitigating practices. The genre of the outbreak report is often ignored in the historiography of epidemiology, which predominantly focuses on the development of statistical methods and mathematical models. However, the narrative form of capturing and classifying epidemic outbreaks was crucial to the broad interdisciplinary nature of epidemiological reasoning at the time. Historically, the genre of the outbreak report exhibited similarities to the clinical case report and its capacity to stitch detailed observations of singular cases to systematic considerations of the characteristics of the disease.[3] Much in the same way, the outbreak report presented a singular outbreak to other epidemiologists to engage debates about common aspects of particular local conditions and to contribute to the production of generalizable characteristics of an epidemic.

The aim of this chapter is to rediscover the outbreak report as a long-overlooked source of fine-grained and systematic epidemiological observations. The texts contain a wide range of valuable information, reaching from individual case reports over dispersed mortality and morbidity statistics to sections about causation theories and observations of treatment and prevention practices.[4] This information is currently not available as structured data and is dispersed throughout the texts in semi-structured formats. The first goal of this paper is therefore to evaluate pathways of extracting this information through text mining. I will present steps and considerations of a thorough analysis of the given structures of the outbreak report and will introduce formalization strategies to arrive at structured datasets, which could eventually

be attached to metadata including the location and dates of outbreaks. While this data might be of interest to epidemiologists, this paper will also provide reflections from the perspective of the historian, who is keen to preserve the value of historical analysis in this process. The guiding concern in the following pages is to design systems for structuring the narrative information that *preserve* difference, local deviation, and conceptual incommensurability within and across the reports. The historical report is not a source that enables us to refine and consolidate accurate epidemiological concepts of bubonic plague; rather, it allows for the epistemological analysis of historical ways of seeing the epidemic.[5]

The second, but by no means secondary, goal of this study is then to draw out feasible methods of extracting the structure and composition of epidemiological argumentation, to understand how epidemics were seen and how they were reasoned about. The reports allow for a careful reconstruction of the interdisciplinary nature of reasoning in pre-formal epidemiology. Historical sections illuminate the use of the natural histories of diseases. Arguments about incidence among different populations enhance our understanding of the anthropological and colonial frameworks through which epidemics were conceived. Considerations about local conditions and speculations about causes provide a basis to reconstruct the ecological and environmental arguments that underpinned much of the understanding of infectious diseases at the time.

Network analysis supported by natural language analysis enables both epidemiological as well as epistemological interests in the history of diseases. Polemically speaking, the "what" of the history of an epidemic outbreak can be brought into a productive relationship with the "how" of its interpretation at the time and place of observation. Building a model for the extraction of data about clinical observations, climatic conditions, or causal relations will have to integrate the structure and form of how these aspects were presented and will lay bare the conventions of the genre of outbreak reports. Reflecting and discussing the conceptual aspects of the

development of a pathway for successful data extraction will thus deliver insights into the structural underpinnings of the complex epidemiological reasoning from a time when epidemiological science was not predominantly perceived as a mathematical exercise.

The pilot study presented in this chapter focuses on a small sample of outbreak reports of one disease and one particular aspect of its epidemiology. I am particularly interested in reports that cover local outbreaks of the third plague pandemic from 1894 to 1950. The return of the disease from the Middle Ages ignited extensive epidemiological interest at the end of the nineteenth century. The disease's global distribution, its challenge to modern institutions of hygiene and sanitary cleanliness, as well as its unexplained dynamics on the heel of the successful identification of its infectious agent makes it an excellent case for the questions outlined above.[6] The reports offer a broad sample of late nineteenth-century conventions of epidemiological reporting as they contain a vast amount of speculations about local influences, causal relations, disease vectors, and the epidemic's containment. Finally, the duration of the third plague pandemic over six decades also bridges a timespan of dramatic epistemological transformation in the field of epidemiology, as formal methods and mathematical models began to take center stage in the 1920s.[7]

Two kinds of networks can be envisioned in this sample. The first network would include the outbreaks of plague structured by arguments made about local conditions. Each report of plague presents a node, associated with an outbreak within the network of the pandemic spanning geographical and historical dimensions. It would be possible to map outbreaks where the authors suggest a strong importance of seasonal influence or to look at those outbreaks emphasizing racial arguments about the incidence of plague. Individual cases could be compared along the global sample and treatment as well as prevention methods could be contrasted with traditional maps of plague incidence. Second, it appears to be possible to trace networks of arguments made within each outbreak

report to better characterize the epidemiological reasoning about plague in Hong Kong or Sydney and to contrast it with other cities around the world. Instruments from epistemic network analysis could be used to visualize the argumentative structures of outbreak reports as well as to visualize the observations and details associated with causality, contrasting them with the argumentative elements essential to historical narratives about plague.[8] However, these visualizations have not yet been made but, rather, stand as the goal of the project, once the structuring has been concluded.

In this chapter I describe some of the early steps necessary to achieve these network visualizations. Then I explain in detail the thought processes I applied to transform a narrative genre into a structured dataset. I focus particularly on one theme that runs through all the reports, across outbreaks in multiple places and periods: namely, the question of cause. Especially in the case of plague, questions of causality exceeded bacteriological findings in the laboratory. Despite the successful identification of *Yersinia pestis* as the infectious agent of plague in 1894, subsequent epidemiological investigation looked at configurations, vectors, and the environmental conditions that could have led the bacteria to cause infections and outbreaks. In other words, one of the most important concerns for epidemiologists working on plague outbreaks was to understand the specific local condition that had caused an unusual amount of cases of plague clustered within a confined space and developed over a short period of time. Network analysis will eventually enable a visualization of the considerations of causes with the expectation to demonstrate clearly the stark variety of identified causes between places and a shifting conceptual focus on causality over time. The first step, however, is to identify sections in the reports that are relevant to the discussion of cause. Then we need to introduce meaningful separations between different concepts of causality. First, though, we need some background.

Early Epidemiology

Modern epidemiology is conventionally considered to have begun in the nineteenth century. With the emergence of modern scientific methods, in addition to the rising significance of population as a calculable entity since the eighteenth century, epidemics became a new object of knowledge. The question that manifested itself quite distinctively in the second half of the nineteenth century was to what extent epidemics could be understood in their own right, differing from singular cases not only in quantitative but also in qualitative terms.[9] How could knowing about populations and their dynamics be exploited to better understand the conditions and laws that seem to govern epidemics? Across Europe, its colonies, and the US, a growing community of physicians, public health officials, and medical officers began to investigate repeating patterns of epidemic outbreaks of cholera, smallpox, tuberculosis, syphilis, or plague.

The epidemiologist Alfredo Morabia has suggested framing the epidemiological practice of the nineteenth century as "pre-formal epidemiology."[10] As an epidemiology void of theory and conceptual underpinning, it lacked the foundations to address its most pressing problems in a formal and systematic way. While this claim surely helped to distinguish the introduction of mathematical methods in the early twentieth-century history of the field, it is the aim of this paper to challenge such diagnostics of the nineteenth-century epistemology of epidemiology. Rather, I suggest to look at early epidemiology as a field that is defined by three distinctive, often loosely defined, but nevertheless constitutive frameworks of analysis. With Andrew Mendelsohn, we can differentiate these into statistical, environmental, and historical approaches.[11] While these three approaches might have lacked an overarching theoretical systematization, each of these frameworks were theorized and conceptualized in their own right.[12]

Perhaps the most visible (and, at least since the mid-nineteenth century, the most important) instrument in epidemiology was statistics. Famously attached to the work of William Farr and John

Snow, statistical analysis of cholera outbreaks had changed the ways in which arguments about epidemics were made. Statistics provided a reliable method of measuring and evaluating the impact of disease on society, while encouraging new ways of questioning society's own involvement in the cause, spread, and exaggeration of diseases.[13] Population was not anymore seen to be an amorphous entity, but could be separated in different populations along a broad line of concepts reaching from habitation, to nutrition, to factors like age and heritage.[14] With attempts to separate populations into affected/non-affected or exposed/non-exposed parts, both Farr and Snow took inspiration from the mathematical work of Laplace, Poisson, and Bernouli. But late nineteenth-century epidemiologists were also influenced by a number of emerging sciences in which the compartmentalization and calculation of populations took on further significance. Quetelet's early approaches to statistical mean values of physiological aspects (such as height, the introduction of evolutionary biology, or the production of economic theory) might have contributed to the attraction of statistical thinking in epidemiology. All of these approaches showed that when looking at complex human events in aggregate forms, even those intentionally and willfully created, they seem to exhibit law-abiding tendencies.[15]

Beyond the calculation of population, the environment was an important object of epidemiological consideration. To many early epidemiologists, the environment provided an ideal vehicle to conceptualize ambitious sanitary reforms merging political and medical motives. Many early epidemiologists continued the traditional skepticism of William Farr about contagion and principles of infection to advance epidemiology as a sanitary science.[16] The environment served as placeholder for a multitude of factors, which influenced the cause, distribution, and exaggeration of diseases. As Anne Hardy has emphasized, this led to the development of a "highly environmentalist, observational tradition" in the conduct of epidemiological analysis.[17] Factors like stench and noxious vapors were considered as much as bad air or emanating influences from the soil.[18] Charged with various theories and

conceptual underpinnings, the environment remained a constant epidemiological concern throughout the nineteenth century and even in the face of reductionist bacteriological aetiologies, providing an open-ended repository for the conceptualization of causation.

Third, traditional epidemiology was indebted to a historical method. Epitomized in the geographical-historical work of August Hirsch, historical narratives of the origin and distribution of epidemics were regularly considered to be of eminent analytical value in the interpretation of occurring epidemics.[19] The history of epidemics, often including their ancient origins, was more than just illustrative contextualization.[20] Instead, the historical narrative was seen as a conceptual element through which epidemics achieved their status of transhistorical entities, and understanding their history enabled diagnosis as much as prognosis. Amassing the historical events of an epidemic, so believed historical geographers like Hirsch, allowed for productive generalizations. Similar to the production of clinical records, it was the identification of series and seriality throughout an epidemic's history that contributed to its understanding in the present.[21]

Without diminishing the significance of statistical methods, it is important to acknowledge that epidemiology of the nineteenth century was fundamentally driven by text-based methods. Assessments of environment relied on refined practices of observation and their empirical, sober reporting, while the building of the historical background of an epidemic was fundamentally an art of storytelling. Although historical geography of disease included the production and invention of new forms of mapmaking, key reference works such as Hirsch's vademecum were exclusively text-based works. The outbreak reports of plague should therefore be considered to offer much more than mortality rates, case numbers, or dates relevant to the outbreak. The reports also provide both interested historians as well as epidemiologists with rich descriptions, detailed discussions, and decisive arguments about

the local environment and its multifaceted relation to the disease. Moreover, each of the reports offers its own version of the long history of bubonic plague.

The Case of the Third Plague Pandemic

This study focuses on the third plague pandemic for various reasons. Usually accredited to an outbreak in 1894 Hong Kong, the third global occurrence of plague was distributed along the trade routes of growing sea commerce and affected almost every port city in the world in the following decades.[22] But outbreaks differed in severity, mortality, and longevity, and prompted a wide range of different measures mounted to halt the epidemic's distribution. Within the first year of the new outbreak of bubonic plague, its bacteriological agent was identified, first by Shibasuro Kitasato and later by Alexandre Yersin.[23] The emerging global crisis, with catastrophic effects especially in colonial India, could not be quickly resolved despite the successful identification of the bacteria. It was rather the sanitarians and their epidemiological expertise, which became of high value to identify and to explain the mechanism through which plague was distributed.[24] Plague became a showcase for early epidemiology to demonstrate that it was the exclusive scientific practice that could explain the prevalence for plague to devastate some port cities while leaving others unharmed.

To epidemiologists in the late nineteenth century, plague must have appeared as a paradigmatic set of questions. With the problem of etiology out of the way and relegated to the laboratory, epidemiologists could demonstrate the capacities of their knowledge practices to explain an epidemic event.[25] Because this plague was a global disease–a pandemic–it also gave ample opportunity to engage with any of the large frameworks of epidemiological reasoning that persisted at the time, including population, environment, and history.

Statistical work was employed to understand precisely how plague's relationship to population differed from the disease appearance in an individual case.[26] The high mortality rate and the quick progression of the disease in individual cases led to the appearance of a slow onset of the epidemic as an aggregate of cases. Moreover, plague was often perceived through racial and ethnic filters, which in turn prompted extensive comparison of populations.[27]

Nevertheless, one of the most fundamental concerns of the plague epidemiologists was the relationship of the disease to its physical environment. This invariably included further concerns about infection pathways and of conditions of the soil or food, which might provide opportunities for bacteria to survive outside of the human host.[28] What kind of surroundings did encourage or diminish the course of the epidemic? Under which conditions did the bacteria thrive, and what contributed to its containment? What emerged was not only a re-fashioning of the old sanitarian's obsessions with cleanliness and hygienic appearances, but a new focus on conditions under which a bacteria's capacity to infect and to lead to the outbreak of a case of plague was increased or attenuated. This subject, often referred to at the time as *virulence*, marked precisely the difference between the observed behavior of a bacteria in the laboratory and the invisible conditions of it leading to a disease on the epidemic streets.[29]

Plague was also widely seen as the return of a historic disease, a disease of the Middle Ages that had been overcome by Western civilization. This history was used as a repository for symptom-based diagnostics, comparing old descriptions to the occurrences in the nineteenth century. But references also were drawn regularly to the epidemic's younger history, comparing outbreak reports from Egypt and Russia with the series of events that characterized the third plague pandemic.

Finally, with the arrival of the third plague pandemic, the transnational dimension of epidemiology would prove to be crucial. Plague was perhaps one of the first epidemics registered by its

contemporaries as a global event. Epidemiologists had to develop a system of accurate comparison that sought to understand the difference in places with regards to all of the factors above. Different populations with varying demographics were subjected to changing climatic conditions, followed different cultural customs, were considered to belong to different racial, ethnic or cultural groups, and had developed different ways of responding to the plague. Outbreaks in cities around the world needed to be compared and discussed along the lines of their statistical significance and the specifics of their environmental conditions to understand how they form an event within the series of outbreaks that formed the pandemic on a global scale. For this purpose, epidemiologists, sanitary officers, local physicians, and national health officers produced accounts of local outbreaks, written up and drawn together in outbreak reports which were then disseminated globally.

The Bubonic Plague Report

Almost every significant outbreak and many minor incidents of plague have been reported in a more or less formalized way since the first outbreak of the third plague pandemic in Hong Kong in 1894. My non-exhaustive list of reports consists currently of about 50 unique entries. For pragmatic reasons, the list is limited to English-language reports.[30] For the purpose of this study, I excluded reports that provided only a general account of the disease as well as those that focused on a single case. All of the reports in the list discuss the specific occurrence of multiple plague cases clustered around a location and occurring within a limited timeframe. While the geographical scope of a report is usually urban, I have also included reports considering nations or regions.

Methodologically, I have considered linguistic approaches to the definition of the epidemiological outbreak report as a genre of communication. The report could then, however anachronistically,

be considered consistent with English for Specific Purposes (ESP).[31] Here, as discussed by Bathia, a definition would apply in which the outbreak report is seen as a "communicative event with a particular purpose which is readily identified by what they refer to as its discourse community (those people who regularly engage in it)."[32] The report achieves its purpose through the realization of a sequence of what Swales and Bhatia have called *moves* and component *steps*. While the sequence may vary–moves and steps might occur in different orders and different realization patterns–each sequence component can, in theory, be isolated and analyzed as a schematic structure. Looking at the epidemic outbreak report, the following questions are essential:

A) What is its communicative purpose?

B) How were these purposes achieved through the schematic structuring of its moves and steps?

C) To what extent can a systematic schematic structure be generalized across the genre?

I assume here that the epidemic outbreak report serves the overarching communicative purpose of describing and explaining the relationship between the disease and the location for which the report is written. This relationship is complex, and its variation from case to case and from report to report is of key interest to this pilot study. My hypothesis is that all reports–despite the multitude of possibilities in which local conditions are described and related to the variable understandings of bubonic plague–follow a fairly conventional way of presenting and structuring their arguments, as they utilize the same *moves* and *steps*. After all, the corpus of reports can be considered a genre because each report tends to follow conventions of reporting that address concerns of the intended audience, usually government officials or fellow epidemiologists.

A first step to zone the documents along the scheme that undergirds the reporting is based on the structures that report authors have applied through headings and sections. Additional to the standard inventory–a preface, an introduction, and occasionally a conclusion–all other sections of the reports appear to repeat

a scheme characteristic for reporting on plague outbreaks across places and time. After the aggregation of all sections from all reports in this sample, 11 categories have been devised to cluster the majority of existing sections. This scheme preserves the *moves* and *steps* of the outbreak reports, and although it doesn't necessarily reflect their original order, it enables comparison of these steps across the reports and thus across outbreaks.

Table 4.1: Sequence titles that represent the scheme of reporting on epidemic events identified across the outbreak reports in the given sample

#	Sequence title	Description of Content
1	Title matter, preface	Title page and letters in the preface
2	Introduction	State of the epidemic at the time of the production of the report, summary of key features, evaluation of significance of the epidemic, history of disease, history of outbreak, short overviews of the epidemic's course
3	History of Disease	General points on the history of the epidemic, origin of outbreak
4	History of Outbreak	Geographical and chronological overview of local outbreak
5	Local Conditions	Descriptions of key elements that are considered noteworthy by the author in relation to plague
6	Causes	Causes identified by the author. Usually points of origin, specific local conditions or descriptions of import, later zoonotic factors
7	Measures	List of the measures undertaken to curb the outbreak, sanitary improvements, quarantines, disinfection or fumigation and rat-catching, poisoning, education, behavioral changes, treatment given as prophylaxis
8	Clinical Appearance	Description of the diseases appearance, its usual course and its mortality
9	Laboratory	Description of bacteriological analysis, other laboratory work
10	Treatment	Description of the treatment given to patients
11	Cases	List of individual cases, usually with age, gender, occupation, course of disease, and time and dates of infection and death

Table 4.1 indicates the sequence titles that I have chosen to apply on the aggregated section titles from the outbreak reports. I added a short description of the expected content of the sequences. Some reports have additional sections, which are concerned with details

beyond this scheme; these will be registered for the time being as "other." Additionally, many of the shorter reports do not have sections, so I have broken up the text where possible into the appropriate categories.

Visualizing Causation : Three Examples from Bubonic Plague

My goal here is to a) consider arguments made in the reports about the causes of bubonic plague in specific outbreak locations, and b) showcase a possible way to structure those arguments. To this end I have identified the sections across the sample that can be identified with the sequence title "Causes" and have transferred them into a discrete dataset for further analysis. After experimenting with various tools and instruments I found simple word counts to be surprisingly accurate to match the arguments presented by the reports. To this end I counted the frequency of significant terms in the sections identified and classified as "Causes." Afterwards, a classification of significant words among the ten most frequent terms provided for a vague, but accurate, identification of argumentative classes. These classes could be translated to match themes or motifs that were considered by the authors of the report when looking into the local causes for an outbreak. I will present here three examples to demonstrate the method.

The first example is taken from a report on Hong Kong's 1894 plague outbreak, the first outbreak in the history of the third plague pandemic. The author of the report is the colonial medical officer James Lowson, and in it Lowson includes a section titled "Causes" in which he discusses his observations and hypothetical considerations of what caused plague to appear suddenly and devastatingly in the district of Taipingshan in Hong Kong.[33] After removing stop-words and standardizing multiple forms, the

resulting list gives a clear picture of Lowson's thinking on what caused plague. I applied a preliminary classification of the terms to quickly visualize the characteristics of causation this report implies.

Table 4.2: Standardized word count for "causation" sequence in outbreak report for 1894 Hong Kong

Count	Term	Class
23	Latrine	Built Environment
14	House	Built Environment
10	Street	Built Environment
8	Case	Condition
7	Epidemic	Condition
6	Disease	Condition
5	Chinese	Population
5	Overcrowding	Population
5	Well	Built Environment
5	Hong Kong	Location

This simple analysis shows that Lowson is focused on the material configurations of the urban environment. "Latrine," "house," and "street" appear as the pivotal points of concern, here classified as aspects of the "built environment." By associating the terms "case," "epidemic," and "disease" with the class "condition" Lowson leads one to expect that at least a number of sentences in this sequence will include strong connections–or at least significant proximity–between terms indicating "condition" and those associated with "built environment." The following two terms ("Chinese" and "overcrowding") further indicate that the built environment is accompanied by the allocation of causes to Chinese aspects, here coded as a qualification of the class "population." This weighted word list demonstrates the sanitary perspective of Lowson, and the order visualized in the table resembles his argument that plague was driven by what he conceived of as an unsanitary state of Chinese life, manifested in the built environment.

The second example is a report written by Ernest Hill from the London School of Hygiene and Tropical Medicine concerning the outbreak of plague in the South African city of Natal in 1902.[34] Two sequences zoned as "Cause" are titled "Relation to Race, Sex, Age, Occupation, and Surroundings of Dwellings" and "The Manner in which the Disease spread." As the title of the first section indicates, Hill did not primarily focus on the urban environment, but rather attributed the causes for the distribution of plague to the question of population.

Table 4.3: Standardized word count for "Cause" sequence in outbreak report for 1902 Natal

Count	Term	Class
52	Case	Condition
17	Infected	Condition
15	Plague	Condition
14	Person	Population
12	Tenement	Built Environment
11	Man	Population
10	Durban	Location
9	Disease	Condition
7	Indians	Population
6	Place	Built Environment

The table shows that terms associated with "condition" rank highest in this chapter. While it is difficult to ascertain why this is so, it might prove interesting to look into the significance of "cases" for the arguments made in this sequence. The association of "infected" and "person" indicates that Hill, in contrast to Lowson in Hong Kong, argued about causation mostly in connection to infected population and perhaps their behavior or their identity. While the "built environment" is not excluded from his considerations, it ranks comparably low, and the usage frequency of both "tenement" and "place" suggests a secondary significance. This ranks on the same

level as the "Indians" designation under "population," which seem to have some, but not much importance to the elaboration of causes for plague in this case.

In this example, decisive limits to this method become quite clear. These limits might be mitigated by integrating further analysis of collocation of terms to identify units of meaning beyond singular terms. However, Hill does indeed state in the text that there seem to have been no indications for a disproportionate distribution of plague cases among people he describes as "Indians." A preliminary conclusion could therefore be that the vagueness of the results listed above is indeed indicative of the vagueness present in Hills writing about causes.

The third and final example is taken from a report about an outbreak of plague in Peru in 1932. The report is written by the American epidemiologist Charles Eskey.[35] Sequences that have been zoned as "Cause" were called "Relation of rat species to plague," "Relation of flea species to plague," and the "Summary" for both of these sections. In this report, published a good three decades later than the other two, a very different picture of epidemiological reasoning about causes for plague has been established.

Table 4.4: Standardized word count for "Cause" sequence in outbreak report for 1932 Peru

Count	Term	Class
62	Rat	Animal
41	Plague	Condition
23	Cheopis	Animal
23	Building	Built Environment
20	Caught	Measures
15	Peru	Location
12	Place	Location
12	Index	Laboratory
11	Human	Population
10	Communities	Population

The word count in Table 4.4 shows a very different picture of the consideration of causes for plague. Both the highest and the third most frequent term are now concerned with animals—"rat" and the rat flea "cheopis"—which were by that time accepted as principal vectors of bubonic plague. The concern over built environment has certainly not disappeared, but in this context it appears as the environment of the principal vector rather than a concern of infection in and by itself. Furthermore, the presence of location as well as population at the end of the list is interesting; it appears almost as if the hierarchy of terms resembles the causal chain identified in the field. The word list delivers a fairly accurate picture of Eskey's perspective, as he believed that plague was indeed driven by rats and fleas and that the considerations of the built environment and geographical aspects had to be undertaken in relation to the zoonotic factors that undergird the propagation of bubonic plague before it affects humans and communities.

These three examples are preliminary. I've included them here to show how one might go about building a structured dataset out of a fairly unstructured list of documents. With the above examples, I've shown that simple word counting, within a carefully zoned sequence of text, yields results that largely match the arguments made by the authors. The word lists deliver obvious hierarchies, which indeed catch the themes and concepts of causation used in various places and times, once they have been classified in a sensible and historically sensitive way. My hope is that by expanding this method to other examples and by integrating the term collocation I will end up with a robust set of classifications useful for network visualizations.

Discussion and Outlook

This method of visualizing the conceptual underpinnings of causality in plague outbreaks is clearly far from satisfying my goal of representing the specific arguments made in each of these reports.

The word lists are useful insofar as they foreground categories and concepts that were indeed significant to the attribution of causes in 1894 Hong Kong, 1902 Natal, and 1932 Peru. The shift from broad considerations of the urban environment to a focus on population to the identification of rats and fleas as principal vectors is well aligned with the arguments presented in the reports (as well as with the historical scholarship) about these outbreaks and their perception at the time.

The method discussed in this paper offers an overview of how causation of bubonic plague was perceived differently in three places. To the historian interested in the epistemology of epidemiology, these abbreviations of the sections might be useful for the construction of concepts assumed to be influential in the production of epidemiological knowledge. Clearly, with the current size of the sample, simply reading the reports will offer deeper insights and more reliable conclusions. But the purpose of the experimental zoning and structuring of the report as discussed above, was not to replace the traditional approach to these historical sources but to outline a method of modeling epidemiological reasoning.

Moving forward, my aim is to refine this method and to train a model that reliably resembles the arguments in reports. This will enable large-scale comparison across all outbreak reports and sections to deliver two modes of network visualization. First, this method allows for a visualization of networks of concepts and theories that structured the epidemiological observation of plague. To historians working on the history of the third plague pandemic, this will be a useful instrument to trace theories and practices along the network of outbreaks. It will be possible to trace networks of expertise through the references included in reports as well as to create an inventory of person names involved in the research on plague on a global scale. Patterns of fumigation practices might follow the political contours of an empire, and patterns of treatment protocols might be indicative for the global reach of the Institut Pasteur. Furthermore, practices of prevention can be compared to

concepts of causation to identify, for example, inconsistencies. Moreover, a plethora of data would be made available for epidemiological analysis, including mortality and incidence rates, dates and individual case descriptions accompanied by detailed datasets to enrich models of the dynamics of bubonic plague.

Second, network visualizations of each report can be created to demonstrate the weight of arguments and concepts in individual texts. Utilizing epistemic network analysis, these networks of epidemiological reasoning will be useful to enhance our understanding of the formal underpinnings of pre-formal epidemiology. The sample of bubonic plague reports, spanning the decades from 1894 to 1950, contains important shifts in the significance of the animal vector, for the role of the laboratory, and for the rising position of mathematical models. The reports offer a rich sample to better understand the role of the environment and its significance for epidemiological arguments. Historical narratives of the plague can be compared over time to gain insight into the role of history for epidemiological analysis.

Once these research practices have been developed and tested, the model can be used far beyond the genre of outbreak reports. It might very well provide us with an instrument to crawl through large collections of digitized works in the history of medicine and public health to retrieve meaningful new information about the history of the third plague pandemic. Important questions about concepts of causes, about the dates and places of specific measures and about the emergence of theories about the vector of the rat could be raised against the entirety of sources available through the Medical Heritage Library. Such efforts promise new research questions and will enrich our understanding of the historical contingency of observing and understanding epidemics.

Appendix: A List of Outbreaks

Ayres, Philip Burnard Chenery, and James Alfred Lawson. *Report on the Outbreak of Bubonic Plague in Hongkong, 1894, to the International Congress of Hygiene and Demography Held at Budapest, 1894*. Hong Kong: China Mail Office, 1894. http://archive.org/details/b24974407.

Baxter-Tyrie, C. C. "Report of an Outbreak of Plague in Queensland during the First Six Months of 1904." *Journal of Hygiene* 5, no. 3 (1905): 311–32.

Blackburne, G. H. S., and T. L. Anderson. *Report on the Outbreak of Plague at Fremantle*. Perth: Wm. Alfred Watson, Government Printer, 1903. http://archive.org/details/b24916614.

Bombay. Municipal Commissioner's Office. *Report of the Municipal Commissioner on the Plague in Bombay for the Year Ending 31st May 1901*. Bombay: Advocate of India Press, 1902. http://archive.org/details/b28037510_0003.

Bombay Plague Committee, James M. Campbell, and R. Mostyn. *Report of the Bombay Plague Committee, Appointed by Government Resolution No. 1204/720P, on the Plague in Bombay, for the Period Extending from the 1st July 1897 to the 30th April 1898*. Bombay: Times of India Steam Press, 1898. http://archive.org/details/b24974535.

"Bubonic Plague." *Journal of the American Medical Association* XXXIII, no. 22 (November 25, 1899): 1366–1366. https://doi.org/10.1001/jama.1899.02450740054010.

"Bubonic Plague in Bombay." *Journal of the American Medical Association* XXXI, no. 1 (July 2, 1898): 29–30. https://doi.org/10.1001/jama.1898.02450010039005.

"Bubonic Plague in San Francisco." *Journal of the American Medical Association* XXXV, no. 19 (November 10, 1900): 1213–14. https://doi.org/10.1001/jama.1900.24620450029001j.

California, and California State Board of Health. *Report of the Special Health Commissioners Appointed by the Governor to Confer with*

the Federal Authorities at Washington Respecting the Alleged Existence of Bubonic Plague in California: Also Report of State Board of Health. Sacramento: A.J. Johnston, Supt. State Print, 1901.

Calmette, Albert. "The Plague at Oporto." *The North American Review* 171, No. 524 (Jul., 1900), 104-11.

Fabela, O. G. "Something about the Bacteriology and Clinical History of Plague." *Public Health Papers and Reports* 29 (1903): 255–58.

Gatacre, William Forbes (Sir), and Bombay Plague Committee. *Report on the Bubonic Plague in Bombay, 1896-97.* Bombay: Times of India Steam Press, 1897. http://archive.org/details/b24974523_0003.

Goff, A. P. "Bubonic Plague in Manila." *Journal of the American Medical Association* 60, no. 26 (June 28, 1913): 2042–43. https://doi.org/10.1001/jama.1913.04340260016009.

Great Britain. Local Government Board, and Royal College of Physicians of London. *Reports and Papers on Suspected Cases of Human Plague in East Suffolk and on an Epizootic of Plague in Rodents.* London: HMSO, 1911. http://archive.org/details/b24976775.

Grubbs, S. B. "The Plague Outbreak in Porto Rico." *Journal of the American Medical Association* LXII, no. 4 (January 24, 1914): 288–89. https://doi.org/10.1001/jama.1914.02560290038013.

Ham, Bertie Burnett, and Queensland. Department of Public Health. *Report on Plague in Queensland, 1900-1907 (26th February 1900 to 30th June 1907).* Brisbane: Department of Public Health, Queensland, 1907. http://archive.org/details/b28039099.

Ham, Bertie Burnett, and Queensland. Department of Public Health. *Report on the Outbreak of Plague in Maryborough, 1905 [Electronic Resource]: May-June 1905.* Brisbane: By authority: George Arthur Vaughan, Government printer, 1905. http://archive.org/details/b21351582.

Ham, Bertie Burnett, and Queensland. Department of Public Health. *Report on the Outbreak of Plague in the State of Queensland, 1903.* Brisbane: George Arthur Vaughan, Government Printer, 1903. http://archive.org/details/b24916602.

Havelburg, W. "BRAZIL. Reports from Rio de Janeiro–Plague Imported from Oporto. April 28, 1900" *Public Health Reports* 15, no. 23 (1896-1970): 1442-44.

Hill, Ernest Edward. *Report on the Plague in Natal, 1902-3 [Electronic Resource]*. London, Cassell, 1904. http://archive.org/details/b2135392x.

James, C. H., Punjab, and Royal College of Physicians of London. *Report on the Outbreak of Plague in the Jullundur and Hoshiarpur Districts of the Punjab, 1897-98*. Lahore: Printed at the Civil & Military Gazette Press, 1898. http://archive.org/details/b24975886.

Liceaga, Eduardo. "The Bubonic Plague in the Port of Mazatlan, State of Sinaloa, Republic of Mexico." *Public Health Papers and Reports* 30 (1905): 226–37.

Lowson, James A. "The Epidemic of Bubonic Plague in Hongkong, 1894." *The Indian Medical Gazette* 32.6 (1897): 207–09.

Lowson, James A., Bombay, and Royal College of Physicians of London. *Report on the Epidemic of Plague from 22nd February to 16th July, 1897*. Bombay: Publisher not identified, 1897. http://archive.org/details/b24974511.

Lowson, James A., and Hong Kong. Colonial Secretariat. *The Epidemic of Bubonic Plague in 1894. Medical Report*. Hong Kong: Noronha & Company, 1895. http://archive.org/details/b24398287.

Mitra, A., and London School of Hygiene and Tropical Medicine. *A Report on the Outbreak of Plague in Kashmir from 19th November 1903 to 31st July 1904*. Kashmir: Central Jail Press, 1904. http://archive.org/details/b2476467x.

Montgomery, Douglass W. "The Plague in San Francisco." *Journal of the American Medical Association* XXXV, no. 2 (July 14, 1900): 86–89. https://doi.org/10.1001/jama.1900.24620280022001f.

Mullowney J. J. "The Plague in North China." *Journal of the American Medical Association* LVI, no. 10 (March 11, 1911): 737–737. https://doi.org/10.1001/jama.1911.02560100029011.

Nathan, Robert. *The Plague in India*, 1896, 1897. Simla: Government Central Printing Office, 1898. http://archive.org/details/ plagueinindia18901nath.

Nathan, Robert. India. Home Department, and Royal College of Physicians of London. *The Plague in India*, 1896, 1897. Simla: Government Central Printing Office, 1898. http://archive.org/ details/b2497528x_0001.

New South Wales. Department of Public Health and Thompson, J. Ashburton. *Report on the Outbreak of Plague at Sydney*, 1900. Sydney: William Applegate Gullick, Government Printer, 1900. http://archive.org/details/b21354704.

New South Wales. Department of Public Health and Thompson, J. Ashburton. *Report of the Board of Health on a second outbreak of plague at Sydney*, 1902. Sydney: Government Printer, 1903.

Philip, W. M., and L. F. Hirst. "A Report on the Outbreak of the Plague in Colombo. 1914-1916." *The Journal of Hygiene* 15.4 (1917): 527–64.

"The Plague at Sydney." *Journal of the American Medical Association* XXXVI, no. 22 (June 1, 1901): 1565–66. https://doi.org/10.1001/ jama.1901.02470220039012.

"A Plague Focus in California." *Journal of the American Medical Association* LIII, no. 25 (December 18, 1909): 2106. https://doi.org/ 10.1001/jama.1909.02550250060009.

"Plague in California and the Anti-Plague Campaign." *Journal of the American Medical Association* LI, no. 12 (September 19, 1908): 1010–14. https://doi.org/10.1001/jama.1908.25410120054004.

"The Plague in China." *Journal of the American Medical Association* XXII, no. 25 (June 23, 1894): 960–960. https://doi.org/10.1001/ jama.1894.02421040028006.

"Plague in South Africa – Report of Brookline, Mass., Board of Health – 'American Medicine': A New Medical Journal – A So-Called Victory for Christian Science – Medical Notes." *The Boston Medical and Surgical Journal* 144, no. 15 (1901): 360–65. https://doi.org/10.1056/NEJM190104411441512.

"A Report on the Last Epidemic of Plague at Hong Kong – Medical Notes." *The Boston Medical and Surgical Journal* 136, no. 22 (1897): 550–52. https://doi.org/10.1056/NEJM189706030002211.

"The Reported Appearance of Plague in Bombay." *British Medical Journal*, no. 2 (1896): 966.

"Reported Case of Plague in Ann Arbor, Mich." *Journal of the American Medical Association* XXXVI, no. 15 (April 13, 1901): 1049. https://doi.org/10.1001/jama.1901.02470150045010.

Roys, Charles K. "Report on the Epidemic of Pneumonic Plague in Tsinanfu 1918." *China Medical Journal* 32, no. 4 (1918): 346–48.

Simpson, W. J. (William John), and Great Britain. Colonial Office. *Report on the Causes and Continuance of Plague in Hongkong and Suggestions as to Remedial Measures [Electronic Resource].* London: Waterlow and Sons, Printers, 1903. http://archive.org/details/b21297496.

Simpson, William John, and Hong Kong. Sanitary Board. *Preliminary Memoranda on Plague Prevention in Hongkong.* Hong Kong: Noronha & Co, 1902. http://archive.org/details/b24975242.

Simpson, William John. *Report by Professor W. J. Simpson on Sanitary Matters in Various West African Colonies and the Outbreak of Plague in the Gold Coast.* London: Printed for His Majesty's Stationery Office, by Darling & Son, 1909. http://archive.org/details/b21365398.

Staff Surgeon Wilm of the Imperial German Navy. "A Report on the Epidemic of Bubonic Plague at Hongkong in the Year 1896." Translated for the Government of Hongkong by Maurice Eden Paul, M.D. *The Indian Medical Gazette* 32, no. 5 (1897): 167–71.

–––. "A Report on the Epidemic of Bubonic Plague at Hongkong in the Year 1896." *The Indian Medical Gazette* 32, no. 6 (June 1897): 207–9.

Thompson, J. Ashburton (John Ashburton), New South Wales. Department of Public Health, Guy's Hospital Medical School former owner, and King's College London. *Report on the Outbreak*

of the Plague at Sydney, 1900 [Electronic Resource]. Sydney: W.A. Gullick, government printer, 1900. http://archive.org/details/b21298968.

Wemple. "Report of a Case of Bubonic Plague." *California State Journal of Medicine* 2, no. 1 (1902): 40–42.

Wu, Lien Teh. "First Report of the North Manchurian Plague Prevention Service." *Journal of Hygiene* 13, no. 3 (1913): 237–90

Endnotes

1. Lorraine Daston, "The Sciences of the Archive," *Osiris* 27, no. 1 (January 1, 2012): 156–87, https://doi.org/10.1086/667826.

2. Anne Hardy and M. Eileen Magnello, "Statistical Methods in Epidemiology: Karl Pearson, Ronald Ross, Major Greenwood and Austin Bradford Hill, 1900–1945," *Sozial- Und Präventivmedizin* 47, no. 2 (March 1, 2002): 80–89, https://doi.org/10.1007/BF01318387.

3. Volker Hess and J. Andrew Mendelsohn, "Case and Series: Medical Knowledge and Paper Technology, 1600–1900," *History of Science* 48, no. 161 (2010): 287–314.

4. Many digital approaches to the history of diseases have been developed in recent years but the overwhelming amount of studies has focused on data sources that already exist in formalized and quantifiable form, such as mortality and morbidity statistics. See "Project Tycho," University of Pittsburgh, 2018, www.tycho.pitt.edu/ or K. Hempel and D. J. D. Earn, "A Century of Transitions in New York City's Measles Dynamics," *Journal of The Royal Society Interface* 12, no. 106 (May 6, 2015): 20150024, https://doi.org/10.1098/rsif.2015.0024.

5. Lukas Engelmann, "The Burial Pit as Bio-Historical Archive," in *Histories of Post-Mortem Contagion*, eds. Christos Lynteris and Nicholas Evans (Palgrave Macmillan, 2018), 189–211, https://doi.org/10.1007/978-3-319-62929-2_8.

6. Myron J. Echenberg, *Plague Ports: The Global Urban Impact of Bubonic Plague, 1894–1901* (New York: New York University Press, 2007); Christos Lynteris, *Ethnographic Plague: Configuring Disease on the Chinese-Russian Frontier* (Palgrave Macmillan, 2016).

7. Alfredo Morabia, "On the Origin of Hill's Causal Criteria," *Epidemiology* 2, no. 5 (1991): 367–69; Alfredo Morabia, "Epidemiology: An Epistemological Perspective," in *A History of Epidemiologic Methods and Concepts*, ed. Alfredo Morabia (Boston: Birkhauser Verlag, 2004): 1-126 ; Alfredo Morabia, *Enigmas of Health and Disease: How Epidemiology Helps Unravel Scientific Mysteries* (New York; Columbia University Press, 2014).

8. "ENA," accessed 07/03/2018, http://www.epistemicnetwork.org. See also chapter 5 in this book by DiMeo and Ruis.

9. This point has been already raised by Crookshank in 1922: F. G. Crookshank, "First Principles of Epidemiology," in *Influenza: Essays by Several Authors*, ed. F. G. Crookshank (London: William Heinemann, 1922), 11–30.

10. Morabia, "Epidemiology: An Epistemological Perspective."

11. J. Andrew Mendelsohn, "From Eradication to Equilibrium. How Epidemics Became Complex after World War I.," in *Greater Than the Parts: Holism in Biomedicine, 1920–1950*, ed. Christopher Lawrence and George Weisz (Oxford University Press, 1998), 303–34.

12. On the persistent lack of theory in the history of epidemiology: Nancy Krieger, "Got Theory? On the 21st Century Rise of Explicit Use of Epidemiologic Theories of Disease Distribution: A Review and Ecosocial Analysis," *Current Epidemiology Reports* 1, no. 1 (March 1, 2014): 45–56, https://doi.org/10.1007/s40471-013-0001-1; Nancy Krieger, *Epidemiology and the People's Health: Theory and Context* (Oxford University Press, 2011).

13. John M. Eyler, *Victorian Social Medicine: The Ideas and Methods of William Farr* (Cambridge: Cambridge University Press, 1979), https://repository.library.georgetown.edu/handle/10822/782369; Kari S. McLeod, "Our Sense of Snow: The Myth of John Snow in Medical Geography," *Social Science & Medicine* 50, no. 7 (2000): 923–935; John M. Eyler, "The Changing Assessments of John Snow's and William Farr's Cholera Studies," *Sozial- Und Präventivmedizin* 46, no. 4 (July 1, 2001): 225–32, https://doi.org/10.1007/BF01593177; John M. Eyler, "The Strange Case of the Broad Street Pump: John Snow and the Mystery of Cholera," *Journal of the History of Medicine and Allied Sciences; Oxford* 63, no. 4 (October 2008): 525–26, http://dx.doi.org/10.1093/jhmas/jrn040; Tom Koch and Kenneth Denike, "Crediting His Critics' Concerns: Remaking John Snow's Map of Broad Street Cholera, 1854," *Social Science & Medicine* 69 (2009): 1246–51; Tom Koch and Ken Denike, "Essential, Illustrative, or ... Just Propaganda? Rethinking John Snow's Broad Street Map," *Cartographica: The International Journal for Geographic Information and Geovisualization* 45, no. 1 (March 2010): 19–31, https://doi.org/10.3138/carto.45.1.19.

14. Jean-Paul Gaudilliére and Ilana Löwy, *Heredity and Infection: The History of Disease Transmission* (New York: Routledge, 2012).

15. John M. Eyler, "William Farr on the Cholera: The Sanitarian's Disease Theory and the Statistician's Method," *Journal of the History of Medicine and Allied Sciences; Oxford* 28, no. 2 (April 1, 1973): 79–100.

16. Erwin H. Ackerknecht, "Anticontagionism between 1821 and 1867," *Bulletin of the History of Medicine* 22 (1948): 562–93.

17. Anne Hardy and M. Eileen Magnello, "Statistical Methods in Epidemiology: Karl Pearson, Ronald Ross, Major Greenwood and Austin Bradford Hill, 1900–1945," *Sozial- Und Präventivmedizin* 47, no. 2 (March 1, 2002): 82, https://doi.org/10.1007/BF01318387.

18. Anne Hardy, *The Epidemic Streets: Infectious Disease and the Rise of Preventive Medicine, 1856–1900* (Wotton-under-Edge: Clarendon Press, 1993); Graham Mooney, *Intrusive Interventions: Public Health, Domestic Space, and Infectious Disease Surveillance in England, 1840–1914* (Woodbridge: Boydell & Brewer, 2015).

19. F. A. Barrett, "August Hirsch: As Critic of, and Contributor to, Geographical Medicine and Medical Geography," *Medical History. Supplement*, no. 20 (2000): 98–117.

20. Volker Hess and Andrew Mendelsohn, "Sauvages' Paperwork: How Disease Classification Arose from Scholarly Note-Taking," *Early Science and Medicine* 19, no. 5 (2014): 471–503.

21. August Hirsch, *Handbook of Geographical and Historical Pathology* (London: New Sydenham Society, 1883); on the role of aleatoric series and the comparability of the spatial arguments attributed to disease in the individual and the social body, see Michel Foucault, *The Birth of the Clinic: An Archaeology of Medical Perception* (New York: Pantheon Books, 1973).

22. Echenberg, *Plague Ports.*

23. D. J. Bibel and T. H. Chen, "Diagnosis of Plaque: An Analysis of the Yersin-Kitasato Controversy," *Bacteriological Reviews* 40, no. 3 (September 1976): 633–51.

24. On the incommensurability of bacteriology and epidemiology at the time, see J. Andrew Mendelsohn, "'Like All That Lives': Biology, Medicine and Bacteria in the Age of Pasteur and Koch," *History and Philosophy of the Life Sciences* 24, no. 1 (2002): 3–36.

25. J. Andrew Mendelsohn, "'Like All That Lives.'"; Andrew Cunningham, "Transforming Plague: The Laboratory and the Identity of Infectious Disease," in *The Laboratory Revolution in Medicine*, ed. Andrew Cunningham and Perry Williams (Cambridge: Cambridge University Press, 1992), 209–44.

26. An argument that was made in particularly strong form by the Pasteurian Calmette: Albert Calmette, *The Plague at Oporto* (The North American Review, 1900).

27. Guenter B. Risse, *Plague, Fear, and Politics in San Francisco's Chinatown* (Baltimore: Johns Hopkins University Press, 2012); Nayan Shah, *Contagious Divides: Epidemics and Race in San Francisco's Chinatown* (Berkeley: University of California Press, 2001); Lynteris, *Ethnographic Plague.*

28. William John Ritchie Simpson, *A Treatise on Plague; Dealing with the Historical, Epidemiological, Clinical, Therapeutic and Preventive Aspects of the Disease* (Cambridge, UK: Cambridge University Press, 1905), http://archive.org/details/treatiseonplague00simp; Christos Lynteris, "A Suitable Soil: Visualizing Plague's Environment during the Third Pandemic" (presentation, The Plague and the City, Cambridge, December 5, 2014); David S. Barnes, "Cargo, 'Infection,' and the Logic of Quarantine in the Nineteenth Century," *Bulletin of the History of Medicine* 88, no. 1 (2014): 75–101, https://doi.org/10.1353/bhm.2014.0018.

29. J. Danysz, "Some Reflections Regarding the Free Use of Bacteriological Cultures for the Destruction of Rats and Mice," *British Medical Journal* 1, no. 2508 (1909): 209; Mendelsohn, "'Like All That Lives.'"

30. This, of course, is problematic; as, for example, many contributions from the Institut Pasteur were written in French, almost all Latin American outbreaks were covered either in Spanish or Portuguese and many Chinese and Russian observations only exist in their respective languages. This bias, which follows the macro-political lines of the British and the American Empire, structures this first pilot study and needs to be taken accordingly into account.

31. John Swales, *Genre Analysis: English in Academic and Research Settings* (Cambridge: Cambridge University Press, 1990); V. K. Bhatia, *Analysing Genre: Language Use in Professional Settings* (New York: Routledge, 2014).

32. John Flowerdew and Alina Wan, "The Linguistic and the Contextual in Applied Genre Analysis: The Case of the Company Audit Report," *English for Specific Purposes* 29, no. 2 (April 1, 2010): 78–93, https://doi.org/10.1016/j.esp.2009.07.001.

33. Lowson, "Medical Report on the Epidemic of Bubonic Plague in Hong Kong," *Chinese Medical Missionary Journal* 9, no. 3) (1895): 141–47.

34. Ernest Edward Hill, "Report on the Plague in Natal, 1902-3" (London: Cassell, 1904), http://archive.org/details/b2135392x.

35. C. R. Eskey, "Epidemiological Study of Plague in Peru," *Public Health Reports* 47 (1932): 2191–2207.

5. Thinking about Sources as Data: Reflections on Epistemic Network Analysis as a Technique for Historical Research

MICHELLE DIMEO AND A. R. RUIS

Network models, in particular social network models, have improved our understanding of a variety of historical phenomena, including correspondence communities, trade networks, citation patterns, dissemination of news, and so on. In many cases, social network analysis has been used to show relationships among people–who corresponded with, traded with, cited, or otherwise interacted with whom? But what if we extended our scope to consider the networks of knowledge created by these individuals? Instead of asking merely "Who was in this network and how were they connected?", we could ask, "How did information move through this network?" Such questions more closely model the qualitative questions that historians concerned with discourse and concepts have traditionally asked and usually try to answer without computational approaches; however, as access to historical data is expanding rapidly due to digitization efforts, it will be useful, if not necessary, to collaborate with machines on our analyses. To do so, we need to think about mixed-methods approaches that integrate the strengths of humans and computers, and network analysis is one methodological approach that could prove helpful in answering the kinds of qualitative research questions often asked by social, cultural, and intellectual historians.[1]

In this chapter we reflect on the use of *epistemic network analysis* (ENA) as a tool for modeling conceptual networks. Because there are a number of resources that explain ENA in great detail as a

technique and a tool,[2] we will not discuss how to use ENA, but rather explore why and how a historian might find the approach useful. Following this, we explore some of the issues with which the historian must engage in order to move from a strictly human, qualitative methodology to a mixed-methods approach that includes ENA. While digital humanities papers commonly include a methods section, these final products tend not to reflect on the complexity of the methodological process that got the authors to that stage, to talk openly about which data models failed, or to reflect on the limitations of tools they previously considered and rejected. This chapter is intentionally focused on this "work in progress" stage that all historians go through, and which newcomers to the digital humanities can find isolating. Using a case study approach–applying ENA to a seventeenth-century archival collection of letters known as the Hartlib Papers–we will consider the kinds of intellectual and theoretical challenges historians may grapple with as they try to think about their source materials as a dataset and supplement their qualitative analyses with quantitative models.

Epistemic Network Analysis: A Brief Introduction

Before we consider the affordances of ENA as a tool for historical research, we will briefly outline ENA as a technique. ENA was originally developed to model *cognitive networks*: the patterns of association between knowledge, skills, decision-making processes, and other elements that characterize complex or collaborative thinking in some domain. However, ENA is a versatile method that can be used to model patterns of association in any system characterized by a complex set of dynamic relationships among a relatively small, fixed set of elements. Thus, ENA is particularly suited to analyzing *discourse*–the actions and interactions of people in some culture–and it is optimized for text data.[3]

To understand the affordances of ENA for historical research, it may help to contrast it with social network analysis (SNA).[4] For our purposes here, there are two key differences. First, where SNA is optimized for exploring the properties of a single large network, ENA is optimized for comparing a number of relatively small networks. Social networks are often too large to visualize usefully, so social network statistics are designed to identify and quantify characteristics of network structure (e.g., structural cohesion, network density) or characteristics of the nodes in the network (e.g., centrality, betweenness). That is, social network statistics are designed to help researchers understand the overall structure and attributes of some network or to identify nodes or edges (i.e., individuals or the connections among them) that are outliers or that have particular effects on the network. Unlike an SNA model, which consists of one large and typically complex network, an ENA model is comprised of dozens or hundreds or even thousands of small networks, which are projected into a metric space that facilitates both visual and statistical comparison of networks. Thus, where social networks contain information about how nodes are connected, epistemic networks contain information about how nodes are connected *and* spatial information that enables both statistical and visual comparison of network structure. Thus, ENA is better suited for exploring how networks change over time or differ across contexts.

Second—and related to the first point—social networks and epistemic networks differ in how they incorporate the key unit of interest. In a social network model, the units are nodes. That is, what we care about are the people (or other entities) in the network and how they are connected. In an epistemic network model, each unit is represented not *in* a network but *as* a network. So if we are modeling cognitive networks, each individual's thinking is represented as a network, where the nodes are relevant elements of cognition (e.g., bits of knowledge, different skills, etc.) and the connections indicate integration of those elements in some context. Thus, a key challenge in developing ENA models is determining what

elements (i.e., what nodes) to include in the model and to define clearly what it means for two elements to be connected. In the next section, we use a specific example to explore this issue in the context of historical research.[5]

Case Study: The Hartlib Papers as a Dataset

Over the last decade, many historians have used network analysis to explore and identify patterns in correspondence communities, as letters exchanged can be readily modeled as networks thanks to having such data as a sender, receiver, date, and place. Impressive, wide-reaching collaborative projects such as "Mapping the Republic of Letters" have exposed otherwise-unknown social networks by using correspondence data, and these projects are a useful starting point for mapping intellectual connections among individuals.[6] The increased use of big data represents a historiographic shift in the discipline, and historians must consider what to do with the vast new amounts of information available. For example, now that an early modernist can put a name into "Six Degrees of Francis Bacon" and quickly see that person's intellectual network (even if it may be incomplete),[7] the next step could be to question what that person was talking about and with whom, how these conversations changed over time, and what such topics of discussion can tell us about their wider intellectual culture. Such a project would require us to engage with the content of the letters and select another technique and tool, such as ENA, to model these intellectual connections.

To explore some of the issues historians need to think about when considering epistemic network models, we will use this section to work through a case study provided by the Hartlib circle. The international correspondence group now known as the Hartlib circle was active circa 1640 to 1660. While based in London and centered around Samuel Hartlib, the network reached across Ireland, continental Europe, and into the American colonies. Hartlib

and his network wanted to seize the opportunities afforded by the breakdown of social order during the English civil wars and interregnum in order to organize and widely distribute all useful knowledge to the public.[8] The Hartlib Papers archive (held at the University of Sheffield Library but now easily accessible online through the University of Oxford's Cultures of Knowledge project) comprises an eclectic mix of letters concerning everything from chemistry to educational and political reform, and from beekeeping to theology and prophecy.[9] The archive holds over 4,000 letters from more than 400 individual correspondents, many of whom do not have records in national or international name authority files because they were merchants, students, and exiles who have been difficult to identify. Practical and theoretical discussions blend as Hartlib and his associates exchange ideas, comment on proposals, and make recommendations for wider circulation and adoption. As such, the Hartlib circle provides an excellent place for the historian to consider structures of knowledge creation and patterns for sharing ideas during a period of rapid intellectual change.

Because the Hartlib Papers have been openly available online for many years, and because projects using this dataset have been the recipients of several grants for improving cataloging, transcription, and access, scholars have already produced valuable network models from it. The most often cited is Scott Weingart's experimental heat map, which uses a modern Google map to show where Hartlib's correspondents lived and visualizes the density of their geographic distribution.[10] More recent projects include the works of Robin Buning and Evan Bourke. Buning used the Hartlib circle's biographies and correspondence to show a prosopographic study of individuals' lives and networks. Bourke considers gender and centrality within the Hartlib circle, making use of Gephi and recent theories concerning early modern social networks to highlight the role of significant female correspondents.[11] These studies have helped us better understand the complexity and diversity of the Hartlib circle as a whole, but they treat the social interactions between individuals as the end point. If, for example,

we wanted to better understand which individuals in the Hartlib circle talked most frequently about religion, and when these conversations verged into discussions of natural philosophy, we might take as a starting point these existing social network models and open datasets, but we would then need to consider how to model not just the exchange of letters but the exchange of knowledge and ideas.

To ground the following discussion in a concrete example, we have included in the Appendix, at the end of the chapter, a transcription of a sample letter from the Hartlib Papers, written in English and Latin by John Winthrop in New Haven, CT, and sent to Samuel Hartlib in London, England, on May 10, 1661. The transcription was done by the Humanities Research Institute at the University of Sheffield, which also provides scans of the original manuscript letter for reference. They expanded abbreviations by using italics to represent letters that were not in the original. Words that were difficult for the transcriber to read are included as possible suggested text in brackets with a question mark. Original spelling and punctuation was retained throughout, with an occasional bracket to indicate where Hartlib edited the original letter he received.

At first glance, this may seem like an ideal set of records with which to take a mixed-methods approach, as the collection is too large for a person to read. However, there are a number of challenges that must be addressed in order to do so. Many letters do not exist as full transcriptions, which means that there are data missing; and of the transcriptions that do exist, there are inconsistencies in the spelling, abbreviations, and names, which makes machine recognition of terms more complicated. While Early Modern Letters Online has improved standardization of catalog information and metadata related to the individuals who wrote and received these letters, the transcription data from the original Humanities Research Institute project still remains imperfect and is not accessible as an open dataset.[12] Additionally, letters in this archive are written in multiple languages, including English, Latin,

and German, and, as the sample letter in the Appendix shows, authors often moved freely among languages within the same letter (sometimes even within a single sentence). Thus, even with access to the complete transcription data, the dataset is difficult to process using techniques from computational linguistics. But let's assume, for the purposes of this discussion, that we had solved these problems by obtaining the full set of transcriptions, standardizing spelling, and so on. Now what?

Theorizing an Epistemic Network Model of the Hartlib Papers

As with any analysis, we need to begin with a research question—in this case, a question about transatlantic discussions of medicine within the Hartlib Papers. If an ENA model would help us answer that question, there are three additional questions we need to address:

1. What are the elements whose association we want to model? That is, what will the nodes of the network be?
2. How do we understand connectivity and operationalize it in the model? That is, what does it mean for two nodes to be connected?
3. What is the unit of analysis? That is, what or whom does each network in the model represent?

The answers to these questions, in turn, guide how we structure and process the data and how we define the parameters of the model. Note that, as in nearly all research endeavors, this process is iterative, as each decision made in the design of a study will potentially affect both subsequent and prior decisions.

Choosing a research question may seem a trivial task, but it quickly becomes non-trivial if a close reading of all or even most of the source material is not feasible. If we take the letter in the

Appendix as a representative example, we can begin to see how time-consuming it would be to read more than 4,000 other letters similar to it, each with its own unique challenges and idiosyncrasies. Furthermore, the encyclopedic range of topics discussed by these correspondents can be challenging for anyone using the Hartlib Papers today, and this has usually resulted in intellectual, cultural, and literary historians asking questions that relate to a subset of the archive and not the Hartlib Papers as a whole. As such, while a question such as "How did discussions of medicine travel internationally among the Hartlib circle?" could be addressed using a network analytic approach, the question is too broad to offer much guidance on model construction. Instead, it would be more manageable to define a narrower scope that still has intellectual value, such as considering only discussions of medicine within the transatlantic correspondence of the Hartlib circle. While key London figures like Samuel Hartlib and John Dury never traveled to the American colonies, they were in conversation with individuals like John Winthrop in Hartford, CT, and Thomas Browne in Barbados (then an English colony). Such a dataset would likely result in several dozens of letters instead of thousands, and among those even fewer would have medical content. We could use this subset of letters to refine our research question and model, then apply what we find to the whole dataset.

Now, however, we must make some important decisions. For a network approach to be useful, we must believe that the connections among elements in the network are more important than the mere presence or absence of the elements in isolation—otherwise, why do a network analysis at all? In this case, a network approach makes sense; we care not only *that* different letters have medical elements (e.g., discussion of illnesses, therapies, regimens, etc.), but also how those elements are associated with one another, and whether changes in the patterns of association may be related to who exchanged correspondence with whom.[13] This leads to the question: Which elements (nodes) should we include, and what does it mean for them to be associated

(connected)? This is where having chosen a reduced dataset with which to develop our model comes in handy. We actually *can* read several dozen letters closely, and we can use that close reading to generate hypotheses—that is, to refine our research question and develop an initial set of candidate nodes whose association we want to model.

There are several different ways that the letter in the Appendix can be modeled, taking us back to our need to refine our research question. Is it important for us to understand the nuances in how John Winthrop's letter related issues of food and diet to medicine? This could be important for an intellectual historian tracing John Winthrop's medical practice and philosophy over time. Or do we want to learn how his recommendations for treatment changed depending on which country he was discussing (as the first paragraph of the letter discussed the American colonies and the second referred to the recipient's experience in England)? This could enhance a cross-cultural comparison, allowing us to see how geographic distribution of local resources shaped plans for healing. The next step is to look more closely at the text and consider how to model the data to answer such questions. Let's take an example toward the beginning of the letter, in which Winthrop notes that "Indian corne" could be "used to make a most ordinary & pleasant food thereof called sampe which easy of digestion & very diuretique & it hath beene observed that whiles people vsed most of that foode it was rare to hear of any troubled with the stone" (Appendix). If we think about this as a (very simple) network, there is an association structure in which "corn" is connected to "nourishment,' "diuresis," and "antilithiasis"; an even simpler network would connect "corn" with "nourishment" and "urinary health."

What this simple example shows is the beginning of the process through which *codes* are developed. Codes—also termed categories, annotations, or labels—are constructs that represent specific interpretations of content in some context. In an operational sense, codes are the elements of our source material that we want to include as nodes in our epistemic network model, and whose

association structure we want to examine. It may be helpful to think of codes as rules for sorting; in taxonomy, for example, if we were coding organisms, we could categorize at the kingdom level (in which case we would have 6 codes), or we could categorize at the phylum level (in which case we would have more than 50 codes), or we could categorize at any other level, with different degrees of granularity. We could also mix and match, and code animals by phylum and all other organisms by kingdom. Note that codes need not be exhaustive; if our dataset contained, say, viruses (which aren't organisms), then they would not be coded for anything. No choice is right or wrong per se, but each choice will afford or constrain different kinds of analysis. The point is that any given organism either is or is not associated with a particular category being used in some analysis. What coding does, then, is allow the researcher to construct standard interpretations across some dataset so that each item in the dataset either is or is not associated with a given code. In other words, coding is a process for converting qualitative interpretations into numbers (1s and 0s) so that computational techniques, such as statistical analyses, can be performed on otherwise non-numeric data.[14]

When coding the letters in our dataset, we must define the types of connections we intend to explore. For the purposes of this case study on the Hartlib circle's transatlantic letters, let's say we want to understand the exchange of medical theories, materials, and practice between the New World and Old World, especially the integration of herbal and chemical remedies. As such, some topics for coding could include references to Education, Equipment, Chemicals, Minerals, Books, and Medical Practice. The dataset would include a column for each of these terms, and the historian could use binary code to say whether each segmented unit presented a reference to each topic. If the research question was focused on a more narrow issue within the history of medicine, then the historian might choose to work with a finer taxonomy. For example, if we wanted a more in-depth exploration of materiality, we might choose to break down the category Equipment into

references to specific kinds of equipment (furnaces, glassware, etc.). Such questions regarding granularity can be seen when considering the letter below: Should we code for Cranberries, or should we include cranberries within the larger category Fruits? The answer to this question depends on the theoretical framing of the historical question being asked. When one begins working on a dataset, it is natural to continue improving the coding as the project progresses. There is a rich body of literature on coding qualitative data for quantitative analysis, and it is beyond the scope of this chapter to discuss the topic in detail.[15] However, when thinking about codes in the context of a network analysis, we also need to think about connections. There are two basic questions that need to be answered: (1) What does it mean for two constructs (i.e., two codes) to be connected? (2) How can we implement this understanding of connectivity in a network model?

There are, of course, many ways to conceptualize connections. For example, causation is a form of connection. In a causal network model, if Code A is connected to Code B, then there is a causal relationship between them. Note that networks like these are usually *directional*, meaning that there is information incorporated into the network model that indicates order. In this case, that information might be that A causes B, but B does not cause A. This could be represented visually, such that the two nodes are connected by an arrow from A to B rather than a simple line. Or it could be that each code is represented by two nodes, a sender node and a receiver node, and A_{sender} is connected to $B_{receiver}$ but B_{sender} is not connected to $A_{receiver}$. As one might imagine, such networks can become complicated very quickly. For many network analyses, however, a simpler concept of connection is often sufficiently powerful. For instance, in Winthrop's reference to the health properties of Indian corn, discussed in the example above, a connection could be simple association: corn is associated with the properties nourishment, diuresis, and antilithiasis; eating corn has these effects, and thus there is an underlying causal relationship, but it isn't necessary to model it that way. In fact, we may care about

the extent to which diuresis and antilithiasis are associated with one another regardless of what causes each effect. Thus, instead of a network model where corn is connected to each of those properties, we could develop a network model where all of those properties are also connected to one another by virtue of the fact that they are discussed in conjunction. This kind of model is often useful when analyzing conversations or other complex forms of communication. These general association structures are embedded in language, and we may not have *a priori* hypotheses about which kinds of association (e.g., causal) are most important.

This raises another issue. How do we operationalize "association" into "connection" in an ENA model? That is, if we don't want to build a network by hand—or if it is unfeasible due to the volume of data, which will almost always be the case—we need to be able to specify rules for determining what counts as association (and thus contributes to connections in the network model) and what does not. In making this decision, we are actually making a decision about how to structure our dataset, as both coding and rules for determining association are based on how we convert our historical sources into machine-readable data.[16]

In thinking about how to structure data for an ENA model, there are two things that are important in this context: (1) Codes are applied to each row in a data table, and codes that co-occur within the same row are considered to be connected; and (2) there are multiple ways to indicate whether and to what extent codes on different rows should be considered connected. Thus, a key decision to be made involves how to *segment* our data into rows. There are three main ways we might segment a letter: each sentence could be a row, each paragraph could be a row, or each letter could be a row. There are, of course, pragmatic issues to be considered. In the Hartlib Papers, the correspondents often used punctuation and paragraph structures loosely and inconsistently, making it difficult to segment letters by sentence or paragraph. This archival collection has the added complication that Hartlib sometimes added or changed punctuation and capitalization once

he received a letter, and some letters only exist as scribal copies that might no longer faithfully represent the original author's epistolary style or structure. However, many of the letters are quite long and cover multiple unrelated topics; if we segmented simply by letter, with each row in the data table containing the entire contents of one letter, everything coded in the letter would be considered connected in the ENA model. As one might imagine, this could produce a very skewed representation of the association structure.

In general, it is desirable to segment at a smaller (e.g., sentence or paragraph) level. In addition to making more sense when it comes to conceptualizing meaningful associations within rows, it is also much easier to aggregate rows than to disaggregate them, and finer-grained segmentation provides more options for defining what counts as a connection in the ENA model. For example, let's assume we segment each letter by sentence. This may be imperfect at times due to the inconsistencies in punctuation usage noted above, but it will at least break up letters into more discrete pieces. By doing this, however, we gain two key advantages. First, we can reasonably assume that codes co-occurring within a given row are actually associated in some meaningful way. Second, we can define association across rows by recent temporal context using a *moving window*. A moving window defines some fixed number of lines within which codes should be considered connected.[17] For example, if we choose a moving window of three rows, then each row in the dataset (corresponding to one sentence in a letter) would be considered associated with the two prior rows (that is, the two prior sentences). There are methods for determining how big this window should be, but the point is that ENA can use some definition of proximity to determine which codes should be connected and which should not.[18] This is useful when working with archival data that may not be cleanly divisible by standard methods (e.g., paragraph breaks), but it also reflects the fact that in conversations and other forms of complex communication, proximity is a good indicator of association. Indeed, if someone wants to make a connection between a new topic and something from much earlier

in a conversation (or essay, or letter, etc.), they will typically restate the earlier point so that it is made proximate with the new contribution.

Now that we have considered how to structure our data, code it, and define connections, there is one final element that is critical to think about early in the process: what or whom will each network in the model represent? In other words, we have to think about what the unit or units of analysis will be. For example, we could set the unit as "letter writer," in which case we would get a network for each author in the dataset, and that network would represent the accumulated connections they made across all of their letters. Or, we could define the unit by "letter writer" and "year," in which case we would get (potentially) multiple networks for each author–one for every year in which that person authored at least one letter. Such an approach could help show changes over the nearly twenty years in which the Hartlib circle was in existence. Of course, we can define the units without reference to authors at all. For instance, we could set the units based on the geographic origin of the letters, in which each network would represent the connections in all the letters that originated in a particular location. This would allow us to compare all of the transatlantic letters that originated in New England with all of the letters written in the Caribbean to track differences in the cultural knowledge being imported into London.

When recording names and places in the dataset, it is important to be consistent and standardize across multiple historical variants for a single name. For example, the letter below includes a reference to "Mr. Davenport" without including his first name, but in another letter in our dataset we learn that his name is John Davenport. Similarly, location data differs between letters across the archive: one might say "London" and another "St. James's, London." Machine-readable unique identifiers are not required for ENA, but the historian should consider using the most granular level of data that is most consistent across the dataset. In these examples, for instance, "John Davenport" gives more information than "Mr. Davenport," and references to the latter can be coded as John

Davenport by using contextual clues to confirm his identity. Since references to neighborhoods within cities were included too infrequently across the Hartlib Papers, coding at the city level seems most appropriate, with all places within London simply being recorded as "London."

As will hopefully be clear at this point, selection of units, segmentation of data, choice of codes, and definition of connections are all interrelated decisions which are ultimately made to address the research question or questions. Of course, there are many other decisions that go into the construction of an ENA model, and it is important to have a clear understanding of both *the historical source material* and *how ENA works* in order to make those decisions well. The latter topic is covered in great detail elsewhere (see note 2), and is thus beyond the scope of this brief reflection on how to think about ENA as an approach to understanding the past. Rather, our goal here is to provide a framework that will help historians new to network analysis begin to think about historical source material as data that can be modeled as an epistemic network, enhancing traditional qualitative analysis with sophisticated quantitative methods. The time-consuming nature of applying ENA to the Hartlib Papers dataset means that we are unable to provide a fully complete example of analysis here. However, readers are encouraged to read A. R. Ruis's essay in this volume, which provides a more polished historical analysis using ENA to show changing definitions of "nutrition" in English-language sources over the nineteenth and twentieth centuries.[19]

Conclusion

By walking through the challenges of modeling the Hartlib Papers as an epistemic network, we hope to have broken down the false dichotomous relationship between qualitative and quantitative methodologies, demonstrating that historians need not abandon qualitative strategies or traditional research questions in order to

embrace new technologies and tools. Rather, the challenge is in learning how to translate the many nuances required in historical research into data that can be processed by a computer. While historians are trained to work in isolation and are inclined to produce single-authored pieces, a mixed-methods approach such as the one outlined here almost necessitates a more collaborative model to achieve success, drawing upon the strengths of theorists and practitioners who have already been using these quantitative methods for decades. Samuel Hartlib himself endorsed the value of network learning, advocating that useful knowledge could only be achieved by drawing upon the collective strengths of diverse individuals each specializing in their own fields. When experimenting with a new technique and tool such as ENA, the historian quickly realizes that there is an entire body of literature that explores many of the challenges that may seem new or foreign, ranging from best practices for coding to accounting for comprehensiveness (or lack thereof). Our advice is to experiment without fear of failure and forge new connections with unlikely partners, some of whom just might be looking for an interesting new dataset or challenging new problem. Through more collaborations between social scientists, data scientists, and humanists, we can continue to improve and expand upon the mixed-methods approaches that have already begun helping us to better understand the connections between various elements in the vast historical record.

Appendix

Letter, John Winthrop to Samuel Hartlib, 10 May 1661. Hartlib Papers 32/1/10A-11B.

Transcription provided by M. Greengrass, M. Leslie, and M. Hannon (2013), The Hartlib Papers. HRI Online Publications, Sheffield. https://www.dhi.ac.uk/hartlib

Much honored Sir.

By my former I mentioned the receipt of your of the 6th of March last w*i*th those several rarities of bookes and Manuscript papers for w*hi*ch I am much obliged and returne you many thankes. I sent you back in my former letter according to your desire a catalogue [*see 32/1/12*] of every p*a*rticular both bookes & papers, & am surprised by this suddain oportunity by a freind going to a place <+ called New london> <*left margin:* + New london is about [50?]miles fro*m* heare, a very brave Harbour & so called by our court here only in memory of that famous citty.>to take shipping for Barbados, who promiseth safe delivery there to a good hand but I have but few hours to write to your selfe & divers other. I have intelligence fro*m* my brother mr John Richards fro*m* Boston that he hath shipped aboard a ship that is bound to London a barrell of the best cranburies could be procured, & directed them to Mr John Harwood who I thinke lives upon tower hill [*H underlines*] neere Savage house, & hath many other goods consigned to him, & writes *tha*t he desired him to take speciall notice of that Barrell of cranburies & that would take speciall care to see them safely delivered to you selfe, mr Harwood is [*H underlines*] a friend of mine who lived also not long since in New England: & I know wilbe very carefull of them: he writes also *tha*t he gave you notice of *the* same by a letter: I wrote to him[*H underlines*] also to put vp for me & ship aboard & direct to your selfe, a barrel of Indian corne, w*hi*ch the season was not to be putt up when the other barrel was shipped, but he writes me word he hath taken special order about the same,[*H underlines*] if athe fraught of the other barrell he writes me he hath satisfied as I directed him & hath ordered the fraught of this also to be paid when shipped [*H underlines*] (For he himselfe is now newly sayled towards Barbados) that sort of corne hath they used to make a most ordinary & pleasant food thereof called sampe w*hi*ch easy of digestion & very diuretiq*ue* & it hath beene observed that whiles people vsed most of that foode it was rare to hear of any troubled w*i*th the stone, & its rare also among the Indians who vse it constantly: mr Harwood or any [*H underlines*] New England man will or woman ca*n* direct the making of & dressing of that sampe or direct to some

New England woma*n that* will doe [*altered from* sh] it & shew your servants to doe it rightly &c: If these barrells come safe to your hands be pleased to accept them as a very small token of greater respects & ingage*m*e*n*ts: I hope they wilbe safely transmitted I could take no greater care about them & I know my said friend there at Boston was very carefull to order the best way for safe transportation. [*catchword:* Sir I thought]

[32/1/10B] Sir I thought fitt to add a word or 2 to what I formerly wrote concerning the vse of minerall waters in reference to your sad afflicted condition (the consideration whereof is really a continuall affliction to my heart Simpathising with you sorrows therein) If you please to make inquiry by your correspondents & friends I doubt not but you will be informed of some fitting waters in some p*art*s of England for such cures, & will heare of many experimentall cases in that kind it may be of some yet living: & will know w*h*i*c*h may be the fittest for your p*art*icular case: & whether they may be transported with their intire virtue fro*m* the place, or whether certius ex ipso fonte bibuntur aquæ. I have great hopes of those waters for your helpe especially often reiterated though possibly with some necessary intermission as those *that* know you will best direct (Gutta cavat lapidem non vi sed sæpe cadendo) the Thermæ Færinæ in Ducatu Witt. Wirtembergico, are said by Andernacus (si memini) aut Rulandus to be et potu insidendo vtiles ad expellendos calculos renu*m*, I have not the bookes at present but find this in some papers w*h*i*c*h I overlooked lately in reference to your trouble as a [*word deleted*] memorandu*m* I had taken, I suppose out of one of those authors my note also speakes*ment*ions De fonte Bollensi ex Fallopia de aquis medicatis In & I thinke Bauhicuss hath something of the same In Regiens*e* agro aput castellu*m* vocatu*m*Brondale est fons aquæ medicatæ quæ sanat vesicæ dolores, et expellit arenulas et lapillos et saniem: & I am not long since now informed of one *that* I know longe tyme to have been troubled with great dolour in the bladder & I heare is cured by a water in those p*art*s where he liveth w*h*i*c*h is much used for other distemp*er*s. I shall inquire further about it it is farr from this place *that* I cannot now have any certaine

inquiry till after winter: I have read over *th at* booke De Societate Christiana, and *that* other you mentioned which I borrowed lately of our worthy friend Mr Davenport (who was last weeke in good health I heard then from him he knoweth not of this oportunity) I meane *that* Cynosura et amussis restaur &c the scope of them is of singular [*word deleted*]<matter> & worthy consideration but whether there be really such a christian society in Germany or else where is worth the inquiry: that booke of a Banke by ingenious Mr Potter I have perused & what your selfe have written about the same subiect in your letter it is certainly a matter of very great consequence & would tend much to the publique good [*catchword: but I doubt*] [32/1/11A] but I doubt whether it wilbe ever atteined because very few wilbe perswaded to ingage their lands though the thing be so rationall that noe obiections but might be answered, & though divers in their owne spirits would be satisfied & willing to it, yet there wilbe so many relations to be satisfied also, wives children *that* are growne vp, parents of some or, their wives parents & kindred or the childrens kindred in pretence of care of them & other friends all must be satisfied, (which is impossible) or it will come hardly of, exept in some few. that friend of whose talents you desired to be informed, hath an other very reall way which may be probably attainnable, without any ingagement of lands, & thereby mony would flow in a abundantly: he had once purposed to promote it in these plantations, but for some reasons hath deferred till he could goe into England finding vpon further consideration *that* it might be better effected with correspondence there though but with some particular company, but much more if a general banke were there setled but the troubles & warres there have [*altered from* hath] diverted his thoughts, of that voyage hitherto, if he hath not prepared or taken any course to have such a stock transferred & at command there, as might defray the charges & [*occurrences? hole in MS*], & consequences of such a voyage, which he thinks he had neede first have a thousand pound or 2 visible estate in some knowne sure hand before he could comfortably adventure vpon such a voyage, which possibly tyme might produce but interim

currant dies, & the work *that* God setts before vs is greate sed vita brevis: this way w*h*ich he intends hath some concomitants w*h*ich would greatly advance co*m*merce & other publiq*ue* concern*m*e*n*ts for the benifitt of poore & rich in great Britaine & the good of these plantations would easily be involved therein [*word deleted*] but it ca*n*not be satisfactorily (so farre as I know of it) declared in a letter, his collections in reference therevnto using of many sheets, neyther may some matters *that* concerne the secretts of some waies of profitt to <in w*h*ich> the vndertakers of such a banke would be invested, be conveniently intrusted in a letter but if he could by any oportunity speake w*i*th you I hope he would make it appeare really: and then he could also best satisfy your question himselfe, what Talents God hath intrusted him &c: w*h*ich I have also in some measure answered in another letter But you may also be satisfied sufficiently by what I have above [*catchword:* mentioned] [32/1/11B] mentioned, concerning his vnpreparedness <for the charges> for such a voyage how farr short his estate is fro*m* what you seeme to hint in your letter to be surmised, he is contented w*i*th a wilderness condition & I beleive ca*n* truly say Fælix cui deus obtulit Parca quod satis est [manu?] yet I know when he can have such a visible stock, is not w*i*thout thought of one voyage more into Europe: I know it is his iudge*m*e*n*t *that* it is not safe for a stranger (for so now he accounts himselfe to his native country having sold all long since there & long absent thence & many knowne old friends gone) to be in an other country without some knowne visible way of supply especially one *that* ca*n*not but spend much, w*h*ich I think hath made him speak of a visible stock as I have me*n*tioned fro*m* his owne expressions: though he might have supply by what traffiq*ue* he might bring over, yet not being knowne as a merchant would not be so convenient as certaine supplies as by bills of exchange to knowne merchants as the manner is in these cases: Sir I should add many other things but tyme cutts me short & therefore with most harty desires to that great phisitian to give you p*er*fect recovery, and my most reall respects presented, I shall take leave to subscribe myselfe

   Honored S*i*r

Hartford Jan: 7: 1660 Youre cordiall friend

in New Engl*and* John Winthrop

Si*r* If you can receive pay for them according to this inclosed letter I desire you to procure me these few bookes: viz: Selenographia

Systema Saturniu*m*

All Glaubers bookes exe in duch or latine

exept his Fur booke of New Furnaces with

appe*n*dices & .. de auro potabili

& his thre books operu*m* mineraliu*m*.

and his Miraculu*m* mu*n*di: for these I have

seene already & have some of then in latine

but none of the rest I have seene

[*left margin, at right angles:*]

a small booke

Vom Weinsteine

printed I think at Ha*m*burg

[Kesler*us*?] Fur auserlegene

process the last edition

I think it is funff Hundred

auserlegene processen

Acknowledgments

Endnotes

1. A. R. Ruis and David Williamson Shaffer, "Annals and Analytics: The Practice of History in the Age of Big Data," *Medical History* 61, no. 2 (2017): 336–339.

2. David Williamson Shaffer, Wesley Collier, and A. R. Ruis, "A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data," *Journal of Learning Analytics* 3, no. 3 (2016): 9–45; David Williamson Shaffer and A. R. Ruis, "Epistemic Network Analysis: A Worked Example of Theory-Based Learning Analytics," in *Handbook of Learning Analytics*, ed. Charles Lang et al. (Society for Learning Analytics Research, 2017), 175–87; David Williamson Shaffer, *Quantitative Ethnography* (Madison, WI: Cathcart Press, 2017).

3. Although ENA is most commonly used to analyze text, it has also been used to analyze video, eye-tracking data, fMRI scans, and other kinds of data. On discourse analysis more generally, see Norman Fairclough, *Discourse and Social Change* (Wiley, 1993); James Paul Gee, *An Introduction to Discourse Analysis: Theory and Method*, 4th ed. (London: Routledge, 2014).

4. Note that despite the term "social" network analysis, SNA techniques are used for a wide range of analyses, including those that have nothing to do with people per se. For simplicity, this paper will assume that social networks are networks of individuals connected through some form of social interaction (letters sent and received, joint attendance at some event, services rendered, etc.). While this is only one use case, the issues we discuss are generic to SNA as a set of techniques, regardless of what kind of network is being modeled.

5. For those who want a deeper dive into ENA, see the citations in note 2, which cover the theoretical and methodological underpinnings of ENA in considerable detail. For a worked example of an epistemic network analysis conducted on historical data, see the chapter by Ruis (this volume).

6. "Mapping the Republic of Letters," Stanford University, accessed January 28, 2018, http://republicofletters.stanford.edu.

7. "Six Degrees of Francis Bacon," accessed January 28, 2018, http://www.sixdegreesoffrancisbacon.com.

8. Mark Greengrass, Michael Leslie, and Timothy Raylor, eds., *Samuel Hartlib and Universal Reformation: Studies in Intellectual Communication* (Cambridge: Cambridge University Press, 1994); Leigh Penman, "*Omnium Exposita Rapinæ*: The Afterlives of the Papers of Samuel Hartlib," *Book History*, 19 (2016), 1–65.

9. Mark Greengrass and Howard Hotson, "The Correspondence of Samuel Hartlib" in *Early Modern Letters Online*, Cultures of Knowledge, accessed January 15, 2018, http://emlo-portal.bodleian.ox.ac.uk/collections/?catalogue=samuel-hartlib; Mark Greengrass, Michael Leslie, and Michael Hannon, "The Hartlib Papers," HRI Online Publications, 2013, http://www.hrionline.ac.uk/hartlib.

10. Scott Weingart, "Experimental Heatmap of Hartlib's Correspondents," accessed December 28, 2017, http://www.culturesofknowledge.org/?page_id=172.

11. Among other conference papers and posters he has given on this topic, see Robin Buning, "Collecting Biographies of the Members of Samuel Hartlib's Circle: A Prosopographical Approach to Networking the Republic of Letters," (presentation at "Reception, Reputation and Circulation in the Early Modern World, 1500-1800," NUI-Galway, March 22, 2017); Evan Bourke, "Female Involvement, Membership, and Centrality: A Social Network Analysis of the Hartlib Circle," *Literature Compass* 14, no. 4 (2017).

12. Greengrass, Leslie, and Hannon, "The Hartlib Papers."

13. For a more detailed discussion of connectivity in historical data, see Ruis (this volume).

14. The coding process described here is known as *binary coding*, where a "1" indicates that a code is associated with some item and a "0" indicates that it is not. It is also possible to use *weighted coding*, in which a non-binary rating scale is employed, but regardless, the researcher must ultimately be able to say that a given code either is or is not associated with a given item in the dataset. Weighted codes simply provide more information about the magnitude or nature of the association in cases where there is one

15. For a good primer on coding written for a broad audience, see Shaffer, *Quantitative Ethnography*, ch. 3. There are, of course, ways to automate some or even all of the coding process—keyword or keyphrase matching is often highly effective, for example—and there are also methods for ensuring that a given automated coding process is reliable and valid.

16. For more information on formatting data for ENA, see the references in note 2.

17. For a more detailed description of moving windows, see Amanda L. Siebert-Evenstone et al., "In Search of Conversational Grain Size: Modeling Semantic Structure using Moving Stanza Windows," *Journal of Learning Analytics* 4, no. 3 (2017): 123–139.

18. For more on determining appropriate window size, see Andrew R. Ruis et al., "A Method for Determining the Extent of Recent Temporal Context in Analyses of Complex, Collaborative Thinking," in *Proceedings of the International Conference of the Learning Sciences (ICLS) 2018* (in press).

19. See Ruis (this volume).

6. Anatomical Reading of Correspondence: A Case Study of Epistolary Analysis Networks

KATHERINE COTTLE

The recent transition from paper to electronic form as the standard means of communication has shifted not only the medium of epistolary expression, but also the networking potential of scholars and historians. Visualizations of networks can no longer rely solely on humanistic expectations of time, space, direction, and location with regards to communication, even when reading and studying text from pre-digital times. As personal print text becomes more and more indistinguishable from public digital communication, we find ourselves at a crossroads in finding appropriate venues for representing words that relate "a momentary experience which incorporates but stands outside orthodox conceptions of material and immaterial existence."[1]

How do we, as current correspondents, scholars, and researchers, imbed standardized networking frameworks, such as traditional mapping, into current and future networking needs and applications? How can data-driven networks help to increase accessibility and knowledge of past figures and texts while simultaneously sustaining humanistic foundations, ethics, and aims? The Viral Networks workshop provided the time, physical and virtual space, guidance, and digital resources for me to explore these questions through networking applications of a recently discovered archive of personal correspondence, "The Esther Richards Letters, 1915–1932," included within my forthcoming book, *The Hidden Heart of Charm City: Baltimore Letters and Lives* (AH/ Loyola University Maryland).

My immediate urge with the project was to map Richards's letters through a network which existed at the time frame of the letters' origination (1915–1932), like this United States Post Office map:

Figure 6.1: Post Office Department map of air mail routes, August 1928[2]

However, it did not take me long to realize that my current students—our future scholars and researchers—already view traditional mapping (and the postal system) as outdated and disconnected from their understanding of communicative networks. My visualizations, to be relevant and engaging to future readers, needed to apply networking in a more presence-centered framework. Therefore, instead of trying to find a compromise—between physical and digital lenses—in networking visualizations of epistolary correspondence, I choose to utilize a hybrid humanistic/data-driven structure for my diagrams. I constructed an anatomical reading networking series—a conceptual reading approach that combines surface-level views of letters with network applications which reach below the surface of text in ways only possible by digital analyses. The letters in the "Esther Richards Letters" archive were ideal for this project, as the correspondence written by Dr. Esther Loring Richards, "psychiatrist-in-charge of the outpatient department of the

Phipps Clinic from 1920 until her retirement in 1951,"[3] contains structural and content patterns reflective of an unorthodox woman utilizing words to find support, companionship, and enlightenment within fields and academic realms often deemed incompatible–approximately one hundred years before I found myself making the same attempts, in the same city.

Richards's letters are addressed to Dr. Abby Howe Turner, Richards's former professor, and these letters are contained within a digital archive devoted entirely to Mount Holyoke College. Richards's letters to Turner have only been accessible to the public since 2005. Due to the personal and voluntary efforts of Mount Holyoke alum, Donna Albino, viewers across the world can now see and read the dedicated and prolific communication of many early women in American science connected to Mount Holyoke College.[4] Albino's online archive showcases the need of women in early American science to find personal and written support and companionship outside of their individual medical communities and higher education institutions.

Correspondence networks, as evidenced in Albino's archive, were the primary communicative routes which enabled pioneering women such as Richards and Turner to endure the isolation, uncertainty, biases, and challenges of higher education institutions and medical communities to become pivotal figures in early American science. The Viral Networks workshop enabled a deeper view of the words, places, and people within these correspondence networks. Through macroscopic and microscopic anatomy readings, we see Richards, and ourselves.

> **Macroscopic Anatomy** *The examination of relatively large structures and features usually visible with the unaided eye, including surface, regional, systemic, and developmental anatomies.*

Attentive readers are quite able to make thoughtful observations and analyses without the assistance of digital enhancement. Correspondence structures which lend themselves to macro-level

networks might include surface-level reading (words and inventories), regional-level reading (locative information to showcase the importance of place), systemic-level reading (societal frameworks), and developmental-level reading (a combination of surface, regional, and systemic reading via developing institutions and histories).

Surface Reading

Figure 6.2: Envelope of letter addressed to Miss Abby H. Turner from Dr. E.L. Richards[5]

A surface, inventory-based reading of the Richards/Turner letters' archive reveals an intimate and long-term epistolary network and relationship which began at Mount Holyoke College, where Richards graduated with an A.B. degree in 1910,[6] and where Turner founded and taught within the physiology department from 1896–1940.[7] Richards's preserved letters to Turner date from 1915–1932, the years during which Richards was a graduate student and then faculty member at Johns Hopkins Hospital.[8,9] Albino has listed each preserved letter by date, with links to digital visuals of available addressed envelopes, partial letter scans, and transcriptions of content. There are a total of 42 letters presented on the webpage

"The Esther Richards Letters, 1915–1920" and 49 letters presented on the webpage "The Esther Richards Letters, 1921–1932." [10, 11] Turner's letters to Richards are not preserved, though hundreds of Turner's letters to other peers/early women in American science are preserved and accessible in the "Abby Howe Turner 1896" section of Albino's website.[12]

The amount and depth of the Richards/Turner letters, viewed within the scope of so many other personal epistolary exchanges of academic women from the late 1800s and 1900s, immediately highlights the prolific writing habits and dedicated unions of these women, especially in providing consistent communication and support across state lines, decades, and career fields. Even without extensive and in-depth critical examination and analysis, a surface reading of the Richards/Turner letters, and the archive as a whole, showcases the role of words as a foundation for correspondence networks which began as academic relationships, yet quickly branched into the lives, places, and projects inspired by Mount Holyoke's early mission to "[g]o where no one else will go, do what no one else will do."[13]

Readers can easily navigate Albino's organized and link-based website: a network of female connections inspired by Albino's own role as an alum, a preserver, and a tributary in sharing access to the behind-the-scenes lives of women in early American science. Surface reading is vital for textual analysis, not only as an inventory-based assessment, but also to establish a set of artifacts, a foundational framework, and an accessible range of material. Albino's website provides these elements for an examination of the Richards/Turners letters; however, immediate voids within surface reading are notable due to missing correspondence (all of Turner's correspondence to Richards and potential missing correspondence from Richards to Turner), human error (in transcription and translation), and accessibility (economic and temporal realities).

Regional Reading

Figure 6.3: 1920 Baltimore City Directory[14]

Just below the surface level of the Richards/Turner correspondence, additional regional networks quickly emerge which strengthen geographical reading connections. Richards writes to Turner at "Mount Holyoke College, South Hadley, Massachusetts"[15] from "Johns Hopkins Hospital, North Broadway, Baltimore, MD."[16] Johns Hopkins Hospital's role in the Baltimore community is notable, beginning with its pronounced return address on Richards's envelope. Early on in her employment at the Phipps Clinic, Richards recounts a local Baltimore preacher's words in her February 27, 1916, letter to Turner, expressing anger at the preacher's doubt of the hospital's psychiatry program legitimacy: "The Rev. said 'If Onesimus had lived in Balt. today people would have considered him the product of his heredity & environment, & sent him to the Phipps Clinic to be investigated.' That made me hot

too."[17] Richards's emotions guide her portrait of Baltimore, painting a combustive picture of a city grappling with poverty, health issues, institutional dysfunction, and cultural shifts.

Due to Richards's regional outsider status, her words depict a different geographical network from that of an insider, especially regarding Johns Hopkins Hospital and its immediate surroundings. "It has been warm here," Richards writes to Turner on August 7, 1917, "but the patients have not minded it much. You see they are southerners."[18] While adjusting to living in a warmer climate than her native New England, Richards's early correspondence to Turner often refers to the humidity and physical drain of Maryland's summer months. Richards's August 7, 1917, letter admits that "[t]he heat is so hard on your spirit, I know from past summers."[19] The mid-Atlantic seasons not only appear in the content of the correspondence, but also in their reflection of a medical career which is consistently and constantly cycling, blurred with the weight of perpetual precipitation, transition, and challenge. Baltimore is a place, Richards reinforces on August 7, 1917, where "the children have suffered fearfully, & their lives are snuffed out easily."[20]

Richards's mapping of Baltimore includes paths into Johns Hopkins Hospital not found on street signs or directories–a preserved region of the children she hears "cry[ing] at night, and in the daytime when they trudge by the clinic over the hot & dusty walk"[21]– transporting routes only revealed in an epistolary key. While regional readings of correspondence help to widen the internal and external geographical networks connected to sender and receiver endpoints, such as Baltimore's Johns Hopkins Hospital and South Hadley's Mount Holyoke College in the Richards/Turner letters, analysis is limited to locative-based markers. Mappings moving into more metaphorical and conceptual frameworks may need to dig deeper into epistolary anatomies.

Systemic Reading

Figure 6.4: "Photograph of Anne Hall, Mount Holyoke College Class of 1910, high jumping on May 11, 1910. The meet was officialled by three men from the Springfield training school."[22]

Uncovering the underlying systems below surface and regional views, then, exposes the people and societal frameworks controlling the words and places of existing texts. For example, Richards's letters regularly critique the gender-biased and elitist medical community in Baltimore, as well as the country at large. Richards's earliest archived letter, sent to Turner on March 10, 1915, while Richards was still a graduate student, describes her displeasure at a conversation at a recent Johns Hopkins Medical dinner, in which the hostess "told [Richards] [h]ow many maids she carried abroad with her when she first went after marriage."[23] This early glimpse of Hopkins society is a bitter pill Richards must swallow in order to carve out her reputation as a woman in early American science. Her correspondence to Turner provides a place for unfiltered venting about Baltimore's upper class, especially those in high ranking positions at Hopkins. Richards's March 10, 1915, letter to Turner ends

with a perfect example of such elitism, a quote from the Hopkins dinner hostess: "She was interested to know how I survived such close & continuous contact with the 'masses.'"[24]

Richards's outsider status, not just in terms of her home region, but also in terms of her gender and class, influences many of her letters to Turner. Richards often relays variations of her message written on September 4, 1920: "[t]he battle with me is pretty much alone."[25] Within this long-term state of isolation, Richards's armor becomes the words and letters exchanged between herself and Turner, in addition to her communication with other female peers and friends, many originating from her time at Mount Holyoke College. Richards's September 4, 1920, letter to Turner is clear in its declaration of the correspondence necessity for her survival: "Please write me often. I need your letters."[26] The network of letters from women provides Richards with the support and validation that she neither receives from Johns Hopkins Hospital, nor from medical communities elsewhere in the nation, even while being one of their pivotal figures.

Richards's words to Turner on September 29, 1924, still ring with her anger: "How slip-shod they do things at the Harvard Medical & that nice discrimination against our sex! Pleasant isn't it. I've often longed to put a bomb under that noble University, blow it sky high, & begin again with something less conservative & aristocratic."[27] Free from career and collegiate restraints and requirements in the epistolary form, Richards can critique the male-dominated, elitist medical field without fear of retaliation.

Ironically, Richards's correspondence to Turner becomes its own medical university curriculum proposal, enabled, because of its unique genre status, to exist separate from the systemic inequities of Richards's and Turner's time. Clearly organized, defended, and debated back and forth across multiple states—for close to two decades, Richards's desired medical university is only found on paper, its "less conservative & aristocratic"[28] elements tucked neatly inside envelopes, its enrollment limited to two corresponding members. While systemic readings unveil larger conceptual

anatomies of text and help to place surface and regional elements into context, they are also filtered through the systemic influences of the reading time period. Current biases and preferred scholarly lenses will look obvious only a few years into the future, and analyses will date themselves almost immediately upon presentation and/or publication.

Developmental Reading

Figure 6.5: Henry Phipps Psychiatric Clinic, Johns Hopkins Hospital[29]

The networks of scientific advancements, psychiatry trends, global military action, and religious and cultural shifts happening in the first decades of the twentieth century provide examples of the fluid nature of epistolary analysis in the Richards/Turner correspondence, showcasing fluctuating views of society that often cannot be seen or found in traditional non-epistolary sources.

Within these macro-levels of reading—surface, regional, and systemic—networks of words, places, and people coincide and are visible in developing institutions and their developing histories.

Richards, the once idealistic pioneering female student, gradually grows disenchanted with her former alma mater, the psychiatry field, and "the masses." Her February 22, 1917, letter admits that "[Mount Holyoke] seemed ideal when I left 7 yrs ago, and now it might suffocate me if I stayed there long enough."[30] Richards's desire for humanistic connection and faith increases as she ages, and Richards often relates her analysis of the current state of the country to Turner, as seen in her February 13, 1932, letter: "Education does not educate emotions of selfishness, & greed & Ego striving. Only the Grace of God does that, & people don't believe in that any more. We are sold to service & culture."[31] Even with Turner's missing correspondence, Richards's portion of the communication exposes a search for identity, meaning, and integrity as the world develops and changes around her and the other women trained and based in late nineteenth and early twentieth century customs and ideologies.

Yet, the developments of Richards's and Turner's epistolary network fostered the communication, analysis, criticism, and growth necessary to directly support them, as well as to indirectly bridge opportunities and advancements to other women in early American science, as noted in many letters in which early American sister schools are referenced. For example, Richards's February 17, 1920, letter updates Turner on a newly formed alum organization at Johns Hopkins and an education rally "in conjunction with Smith, Goucher, Mt. H., Bryn Mawr for endowment campaign interest."[32]

Over a century later, Richards's preserved personal correspondence to Turner (and Turner's unpreserved personal correspondence to Richards) remains the clearest evidence of their personal relationship and the communicative support necessary for them to sustain long-term careers as women in early American science, yet their account remains missing from standardized histories and publications, as it does for so many other women,

unless voluntarily brought to the surface. Macro-level reading and analysis provide further evidence of this neglect; however, this analysis often stops just below the surface, due to humanistic limits. Through the use of data-driven visual networks, further views of words, places, and people are better able to be revealed, helping to widen the scope of perspective, proof, and connection.

> **Microscopic Anatomy** *The examination of structures involving the use of optical instruments, including [histology](#) (the study of tissues), and **embryology** (the study of an organism in its immature condition).*

Through digital networks, readers may identify layers incapable of being penetrated by humanistic practices and utilize visuals to further support, refute, or develop existing analyses. As with any anatomical surgery, expectations are often shifted and/or transformed with surprising discoveries and co-morbid findings. By combining micro-level digital analysis with macro-level critical analysis, correspondence reading becomes not only an accountable set of word, place, and people networks which connected via the postal system, but the correspondence also forms an intricate network of literary tissues which document and connect underlying and preferential choices, topics, and relationships.

Embryology Reading

An embryology reading presents the opportunity to break down the correspondence to its most immature condition: a list of individual words. The process of creating a word inventory for any large set of text—without digital support—is undesirable for most readers. The time, effort, and consistency needed to count and chart the words contained within the 91 letters in the Richards/Turner correspondence archive is daunting and out of reach for most readers. Data analytics, however—and word cloud diagrams in particular—provide not only an accurate and speedy inventory

count of words, but also the potential for visual representations which can quickly expose the frequency of words in a comparative structure.

Figure 6.6: 200 Most Common Words in Letters from Dr. Esther Loring Richards to Dr. Abby Howe Turner, 1915-1932

Full text of all of the letters in the Richards/Turner correspondence were downloaded digitally and processed using Python.[33] A word cloud (above) was generated based on word frequency in the entire corpus of letters.[34]

Immediately, readers can see patterns in the frequency of words in the Richards/Turner correspondence, especially concerning time and actions. A quick glimpse at the Richards/Turner correspondence high-frequency word cloud reveals "year," "day," "time," "week," "till," and "first" to be dominating words within the

correspondence. While date-related references surely do not surprise in postal correspondence, the frequency and range of such words clearly emphasizes the important role of time in the letters and Richards's and Turner's lives. Short-term and long-term temporal qualifiers are matched in their usage and importance throughout the correspondence.

Action words are also frequently utilized, such as "see," "know," "work," and "think." "See" is Richards's most repeated word, incorporated in her letters to Turner as a physical-based desire for vision, as noted in her February 27, 1916, letter ("I wish you were nearer that I might see you once in awhile"); an observation of condition, as expressed in her May 13, 1920, letter ("Whereas I see in patients & people at large a dozen other twists of personal behavior that are just as & even more serious in their results"); and an understanding of situation, as shown in her December 23, 1919, letter ("It is easy to see why she has been discriminated against").[35]

The frequency of "one," is also quite notable—as a number, as evident in Richards's September 15, 1922, letter ("We have on our wards one of Mildred Gutterson's sisters – a Mrs. Smith"); as a nonspecific person, as seen in her October 20, 1921, letter ("One must consider not only the 4 years of confining study, but also the 4 more years of hospital apprenticeship, after which one enters the field of practise to begin the real struggle in competition"); as a societal entity, as viewed in her May 31, 1922, letter ("Caring is a quality that one cannot put into a human being"); and as a pronoun referent, noted in her March 21, 1915, letter ("Ruth Guy has one [a cold], as well as [a] girl in my own class").[36,37]

An embryology reading's strengths rely on the presentation of high frequency words through digital analytics. The ability to quickly and accurately compile word frequency lists in visual format is invaluable when a reader is interested in confirming a critical analysis assumption. As with any inventory-based analysis, an embryology reading's strengths rely on the presentation and the histories, preferences, experiences, and desires of the reader. High frequency count signals repetition, but that repetition does not

necessarily represent content or analytical significance, as was noted from the need to remove non-stop words before performing the data analytics necessary to make a meaningful word cloud and the range of meanings and/or parts of speech for any individual word.

Histology Reading

By using computer algorithms to detect underlying topics in a corpus of work and cluster words based on their association with each topic, readers can view unpreserved movements and correlations between words, similar to the unpreserved motions between mailed letters, time spans between correspondence receipts, and actions between communications. An histological reading, only possible through the micro-level ability of network data processing, starts to reveal the forces supporting the words in preserved correspondence: the tissues holding a large body of work together.

Figure 6.7: 200 Network of Topics and High-Importance Words by Topic in Richards/Turner Letters

Topics within the Richards/Turner correspondence were inductively detected using a technique known as Latent Dirichlet Allocation (LDA). LDA groups words that frequently appear together in the same sources (e.g. letter) and are less frequently paired with other words.[38] Topic Frequency-Inverse Document Frequency (TF-IDF) weighting was used prior to constructing the topic model to increase the relative weight of words in documents where they appear most frequently. The network and visualization were constructed in Cytoscape. Larger nodes represent distinctive topics, whereas the words in smaller nodes are spread fairly evenly throughout the sources. The thickness of each edge is based on how closely the pair of topics are connected by occuring in similar sets of sources.

Immediately, an amplified connection is apparent between the topical groupings revolving around "home" (including "father," "come," "good," "mother," "hard," "sept", "weeks") and "dr" (including "work," "day," "year," "course," "miss," "nurses," "people," "life," and "chief"). The role of time—through "days," "weeks," or "years"—is revealed to be a common thread in both of the largest distinctive topics, whether private or public in their focus.

Other secondary-level distinctive topic tissues include strong relationships between the topics "speak"/"state"/"10″/"times" and "days"/"nursing"/"matter"/"better." Topical grouping around "holyoke" and "hopkins" are not central in this networking visual, but rather secondary and tertiary in their placement. "Hopkins" is viewed, in small significance of high frequency topic connection, in several of the groupings, while "holyoke" stands out as highly frequent and closely connected to "dr" and "home."

Strengths of using histological networking for topic analysis are evident in the visual's ability to demonstrate relational connections and influence both within and across topics. Degrees of connection and force are capable of being perceived and recorded as part of a larger picture of others' writing processes and products. Yet, human assumptions are still inevitable in our own documenting processes to create these products, and individually-preferred choices and

limits must be made when selecting data for entry and exit analysis. Still, this type of micro-level networking provides a cohesive view of long-term correspondence which has been previously impossible to capture—a view which documents the people and places between the words.

Conclusion

Anatomical networks provide surface, underlying, and data-driven views of words, places, and people which expose multiple layers of human experience. As with any series of analyses, including those that are medical based, multiple scans are often necessary to see external and internal components; layered views enhance readings, analyses, and networks of historical text. Macro- and micro-level readings, therefore, need not be performed in exclusion of one another, especially when analyzing personal correspondence. As network technology and humanity continue to advance, so do developments and options for further study, identification, connection, and understanding between words, places, and people. Yet, as Richards herself warns Turner in her November 26, 1917, letter, we must not devalue the human spirit and vision in this process: "The great trouble with many scientific giants today is that they grow enslaved by what they can grow in a test tube, by what they can see thru' a microscope, or do with electricity."[39]

Another major challenge of validating correspondence-based anatomical networks is that the majority of personal communication is not, nor will it ever be, digitized, transcribed, or accessible to the public. We are also still in the early stages of archiving epistolary texts, due to the relatively recent partial extinction of the print letter, new standards of communication modes, the time-consuming and costly transfer of private letters into publicly accessible digital archives, and the necessary but difficult conversations about the most appropriate and ethical methods for representing past networks in present visuals. Still,

as Anais Nin famously noted, "we [continue to] write to taste life twice: in the moment and in retrospect,"[40] and in parallel effort, we must continue to utilize unfolding technologies to create multiple networks to simultaneously view the past and the present–words and patterns that need the eye and the equation to more fully and accurately "see" the bodies of our epistolary selves.

Acknowledgments

Endnotes

1. Dorothy A. Lander, "Love Letters to the Dead: Resurrecting an Epistolary Art," *Omega: Journal of Death & Dying* 58, no. 4 (2008/9): 314.

2. "Postal Office Department Map of Continental U.S. Air Mail Routes," National Archives Catalog, accessed April 15, 2018, https://catalog.archives.gov/id/6857715.

3. "The Esther L. Richards Collection," The Alan Mason Chesney Medical Archives of the Johns Hopkins Medical Institutions, accessed April 10, 2018, http://www.medicalarchives.jhmi.edu/papers/richards_el.html.

4. "A Postcard Collection of Mount Holyoke College," The American Genealogy and History Project, Mount Holyoke College History, https://www.mtholyoke.edu/~dalbino/index.html.

5. Esther Richards, "The Esther Richards Letters, 1915–1920," The American History and Genealogy Project, August 8, 2013, https://www.mtholyoke.edu/~dalbino/letters/erichards1.html.

6. "Esther L. Richards Collection," The Alan Mason Chesney Medical Archives of the Johns Hopkins Medical Institutions, accessed April 10, 2018, http://www.medicalarchives.jhmi.edu/papers/richards_el.html.

7. "Abby Howe Turner 1896," Mount Holyoke College, accessed April 14, 2018, https://www.mtholyoke.edu/~dalbino/women19/abby.html.

8. Richards, "The Esther Richards Letters, 1915–1920."

9. Esther Richards, "The Esther Richards Letters, 1921–1932," The American History and Genealogy Project, August 8, 2013, https://www.mtholyoke.edu/~dalbino/letters/erichards2.html/.

10. Richards, "The Esther Richards Letters, 1915–1920."

11. Richards, "The Esther Richards Letters, 1921–1932."

12. "Abby Howe Turner 1896."

13. "A Detailed History," Mount Holyoke College, accessed April 16, 2018, https://www.mtholyoke.edu/about/history/detailed.

14. "1920 City Directory," Special Collections, Johns Hopkins University, accessed April 8, 2018, https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/33836/1920%20City%20Directory%20.jpg?sequence=2.

15. Richards, "The Esther Richards Letters, 1915–1920."

16. Ibid.

17. Ibid.

18. Ibid.

19. Ibid.

20. Ibid.

21. Ibid.

22. "File:Anne Hall, Mount Holyoke College Class of 1910, high jumping..jpg," Wikimedia Commons, October 2, 2013, https://commons.wikimedia.org/wiki/ File:Anne_Hall,_Mount_Holyoke_College_Class_of_1910,_high_jumping..jpg.

23. Richards, "The Esther Richards Letters, 1915–1920."

24. Ibid.

25. Ibid.

26. Ibid.

27. Ibid.

28. Ibid.

29. "The_Henry_Phipps_Psychiatric_Clinic.jpg," Wikimedia Commons, June 21, 2016, https://commons.wikimedia.org/wiki/ File:The_Henry_Phipps_Psychiatric_Clinic.jpg.

30. Richards, "The Esther Richards Letters, 1915–1920."

31. Richards, "The Esther Richards Letters, 1921–1932."

32. Richards, "The Esther Richards Letters, 1915–1920."

33. Processing of the correspondence, topic modeling, and word cloud visualizations were created by Nathaniel Porter, in consultation with the author, to explore and demonstrate additional analytic options. Data files and code used in analysis available at doi:10.7294/284t-bf10.

34. English stopwords (very common words with minimal semantic value, such as "the") were removed before creating topic models and word clouds.

35. Richards, "The Esther Richards Letters, 1915–1920."

36. Richards, "The Esther Richards Letters, 1921–1932."

37. Richards, "The Esther Richards Letters, 1915–1920."

38. LDA does not label topics in any way; it is strictly inductive and leaves interpretation to the user, based on commonalities among the words most closely associated with each topic.

39. Ibid.

40. Anais Nin, "The New Woman," *In Favor of the Sensitive Man and Other Essays* (Orlando, Florida: Harcourt Brace and Co., 1994), 13.

7. The "First Mortality" as a Time Marker in Fourteenth-Century Provence

NICOLE ARCHAMBEAU

My main research question in this project is to explore how people understood and reacted to the first two waves of plague in 1348 and 1361 by looking at how they talked about the events. Specifically, I analyzed how a group of people who all testified in one canonization inquest used–or did not use–the word "mortality" in reference to waves of plague. A canonization inquest was a large-scale legal procedure sanctioned by the papacy that explored the life events and reputation of a candidate for canonization, primarily by interviewing witnesses to the proto-saint's life and miracles. This particular inquest took place in Provence in 1363, which means that I can date it to a moment after the second wave of plague in 1361 but before the third wave in 1370. The source is especially useful because it includes descriptions of events during both the first and second waves of plague.

Overall, I found that by 1361, some people in this source spoke of a "first mortality" (meaning the first wave of plague in 1348) as a fixed moment around which to date other events. This was not true of everyone in the source, however. For example, many people did not mention the "first mortality" at all, even when it would have made sense to do so. My focused study makes the small, but significant, point that the ways people spoke about catastrophic epidemics could vary, even within a group of people who lived in the same geographic region and shared other characteristics, like religion and affiliation with a proto-saint.

I used network analysis in multiple ways in this project. First, I looked for characteristics that might connect the people who used the term mortality and perhaps suggest a network that was not

clear on the surface of the source. Second, and more importantly, I used network analysis as a way to push against my own assumptions about how people responded to, especially how they spoke about, the first waves of plague. As I constructed network visualizations, I realized that I had assumptions that were not borne out. As a result, the network visualizations prompted me to generate new questions about this data.

Plague and Saints in the Fourteenth Century

Modern and medieval scholars have shown how "the last past plague" can shape expectations of and responses to an emerging epidemic.[1] But from 1347 to 1351, an epidemic spread that had no ready comparison for people at the time. In Europe, it killed "an estimated 40%-60% of the population."[2] Although late medieval Europeans experienced epidemics with some regularity, *this* epidemic was different. As Ann Carmichael writes, "[W]ithin some finite period of time after the great mortality became part of their past, survivors began to characterize its distinctiveness from other epidemics."[3] But they did not have a last past plague to compare it to.

In 1361, however, a second wave of this plague moved through Europe. The epidemic was no longer a unique catastrophe that people had to understand in a world without that disease. For these people, there had been a last past plague. Everyone over the age of 15 had now lived through two waves of plague. People over 20 to 25 years old could remember both. And people of every age group and social group spoke to each other, in some cases shaping their experiences around these two moments of high mortality. In 1361 they could use the last past plague to understand their experiences.

These canonization inquest documents bring together a group of 68 witnesses who had all lived through two waves of plague. This particular inquest took place in 1363, which meant the second wave was fresh in their minds, but the first wave of plague was

not in the distant past. In terms of network analysis, a group of witnesses in a canonization inquest is a de facto network of sorts. All of the witnesses shared a faith in the holy person's sanctity and had been gathered by local inquest organizers to testify. This was not a random group of people.

The faith they shared reflects the medieval culture surrounding sainthood, which was an institution that people used to solve personal problems, deal with changing environment and political situations, or manipulate the physical world. Late medieval sainthood was also an institution that generated extensive written documents. This canonization inquest fits into a larger branch of research on medieval plague that uses surviving written legal sources, like wills and court cases, to see the impact of plague on daily life and family choices.[4] These kinds of legal sources allow modern scholars to see reactions to plague beyond the more famous literary and medical sources.

The Canonization Inquest for Countess Delphine

I am using the canonization inquest for Countess Delphine de Puimichel, which took place in Apt and Avignon, Provence, which was then a county in the Kingdom of Naples.[5] By the mid-fourteenth century these inquests were elaborate legal procedures with extremely high standards and high stakes.[6] Like all fourteenth-century canonization inquests, the organizers of Delphine's inquest gathered evidence to see whether or not this local holy woman should be considered an official saint of the Catholic Church.[7]

Great prestige and potentially great profit could come from having an official Catholic saint in one's community, so the process was taken very seriously. During the inquest into Delphine's sanctity, two papal commissioners and at least one official papal notary traveled to the place where Delphine had lived. They joined local organizers, most importantly a local notary named Master Nicholas Laorench, who acted as proctor of the inquest. Master

Nicholas gathered witnesses and wrote the 98 articles of questioning. The joint papal and local group interviewed people who had known Delphine or experienced miracles by praying to her. The local and papal notaries then collected the written testimonies and other materials and gave them to the papal court.

The final document produced for Delphine's inquest was a 204-folio collection of official papal letters, opening statements, a list of witnesses, a summary of daily events, 98 articles of questioning, 68 witness testimonies, supplementary materials provided by the local organizers, and closing statements by the two official notaries.[8]

The document was for internal use within the papal curia. It would be used by a small number of papal officials as they considered Delphine's canonization. Most of these officials would never read the text, however. Instead they would read a summary of the inquest produced by a papal notary. They would likely only read the inquest documents if a debate arose about a specific miracle or event.[9] The audience is important here. This was primarily an internal document, not a didactic document, like a saint's life (also called a *vita*) meant for a wide readership. Therefore the witness testimonies did not have to be deleted, screened, or reconstructed in order to teach people how to be better Christians.

The most useful parts of the inquest for this study are the witness testimonies and articles of questioning. Each witness was interviewed individually. The testimonies were written down by two notaries, a local notary and the papal notary. In Delphine's inquest (as in most other inquests), each witness testimony starts with the statement of swearing in. Some testimonies include a statement about the witness speaking their maternal tongue; for this group, that language was Provençal. The notaries translated the testimonies into Latin, which was the common language of the papal court. The testimonies were also written down in the third person, rather than the first person.

Each witness was given the opportunity to speak to all 98 articles of questioning. These articles were statements about Countess

Delphine's life events and miracles and were produced uniquely for this inquest. They were written by a local notary, Master Nicholas Laorench, who had been part of Countess Delphine's entourage since 1351. There is evidence that Master Nicholas wrote the articles of questioning based on stories told to him by various people chosen to testify in the inquest.[10]

Master Nicholas also wrote an open-ended article of questioning—Article 1—that asked witnesses to describe anything they knew about Countess Delphine. The witnesses and papal commissioners took advantage of this article. In response to it, witnesses told stories about Delphine, themselves, and others that appear nowhere else in the inquest. The papal commissioners frequently asked follow-up questions to responses to Article 1, including questions along the line of "What else do you know?" Since Countess Delphine's inquest happened less than three years after her death, this is not surprising. There had not been much time for a local following to emerge, and the local officials and papal commissioners needed every story they could get to show that local people did or did not consider Countess Delphine a saint.

During questioning, as the witnesses responded to articles of interrogation, they described events, agreed or disagreed with the articles, or told their own stories related to the articles. In other words, they did not strictly repeat information in the article, nor were they limited by the language of the article.[11] Each testimony also included information about age, sex, social status, clerical status, and where the witness was from.

These testimonies are a useful source for reaction to the two waves of plague.[12] Although there were no articles of interrogation about plague, witnesses used phrases that included the term mortality, which was how they referred to the waves of plague. (No one used a word like pest, pestilence, or plague.) And witnesses did talk about the two waves, particularly in response to the open-ended Article 1. Some witnesses made requests for miraculous healing. Although learned medicine was increasingly popular and available by the mid-fourteenth century, most of Europe still

considered an appeal to God's grace through a holy person as a valid healing option.[13] People appealed to saints on their own and others' behalf for healing from many injuries and illnesses, including plague.

These testimonies are also a robust resource because they include a diverse group of people. Canonization inquest testimony included people often left out of the historical record because they did not write. As Michael Goodich puts it, "The details provided in miracle stories—the who, what, when, where, why and how of any inquiry —especially those reported in the framework of a papal canonization process, which demanded high judicial standards, may assist us in recapturing the voices of otherwise inarticulate folk."[14] While most of the witnesses in Delphine's inquest were educated, relatively wealthy, and well traveled, it still included many people whose voices would usually not be heard, especially women. Their individual testimonies were required for a successful canonization, so clergy copied their words carefully. Organizers did not want the inquest to fail because there was not enough local support or the testimonies were too homogeneous.[15]

Through word choices and witness characteristics, therefore, I hoped to uncover networks within this group of witnesses who were already under the umbrella network of Delphine's canonization.

Methods of Analysis

Testimonies like these are a potentially robust resource for network analysis. First, as I pointed out above, this group of witnesses is, in many ways, a network already. The witnesses shared the common link of belief in and use of the same proto-saint. Also, in this inquest, the majority came from the same geographical region—southeastern Provence—so they shared similar experiences and cultural expectations. It is also clear from witness testimony that many of these people knew each other. In

other work, I have used network analysis and visualization to explore how the witnesses referred to each other and people outside the inquest in their testimonies.[16]

With this project, I knew that I wanted to see if there were patterns within this group concerning how people spoke about the waves of plague. All 68 witnesses had lived through both waves of plague–one in 1348 and one in 1361. The youngest witness might not have remembered the first wave all that well (he would have been five), but the average witness age was roughly 35 at the time of the inquest, so most would remember both.

I used network analysis and the visualization tool, Cytoscape, in the hopes of revealing a group of witnesses who all spoke of plague a certain way and shared identifiable characteristics, like sex, age, or clerical status. This might indicate a group of people connected to one another in a way not clear on the surface of the inquest.

I analyzed the testimonies to find people who spoke about events in 1347–1349–dates that could be associated with the first wave of plague–and who spoke about 1361, which was associated with the second wave of plague in Provence. I assembled a table which included all of the witnesses, what phrase they used, and the article of interrogation they were speaking about.[17] I then created three tables that broke down the witnesses into groups of whether they mentioned the word "mortality," did not mention it, or used multiple methods to refer to these time periods. In these tables, I included personal information for each witness.

The tables were useful, but it was not easy to see patterns of how people spoke of the plague or if certain groups of people spoke in certain ways. So I used Cytoscape to create different visualizations of the various data points in order to see if patterns or a network emerged within the network of Delphine's witnesses. I was particularly interested in any networks emerging around sex, age, or clerical status. Because I found that the ways people spoke about 1347–1349 differed significantly from the way they spoke about 1361, I created different visualization sets for the two waves of plague.

For both sets of visualizations, I attempted to find all of the different ways that people referred to the same moments in time. I found four main methods for 1347–1349, including specific dates, years ago, a reference to mortality, or multiple methods at once. These appear in figure 7.1.

Figure 7.1: The number of references each witness made (indicated by number of lines)

As this visualization shows, the time references for 1347–1349 were diverse. The majority of witnesses used one method, but not all did so. Some witnesses, like Lady Raynauda Laugeri, used multiple methods of marking time. She used the phrase "the first mortality" for one event in 1348, but dated another event as happening 15 years ago. One of the only patterns to emerge was a group of three nuns who combined time references. They said that an event happened "after the time of the first mortality, around 14 years ago."[18] However, at least one of those nuns also referred to something only by using years ago, so this is not a strong pattern.

In contrast, for 1361 I found only two methods—a reference to mortality or years ago. These appear in figure 7.2. Unlike for 1347–1349, in which everyone who referred to mortality used the phrase "first mortality" in some way, the references to mortality in 1361 were diverse. Witnesses used phrases like "the most recent mortality" or just "the mortality."[19]
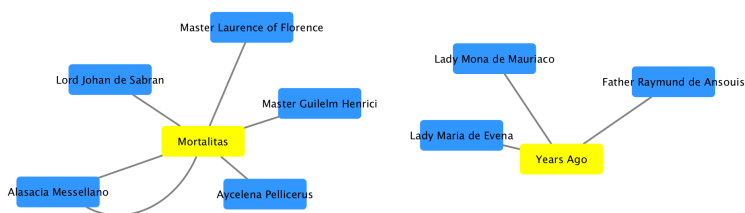
Figure 7.2: The number of references each witness made (indicated by number of lines)

Figures 7.1 and 7.2 establish that people used different methods of referring to these two time periods. This speaks strongly against homogenization of witness testimony by the notaries copying the testimony and translating it into Latin. I am making the assumption here that if the notaries had homogenized the testimony, they would have chosen one or maybe two methods for marking time rather than four. Therefore, looking at these witness testimonies can reveal how people spoke about the waves of plague. These visualizations, however, did not reveal any obvious patterns that would suggest networks within the inquest.

Finding multiple methods of marking time, I looked for patterns in who used which methods. Overall, I looked at sex, age, and clerical status. Surprisingly, I did not find significant networks or patterns emerging around any category. In terms of gender, the witnesses who spoke about 1347–1349 included 6 men and 13 women, seen in figure 7.3. While there are more women, these women did not all talk about the same event nor use the same phrases, so there was not a strong pattern.

Age also did not reveal any clear patterns. The witnesses' ages ranged from 28–65, but no one group used a specific method of referring to 1347–1349. In figure 7.4, I gave each decade a different color, but found no significant patterns emerging among thirty-year-olds or fifty-year-olds, for example.

Figure 7.3: Gender of witnesses (green indicates female witnesses, blue indicates male witnesses)

Figure 7.4: Ages of witnesses (pink indicates 20s, orange indicates 30s, blue indicates 40s, lavender indicates 50s, light green indicates 60s, and dark green is unknown)

The witnesses came from diverse backgrounds. One main division was religious vs. lay people (figure 7.5). The religious included six individuals from four institutions. Lay people included 13 individuals, including four members of the aristocracy, a lawyer from the royal court in Aix-en-Provence, two merchants, three diverse female inhabitants of Apt and Ménerbes, and two of Delphine's long-term companions Bertranda Bartholomea and Catherine de Pui.[20]

Figure 7.5: Religious vs. lay people (violet indicates the witness was a religious)

Although I did not find significant patterns in the categories of sex, age, and clerical status that I had expected, these results helped me ask new questions. These new questions emerged from two strong patterns in how people spoke about 1347–1349. First, although people used the phrase "first mortality," they rarely talked about plague. Only one of the 19 witnesses described someone suffering from the illness that caused the first mortality (see figure 7.6). Instead, witnesses used it as a time marker for something else.

Figure 7.6: How witnesses spoke about 1347–1349 (red indicates the witness spoke of plague)

This contrasts to how people spoke about 1361. Out of eight witnesses who spoke about this time period, four spoke about their own or another's experience of the epidemic illness in 1361 (see figure 7.7).

Figure 7.7: How witnesses spoke about 1361 (red indicates the witness spoke of plague)

Figures 7.3–7.7, however, did not produce a clear group of people (based on age, sex, status, or location) who used references to plague. This was a surprise for me, and was a worthwhile use of network visualization. Although I did not find the patterns I expected, I realized that I had assumed patterns were there, but I just was not seeing them in the tables. Seeing the information in different ways, pushed me to reassess my expectations.

Figure 7.8: How witnesses referenced time (green indicates a time reference before 1348, orange indicates a time reference of 1348, yellow indicates a time reference after 1348, and grey indicates a time span that included 1348)

Since witnesses used references to the first mortality as time markers for other events, I decided to look for patterns and perhaps

networks in what they dated using the different methods. Sometimes they used references to plague as a time marker for events happening during 1348–1349, but they also referred to events before and after. Or they referred to a span of time (see figure 7.8).

I focused my analysis on people who used the phrase "first mortality." For these witnesses, the first wave of plague was a fixed point in relation to which they remembered other events.[21] Considering the general categories of before, during, after, or a span did not reveal any kind of pattern or network, however.

Finally, I tried to map what specific events witnesses dated with references to the plague. No clear network emerged. Again, this was a surprise—even more of a surprise than the lack of connections or networks based on witnesses' personal information. Witnesses dated all kinds of events with references to the plague, which my rather wild figure 7.9 shows. In this visualization, I link witnesses who mentioned either the first or second mortality to the articles of interrogation they were responding to. As stated above, there were roughly 100 articles of interrogation and witnesses referred to mortality in response to roughly a quarter of them.
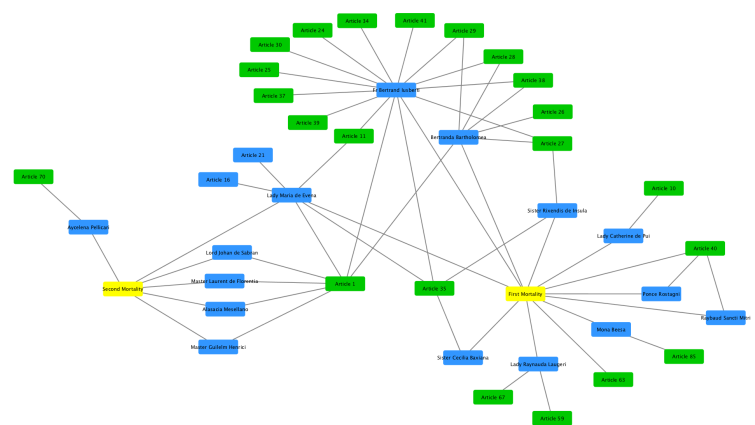

Figure 7.9: How witnesses dated events with plague references (blue indicates a witness, green indicates an article)

Through visualizations like this, I understood that witnesses did not associate one particular event or characteristic of Delphine's sanctity with plague. Different witnesses associated the plague in their memories with a wide variety of things, represented by the many different articles (in green) in the visualization.

Conclusions

Overall, network visualization allowed me to look at information that I am very familiar with in a new way. In particular, I did not find the networks or patterns I expected. Instead, unexpected patterns—like the fact that while many people used the phrase "first mortality," only one person actually spoke about the first epidemic illness—seemed important, but did not reveal a network. Seeing this in the visualization pushed me to reconsider how witnesses understood the first mortality as part of their lives.

Once I saw the lack of clear networks based on witness characteristics or with what witnesses associated the first mortality, I knew I needed to reconsider my assumptions about witness testimony. These witnesses not only had freedom in their word choices about this time period, they in fact made different choices about words to use. This spoke strongly to individual autonomy of the witnesses. It was clear that the years 1347 to 1349 stood out in many people's minds, but not everyone spoke about them the same way.

A specific example will help us see those individual choices. Friar Bertrand Iusberti used the phrase "first mortality" 16 times to date events before, during, and after 1348, and he used it to mark the span of time between 1348 and Delphine's death. In contrast, Lord Aycardus Boti never used the phrase "first mortality," even though he spoke of events in 1349 five times. For one of these events, he refers to hearing about it from Friar Bertrand Iusberti, who may have used the phrase "first mortality" in his hearing.[22]

Both men held positions of influence in Apt, Provence, and were roughly the same age. While I cannot know exactly why Lord Aycardus did not use the phrase and Friar Bertrand did, I can see from these visualizations that they both had the option, and they both made a choice.

The striking difference in the ways witnesses spoke about the second wave shows that they thought about it differently from the first wave. Even though far fewer witnesses mentioned the second wave, four times as many spoke about the epidemic illness. It was as if having a last past plague, or in this case a "first mortality," allowed them to talk about the illness itself. This moment was used far less frequently to refer to other events, however. In 1363, it did not have the cultural resonance of the first mortality—there was no one phrase everyone used, people did not use it to reference significantly earlier events—and was not as robust of a term.

Appendix

Table 7.1: References to the first mortality, second mortality, dates, and years ago[23]

	Article or Witness	Page #	First or Second	Word or phrase	In relation to Countess Delphine's life or miracles
1	Article 40	56	F	"in hospicio pontem staret citra primam mortalitatem quasi per duos annos"	Time reference to wondrous light seen in her room when she stayed near the bridge in Apt
2	Article 63	75	F	"quod dum semel post primam mortalitatem"	Time reference to healing of a woman named Saura when Delphine went to Cavaillon to negotiate peace between warring lords
3	Article 70	79	S	"generali mortalita"	Time reference for the death of the recipient of a miracle
4	Noble Lady Mona de Mauriaco	145	S	"dixit quod erant in proximo mense Augusti duo anni"	Time reference to a miraculous healing
5	Fr. Bertrand Iusbert	205	F	"videlicet a tempore mortalitatis prime usque ad diem obitus sui"	Time reference for how long he had observed Delphine's life
6	–	207	F	"dixit quod a prima mortalitate citra"	Time references for when he had spoken to Delphine about her virginity (roughly article 11)
7	–	–	–	"fuit infra primum annum post dictam mortalitatem"	Time reference for when Delphine made a full, general confession to him (roughly Article 30)
8	–	207-208	–	"citra tempora dicte prime mortalitatis et per aliquos annos ante dictam mortalitatem"	Time reference for when he had heard from lord Guido and others about Delphine (article 1)

Table 7.1 *(continued)*

9		208	F	"per aliquos annos ante mortalitatem predictam"	Time reference for when he heard and saw people talking about Delphine's conversing and praying (roughly article 25)
10	–	216	F	"vidit ante mortalitatem primam"	Time reference for when he saw her evading worldly honor (roughly article 24)
11		225	F	"dixit p- mortalitatem primam et citra:	Time reference for Delphine's tears and consumption of brain (article 27)
12	–	226	F	"dixit quod post primum mortalitatem,"	Time reference for article 28
13	–	–	–	"dixit quod quadam vice ante mortalitatem primam"	Time reference: when he saw and heard about the events of article 29
14	–	230	F	"videlicet ante mortalitatem primam et post"	Time reference for article 34
15	–	231	F	"sed a tempore mortalitatis prime quo fuit moratus cum dicta domina"	Time reference for article 35
16	–	232-3	F	"quod quadam die ante mortalitatem primam"	Time reference for article 37
17	–	233	F	"quod quadam vice circa magnam mortalitatem"	Time reference for article 38, esp the problems between Raymund Agoult and Hugo of Baux
18	–	234	F	"a tempore prime mortalitatis citra"	Time reference for article 38 about the dissention between the counts
19	–	235		–	Time reference for article 39
20		236	F	"dixit quod a tempore prime mortalitatis citra, quo morabatur cum ipsa domina Dalphina"	Time reference for article 41
21	Maria de Evena	281	F	"anno prime mortalitatis"	Time reference for when her husband was greatly ill and no one believed he would live (Article 1)

<div align="center">

Table 7.1 (continued)

</div>

22		282	S	"vidit post mortem dicte domine Dalphine circa Quadragesimam, et sun elapsi duo vel tres anni aut circa" (+ footnote)	Time reference for illness of boy, Franciscus, who had fever and stomach flux (Article 1)
23	–	283	F	"ante mortalitatem primam, per unos vel duos annos"	Time reference for hearing about Delphine's virginity (roughly Article 11)
24	–	–	–	"infra annum dicte prime mortalitatis"	Time reference for when she began to notice what Delphine wore (roughly Article 21)
25	–	285	F	"ab anno sequenti proxime post mortalitatem primam usque ad diem obitus sui"	Time reference about Delphine as a faithful Catholic and how long she listened to the good words of Delphine (Article 16)
26	–	287	F	"anno sequenti proxime post primam mortalitatem usque ad tempus obitus dicte domine Dalphine"	Time reference for article 35 – about how long she had been hearing Delphine speak to groups and transform and console them
27	Aycardus Boti local official of Apt	294	F	"XIV anni elapsi"	Witness's fever
28		296	F	"XIV anni elapsi"	Niece becoming a nun
29		297	F	"sunt bene XIV anni elapsi vel circa"	Article 16
30		298	F	"bene sunt XIV anni elapsi, vel circa"	Article 35 (spoke to Bertrand Iusbert)
31		299	F	"dixit quod sunt bene XIV anni"	Article 35
32	Bertranda Bartholomea	328	F	"per unum annum ante primam mortalitatem"	Time reference for Article 26
33	–	329	F	"a tempore prime mortalitates citra pluribus et diversis vicibus usque ad obitum ipsius"	Time reference for Article 27 – about Delphine's illnesses including her tears
34	–	–	–	"a XII annis ante primam mortalitatem citra usque ad obitum dicte domine Dalphine"	Time reference for Article 28
35	–	330	F	"a XII annis ante primam mortalitatem citra"	Time reference to Article 29

Table 7.1 (continued)

36		337	F	"dixit quod post primam mortalitatem" (+footnote)	Time reference to Article 38
37	Johan de Sabran	347	S	"tempore mortalitatis prime proxime preterite"	Time reference to a girl who was ill, but not with plague (Article 1)
38	Laurence of Florence	359	S	"de anno Domine MCCCLXI, et de mense Maii vel Iunii, de die tamen non recordatur, quo tempore vigebat magna mortalitas Aquis	Time reference to his own illness and recovery through a vow to Delphine (Article 1)
39	Guillem Henrici	363	F	"in anno Domini MCCCXLIX"	In reference to hearing about the public fama of Delphine's virginity in Article 1
40		366	F	"in anno Domini MCCCXLIX"	In reference to hearing Delphine speak words of God in Article 1
41		370	S	"in civitate Aquensi magna mortalitate vigente"	Time reference to Laurence's illness and recovery (also calls it lo cat) (Article 1)
42	Raybaud Sancti Mitri	378	F	"cum quadam vice citra primam mortalitatem quasi per duos annos"	Time reference for seeing light in Delphine's room (Article 40)
43	Sister Cecilia Baxiana	384	F	"post mortalitatem primam, sunt bene XIV anni elapsi vel circa"	Time reference for her widowhood and her transformation recalled in her testimony to Article 35
44	Catherine de Pui	388	F	"a tribus annis ante primam mortalitatem, et possunt bene esse XVIII annis"	Time reference for her speaking to Delphine's sister about Delphine's marriage (Article 10)
45		396	F	"dixit quod sunt bene XV anni elapsi vel circa"	Time reference for Delphine in Cabrieres, (Article 26)
46	Lady Grossa Autriga	419	F	"audivit a XVI annis et citra"	Time reference for hearing about Delphine's public fama in Article 1
47		420	F	"sunt bene XV anni elapsi vel circa"	Time reference for healing of her mother, Bauda de Rellania's, healing of a continual fever – face to face with Delphine and whispered words
48	Aycelena, wife of Petrus Pellicerus	422	S	"dixit quod ex tunc usque as mortalitatem proxime preteritam	Time reference to Article 70

Table 7.1 *(continued)*

49	Alasacia Messellano	432	F	"sunt bene XIV anni elapsi vel circa"	In reference to a miraculous healing after a fall
50		435	S	"subtus aurem tempore mortalitatis"	Time reference for the illness of her grand-daughter (Delphine's goddaughter) Delphina, who had fever and tumor (Article 1)
51	–	436	S	"quod tempore mortalitatis ultime et proxime preterite, de anno et mense Iulii proxime nominatis"	Time reference for her own fever and tumor (which everyone who had it died); she was given extreme unction, but was speaking as if demented and not in "bona memoria" (Article 1)
52	Bartholomea Macella of Cabrieres	454	F	"sunt bene XVI anni elapsi"	In regard to Article 58
53	Raynauda Macella of Cabrieres (widow)	456	F	"XVI anni sunt elapsi"	In regard to Article 58
54	Mona Beesa	457	F	"per unum annum post mortalitatem primam; et sunt bene XIV anni elapsi, ut sibi videtur, vel circa, et de mense Septembris"	Time reference for a fever she had for six months, Article 85 (not plague)
55	Aycelena de Apta (Abbess holy cross)	481	F	"audivit a XVI annis citra"	Time reference for Article 35
56		484	F	"a XVI annis citra"	Time reference for Article 35
57		484	F	"a XV annis citra"	Time reference for Article 35
58	Sister Rixendis de Insula (nun Holy Cross Convent)	486	F	"tempore mortalitatis prime, et sunt bene XV vel XVI anni elapsi, ut sibi videtur"	Time reference for Article 27
59		488	F	"dixit quod a XVI annis citra"	Time reference for Article 35
60	–	489	F	"anno predicte mortalitate sunt bene XVI anni elapsi, ut sibi videtur"	Time reference for widows transformations in Article 35
61	Sister Maybilia Raymunda (nun Saint Katherine's)	501	F	"erunt XVI anni elapsi"	Article 60

Table 7.1 (*continued*)

62	Raynarda Laugeri	510	F	"dixit quod a XV annis citra"	Time reference for hearing about Delphine's virginity (Article 1)
63		511	F	"post primam mortalitatem, et sunt XV anni elapsi vel circa"	Time reference for Francisca's fever (Article 59)
64	–	–	–	"anno prime mortalitatis infra XV dies post festum nativitatis sancti Iohannis Baptiste, vel circa"	Time reference for her own fever (Article 67)
65	Raymond of Ansouis (priest)	516	S	"fuerunt duo anni elapsi"	Time reference for infirmity with fever and bossa (515)
66	Philippe Cabassoles (bishop)	542	F	"quod bene sunt XIV anni elapsi"	Time reference for Article 38
67	Ponce Rostagni	546	F	"sicut in articulo continentur"	Time reference for seeing light in Delphine's room (article specifies primam mortalitatem) (Article 40)

Table 7.2: Witnesses Referring to Mortality.

	Witness	Page #	First or Second	# of Mentions	Title	Sex	Age	Information
1	Fr. Bertrand Iusberti	205-236	F	16	Franciscan Friar	M	40	Guardianus of the Friars in Apt, close associate and confessor of Countess Delphine for 15 years.
2	Noble Lady Maria de Evena	281	F	5	Noble	F	28	Noble wife of Lord Giraud de Simiana, Lord of Apt and Casaneuve
3	Bertranda Bartholomea	328-337	F	5	Maid	F	60	Delphine's maid for almost 50 years
4	Noble Lord Johan de Sabran	347	S	1	Noble	M	23	Relative of Countess Delphine by marriage
5	Master Laurence of Florence	359	S	1	Court Official	M	29	Legal Official in the Queen's court in Aix
6	Master Guillelm Henric	370	S	1	Court official	M	65	Senior legal official in the Queen's court in Aix
7	Raybaud Sancti Mitri	378	F	1	Merchant	M	50	Draper of Apt
8	Sister Cecilia Baxiana	384	F	1	Nun	F	35	Nun in the Holy Cross Convent
9	Aycelena Pelliceri	422	S	1	Merchant	F	30	Wife of local merchant Petrus Pelliceri

Table 7.2 (*continued*)

10	Alasacia Messellano	435–436	S	2	Merchant	F	50	Widow of Johan Messellano, draper of Apt
11	Noble Lady Raynauda Laugeri	511	F	1	Noble	F	50	Widow of Noble Lord Guillermi Laugeri of Apt
12	Mona Beesa	457	F	1		F		

	Witness	Page #	First or Second	# of Mentions	Title	Sex	Age	Information
1	Noble Lady Mona de Mauriaco	145	S	1	Noble	F	30	Noble widow of Rigonis de Mauriaco, militis, of Paternis, vicar of Malausana for Pope
2	Noble Lady Maria de Evena	282	S	1	Noble	F	28	Noble wife of Lord Giraud de Simiana, Lord of Apt and Casaneuve
3	Lord Aycardus Bot	294-299	F	5	Local Noble	M	44	Member of a powerful local family of Apt
4	Master Guillelm Henric	363-366	F	2	Legal Official	M	65	Senior legal official in the Queen's court in Aix
5	Lady Catherine de Pui	396	F	1	Local Noble	F	35	Member of a powerful local family in Bonnieux; Countess Delphine's close associate
6	Lady Grossa Autriga	419-420	F	2	Local Noble	F	28	Widow of Lord Boniface of Vaqueri
7	Alasacia Messellano	432	F	1	Merchant	F	50	Widow of Johan Messellano, draper of Apt
8	Bartholomea Macella of Cabrieres	454	F	1		F	50	Inhabitant of Cabrieres

Table 7.3 *(continued)*

9	Raynauda Macella of Cabrieres	456	F	1		F	28	Widow of Raymund Macelli of Cabrieres
10	Abbess Aycelena de Apt	481–484	F	3	Abbess	F	40	Abbess of the Holy Cross Convent
11	Sister Rixendis de Insula	488	F	1	Nun	F	37	Nun in the Holy Cross Convent
12	Sister Maybilia Raymunda	501	F	1	Nun	F	35	Nun in St. Catherine's Convent
13	Noble Lady Raynauda Laugeri	510	F	1	Noble Lady	F	50	Widow of Noble Lord Guillermi Laugeri of Apt
14	Father Raymund of Ansouis	516	S	1	Priest	M	28	Priest in Marseille
15	Cardinal Philippe Cabassoles	542	F	1	Cardinal	M		Bishop of Cavaillon during Countess Delphine's life

Table 7.4: Witnesses using multiple reference methods at the same time

	Witness	Page #	First or Second	# of Mentions	Title	Sex	Age	Information
1	Master Laurence of Florence	359	S	1	Court Official	M	29	Legal Official in the Queen's court in Aix
2	Sister Cecilia Baxiana	384	F	1	Nun	F	35	Nun in the Holy Cross Convent
3	Lady Catherine de Pui	388	F	1	Local Noble	F	35	Member of a powerful local family in Bonnieux; Countess Delphine's close associate
4	Mona Beesa	457	F	1	Townsperson	F	40	Townsperson in Ménerbes
5	Sister Rixendis de Insula	486–489	F	2	Nun	F	37	Nun in the Holy Cross Convent
6	Noble Lady Raynauda Laugeri	511	F	1	Noble	F	50	Widow of Noble Lord Guillermi Laugeri of Apt
7	Ponce Rostagni	546	F	1	Merchant	M	30	Merchant of Apt

Endnotes

1. Theresa MacPhail, *Viral Network: A Pathography of the H1N1 Influenza Pandemic*, (Ithaca: Cornell University Press, 2014); Ann Carmichael, "The Last Past Plague: The Uses of Memory in Renaissance Epidemics," *Journal of the History of Medicine* 53 (1998): 132–160.

2. Monica Green, "Editor's Introduction to Pandemic Disease in the Medieval World: Rethinking the Black Death," in *The Medieval Globe*, vol. 1 (2014): 9. For estimates of the impact on Provence of the first wave of the Black Death, see Ole Benedictow, *The Black Death 1346-1353: The Complete History*, (Suffolk: The Boydell Press, 2004), 308–315.

3. Ann Carmichael, "Universal and Particular: The Language of Plague 1348–1500," in *Pestilential Complexities: Understanding Medieval Plague*, ed. Vivian Nutton (London: Wellcome Trust Centre for the History of Medicine at UCL, 2008), 19.

4. Daniel Lord Smail, "Accommodating Plague in Medieval Marseille," *Continuity and Change* 11 (1996): 11–41 and Shona Kelly Wray, "Boccaccio and the Doctors: Medicine and Compassion in the Face of Plague," *Journal of Medieval History* 30 (2004): 301–22.

5. I am using the critical edition by Jacques Cambell, OFM, *Enquête pour le procès de canonisation de Dauphine de Puimichel, Comtesse d'Ariano* (Turin: Bottega d'Erasmo, 1978). Page references throughout refer to Cambell's critical edition. For an initial study of the witnesses in Delphine's inquest, see Pierre-André Sigal, "Les temoins et les temoignages au procès de canonisation de Dauphine de Puimichel (1363)," *Provence Historique* 195–196 (1999): 461–71.

6. André Vauchez, *Sainthood in the Later Middle Ages*, trans. Jean Birrell (Cambridge: Cambridge University Press, 1997), 33–57.

7. Vauchez, *Sainthood in the Later Middle Ages*, 33–57.

8. For the influence of notaries in canonization inquests, see Jacques Dalarun, *La Sainte et la Cité: Micheline de Pesaro (1356) Tertiaire Franciscaine* (Rome: Ecole Francaise de Rome, 1992). For a broad overview, see Raimondo Michetti, ed., *Notai, Miracoli e Culto Dei Santi: Pubblicità e Autenticazione del Sacro tra XII e XV Secolo* (Milan: Dott. A Giuffrè Editore, 2004).

9. Michael Goodich, *Miracles and Wonders: The Development of the Concept of Miracle, 1150–1350* (Aldershot: Ashgate, 2007), 88–99. There is also some evidence that authors of saints' lives might have access to the canonization inquest. The friar who wrote Delphine's life may have had access to these documents or just the summary; it is not clear. Jacques Cambell, *Vies Occitanes des Saint Auzias et de Sainte Dauphine* (Rome: Bibliotheca Pontificii Athenaei Antoniani, 1963).

10. Nicole Archambeau, "His Whole Heart Changed: Political Uses of Mercenary's Emotional Transformation," in *Politiques des émotions du Moyen Âge*, eds. Damien Boquet and Piroska Nagy (Florence: Sismel, Edizione del Galluzzo, 2010), 69–90.

11. Laura Smoller, "Miracle, Memory, and Meaning in the Canonization of Vincent Ferrer, 1453–1454," *Speculum* 72 (1998): 429–54; Michael Goodich, "*Mirabilis Deus in Sanctis Suis*: Social History and Medieval Miracles," in *Sign, Wonders, Miracles: Representations of Divine Power in the Life of the Church*, ed. Kate and Jeremy Gregory Cooper (Suffolk: The Boydell Press, 2005), pp. 135–56, 143–44.

12. Nicole Archambeau, "Healing Options during the Plague: Survivor Stories from a Fourteenth-Century Canonization Inquest," *The Bulletin of the History of Medicine*, 85 (2011): 531–59.

13. Michael McVaugh *Medicine before the Plague: Practitioners and Their Patients in the Crown of Aragon 1285–1345* (Cambridge: Cambridge University Press, 1993), 78–87; Joseph Ziegler, "Practitioners and Saints: Medical Men in Canonization Processes in the Thirteenth to Fifteenth Centuries," *The Society for the Social History of Medicine* 12 (1999): 191–225; Iona McCleery, "Multos ex Medicinae Arte Curaverat, Multos Verbo et Oratione: Curing in Medieval Portuguese Saints' Lives," *Studies in Church History* 41 (2006): 192–202.

14. Goodich, *Miracles and Wonders*, 4.

15. Smoller, "Miracle, Memory, and Meaning," 429–54.

16. This study is forthcoming in a book-length project.

17. All four tables appear in an appendix in the digital version of this publication.

18. For example, Sister Rixendis de Insula states "a tempore mortalitatis prime, et sunt bene XV vel XVI anni elapsi." Cambell, *Enquête*, 486.

19. See Table 1 for the different phrases witnesses used.

20. There are no representatives in this list from the very poor—laborers, artisans, or farm workers—that appear in other canonization inquests. We don't know exactly why, but the two likeliest reasons are, first, the fact that Delphine had only been dead for two and a half years. There had been no time for the slow process of building a local cult at her tomb. Second, the impact of several mercenary invasions made it difficult to travel, especially for the poor.

21. This did not happen with witnesses speaking about the epidemic in 1361. They did not use the second wave of plague as a time marker for earlier or later events, nor did they use it to mark a span of time.

22. Cambell, *Enquête*, 298.

23. Names in red did not refer to a first or second mortality even though they spoke about events that took place at the time of one of the mortalities.

8. "Trois Empreintes d'un Même Cachet": Toward a Historical Definition of Nutrition

A. R. RUIS

"There is no subject of more interest to the physiologist, of more practical importance to the physician, or that more urgently demands the grave consideration of the statesman," wrote the English physician George Budd in 1842, "than the disorders resulting from defective nutriment."[1] This assertion proved no mere hyperbole. Over the following century, concern about the pernicious effects of malnourishment only became more widespread, and the study of human nutrition expanded from a minor branch of physiological chemistry to a major domain of biomedical science. Yet as Budd's claim implies, it is overly simplistic to understand human nutrition (or malnutrition) as merely a physiological process, however complex. Nutrition was less a rigorously defined scientific concept than a flexible semiotic device that provided intelligible and actionable explanations for many complex, elusive, or otherwise intractable problems of clinical medicine, public health, and political economy. "Medicine has recently and rapidly developed a keen nutrition consciousness," wrote the American chemist Henry Sherman a century later. "It is finding in nutrition the solutions of many of its most baffling problems."[2]

By the twentieth century, the concept of nutrition—and by extension, the discipline of nutrition—had become deeply entangled with a range of issues: agriculture, health, economics, defense, labor, education, and national identity, among others. Yet as scientists and physicians were extolling the importance of nutrition to just about everything, they increasingly struggled to articulate

just what "nutrition" was. The American physician and nutrition expert George Palmer, for example, noted in 1930 that nutrition "is an ambiguous term. It awaits a specific definition."[3] It is, by and large, still waiting.

Since the early nineteenth century, scientists and health experts have continuously refined and renegotiated the meaning of nutrition, a construct which became ever more important but also ever more amorphous.[4] For many nutrition experts, this expansiveness simply made the term an empty vessel into which anything could be poured. "The word nutrition covers a multitude of sins, gross exaggerations, and misconceptions," wrote the American physician George Dow Scott in 1942. "Its interpretation is quite at odds among varying groups of peoples, and misconceptions, ignorance, the pseudo sciences, tribal, racial, and religious conceptions, all enter into its meaning."[5] Yet others argued for a necessarily broad perspective, as a definition restricted to biochemical or physiological aspects omitted key ways in which nutrition represented a complex set of interactions between an organism and its environment. In this view, as the British nutritionist Christine Rossington put it in 1981, nutrition was best defined as "the outcome of interplay between, and integration of, two dynamic ecological systems, the human internal bio-physical environment, and the external physical, economic and socio-cultural settings in which man lives."[6]

The conceptual plasticity of nutrition was by no means unique among scientific concepts, but it was remarkably broad and enduring. It seemed to many that there was no science unutilized in the exploration of nutritional function, no state of health or disease in which nutrition did not play a contributive or ameliorative role, and no grave social or political matter in which the nutrition of the population was not implicated. "The science of nutrition . . . utilizes the combined knowledge of all fundamental and applied sciences," wrote the nutritionists Kirsten Toverud, Genevieve Stearns, and Icie Macy in a report prepared for the U.S. National Research Council in 1950. "Even sciences such as theology, philosophy, and psychology

are intimately involved in nutrition, owing to their involvement in psychosomatic relationships in the body. . . . Nutrition has been approached from many directions–the bioenergetic, the anatomical, the statistical, the social, and the mental points of view, in addition to those of the physician, biologist, and chemist."[7] Indeed, this fluidity only made nutrition a more powerful concept, as it could be readily adapted to a wide range of contexts, problems, and agendas.

Yet this very fluidity vexed many nutritionists, who regarded it as a lack of intellectual rigor with real-world consequences. The meaning of medico-scientific concepts like nutrition was continually debated and refined in part because definitions matter beyond the realm of theory or semantics. Policy, research, product development, and regulation–and allocations of money and resources in each of those areas–are influenced significantly by fundamental understandings of core concepts and how they are organized. There is a rich literature on the ways in which definition and classification shape, or even engender, the most fundamental features of social action and interaction, and on how such discursive practices can be analyzed and modeled to understand the underlying culture that produced them.[8] In this paper, I argue that conceptual models of a discourse can be abstracted from textual or other evidence as networks of relations among constructs, and that these models can help identify larger patterns in the evolution of such discourses over time.[9] Nutrition, a heavily contested concept imbued with a wide range of meanings across numerous domains, provides a particularly useful case for exploring the affordances of this approach.

This aim arises from two related challenges that historians increasingly face. First, the volume of historical data is large and continuing to grow, and the sheer quantity of available sources–what William Turkel terms the "infinite archive" of digital materials–cannot be processed using traditional methods alone.[10] Second, traditional methods of historical research are typically based on deep and often solitary human engagement with the

relevant materials, an optimal approach for microhistorical analysis. But historians who want or need to engage with macrohistorical questions require a different methodological toolkit, and, in many cases, an entirely different perspective on the research process. In other words, there are important historical questions that cannot be answered solely through close readings of texts.[11]

Of course, good macrohistorical work typically requires considerable microhistorical sophistication. It is facile to assume that more or more accurate data will automatically lead to better understanding, or that broad patterns can be understood without close attention to the underlying source material. The view that computers can take massive amounts of information and do most of our analytic thinking for us, a belief embraced by many data miners and glorified by tech evangelists, more often than not yields statistically significant but conceptually meaningless results. We can and should outsource *some* of our thinking to smart machines, much as we have outsourced some of our memory to books and other media for thousands of years. But to do this well is to understand the limitations and leverage the affordances of different approaches to processing and analyzing information, both human and machine. The practice of historical research stands to benefit considerably from, and may even require, a mixed-methods approach that combines the qualitative and the quantitative and incorporates the analytic strengths of human interpretation and computational processing.

In what follows, I attempt to model the concept of "nutrition" in English-language sources from the nineteenth and twentieth centuries using *epistemic network analysis* (ENA), a set of techniques for measuring, visualizing, and comparing patterns of association among conceptual elements.[12] In doing so, I argue that conceptual networks can help us understand macrohistorical patterns in discourses—in this case, discourses of nutrition—without sacrificing microhistorical rigor. Specifically, I will describe an

approach in which microhistorical analyses inform the development of macrohistorical models that in turn suggest new avenues for microhistorical investigation.

Conceptual Networks

Definition, and the taxonomic practices that attend efforts to delineate knowledge, is the subject of considerable research in the history and philosophy of medicine and biomedical science.[13] Critically, definitions of concepts are rarely simple, stable, or uncontested. How something is defined—and who has the power to define it—often has significant and far-reaching consequences. For example, what counts as a "true" food allergy, or where the line is drawn that distinguishes the obese from the merely overweight, affects everything from patient care and research funding allocations to politics and everyday social interactions. Yet it can be challenging to characterize how complex concepts are defined, especially when the goal is to understand how those definitions change across contexts or over long periods of time.

Conceptual complexity stems in part from the relationship between concepts and the language used to denote them. The French chemist Antoine Lavoisier argued that science consists of three things: the series of facts that constitute the science, the ideas that represent those facts, and the words that express those ideas. The word, he argued, should awaken the idea, and the idea portray the fact, like three impressions of the same seal. It is thus impossible, according to Lavoisier, to separate language from science.[14] In other words, *concepts* (facts) are ultimately represented by *tokens* (words and other symbols). But where tokens are generally static, varying relatively little over time, concepts are both abstract and dynamic; what grounds them in some context is a complex set of interactions among other concepts, and that set of interactions—that *conceptual network* (idea)—is what links a token and a concept. Put another way, as the anthropologist Terrence

Deacon explained, "the pairing between a symbol (like a word) and some object or event is . . . some complex function of the relationship that the symbol has to other symbols."[15]

Importantly, concepts are not immutable, like Platonic forms, but evolve along with the ways of thinking in which they are embedded. Medico-scientific concepts are part of the grammar of some *community of practice*, what Ludwik Fleck termed a "thought collective" (*Denkkollectiv*): "a community of persons mutually exchanging ideas or maintaining intellectual interaction."[16] Through these interactions, a thought collective develops a particular "thought style" (*Denkstil*), a system and set of rules for knowledge production and organization in that culture—that is, a *discourse*. The result, Fleck argued, is that concepts have no abstract meaning; they have meaning only insofar as they are embedded in some thought style, which is, in turn, associated with some thought collective. "The statement, 'Schaudinn discerned *Spirochaeta pallida* as the causative agent of syphilis,' is equivocal as it stands," Fleck reasoned, "because 'syphilis as such' does not exist. There was only the then-current concept on the basis of which Schaudinn's contribution occurred, an event that only developed this concept further. Torn from this context, 'syphilis' has no specific meaning."[17]

Concepts cannot be abstracted from their context in part because they are deeply interconnected with other concepts within the discourse of some community of practice. Disease, for example, is not simply a pathophysiological process; as Charles Rosenberg has argued, it is "a biological event, a generation-specific repertoire of verbal constructs reflecting medicine's intellectual and institutional history, an aspect of and potential legitimation for public policy, a potentially defining element of social role, a sanction for cultural norms, and a structuring element in doctor/patient interactions."[18] To understand disease as a concept is thus to understand the interrelations among all these dimensions—in other words, to see it as a complex network of associations among biological,

interpersonal, social, cultural, political, institutional, and historical factors, all of which are grounded in particular discourses and communities and in particular times and places.

Yet in arguing that concepts cannot be abstracted from their context, I am not suggesting that concepts cannot be abstracted at all. In his work on abolitionist arguments in nineteenth-century newspapers, for instance, Timothy Shortell argues that "the sociocognitive structure of a discourse" can be modeled "as a networked field of concepts from which arguments are fashioned."[19] That is, conceptual networks, appropriately contextualized, can provide a means not only for characterizing the structure of a discourse but also for making comparisons across discourses and over time. In what follows, I explore ways to understand changes in nutrition as a concept over the nineteenth and twentieth centuries.

Nutrition as Word, Idea, and Fact

There are a number of powerful tools available for analyzing language usage, such as changes in word frequencies over time. Google's *Ngram Viewer*, for example, can plot the relative frequency of some *ngram*, a particular string of continuous characters such as a word or phrase, over time.[20] Figure 8.2 shows the *Ngram* graph for the word "nutrition," broken out by case, from 1800 to 2000 in the English language corpus (i.e., English-language books digitized by Google Books). The graph represents, for each year, the relative proportion of all one-grams that were "nutrition" or "Nutrition." As figure 8.1 shows, use of the term was relatively rare until about 1840. Between 1840 and 1870, usage more than doubled. While the fluctuation in relative usage was greater over the twentieth century, the overall trend remained one of increasing frequency. Interestingly, "Nutrition" (with a capital N) was very uncommon until the twentieth century. Starting around 1930, its relative frequency has almost the same pattern as that for "nutrition" (with a lower-case *n*). Because the most likely reason for capitalization in English

is if a term appears as the first word in a sentence—which, when that word is a noun, generally indicates that it is the subject of the sentence—this suggests that "nutrition" became commonly used as an abstract noun only after the turn of the twentieth century.
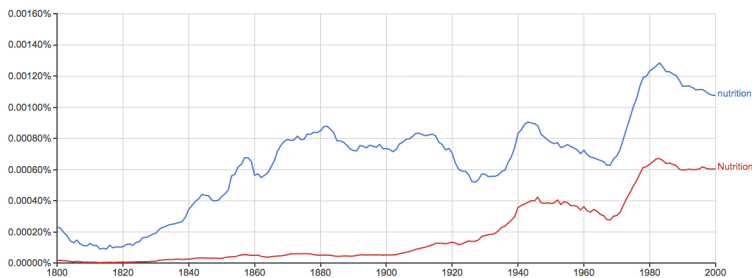

Figure 8.1: Google Ngram graph showing the frequency of the terms "nutrition" and "Nutrition" in the Google Books English language corpus from 1800–2000[21]

Analysis of usage in academic journals shows a similar pattern. The graph in figure 8.2 plots the number of articles in the JSTOR database containing the word "nutrition" or "Nutrition" from 1800 to 2000. As in the Google Books data, use of the term is rare until 1840. While the JSTOR data show what appears to be a steeper increase during the twentieth century, note that figure 8.2 depicts raw data, which haven't been normalized (e.g., to account for overall increases in the number of academic articles published). Nonetheless, it is clear that usage of the term "nutrition" in academic work increased significantly after about 1930.

While these analyses are helpful for understanding changes in word usage and identifying key points in time for more focused investigation, they do not give any indication of what people *meant* when they used the term "nutrition." That is, they are lexical rather than semantic analyses. In the case of nutrition, as noted above, the gap between the two types of analysis is particularly broad, as the term was used in remarkably diverse and, at times, mutually inconsistent ways.
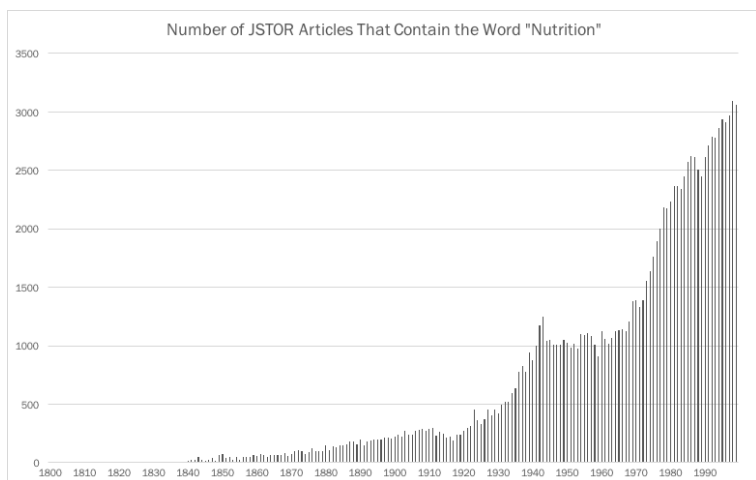
Figure 8.2: *Total number of articles in the JSTOR database published between 1800 and 1999 that contain the word "nutrition" or "Nutrition" (data obtained in January 2018)*

Many scientists and physicians in the nineteenth century described nutrition in almost poetic terms. The eminent physiologist Claude Bernard defined nutrition as "organic creation": "La nutrition et le développement ne sont rien autre chose . . . qu'une *création* organique."[22] Referencing Aristotle's designation of the nutritive soul (θρεπτική ψυχή) as the foundation of all life, such definitions located nutrition among the most basic processes that distinguish living organisms from inert matter.[23] Nutrition was, according to various experts, "the cardinal function of organic life,"[24] or "the great function by which life is sustained—in fact, it is life itself."[25] Yet when it came to defining nutrition in more concrete terms, most nutrition experts in the early to mid nineteenth century regarded nutrition as a specific physiological process through which food is ingested, digested, absorbed, and assimilated into the body. "Nutrition may be considered the completion of assimilating functions," wrote one physiologist in the first decade of the nineteenth century. "The food, changed by a series of decompositions, animalized and rendered similar to the being

which it is designed to nourish, applies itself to those organs, the loss of which it is to supply, and this identification of nutritive matter to our organs constitutes nutrition."[26]

By the turn of the twentieth century, professional definitions of nutrition were starting to become more holistic, reflecting the expansion of nutrition beyond the domain of physiological chemistry. The evolution of the concept into an abstract noun was one marker of this change, as nutrition came to encompass not only the "assimilating functions" but also their end result: the state of health arising from nutritional processes. Nutrition was particularly embraced by pediatricians, both as part of the emerging practices associated with well-child care and as a powerful explanatory element of pathography.[27] "Pediatrics," the German physiologist Franz Knoop wrote in 1913, "has become largely a study of the chemical pathology of nutrition."[28] This broadened use of nutrition led to broader definitions. In the 1921 article "What Do We Mean by Nutrition?" American pediatrician Ira Wile wrote: "One recognizes that in the consideration of nutrition there are involved problems of activity and rest, digestion, mental attitudes, moral entanglements, as well as over-feeding, under-feeding, and unsuitable feeding, inadequate digestive organs or disorders that may affect digestion or assimilation but are dependent upon underlying pathological states such as tuberculosis or syphilis."[29] For pediatricians and public health workers, considering nutrition in the strictly biochemical sense was unhelpful. Whether assessing children's growth and development, diagnosing and treating illnesses, or developing community-based interventions, nutrition had to be considered in a broader socio-medical context. "While there may be normal nutrition without health," wrote the eminent pediatrician L. Emmett Holt, "there cannot be health without normal nutrition."[30]

Pediatricians and dietitians in particular, and health professionals more generally, thus took an ever broader view of nutrition in attempts to understand the role of nutrition in health and disease. Nutrition scientists, too, began to look beyond the organism to understand nutrition, increasingly seeing it in ecological rather than

strictly physiological terms. For example, when *Nutrition Today* published an essay in 1968 by the eminent diabetes researcher Harold Himsworth entitled, "What 'Nutrition' Really Means," it sparked a debate about what the study of nutrition encompassed. Himsworth defined nutrition simply as "the analysis of the effect of food on the living organism." For Himsworth, this wasn't merely an issue of definition, but of professional identity. "As long as nutrition holds firm to that as its *raison d'être*," he argued, "its continued identity is assured. . . . Let it once lose sight of this, however, and then it will lapse back into its component subjects."[31] In the subsequent issue, Ancel Keys wrote in support of this simple statement, but several other nutrition experts took issue with its restricted perspective. D. Mark Hegsted, for example, found it "much too narrow," arguing instead that "nutritionists must be concerned with the entire process" by which food is ingested and utilized. "This means," he argued, "concern about things such as agricultural policy and what foods are produced; processing which may enhance or detract from food's nutritional value and make it more or less acceptable to the consumer; the distribution process which determines food availability to the consumer; and cultural, educational, and financial factors which determine what is actually chosen and eaten."[32]

This expansion of nutrition as a concept in Europe and the United States was due not simply to changes in medical and public health practice, but rather reflects larger changes in state concern about food and health. By the early twentieth century, the once perennial challenge of sufficient production and efficient distribution of foods became increasingly solvable due to improvements in agriculture, surplus management, food processing and preservation, and distribution. With these improvements came a gradual lessening of concern about widespread hunger and a commensurate increase in concern about widespread malnourishment. Consequently, governments began to focus more and more on the complex questions of how best to ensure diets that were optimal not only in food quantity but also in nutritive quality. At the same time,

the tailoring of diets to maintain and restore health in individuals, a central element of medical practice from antiquity, gradually accommodated dietary theories based on universal human requirements for various chemical substances. As scientists increasingly specified human food needs in quantitative terms, nutrition, once a predominantly individual concern, became a population-level issue. Thus, both biomedical research on nutrition and individual self-management of diets became issues of political economy.[33]

Yet, as definitions shifted from the more narrowly physiological to the more expansively ecological, ontological uncertainty remained relatively high. "There is so much ignorance of the fundamental facts which lie behind the science of nutrition," wrote the Scottish physician and physiologist E. P. Cathcart in 1928, "if one can venture to call nutrition a science when so much yet remains obscure."[34] This sense that nutrition was less a body of defined knowledge than a black box with a wide range of functions remained common throughout the twentieth century. "Nutrition science," as the nutritionist Jean Mayer put it in 1986, "is not a discipline, it is an agenda."[35]

A key part of understanding professional discourses on nutrition, then, is understanding how nutritionists and other nutrition experts thought about nutrition as a core concept in their work. However, it is difficult to identify broader trends across long spans of time solely through close readings of texts. Even when it is possible to understand some of the broader macrohistorical trends from a careful microhistorical analysis, it can be helpful to test those theories using a different method, triangulating understanding across modes of knowing. In what follows, I describe a process for modeling the development of nutrition as a concept and present preliminary results that provide a macrohistorical perspective on professional nutrition discourse over two centuries.

Modeling Nutrition as a Conceptual Network

Data Collection

To build a dataset of nutrition definitions published in or translated into English in the medico-scientific professional literature between 1800 and 2000, I searched (a) full-text databases for journal articles, books, reports, and reference materials written on the topic of nutrition by scientists, physicians, and other health professionals, as well as (b) physical copies of books, reports, and reference materials on food and nutrition or on topics likely to contain discussions of nutrition, including physiology, dietetics, medicine, and public health.[36] Works on animal nutrition (or physiology, etc.) were included as long as "animal" was used as a category that incorporates humans; thus, works on veterinary nutrition were excluded. Different editions of the same book or reference work were included.

What counts as a "definition" is, of course, a matter of interpretation; while many writers were explicit in their definitional goals, it was necessary in other cases to determine whether a given discussion of nutrition represented an attempt at definition. To make this determination in ambiguous cases, context and professional judgment were used. Only definitions of nutrition without qualifications were included. Thus, definitions of "good nutrition," "cellular nutrition," and so on were excluded on the grounds that these concepts were explicitly defined as some part or subset of nutrition more generally.

The dataset used in the present analysis contains 226 definitions of nutrition. Figure 8.3 shows the number of definitions from each decade.
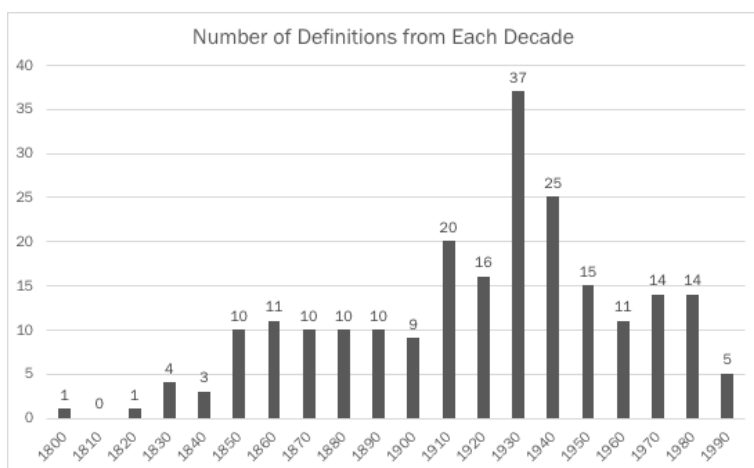
Figure 8.3: Histogram showing the number of definitions from each decade included in the dataset

Importantly, the data collection for this project is an ongoing process, and so this sample is perhaps more haphazard than many historical datasets. In particular, materials that have been digitized and are full-text searchable are over-represented in the dataset, as are physical materials that are easily accessed. The 1930s are also somewhat over-represented as well, though that may be due to an actual uptick in publishing on nutrition, as discussed above; beginning in the 1920s, the discovery of vitamins and other micronutrients and the subsequent construction of the "newer knowledge of nutrition" marked a significant expansion in and alteration of nutrition discourse.[37] All that being said, the dataset is sufficiently representative to warrant analysis, though results should be considered suggestive rather than definitive due to the possibility of significant sampling bias.

Coding

There are many ways to create network models of qualitative data. Perhaps the simplest (conceptually) is to construct a lexical network

of connections among the key words and phrases in the dataset.[38] In this case, for example, one could create a network where each node is a unique word or phrase, and the connections among the nodes are defined by whether or not any two words or phrases appear in the same definition of nutrition. These unique connections could then be summed over some period of time to produce a weighted lexical network model of the definition of nutrition in that period, where the thickness of each line would correspond to the frequency with which the two connected words co-occurred.

Figure 8.4, which shows a simplified example of this kind of network, represents connections from nutrition to other key words and phrases in four definitions published during the 1830s.[39] Thicker lines indicate connections that occurred in more than one definition, with the thickness proportional to the number of definitions in which the two terms co-occurred.
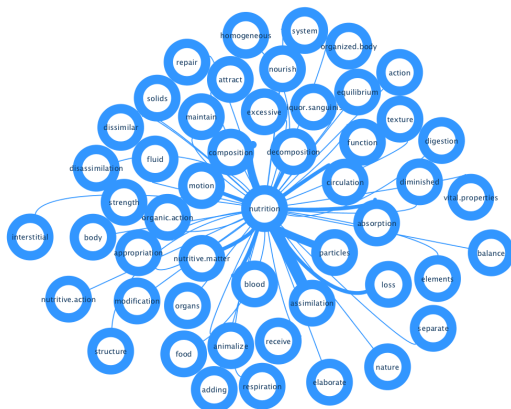

Figure 8.4: Network diagram showing connections between "nutrition" and other key words or phrases in four definitions of nutrition published during the 1830s

On one hand, this network provides some useful information about how nutrition was defined in the 1830s. We can see that assimilation was a key concept, and the only one to appear in all four definitions. Other key concepts include composition and decomposition, absorption, circulation, and particles, but there are a large number

of technical terms that occurred in only one of the four definitions. As a whole, the network indicates that the definitions privilege the physiological, and many of the terms denote actions or processes.

On the other hand, this approach has a number of limitations. If the number of definitions being modeled were larger, the visualization would quickly become nearly impossible to interpret; this would be true even in this very small model if connections among all the terms were included, which may be needed. For example, one might want to know not only the extent to which "nutrition" and "assimilation" are connected, but also the extent to which "assimilation" is connected to other key words or phrases in definitions of nutrition. While there are many sophisticated statistical techniques that could be used to obtain this kind of information from networks too complex to visualize, the network model would quickly become challenging to interpret. This is compounded further if we want to compare the networks of nutrition definitions from different contexts or different points in time. But perhaps most importantly, this network was constructed simply based on the presence or absence of words–that is, it is not based on any interpretation of the definitions. Thus the only way to make meaning is by interpreting the network model itself, but the words in the model have all been abstracted from their context, making that difficult. For example, what are "particles" in this case? Does the term mean the same thing in each of the three definitions in which it occurred? And so on.

One way to overcome these challenges is to construct a network model not with the raw data but with coded data. Within the discourse of some culture, *codes* are symbols or concepts that have meaningful interpretations.[40] Thus, a researcher familiar with a given context can interpret the discourse in terms of codes. For example, Glesne describes coding as "a progressive process of sorting and defining and defining and sorting those scraps of collected data (i.e., observation notes, interview transcripts, memos, documents, and notes from relevant literature) that are applicable to our research purpose. By putting like-minded pieces together

into data clumps, we create an organizational framework."[41] In other words, while coding is a deliberate process of simplification, it is one based on interpretation, providing a method for condensing the messiness of the raw data into a discrete set of key elements that can be quantified to identify larger patterns, patterns which may not be apparent based only on close reading of the materials. In building a network model of the coded rather than the raw text data, the model is based on an interpretation of the texts, not simply on some explicit attribute of them, and thus the larger patterns identified are more likely to be meaningful.

To construct network models using this approach, each definition in the dataset was coded for 14 elements commonly related to concepts of nutrition.[42] The codes, which are summarized in table 8.1, fall into three main categories: (1) *physiological* elements are the internal mechanisms by which foods are processed and used in the body; (2) *adaptive* elements are individual actions or conditions that are related to nutritional processes or outcomes; and (3) *ecological* elements are systemic or structural elements that are related to nutritional processes or outcomes. Thus for each definition in the dataset, there is corresponding information that indicates whether each code is present or absent; that is, each definition is interpreted and categorized according to these concepts.

This raises, however, a key challenge for understanding conceptual change over time, and in particular over long periods of time. As concepts change—that is, as the structure of associations that characterizes a concept in some context changes—so do all of the related concepts in that culture. For example, part of understanding the discourse on nutrition may involve understanding the concept "food" and how it is related to "health." Yet while the concept of "food" in one context was something like aliment or nutritive matter which can be ingested and assimilated into an organism, "food" in another context was also a substance composed of one of more chemical constituents: fats, carbohydrates, proteins, vitamins, minerals, and water. To address

this issue, all codes included in the analysis were applicable across the full time period. The tradeoff in taking this approach, of course, is that each code represents a relatively broad concept.

Table 8.1: Coding scheme used in epistemic network analyses

	Code	Definition	Example
Physiological	Assimilation	The process of making food or nutrients part of the self	"that function by which the nutritive matter already elaborated by the various organic actions, loses its own nature and assumes that of the different living tissues"
	Excretion	The elimination of waste products that arise from the bodily processing of ingested food	"the relative balance and co-ordination of the functions of digestion, absorption, and assimilation of food as well as the excretion or waste products"
	Maintenance	The process of sustaining bodily processes, including generating heat; the process of repairing damage, waste, or loss	"to rebuild body substance and to create heat"
	Energetics	The provision of energy for physiological processes or work	"process by which food is...utilized for body energy"
	Growth	Growth or development of cells, tissues, or the whole organism	"the conversion of the nutrient matter into living matter, ...which may increase that which has been already produced (growth of formed material)"

Table 8.1 (*continued*)

Adaptive	Food & Diet	Aliment, or any of its constitutive elements (e.g. nutrients); diet or consumption habits or patterns at the individual or population level	"food has been defined as a well-tasting mixture of materials, which, when taken in proper quantity into the stomach, is capable of maintaining the body in any desired state"
	Behavior	Mental, emotional, or behavioral processes or states	"the term 'nutrition' should be retained for a wide conception of the state of well-being which characterizes the individual who is both physically and psychically sound"
	Activity	Physical activity, exercise, or work, or consideration of strength, stamina, or vigor	"external work of the body"
	Sleep	Sleep, rest, or fatigue	"body and mental rest"
	Health & Disease	State of health or illness, or reference to specific aspects of health, hygiene, illness, or disease	"bringing about better health and...prolonging life"
Ecological	Environment	One's physical context or surrounding, whether natural or built	"nutritional needs of body tissues vary with such things as climate"
	Economics	Economic aspects of nutrition, financial factors, or socio-economic status	"financial factors which determine what is actually chosen and eaten"
	Education	One's understanding of nutrition or educational processes for teaching or learning about nutrition	"proper education, technical expertise, and the use of resources in applied nutrition and food technology"
	Food System	The production, processing, and distribution of food	"food production and food supplies, including processing, preservation and preparation"

Epistemic Network Analysis

There are a number of publications that describe in detail the method with which ENA constructs network models,[43] but in brief,

ENA creates for each unit a table (adjacency matrix) that quantifies the co-occurrence of coded elements for all lines in the dataset associated with that unit. In this case, each unit is a unique source (i.e., a book, article, reference work, or report); though most sources contain only one definition of nutrition, some contain multiple definitions, and each unique definition was entered on its own line in the dataset. In cases where definitions extend to multiple paragraphs, each paragraph is entered on its own line. This was done so that co-occurrences that were present in multiple definitions from the same source or in multiple paragraphs within the same definition would be modeled as stronger connections.

The resulting co-occurrence matrices were normalized (to model relative rather than absolute differences in connection strength) and embedded in a high-dimensional space, where each dimension represents a unique co-occurrence of codes. To create an ENA model, a dimensional reduction is performed (in this case, a singular value decomposition, or SVD), and the nodes of the network model—the coded elements—are placed in a metric space formed by the reduced dimensions using an optimization algorithm, such that the centroid of each network corresponds to the location of the network in the dimensional reduction. The result is two coordinated representations: (1) the location of each network in a *projected metric space*, in which all units included in the model are located, and (2) a *weighted network graph* for each network, which explains why the network is positioned where it is. An ENA model thus enables comparison of networks both visually and statistically, and every connection in the model is linked to the coded data that the connection represents, facilitating qualitative validation of the quantitative model.

Results

To examine how the discourse of nutrition changed over the nineteenth and twentieth centuries, I constructed an ENA network

model containing a network for each unique source in the dataset, and computed mean networks for four time periods. The divisions between periods reflect points in time when changes in nutrition discourse appeared to be relatively stark based on quantitative (Google *nGram* and JSTOR) and qualitative analysis of the nutrition literature. Figure 8.5 shows the mean ENA network for each of the
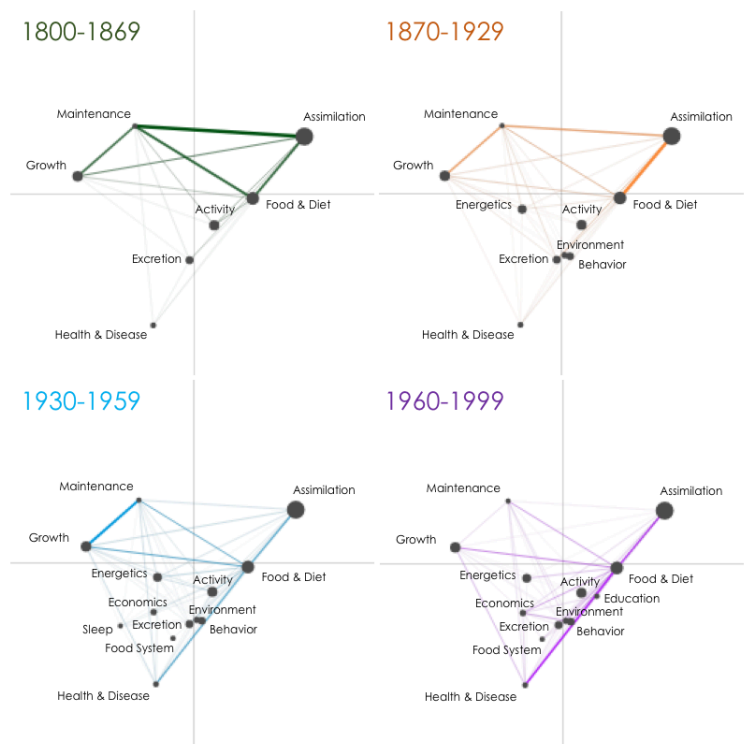

Figure 8.5: Mean ENA networks of nutrition definitions from four time periods

four time periods. Thicker, more saturated edges indicate stronger connections. The mean networks show a general evolution in the definition of nutrition from a largely physiological concept (1800–1869) to one that includes both physiological and adaptive elements (1870–1929), and ultimately one that is more holistic,

balancing physiological, adaptive, and ecological elements (1930–1999). Note, too, that issues of health and disease continued to become more important over time, particularly as they relate to food and diet.

Figure 8.6 shows the mean network locations of each time period, along with the 95% confidence intervals (the individual network locations are omitted for legibility). The location of a network or
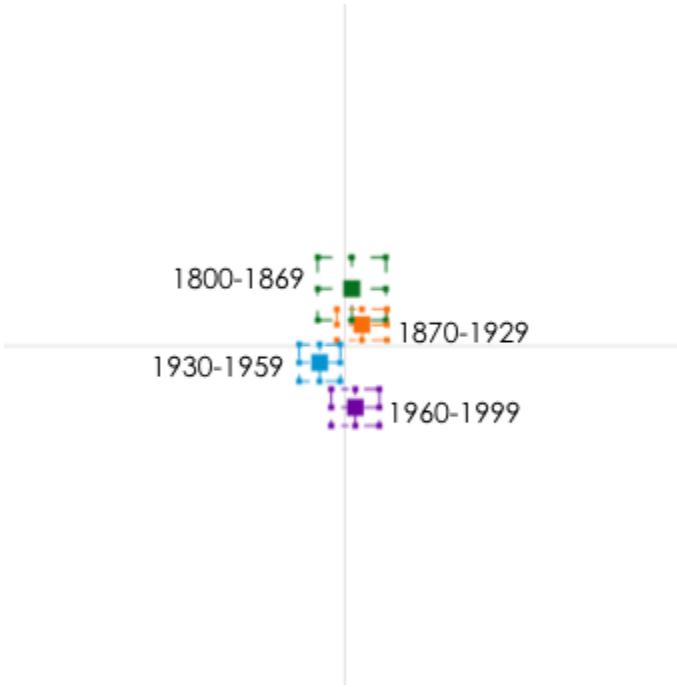

Figure 8.6: Mean ENA network locations of nutrition definitions from four time periods, with the corresponding 95% confidence intervals

mean network in ENA space indicates which connections were strongest in the network. Thus, a network that appears in the upper part of the space (i.e., a network with a high y-value) has stronger connections among the physiological elements, while a network that appears in the lower part of the space (i.e., a network with a low y-value) has stronger connections among the adaptive or

ecological elements. Because the networks are all projected into a metric space, it is possible to compute descriptive statistics and conduct null hypothesis significance tests (see table 8.2). All means are statistically significantly different on the second (y) dimension ($p < 0.05$) with medium effect sizes ($r \approx 0.30$).[44]

Table 8.2: Statistical measures of the differences between mean networks on the second (y) dimension. All differences are statistically significant (p < 0.05) with medium effect sizes (r ≈ 0.30)

	Mann–Whitney U	p	r
1800–1869 vs. 1870–1929	816	0.03*	0.27
1870–1929 vs. 1930–1959	1846	< 0.01*	0.32
1930–1959 vs. 1960–1999	779	0.01*	0.32

Once an ENA model has been constructed, it can be used to explore other phenomena of interest. In this case, for example, networks can be constructed by type of source across the whole time period. As figure 8.7 shows, each type of source tends to favor a different kind of definition. Unsurprisingly, reference works, which tend to have the shortest definitions of nutrition, focus primarily on the physiological elements. But monographs also differ from articles and book chapters, with the latter containing more holistic definitions. This may be because monographs, many of which are textbooks or works designed for broader audiences, are more likely to represent consensus within a field. In contrast, articles and book chapters are more likely to present novel, preliminary, or contrary thinking on a topic, and, perhaps most importantly, they are more likely to be directed at other professionals in the same field rather than learners within those fields or adjacent professionals.
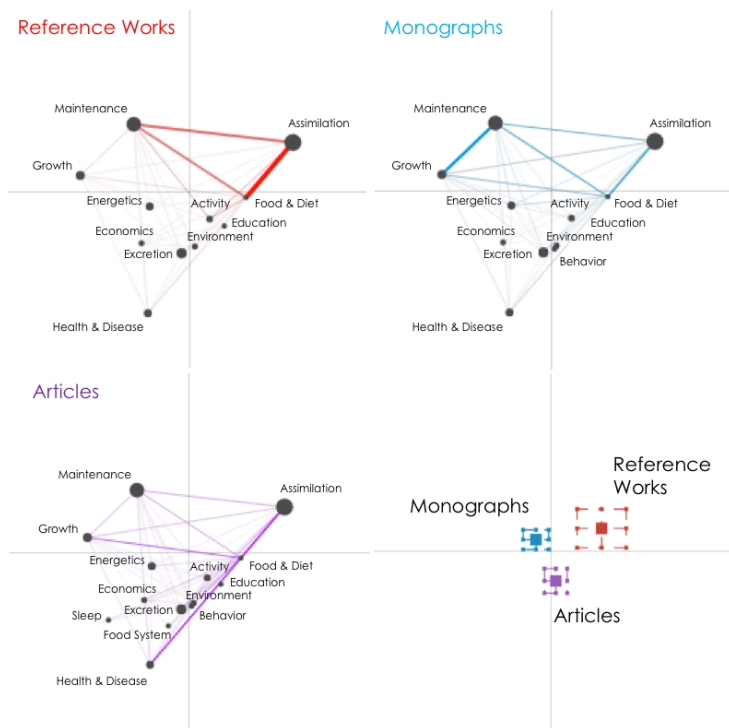
Figure 8.7: Mean ENA networks of nutrition definitions by type of source, and the mean ENA network locations with the corresponding 95% confidence intervals. All means are statistically significantly different (p < 0.05) with moderate-to-large effect sizes (r > 0.40).

In addition, the ENA model can be used to explore the impact of a particularly influential individual. In 1909, the American chemist Graham Lusk published the second edition of *The Elements of the Science of Nutrition*. In it, he defined nutrition as "the sum of the processes concerned in the growth, maintenance, and repair of the living body as a whole or of its constituent organs."[45] This was the most commonly cited definition of nutrition in the English-language literature. In the dataset analyzed here, 17 (11%) of the 155 definitions published between 1910 and 1999 referenced Graham's definition, even when proposing a broader one. Figure 8.8 shows

the ENA difference graph–which is produced by subtracting one mean network from another–for sources that cited Graham and those that did not. Connections shown in blue were stronger among the sources that cited Graham, while connections shown in red were stronger among the sources that did not cite Graham. As the difference graph indicates, the connection between *growth* and *maintenance* was far more common in definitions that cited Graham's

Figure 8.8: ENA *difference graph showing the differences between the mean networks of nutrition definitions that cited Graham Lusk (blue) and those that did not (red). The means are statistically significantly different (p < 0.01) with a large effect size (r = 0.86).*

definition, while most other connections, with the exception of the connection between *assimilation* and *food and diet*, were relatively

similar in both. The difference is statistically significant on the first (x) dimension with a large effect size: Mann-Whitney's $U = 3702$, $p < 0.01$, $r = 0.86$.

Thinking about the Past as a Dataset—A Reflection on Historical Research Methods

The goal of this exploratory study is not to provide a definitive analysis of the meaning of nutrition over 200 years. Neither is it to suggest that a mixed-methods approach to historical research is necessarily better than an exclusively qualitative approach, nor even to argue that all historical research would benefit from the incorporation of modeling or quantitative methods. Rather, because a mixed-methods approach provides additional tools with which to explore historical sources, it can be a very useful way to expand what historians can do to understand the past.

   In this case, the study suggests that ENA models can provide several advantages over qualitative analysis alone. As the initial results illustrate, the models can be used to provide quantitative support for a hypothesis developed qualitatively. I had always believed, based on years of studying the topic, that nutrition as a concept became more holistic and ecological over time, and that this was part of why so many nutritionists expressed varying levels of concern about the nebulous identity of the field. It also fit with the ever expanding list of professionals who considered nutrition a core area of focus; as more and more groups claimed nutrition as part of their purview, it is only natural that nutrition itself would expand to accommodate the wider range of interests. But given the timespan over which these developments took place, it was difficult to know whether these impressions resulted from my idiosyncratic engagement with the material, which was mostly through the literature on public health nutrition, and it was equally difficult to know whether this impression would actually stand up to a systematic approach to the question.

In addition to hypothesis *testing*, where ENA models can be used to confirm (or at least provide additional support for) theories generated by qualitative analysis, hypothesis *generation* is another affordance of mixed-methods approaches. Once an ENA model is created, for example, it can be used to quickly explore a range of relationships, generating new questions for further qualitative and quantitative analysis. In this case, the model can enable rapid exploration of differences in definitions across media, or examination of the effect on the community of a particularly influential member. Conducting these analyses qualitatively would be far more labor intensive. Thus, these exploratory uses of ENA (or other quantitative models) can be used to identify questions that are likely to be worth further examination. For example, the code *sleep* appears only in the network for 1930–1959. This raises an obvious question: why was sleep seen as an important component of nutrition in that period, but not in any of the others? A similar question could be asked of *education*, which appeared in definitions published only in 1960–1999.

Of course, it is important to understand not only the affordances but also the limitations of network analysis. One key limitation is that a network model cannot show you what isn't there. In the case of nutrition, for example, one code that is not part of the model is body weight. Although weight has become increasingly prominent in discussions of nutrition over the course of the twentieth century, and especially in the early twenty-first century, it appeared in only 5 of the 228 definitions analyzed. Discussion of race and gender were even more rare in nutrition definitions, but as anyone who has studied the history of nutrition can attest, both race and gender were frequently invoked concepts in nutrition discourse more broadly. The fact that these concepts do not frequently appear in *definitions* is provocative in and of itself, but further work is needed to understand how they function in nutrition discourse. Thus, while analyses such as the one presented here can provide considerable insight, they can also render invisible anything not included in the model.

That being said, models can be extremely useful for both exploring historical materials and for constructing arguments about the past. Historical research can certainly benefit from—and in a growing number of cases may even require—an approach that combines traditional analysis with computational models. ENA is, of course, only one example of an approach to modeling historical material, and there are certainly more aspects of network analysis worthy of serious discussion by historians. It is my hope that this paper, and the other papers in this volume, will stimulate further discussion about how we can incorporate new approaches and tools into our historical toolkits in order to better understand the past.

Acknowledgments

Endnotes

1. G. Budd, "Lectures on the Disorders Resulting from Defective Nutriment," *London Medical Gazette* (July 22, 1842): 632.

2. Henry C. Sherman, "Adequate Nutrition and Human Welfare," in *Proceedings of the National Nutrition Conference for Defense* (Washington, D.C.: U.S. Federal Security Agency, 1942), 31.

3. George T. Palmer, "The Measurement of Nutritional Status," *Child Health Bulletin* 6, no. 2 (1930): 47.

4. While interest in nutritional processes goes as far back as written records in most cultures, this project focuses on nutrition in the nineteenth and twentieth centuries, which marked a shift in nutrition discourse. Nutrition as a concept distinct from metabolism emerged only around the turn of the nineteenth century in Europe. Moreover, the Hippocratic-Galenic dietetic tradition and use of analogical reasoning to understand the relationship between diet and health had largely transitioned to an experimental and universalizing epistemology by the nineteenth century. See, for example, Frederic L. Holmes, "The Transformation of the Science of Nutrition," *Journal of the History of Biology* 8, no. 1 (1975): 135–144; Steven Shapin, "'You Are What You Eat': Historical Changes in Ideas about Food and Identity," *Historical Research* 87, no. 237 (2014): 377–392.

5. George Dow Scott, *Heredity, Food, and Environment in the Nutrition of Infants and Children* (Boston: Chapman and Grimes, 1942), 320.

6. Christine E. Rossington, "Environmental Aspects of Child Growth and Nutrition: A Case Study from Ibadan, Nigeria," *GeoJournal* 5, no. 4 (1981): 347; cf. Howard A. Schneider, "Toward a Philosophy for Nutrition," in *Human Nutrition Historic and Scientific*, ed. Iago Galdston (New York: International Universities Press, 1960), 225–232.

7. Kirsten Utheim Toverud, Genevieve Stearns, and Icie G. Macy, *Maternal Nutrition and Child Health: An Interpretative Review* (Washington, D.C.: National Research Council, 1950), 3.

8. Classic studies of definition, classification, and discourse in the history of science and medicine include Ludwik Fleck, *Entstehung und Untwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv* (Basel: Benno Schwabe and Co., 1935); Michel Foucault, *L'Archéologie du Savoir* (Paris: Gallimard, 1969); Ian Hacking, *The Taming of Chance* (Cambridge: Cambridge University Press, 1990). On discourse analysis, see Norman Fairclough, *Discourse and Social Change* (Wiley, 1993); James Paul Gee, *An Introduction to Discourse Analysis: Theory and Method*, 4th ed. (London: Routledge, 2014).

9. In this paper, I am elaborating ideas first outlined in A. R. Ruis and David Williamson Shaffer, "Annals and Analytics: The Practice of History in the Age of Big Data," *Medical History* 61, no. 1 (2017): 336–39.

10. William J. Turkel, *Digital History Hacks (2005-08): Methodology for the Infinite Archive* (blog), http://digitalhistoryhacks.blogspot.com.

11. Shawn Graham, Ian Milligan, and Scott Weingart, *Exploring Big Historical Data: The Historian's Macroscope* (London: Imperial College Press, 2016).

12. David Williamson Shaffer, Wesley Collier, and A. R. Ruis, "A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data," *Journal of Learning Analytics* 3, no. 3 (2016): 9–45; David Williamson Shaffer and A. R. Ruis, "Epistemic Network Analysis: A Worked Example of Theory-Based Learning Analytics," in *Handbook of Learning Analytics*, ed. Charles Lang et al. (Society for Learning Analytics Research, 2017), 175–87; David Williamson Shaffer, *Quantitative Ethnography* (Madison, WI: Cathcart Press, 2017).

13. On the importance and challenges associated with definitions of health and disease, see, for example, Georges Canguilhem, *Essai Sur Quelques Problèmes Concernant Le Normal et Le Pathologique* (Clermont-Ferrand, 1943); Gretchen A. Condran and Jennifer Murphy, "Defining and Managing Infant Mortality: A Case Study of Philadelphia, 1870–1920," *Social Science History* 32, no. 4 (2008): 473–513; Jenny Doust, Mary Jean Walker, and Wendy A. Rogers, "Current Dilemmas in Defining the Boundaries of Disease," *Journal of Medicine and Philosophy* 42 (2017): 350–66; Jeremy A. Greene, *Prescribing by Numbers: Drugs and the Definition of Disease* (Baltimore: Johns Hopkins University Press, 2006); Maël Lemoine, "Defining Disease beyond Conceptual Analysis: An Analysis of Conceptual Analysis in the Philosophy of Medicine," *Theoretical Medicine and Bioethics* 34, no. 4 (2013): 309–25; Wendy A. Rogers and Mary Jean Walker, "The Line-Drawing Problem in Disease Definition," *Journal of Medicine and Philosophy* 42 (2017): 405–23; A. R. Ruis, "'Children with Half-Starved Bodies' and the Assessment of Malnutrition in the United States, 1890–1950," *Bulletin of the History of Medicine* 87, no. 3 (2013): 380–408; Matthew Smith, *Another Person's Poison: A History of Food Allergy* (New York: Columbia University Press, 2015).

14. "L'impossibilité d'isoler la Nomenclature de la science et la science de la Nomenclature, tient à ce que toute science physique est nécessairement formée de trois choses: la série des faits qui constituent la science; les idées qui les rappellent; les mots qui les expriment. Le mot doit faire naître l'idée; l'idée doit peindre le fait: ce sont trois empreintes d'un même cachet." Antoine Lavoisier, *Traité Élémentaire de Chimie*, 2nd ed. (Paris, 1793), I.vi.

15. Terrence W. Deacon, *The Symbolic Species: The Co-Evolution of Language and the Brain* (W. W. Norton, 1998), 83.

16. Ludwik Fleck, *Genesis and Development of a Scientific Fact*, eds. Thaddeus J. Trenn and Robert K. Merton, trans. Fred Bradley and Thaddeus J. Trenn, (University of Chicago Press, 1979), 39. On communities of practice more generally, see Etienne Wenger, *Communities of Practice: Learning, Meaning, and Identity* (Cambridge University Press, 1998).

17. Ibid.

18. Charles E. Rosenberg, "Disease in History: Frames and Framers," *Milbank Quarterly* 67, no. S1 (1989): 1.

19. Timothy Shortell, "The Rhetoric of Black Abolitionism: An Exploratory Analysis of Antislavery Newspapers in New York State," *Social Science History* 28, no. 1 (2004): 77.

20. Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331, no. 6014 (2011): 176–182.

21. Google Books Ngram Viewer, Retrieved from http://books.google.com/ngrams.

22. Claude Bernard, *Rapport sur la Progrès et la Marche de la Physiologie Générale en France* (Paris, 1867), 93.

23. Aristotle, *On the Soul, Parva Naturalia, On Breath*, trans. W. S. Hett (Harvard University Press, 1957), 434a22 ff.

24. E. Leigh, *Respiration Subservient to Nutrition: A Thesis Presented to the Medical Faculty of Harvard University, March, 1850* (Boston: Ticknor, Reed and Fields, 1853), 1.

25. Theo. L. Hatch, "Nutrition, with a Report of Some Cases of Mal-Nutrition," *Northwestern Lancet* 9 (1889): 158.

26. A. Richerand, *The Elements of Physiology*, trans. Robert Kerrison (Philadelphia: Hopkins and Earle, 1808), 194.

27. See Ruis, "'Children with Half-Starved Bodies.'"

28. Franz Knoop, "Some Modern Problems in Nutrition," *Johns Hopkins Hospital Bulletin* 24, no. 268 (1913): 175.

29. Ira S. Wile, "What Do We Mean by Nutrition?" *Hospital Social Service Quarterly* 4, no. 3 (1921): 111.

30. L. Emmett Holt, *Food, Health and Growth: A Discussion of the Nutrition of Children* (New York: The Macmillan Co., 1922), 4.

31. Harold Himsworth, "What 'Nutrition' Really Means," *Nutrition Today* 3, no. 3 (1968): 20.

32. "Nutrition Definition," *Nutrition Today* 4, no. 1 (1969): 26.

33. On nutrition science and policy during this period, see Kenneth J. Carpenter, "A Short History of Nutritional Science: Part 1 (1785–1885)," *Journal of Nutrition* 133 (2003): 638–45; "A Short History of Nutritional Science: Part 2 (1885–1912)." *Journal of Nutrition* 133 (2003): 975–84; "A Short History of Nutritional Science: Part 3 (1912–1944)," *Journal of Nutrition* 133 (2003): 3023–32; "A Short History of Nutritional Science: Part 4 (1945–1985)," *Journal of Nutrition* 133 (2003): 3331–42; Holmes, "The Transformation of the Science of Nutrition"; Molly S. Laas, "Nutrition as a Social Question: 1835–1905" (Ph.D. Thesis, University of Wisconsin–Madison, 2017); Elizabeth Neswald, David F. Smith, and Ulrike Thoms, eds., *Setting Nutritional Standards: Theory, Policies, Practices* (University of Rochester Press, 2017); Aleck Samuel Ostry, *Nutrition Policy in Canada, 1870–1939* (Vancouver: University of British Columbia Press, 2011). The quantitative turn engendered what some scholars have termed "nutritionism" or "hegemonic nutrition," the construction of dietetic self-management as a performance of one's moral rectitude or social fitness. See, for example, Charlotte Biltekoff, *Eating Right in America: The Cultural Politics of Food and Health* (Durham: Duke University Press, 2013); Jessica J. Mudry, *Measured Meals: Nutrition in America* (Albany: State University of New York Press, 2009); Gyorgy Scrinis, *Nutritionism: The Science and Politics of Dietary Advice* (New York: Columbia University Press, 2013).

34. E. P. Cathcart, *Nutrition and Dietetics: Our Food and the Uses We Make of It* (London: Ernest Benn, Ltd., 1928), 3.

35. Laas, "Nutrition as a Social Question: 1835–1905," 3.

36. Databases searched include: Google Books, the Hathi Trust, the Home Economics Archive (HEARTH), the Internet Archive, JSTOR, and the Medical Heritage Library. For online databases, I did not use a standardized set of search terms or phrases, as searches were tailored to the size and composition of the database. For physical texts, I used indices and tables of contents, when available, to identify sections of longer works where definitions would be most likely found.

37. Elmer Verner McCollum, *The Newer Knowledge of Nutrition* (New York: The Macmillan Co., 1918).

38. For example of this kind of network analysis, see Alix Rule, Jean-Philippe Cointet, and Peter S. Bearman, "Lexical Shifts, Substantive Changes, and Continuity in State of the Union Discourse, 1790–2014," *Proceedings of the National Academy of Sciences* 112, no. 35 (2015): 10837–44.

39. Note that this network contains less than half of the words contained in these definitions. Common words, such as articles, prepositions, and words with no technical meaning, were omitted. In addition, multiple forms of the same word (e.g., "circulate," "circulating," and "circulation") were combined.

40. Jenny Hyatt and Helen Simons, "Cultural Codes–Who Holds the Key? The Concept and Conduct of Evaluation in Central and Eastern Europe," *Evaluation* 5, no. 1 (1999): 23–41.

41. Corrine Glesne, *Becoming Qualitative Researchers: An Introduction* (New York: Longman, 1999), 133.

42. For more on coding qualitative data for quantitative analysis, see Michelene T. H. Chi, "Quantifying Qualitative Analyses of Verbal Data: A Practical Guide," *Journal of the Learning Sciences* 6, no. 3 (1997): 271–315; Glesne, *Becoming Qualitative Researchers*; Williamson Shaffer, *Quantitative Ethnography*.

43. See note 12 for a list of publications that describe ENA methodology.

44. There is no significant difference on the first dimension (x axis). The axes are produced by the SVD, which constructs dimensions that maximize the variance in co-occurrences across the dataset. In this case, the largest source of variance seems to be the difference between definitions that included connections to food and diet or assimilation (the nutritional "inputs") and those with stronger connections to growth and maintenance (the nutritional "outputs"); this difference, however, was not related to time period. The second dimension, which maximizes the variance in co-occurrences not captured by the first dimension, reflects changes that occurred in nutrition definitions over time, separating more explicitly physiological definitions from those that were more holistic.

45. Graham Lusk, *The Elements of the Science of Nutrition*, 2nd ed. (Philadelphia: W. B. Saunders, 1909), 54.

9. Networks of Statisticians and the Transformation of Medicine

CHRISTOPHER J. PHILLIPS

There is a statistical paradox at the heart of twentieth-century medicine.

In 1900 physicians largely ignored the tools of statistical analysis. Clinicians and laboratory researchers saw themselves as fundamentally opposed to the burgeoning field of academic statistics: they were interested in biomedical causation, statisticians were focused on numerical correlation; they were focused on exceptions and idiosyncrasies, statisticians were focused on norms and averages; they were determinists, statisticians were probabilists. There were essentially no statistical articles in medical journals, no statistical training required for the M.D., no well-known statistical interpretations of laboratory experiments. The American Medical Association lamented that questions about therapeutic efficacy were largely addressed by anecdotal accounts from influential physicians (and drug companies themselves).[1] The burgeoning field of public health (sometimes under the title of "sanitation" or "hygiene") drew on epidemiological measures of disease, and questions of inoculation and epidemic infection had long been resolved with statistical calculations.[2] But these were seen as limited to large outbreaks where people could be treated as interchangeable; in the clinic, the opposite was true. Patients were unique and the aggregative methods of epidemiology irrelevant.[3]

By 2000 the situation was seemingly reversed. A statistically significant randomized clinical trial was the gold standard of therapeutic efficacy, and such proof was required by the Food and Drug Administration (FDA) prior to licensing drugs.[4] Reformers now promoted "evidence-based" medicine (as if medicine had never before been based on evidence), an initiative which claimed best practices should be determined solely on the basis of statistically

rigorous experiments and meta-analyses of past clinical trials.[5] Pre-diabetes, pre-hypertension, and similar threshold-based diagnoses were now determined on the basis of large studies of correlation and risk factors.[6] The patient experience itself had also been transformed into what Robert Aronowitz termed "risky medicine": those at risk of disease and those suffering from chronic conditions looked increasingly alike.[7] A range of factors–exercise, diet, environmental exposure–were now linked to an increasing or decreasing probability of disease.[8]

   How could the role of statistical practice in clinical medicine have been altered so dramatically? Normally explanations of fundamental change in scientific practice–whether considered as paradigm shifts, revolutions, or otherwise–fall into a few categories.[9] There is the shifting role of schools of thought and training. This doesn't seem adequate here; the significance of statistics in physicians' training has not changed dramatically and there are no clearly defined "schools" on the proper role of statistics in medicine. Likewise, the practices within teaching hospitals have remained remarkably stable. Other explanations might rely on the role of charismatic leaders, but again there are no real figureheads, or at least well-known leaders, of any such statistical movement. Some explanations might emphasize powerful new measures that enabled new ways of thinking about the world. There is some of that here–statistical measures largely matured and flourished in the twentieth century–but there is no one measure that was essential or fundamentally transformative. Other explanations rely on high-stakes and visible moments when statistics might prove themselves useful to resolving disputes. Indeed, there is a contender: the use of odds ratios and similar concepts to link smoking to lung cancer in the 1964 Surgeon General's report on *Smoking and Health*. But there are no clear pre- and post- distinctions centered around 1964; the report itself does not attribute its findings primarily to new statistical measures; and opponents quickly condemned the report as inadequate.

In this chapter I want to suggest another way of explaining the seeming paradox of medical statistics: the increasing use of statistics in clinical medicine was largely invisible because it was accomplished by a *network* of unknown people deep within the federal bureaucracy. Specifically, I will highlight a group of biostatisticians at the National Institutes of Health (NIH) who from the late 1940s pioneered new uses of statistical concepts both by publishing research articles showing possible medical applications and by serving as consultants on projects seeking NIH financial support. Hired by Harold Dorn in 1947–1948 in the "methods" division of the Public Health Service (and soon incorporated into the NIH proper), these biostatisticians showed how formal statistical analysis provided powerful tools for determining efficacy, modeling dose-response curves, and evaluating therapies.[10] As the NIH became the dominant funder of medical research (and science generally) in this period, its model gradually became the dominant mode by which new discoveries in medicine were announced and new practices were established.

Parts of this story are easy to support. The NIH was certainly the dominant funder and gradually became the central organ for American biomedical research in the decades after 1950. Nearly all major medical research went through the institutes and their grant evaluators.[11] Moreover, NIH statisticians were deeply involved not just with the 1964 Surgeon General's report, but also with the long-running Framingham Heart Study, another crucial site for promoting statistics-based measures of what constitutes health and disease, as well as with the evaluation of drug efficacy and safety through the FDA.

Other aspects are more difficult to track. The statisticians were not well known outside the field of biostatistics, let alone in medicine. The first generation—Jerome Cornfield, Samuel Greenhouse, Max Halperin, Jacob Lieberman, Nathan Mantel, and Marvin Schneiderman—were self-trained (none initially had doctoral degrees in statistics) and mastered the relevant statistical tools on the job. Though initially based in a single office, after the mid 1950s

they spread out into a variety of new Biometrics Research Branches or Biometric Offices across the NIH.[12] They published prolifically (approximately 650 articles through the 1970s), but remained largely behind the scenes as co-authors, statistical consultants, and advisors, though by the late 1970s had come to assume positions of prominence (head of the American Statistical Association, chair of university departments, etc.).

It is not obvious how to establish an historical argument for the group's influence. No one person or project was responsible for the quantification of clinical medicine. The field and its practices were too diverse and diffuse. We might think of the NIH as causing change, or bureaucratic rule-makers at the FDA as shifting practices, but both claims beg the question of who or what was ultimately responsible, even if it is sensible to focus on the NIH's rules for grant applications or the FDA's regulations for drug approval. Likewise, I'm hesitant to point to the development of odds ratios, Bayesian inference techniques, and the spread of null-hypothesis tests as explanations. Tracking the "successful" concepts on the basis of what turned out to be important risks obscuring what made them attractive in the first place. To twenty-first century observers, it seems obvious that statisticians who developed new measures of efficacy and causality in medicine would be influential. It was not clear in 1946.

I instead want to suggest one way to understand this transformation is to take seriously the way this group functioned as nodes within a network based largely (but not exclusively) at the NIH, and how participants collectively managed to transform standards of practice and spread statistical tools as new ways of defining proof and causality in medicine. I suspect that it is through their research collaborations–often resulting in published papers–that we might look for their influence. Portraying themselves initially as advisers for the design and interpretation of medical experiments and observational studies, they soon showed the worth of their methods. I see them as establishing a network, with people as nodes connected by the projects and papers they

worked on together. Though my use of network tools in this chapter is ultimately more exploratory than conclusive, the reliance on network analysis has the felicitous side-effect that I will study their work using numerical analysis rather than anecdote—precisely the way statisticians thought medical interventions should be assessed.

Thinking of the biostatistics group as a network isn't a replacement for close reading of published materials or deep dives into archival holdings. Rather, thinking in networked terms allows us to take advantage of the ways that researchers and institutions were connected through their projects and papers. This was the era that Derek J. de Solla Price referred to as the dawning of "big science," and the biostatisticians at the NIH were integral to the rapid expansion of biomedical research, as well as the shift from individual researchers to large teams and collaborations.[13] Both the inclusion of new kinds of experts on projects and the use of ever
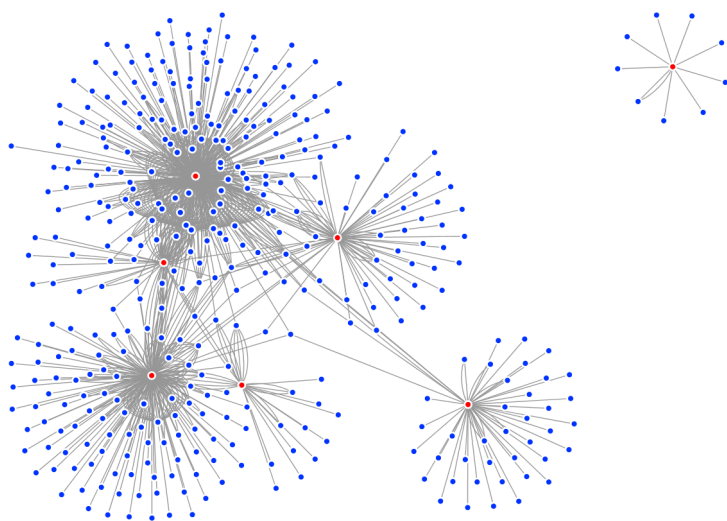

Figure 9.1: Overall publication network, 1930–1980

larger sample sizes in clinical studies in order to establish statistically significant effects often necessitated extensive

collaboration. Mid-century "big science" was not just about giant cyclotrons but also about multicenter studies of therapeutic interventions.

I initially created a network out of every published piece authored or co-authored by one of the first seven members of the NIH's statistical group. Limiting to publications from 1930–1980 (the key timeframe for the spread of statistical ideas), I found 653 unique articles, abstracts, letters, notes, and reviews. By treating these articles as "edges" and the authors and co-authors as "nodes" I created the network shown in figure 9.1.[14]

The red nodes in figure 9.1 are the seven members of the group, with blue nodes indicating co-authorship. (Clockwise from upper right-hand red node: Dorn, Lieberman, Halperin, Cornfield, Greenhouse, Mantel, and Schneiderman.) Each edge in this image represents a single co-authorship relation, so one article by a member of the statistical group with two co-authors would be represented by two different edges.

Some interpretations are immediately apparent. Dorn is entirely isolated, whereas Lieberman shares only a few edges with the main cluster. Indeed, Dorn was head of the group, but was trained as a sociologist and never published extensively in biostatistics (though he did have an ongoing role managing surveys of the prevalence of cancer across the country). Lieberman also had relatively few connections because he did not co-author any articles with other members of the initial group. Among the remaining five statisticians, Mantel and Cornfield have by far the most publications (over 250 and 150 unique publications, respectively) and the largest number of connected edges. Greenhouse, interestingly, is far more connected to Mantel and Cornfield as a co-author (and in the visualization appears directly between them), than to either Schneiderman or Halperin.

Different visualizations of the network can help refine different aspects of the group's influence. First, by dividing the data into two temporal groups (1945–1960 and 1961–1975), it is clear that there is
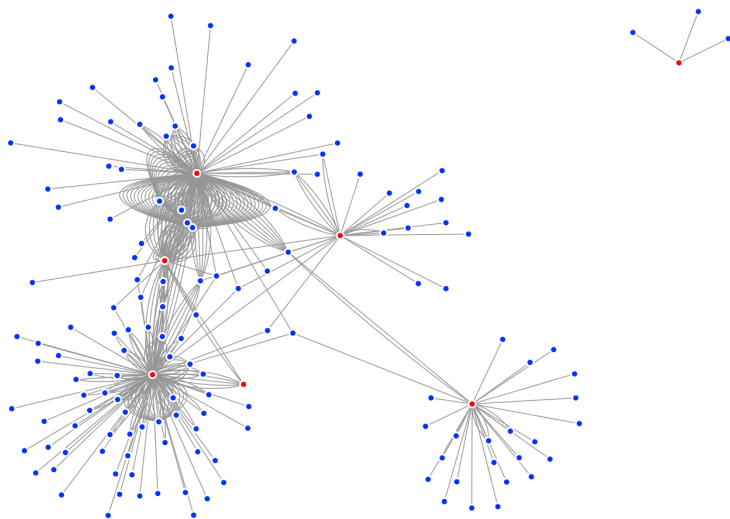
Figure 9.2: Publications, 1945–1960

little difference in publication practice (with the exception that Dorn's untimely death in 1963 removes him). Figures 9.2 and 9.3 portray the networks created respectively by this temporal division.

Rather than dividing by time, it is also possible to look at the entire timespan, labeling edges by the discipline of the publication's journal. This gives a quick estimate of the various fields in which the group was publishing.

The group was publishing widely, with the greatest number of publications in the fields of cancer research (edges colored light green, ~130 publications), medicine (blue, ~150), and statistics (orange for biometrics journals, ~75; pink for general statistics journals, ~125). There were also publications in general biology and chemistry (white, ~55), social science (purple, ~35), and epidemiology and public health (red, ~60).[15] Essentially every member of the group was publishing in both statistics and medical journals, serving as intellectual links between the disciplines. Each author had different disciplinary emphases, but it was not the case
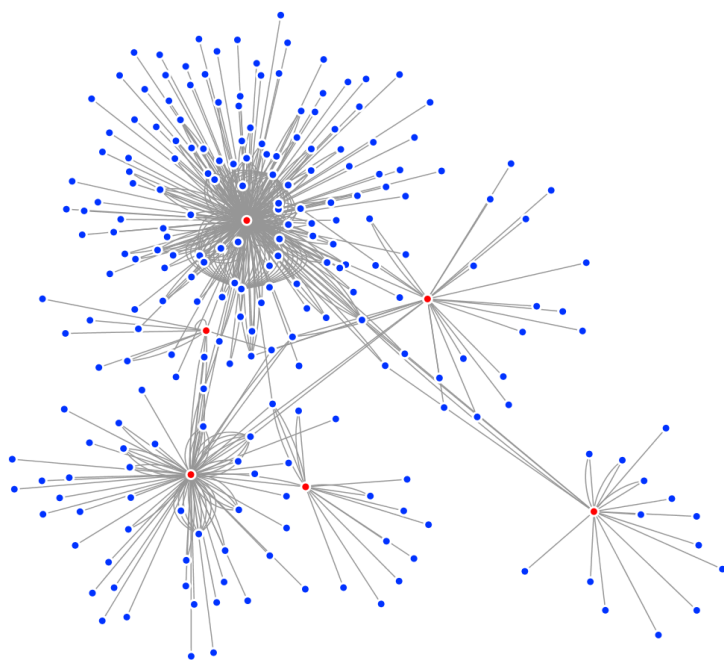
Figure 9.3: *Publications, 1961–1975*

that authors started publishing in statistics journals and then transitioned to medical journals. The entire group published widely across disciplines over time. Moreover, the relative lack of publication in epidemiological journals (the traditional locus of numerical analysis within medicine) suggests an explicit attempt to popularize statistical methods in medicine, and particularly in cancer research. Even as biostatistics and epidemiology were finding more established institutional homes in medical and public health schools in these years, early practitioners were establishing the field's prominence by publishing *elsewhere*.

Because this network was constructed by taking the publications of members of the group, it naturally places them at the center of the graph;
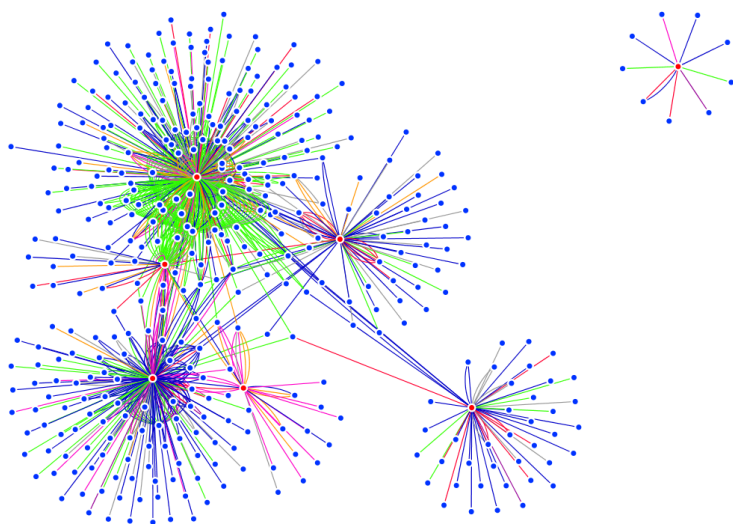
Figure 9.4: Publications (edges) labeled by discipline

a research collaboration that didn't involve one of them is simply missing. To get a wider sense of their influence, we need to situate their work within that of the biometrics and biomedical community. This is not easy, however, as the number of medical articles in this period quickly overwhelm most statistical software packages or network visualization tools. There are nearly six million articles in the PubMed collection between 1930 and 1980, and even when limited to topics involving cancer (using the Medical Subject Headings [MeSH] "neoplasm"), there are still a half-million articles. Given that many of these were co-authored, creating a network of co-publication would quickly make an unwieldy mess.

As a preliminary approach I took what I understood as one key case for the group's influence, namely epidemiological studies of cancer between 1950 and 1965. (A similar claim could be made for influence upon studies of heart disease with slightly later dates, but this search is at least consistent with the group's original location in the National Cancer Institute.) By limiting the articles to those

in English labeled with the MeSH terms "neoplasm" and "epidemiologic methods" between 1950 and 1965, I produced a network with 7585 nodes (authors) and 9116 edges (articles).[16]

Figure 9.5: Cancer and epidemiological methods articles, 1950-1965

There is one large and well-connected network of articles in the upper left hand of the image, and then decreasingly small networks until at the bottom we see many articles with two co-authors who never published with anyone else. If the group I'm looking at had

influence, surely they'd be in the main network in the upper left *and* would be, statistically speaking, important or central members of that network.
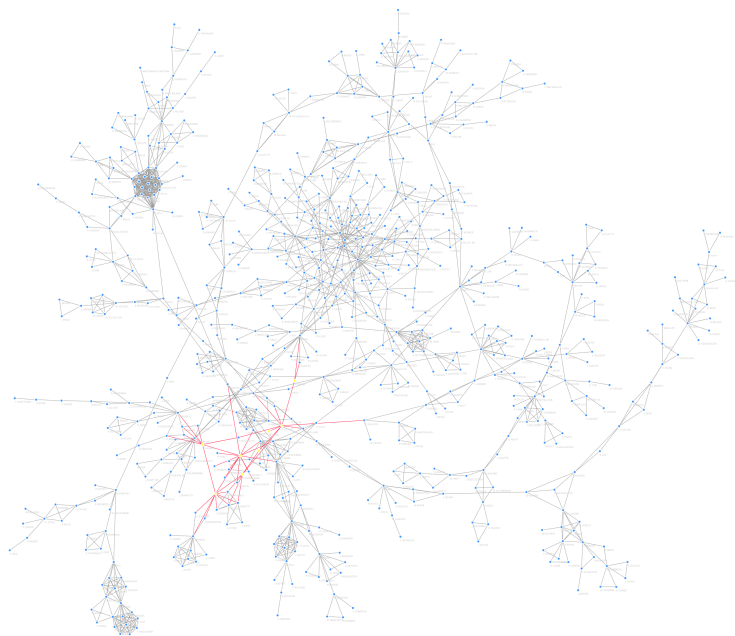

Figure 9.6: Sub-network of articles on cancer and epidemiological methods

Figure 9.6 shows the main sub-network (including 771 authors) from the upper-left corner of figure 9.5, with NIH statisticians listed as yellow nodes. Indeed, by taking statistics of only this sub-network, we can see how important the NIH group was to the publication of articles. If we take the "closeness centrality" or "shortest path length," then out of these nearly 800 authors, Greenhouse has the fourth highest value, Mantel the twelfth, Schneiderman the thirteenth, and Dorn the twenty-eighth. While the "closeness" metric looks at shortest paths within the whole network, "betweenness" looks also at subgroups within the network, and for this latter measure, Mantel's value ranks 21st, Greenhouse 26th, Dorn 272nd, and Schneiderman 265th. (As noted earlier, one problem

of this smaller network is the elimination of other members of the group despite their contribution to the topic of cancer epidemiology; nevertheless at least this gives a first approximation assuming that the other statisticians would have only increased the group's influence.) If we include two members who joined the statistical group slightly later, Sidney J. Cutler and Fred Ederer, the influence is even more impressive. Of the nearly 800 authors, Cutler had the highest score for "closeness" and the second highest for "betweenness" while Ederer had the third-highest score overall for both. Even with the obvious simplifications such an analysis entails, this is rather clear-cut evidence for the influence of the NIH group within the larger publication network concerning cancer and epidemiological methods.[17]
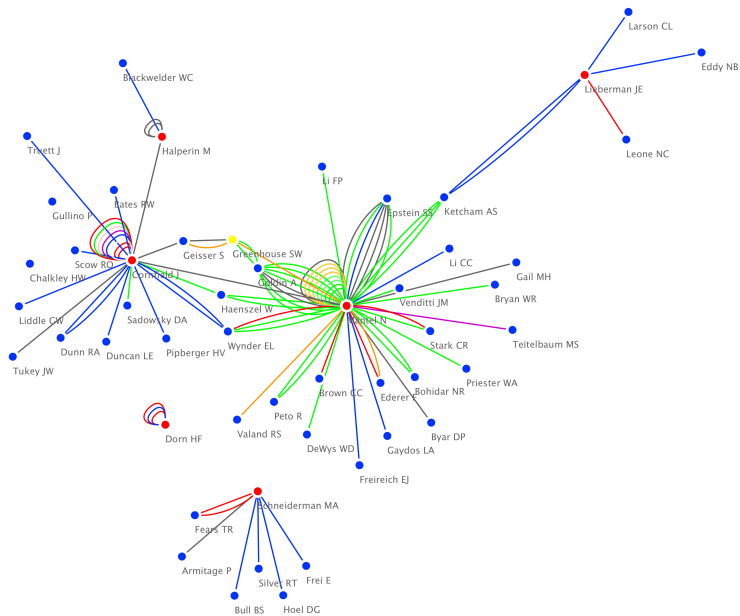

Figure 9.7: Edges and nodes that correspond to publications with over 50 citations

Another measure of influence would be to simply examine whether and how the initial group's publications were cited. Returning to

only those articles that had one of the original members as an author, we can also visualize only articles with substantial numbers of citations.

Figure 9.8: Edges and nodes that correspond to publications with over 100 citations

Some of these publications were certainly widely influential. There are about 100 articles with more than 50 citations, and about 50 of those articles have more than 100 citations. About 10 articles have more than 500 citations, according to the Web of Science citation index. On one level, this is to be expected; the articles are of interest precisely because they were influential. But it does also reveal the nature of their influence, and perhaps explain the network's relative invisibility. There was no one article or author

among this group that took the lead in establishing the field; rather, as the visualization suggests, their efforts were distributed. This is unlike, for example, a traditional laboratory model in the sciences in which publication authorship reflects institutional hierarchy. Furthermore, the majority of highly cited publications were in cancer and heart disease research, suggesting that it was the study of those fields in which the relevance of statistical analysis became most widely visible. The highly cited papers also range from the 1950s through the 1970s, suggesting that there was not one moment of influence, but rather a sustained program of interest to colleagues.

It is also possible, using Clarivate Analytics's Web of Science citation indexing service, to track all the articles which cited those initial publications. Cornfield's work, for example, has been cited in 4889 papers, with the peak of citation occurring around 1980. Cornfield's most cited article (over 750 times since its publication) is on the analysis of patients enrolled in the Framingham Heart Study, a paper which in turn became a central model of the methodological basis of the "risk factor" approach.[18] Similar analyses can be made for the other members of the group: This data, however simplified,

Author	Cited by[19]	Peak of citations	Topic of most cited paper
Cornfield, Jerry	4889 papers	Late 1970s	Multivariate risk analysis of observation study (762 citations, published 1967)
Dorn, Harold	573 papers	mid-1960s	Cancer mortality (227 citations, publ. 1959)
Greenhouse, Samuel	4348 papers	Continuing to grow	Methods for analyzing profile data, such as tests given to individuals (3065, publ. 1959)
Halperin, Max	2812 papers	Around 1980	Estimating risks of diseases (218, publ. 1971)
Lieberman, Jacob	645 papers	Late 1970s	Testing of synthetic analgesics (212, publ. 1950)
Mantel, Nathan	35,724 papers	In late 1980s, then again around 2014	Statistical analysis of data from retrospective studies of disease (11,584, publ. 1959)
Schneiderman, Marvin	1417 papers	Around 1980	Methods of counting platelets (431, publ. 1965)

does suggest some clear aspects of the influence of these original seven members of the group. Their most cited work was originally published between 1950 and 1971, with the peak of citations of the group around 1980. This would be consistent with a general timeline of work in the 1950s and 1960s establishing the basic research that would coalesce in the 1970s into the established role of statistical methods in clinical work. Also, though it is somewhat arbitrary to focus only on the most cited paper by each author (because in some cases that particular paper was not much more cited than others), it is indicative that their most cited work was in interpreting observational data, particularly data around cancer and heart disease. This was indeed how this group was seen. They were known to have invented new measures for making causal claims about complex diseases of unknown origin. Future research might explore whether tools that focus on the content of their papers—epistemic network analysis, for example—might reveal the ways they shifted the conversation on a more granular level.[20]

There are some obvious problems with the network approach. Citation analysis is susceptible to criticism given the possibility of unreliable metadata, as well as the presumption that citation is a direct measure of influence. In addition, it ignores connections and collaborations that did not result in co-authorship. Other influential biostatisticians (including Donald Mainland and A. B. Hill) were in dialogue with this group (we know this because there is correspondence in their papers, as well as many citations in their published papers), but they were not co-authors and so are absent in the network. Moreover, by "flattening" collaborations into nodes and edges, nuances are erased, not least of which is the fact that there are many reasons for including (or excluding) another scholar as a co-author. Co-authored articles may reflect genuine collaboration or may simply reflect a primary author giving credit to others who made minor contributions to the project. Such distinctions are ignored when all co-authors are treated symmetrically.

There is, however, good evidence that co-authorship *was* precisely how the statisticians thought about their work. They initially functioned as a single group in a large office on the NIH campus, and when a call for statistical advice came into the office, whoever answered the phone would take on the consultation.[21] Though at the time they were not very concerned about turning every project into a published article, the group quickly realized that the statistical tools and techniques deployed in their consultations could be published to allow others to know how to approach this kind of problem. In this sense the diagram captures an essential feature of these statisticians' practice–that they served as physical and intellectual links from the NIH out into other researchers' labs (and into other institutes of the NIH). The edges here are not just articles, but true connections between statisticians and the wider biomedical, scientific, and statistical worlds. By setting themselves as an "on call" service, the group's publications serve as a written legacy of the projects to which they contributed.

There are many ways to expand this preliminary work. Some of the most important early clinical trials were conducted abroad, particularly in Great Britain, and it might also be worth trying to analyze more precisely how nodes within this network might be connected in other ways to co-authorship networks based in other nations. Perhaps a particular member of the NIH group served as a conduit to statistical researchers abroad, or perhaps there were many connections across multiple people. It would also be useful to label not just publications by discipline but also nodes by institutional affiliation. This would require a great deal of time, because institutional affiliations shift over a half-century (and some research projects might span multiple affiliations, etc.), but this might also help reveal the pathway of influence out from this initial group. Alternatively, nodes might be institutions rather than authors, and alternative network constructions would certainly provide different views of the phenomena. Moreover, I might include statisticians who joined the NIH after these first seven, or see if new hires changed the direction of the publishing effort.

There is also much to be done to clean up the data. I have checked Cornfield's publication list against a bibliography compiled late in his life, for example, but have not tried to do this yet for any of the other primary nodes.

In the end this analysis is preliminary, both in the sense that the corpus of medical documents is too big a network to examine easily and in the sense that it is still not obvious how, precisely, to add network analysis to traditional archival work. Nevertheless, given the way in which statistical ideas spread at mid-century, changing the entire way medicine is conducted without a clear person or reason driving the transformation, publication networks are useful tools for thinking about how research practices change. We have long known about the key role scientific journals played in the dissemination of research, and that played by funding agencies like the NIH in medicine, but there is surprisingly little historical analysis of how, precisely, novel methods and techniques spread. This chapter, at a minimum, suggests ways that a small group of statisticians hidden away at the NIH could still have an outsized and visible presence in the literature, introducing novel methods for analysis which connect medicine, statistics, and the physical and social sciences.

Endnotes

1. Harry M. Marks, *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900–1990* (Cambridge: Cambridge University Press, 1997), 15-70.

2. William G. Rothstein, *Public Health and the Risk Factor: A History of an Uneven Medical Revolution* (Rochester: University of Rochester Press, 2003); Gérard Jorland, Annick Opinel, and George Weisz, eds., *Body Counts: Medical Quantification in History and Sociological Perspectives* (Montreal: McGill-Queen's University Press, 2005).

3. Of course, there were exceptional clinicians who used numbers—e.g., the so-called "Paris School" of the mid-nineteenth century focused on comparing treatments through the measurements of responses. But these were exceptions that proved the rule (and which emphasized the application of epidemiological tools precisely in large hospitals and charity wards wherein people could be treated as interchangeable). See J. Rosser Matthews, *Quantification and the Quest for Medical Certainty* (Princeton: Princeton University Press, 1995), 1–61.

4. Laura Bothwell, "The Emergence of the Randomized Controlled Trial: Origins to 1980" (PhD diss., Columbia University, 2014); Daniel Carpenter, *Reputation and Power; Organizational Image and Pharmaceutical Regulation at the FDA* (Princeton: Princeton University Press, 2010), 269–280.

5. Jeanne Daly, *Evidence-Based Medicine and the Search for a Science of Clinical Care* (Berkeley: University of California Press, 2005).

6. Jeremy A. Greene, *Prescribing by Number: Drugs and the Definition of Disease* (Baltimore: Johns Hopkins University Press, 2007); David Shumway Jones and Gerald M. Oppenheimer, "If the Framingham Heart Study Did Not Invent the Risk Factor, Who Did?" *Perspectives in Biology and Medicine* 6, no. 2 (Spring 2017): 131–150.

7. Robert Aronowitz, *Risky Medicine: Our Quest to Cure Fear and Uncertainty* (Chicago: University of Chicago Press, 2015).

8. The penetration of statistics into everyday clinical medicine has not meant that physicians are deeply trained in statistics, or that physicians and patients understand the statistical measures that underlie their interactions. See the work of Gerd Gigerenzer, including his *Calculated Risks: How to Know When Numbers Deceive You* (New York: Simon and Schuster, 2002).

9. This is not the place to review the vast philosophical and historical literature on how and why science changes, but some of the classic conceptual arguments about influence include that of research schools: Gerald L. Geison and Frederic L. Holmes, eds., "Research Schools: Historical Appraisal," *Osiris* 8 (1993): 1–248; pedagogy: David Kaiser, ed., *Pedagogy and the Practice of Science* (Cambridge: Massachusetts Institute of Technology Press, 2005); thought collectives: Ludwik Fleck, *Genesis and Development of a Scientific Fact*, trans. Fred Bradley and Thaddeus J. Trenn (Chicago: Chicago University Press, 1979 [1935]); and paradigm shifts and revolutions: Ian Hacking, ed., *Scientific Revolutions* (Oxford: Oxford University Press, 1981) and Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).

10. Though there is little published information on the group, there is a brief discussion of their origins in Jonas H. Ellenberg, Mitchell H. Gail, and Nancy L. Geller, "Conversations with NIH Statisticians: Interviews with the Pioneers of Biostatistics at the United States National Institutes of Health," *Statistical Science* 12, no. 2 (May 1997): 77–81, and of their influence in Sejal Patel, "The Benevolent Tyranny of Biostatistics: Public Administration and the Promotion of Biostatistics at the National Institutes of Health, 1946–1970," *Bulletin of the History of Medicine* 87 (2013): 622–647.

11. For a retrospective on the influence of the post-war NIH, see a special issue of *Science* upon the institution's centenary: *Science* 237, no. 4817 (August 21, 1987), especially two retrospective articles in which NIH directors surveyed the institution's growth leading up to the 1980s: James A. Shannon, "The National Institutes of Health: Some Critical Years, 1955–1957," *Science* 237, no. 4817 (August 21, 1987): 865–868 and James B. Wyngaarden, "The National Institutes of Health in its Centennial Year," *Science* 237, no. 4817 (August 21, 1987): 869–874.

12. Ellenberg, Gail, and Geller, "Conversations with NIH Statisticians."

13. Derek J. de Solla Price, *Little Science Big Science* (New York: Columbia University Press, 1963); more broadly, Peter Galison and Bruce Hevly, eds., *Big Science: The Growth of Large-Scale Research* (Stanford: Stanford University Press, 1992). Most of the literature focuses on "bigness" in physics.

14. I primarily used Clarivate Analytics' Web of Science collection, including Medline, Social Sciences Citation Index, and Biosis Citation Index. The network visualization was produced with Cytoscape, using tables created with OpenRefine.

15. Publication numbers are given as approximate because choices about "unique" publications are difficult—an abstract for a project also later published as a research article counts twice, sometimes multiple letters to the editor are grouped and only count once, etc. All such problems seem to be evenly distributed across disciplines, suggesting that while the absolute numbers are approximate, the *relative* distribution is largely stable.

16. Because of the size of the network, I didn't download and edit the data myself but rather used the "Social Network" plugin for Cytoscape, which auto-populates authorship networks directly from PubMed data. This may introduce problems with metadata (including the duplication of names or misspellings, for example) but I ignored them given the size of the data set and its role in my work primarily for situating the main co-authorship network. There are also, of course, substantial complications using MeSH terms given that (1) they were often introduced at different times by the National Library of Medicine and applied retroactively to articles (epidemiologic methods and neoplasm were both introduced in the mid-1960s) and (2) they may not capture what contemporary actors considered under those terms. This has unfortunate narrowing effects: Jerome Cornfield, despite authoring an influential article on using statistical methods (odds ratios in particular) to examine cancer rates in the 1950s, was filtered *out* using these search terms. Either one would have to go through by hand, or simply expand the network finding other more inclusive MeSH terms.

17. One suspects, in fact, that the large, interconnected network under discussion is in fact just the NIH itself. That's rather to be expected, as the first and largest single funder of cancer research. However, there was no guarantee that the statistical group would be publishing in that particular network, let alone have such central places within it, suggesting that their influence was still impressive.

18. Patel, "The Benevolent Tyranny of Biostatistics," 629–630.

19. All numbers are approximate given that new citations are still appearing.

20. David Williamson Shaffer, Wesley Collier, and A.R. Ruis, "A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data," *Journal of Learning Analytics* 3, no. 3 (2016): 9–45; see also Ruis, this volume.

21. Samuel W. Greenhouse, "Some Reflections on the Beginnings and Development of Statistics in 'Your Father's NIH,'" *Statistical Science* 12, no. 2 (1997): 82–87, on p. 84.

10. Using Data and Network Analysis in Humanities Research: A Guide to Getting Started

NATHANIEL D. PORTER

Network thinking and analysis are now widely used in diverse disciplines throughout the academy. In this chapter I will offer a brief primer on network analysis, aimed specifically at understanding the methods and principles used by the authors in this volume, all of whom participated in the Viral Networks workshop. I will begin by explaining basic terminology and models commonly used in network analysis, which should be valuable to anyone thinking of using network analysis or visualization in their own work. Then I will outline a typical network analysis workflow and offer tips on getting started in network analysis as a traditional humanist, based on my observations from helping workshop participants. This chapter will be most useful to those considering using network analysis for the first time. Those looking for more information or inspiration on network analysis and what it can accomplish can find resources in the book's glossary and this chapter's references.

First, let's clarify what we mean by the terms network thinking and network analysis. Chances are, even if you have never engaged in statistical analysis or other structured, formal types of data analysis, at some point you have used network thinking. Take, for instance, surveys. Traditional surveys and vital statistics , such as measures of victims of a disease reported by physicians or hospitals, are typically used to gather and analyze data about distinct and separable individuals or groups. The gold standard is a population-representative sample that reflects, as closely as possible, the characteristics of individuals in an entire group, so that you can

answer questions such as, "Who is most susceptible to a particular disease?" or "How do disparities in health outcomes compare to race, poverty, or age?" The underlying assumption is that people act somewhat independently and that a good way to understand social patterns is to look at the distribution of people with different characteristics.

In contrast to traditional surveys, *network surveys* start with the assumption that social environment (family, friends, school peers, fellow group-members, etc.) is an integral part of who people are and how they make decisions. Instead of asking, for example, "Are young people most likely to contract sexually transmitted diseases?" a network approach might ask, "Does having strong relationships with family, friends, or co-workers affect the likelihood of contracting a sexually transmitted disease?" In both ways of thinking, questions can be quite nuanced, but a traditional survey is more about individuals, regardless of any ties among them, whereas, a network survey intentionally collects and draws on information specifically about the ties between and among individuals in a given environment.

In many ways, this distinction is not new to the humanities. The clearest parallel is the distinction between case study methods and comparative methods. Scholars use case studies to understand the distinctiveness and character of a single category or entity, be that an author, national or local context, time period, etc., in as much detail as possible. A comparative study focuses principally on defining a set of characteristics that can be compared or contrasted to provide insight into how these characteristics are associated with specific historical factors or outcomes. Case studies help us understand exemplary individuals, communities, or businesses, and yet the subject of a case study (e.g. Florence Nightingale, Detroit, or IBM) is rarely isolated entirely from the influence of contextual factors. Network analysis formalizes the contextual factors and relational thinking already embedded in comparative approaches to treat those very relationships as items of interest, whether as causes, effects, or simply patterns to be studied.

Formal network analysis can, no doubt, be intimidating. Many of the authors in this volume, despite having self-selected into a workshop on historical networks, initially expressed concern at the prospect of moving from close reading of specific events, actors, and processes towards coding data and producing truly relational models. With help, however, all authors came to appreciate both how coding data can produce a disciplined form of reflection and how network analytics can enhance or complement other approaches. It was not the goal of the workshop—nor is it the goal of this volume—to transform traditional historians into network scientists or data scientists, although, frankly, both network and data scientists would benefit from more of the probing attention to detail that is inherent to humanistic inquiry. Instead, the goal for both workshop participants and readers is that they be inspired to new ways of organizing and thinking about evidence and analysis, both as producers and as consumers of knowledge. Now let's delve into basic terminology and models commonly used in network analysis.

Terminology and Models

"What is a network?" The answer to this question is more complicated than it might at first seem. In the broadest sense, a network is any group of entities (people, places, words, ideas, computers, topics, institutions, etc.) that are tied to each other in one of two ways: first, through direct relationships like friendship, partnership, genealogy, or communication; and second, possession of similar characteristics, such as attending the same event or working for the same employer, words or topics that appear in the same corpus of texts, or multiple non-exclusive treatments for the same disease. In many of these cases, a network could just as easily be considered only a collection of similar items; the difference is in the importance placed on the ties. For example, a study of word usage in the works of Shakespeare might ask how the frequency

of specific words changed over time or differed between plays and sonnets (non-network questions); or, instead, such a study could look for clusters of words that tend to appear together across his works and analyze the characteristics of those clusters and/or common language that spans multiple clusters (network questions). It is important to recognize that network and non-network analysis may overlap, intersect, or appear indistinguishable because, as alluded to above, it is a rare analysis that ignores context and relationships entirely. We will return below to the question of what exactly a network is, after exploring network terminology, in order to build a more technical definition that can prepare for the transition from network thinking to network analysis, which requires a clearly-defined network and explicit specification of relationships.

Network Data and Hypotheses

Two elements are basic to any network: *nodes* and *edges*. Nodes are the entities that are connected. In social analysis, nodes are often individual people or organizations. For example, consider the question of peer influence on delinquency and substance abuse among high school students. In this case, the nodes are individual high school students and possibly other important people in their lives such as parents and teachers. Edges are any relationship that ties the nodes together. In delinquency studies, the edge is often friendship, but it could equally be liking or disliking someone, being in the same class or belonging to the same sport team, working on projects together, or sitting at the same lunch table.

Some of these edges are *symmetrical ties*, meaning that both nodes connected by an edge are connected to each other in the same way. Being in a class together is such a symmetrical tie: if person A is in class with person B, person B is also in class with person A. A symmetrical tie that consists of sharing some common characteristic, rather than a mutual relationship, is called an

affiliation tie. Others types of edges, such as friendship, can be *asymmetrical*: person A can consider person B a friend, regardless of whether it is reciprocated from B to A. Another important type of *asymmetrical tie* is network flow: if person A gives advice to person B, the relationship between them is asymmetrical, as person B is receiving advice. Certain types of network properties and hypotheses are only relevant to asymmetrical relationships.

In addition to nodes and edges, the other fundamental type of network data are *attributes*. An attribute is simply a characteristic of a node or edge. N*ode attributes* provide more information about the members of a network: a person's race or age, a place's population, mortality rates, or climate. E*dge attributes* provide information specifically about a tie: strength of friendship, frequency of communication, how commonly words occur together, the date of a connecting event. Many types of edges possess both *sign* (positive or negative, such as like/dislike) and *weight*, which is a special type of edge attribute often used in *network statistics* that represents the strength of a relationship (best friends vs. casual acquaintances).

Network analysts consider a variety of different types of properties, each of which has its own ensemble of language used to describe it. I attempt here to introduce some of the most important network properties pertaining to both whole networks and individual nodes, as well as a few typical types of arguments and the language commonly used to make them. That said, network analysis terminology varies substantially between disciplines, and it may be necessary to consult introductory or reference works within an individual discipline or subdiscipline to understand the specific language you encounter there. This is particularly true for those moving between STEM fields and the humanities and social sciences. Each property will be illustrated with example network visualizations. In general, nodes are represented by points on network visualizations and edges by lines, although there are some variants that will be discussed below.

Properties of Networks and Nodes

The first type of properties to consider are those that apply to the whole network (also called the *graph*). Figure 10.1 shows a sexual contact network of early U.S. patients diagnosed with AIDS. Node labels reflect both the state or city where the diagnosis took place and the order of AIDS diagnosis within a location, which is not identical to the likely order of HIV transmission. Edges represent sexual contact (symmetric), with arrows indicating potential transmission vectors (asymmetric) for the disease. P0 is the person believed to be the initial point of entry for the HIV virus into this contact network. Node color represents the condition(s) with which a person was diagnosed.
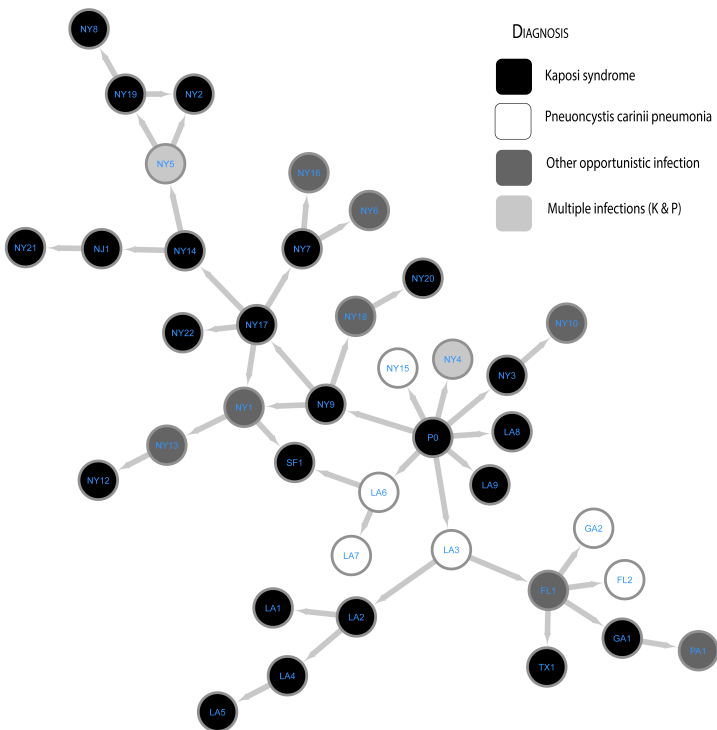

Figure 10.1: Sexual Network of Early Individuals Diagnosed with AIDS

At the most basic level, *density* measures the proportion of possible ties in the network. At one extreme, a fully connected (density = 1) network means that every node has a relationship (edge) with every other node, like a small group of close friends or collaborators. The subnetwork of NY2, NY5 and NY19 near the top of figure 10.1 has density 1. In most cases, however, graphs are *sparse* (density close to 0), particularly larger networks like collaboration across an entire discipline, friendship across a school, or partnerships between physicians licensed to practice in a state. The network in figure 10.1 has a density of 0.053. Each *isolate* (node with no adjacent edges) or *disconnected subgroup* is called a network *component*. *Centralization* measures the extent to which a small group of highly-connected nodes accounts for many of the paths between other nodes, while *clustering* measures the extent to which network components are broken into distinct, loosely connected subgroups.

Specific combinations of these network properties are tied to distinct types of network structures. The most basic structure is a *random network*. Random networks are often used for examples, simulations, or comparison standards, and occur when each edge has a similar or identical probability of being active. They are empirically rare because very few circumstances arise when context or shared characteristics have no relationship to the probability of a tie existing. *Scale-free networks* provide a closer idealized network structure, where the number of nodes with at least X edges follows a *power-law (exponential) distribution*. That is, most nodes have a small number of ties, and the proportion of nodes with at least X ties shrinks rapidly as X grows. Most empirical networks consist of a number of relatively highly-connected subgroups with a few individual nodes *bridging* subgroups to each other. Often, these bridge nodes are of high theoretical importance, for example, as key transmission vectors in the spread of disease or choke points in the diffusion of information. Cohesive subgroups or communities within a network can be distinguished by specific technical variations. The most restrictive type of subgroup is a *clique*, in which every group

member shares an edge with every other; the least restrictive is a *component*, in which every member need only be reachable by tracing edges from every other.

Like networks, individual nodes can be evaluated and scored on a variety of network characteristics. Many are forms of centrality, the importance, however defined, of a given node within the network. The most basic type of *node centrality* is *degree*; that is, the total number of edges it shares with other nodes. *Out-degree* and *in-degree* provide analogues to total degree for *asymmetric* or *directed networks*. In figure 10.1, P0 has a degree (and outdegree) of eight but an indegree of zero. The geodesic distance between two nodes is the minimum number of edges that it takes to connect them. For example P0 had contact with NY9 and NY9 had contact with NY1; the geodesic distance from P0 to NY1 is therefore two.

An individual node has high *closeness centrality* if the average distance to other nodes in its network component is low. However, in many cases, such as diffusion networks, *closeness* is less important than *betweenness*–the proportion of shortest paths (geodesics) a node is on. A node connecting two otherwise separated subgroups is sometimes called a *cutpoint* because if it weren't in the network, the components would be disconnected. Cutpoints have high betweenness. To understand the importance of cutpoints in medicine and epidemiology, consider NY17 in figure 10.1. Without NY17, transmission of HIV from NY9 and NY1 to the top section of the graph could not have occured, at least through this network. A final major concept of node centrality, *prestige centrality*, applies mainly to asymmetric networks. There are many types of *prestige centrality* measures, but all take into account the centrality of nodes tied to each node, rather than simply degree or geodesic distances, in assigning centrality scores.

Networks, nodes, and edges can have many more distinguishable properties. Often they are specific to particular disciplines or substantive research areas. Now let's consider how to assess if network analysis might be useful in your research and, if the answer is yes, how to design the early stages of a network study.

What is My Network?

Every participant in the Viral Networks Workshop was fortunate to have entered with a research project that was in some way "network" oriented. Perhaps it is surprising, then, that the most challenging question that I, as the data and visualization consultant, posed to many of them was, "What is the network you are studying?" It is encouraging, by the same token, that many participants remarked that being forced to answer this question up front was one of the most valuable technical elements of the workshops.

When trying to define a network, it is important to first consider three elements: the network's nodes, edges, and research context. Each of the three, at least in relation to an analytic project, hinges on two questions: what matters and what is measurable. In practice, the step of defining the network is often an iterative process: start with general ideas, try to define a network, check what you might actually be able to do in terms of finding and analyzing data, then refine the general ideas and try again until something workable coalesces.

I often recommend to people that they start the process by thinking about a hypothetical report on their research and drafting a title for the report that incorporates all three elements—e.g. "The Network of [edge relationship] between [nodes] in [research context]." When considering possible nodes, it is important that they share some common characteristic(s). In the early stages of their projects, a number of participants struggled with this because they tended to think of networks more like flow-charts, where anything could qualify as a node and any relationship as an edge. In principle, there is no problem with this; networks can be quite complex as long as the nodes and relationships are clearly defined. However, each additional type of node tends to limit network analysis' potential to serve as more than a glorified concept map. In some cases, more complex projects may involve constructing multiple related networks that can be compared or combined. It

is usually helpful, therefore, particularly in the early stages of definition, to draw a mock-up of the network or networks of interest and think about how they might be analyzed.

The situation is slightly different for affiliation networks, which have two distinct types of nodes rather than one. These nodes are often called *actors* and *events* because early affiliation networks were based on co-attendance at specific events. I often find it helpful to think of them instead as *topics* and *ties.* For example, in an affiliation network of doctors and hospitals, where an edge represents having worked in a particular hospital, a scholar might be interested in understanding how doctors (topics) are connected by hospitals (ties) over time. Or, another scholar might be interested in how hospitals (topics) are connected across locations (attribute) by doctors (ties). In other words, in an affiliation network, the node that is the topic and the node that is the tie *is entirely dependent on the research question.* Thus, one hypothetical title for research on an affiliation network of doctors and hospitals might be: "The Network of Shared Doctors Between Army Base Hospitals during World War One."

Edges are the second element to be considered when trying to define a network. The edges of a network provide the relationship(s) of interest. Like nodes, the more comparable and clearly-defined the content of an edge is, the more likely the analysis is to be meaningful and understandable. Networks of scientific researchers, for example, can be constructed in a variety of ways. Some common examples include collaboration networks (A writes with B or is co-investigator on a grant with B), citation networks (A cites B), co-citation networks (A and B cite C), supervision networks (A served on doctoral committee of both B and C), and institutional affiliation networks (A and B were both at institution D at the same time). Each of these types of relationships is likely to be important in understanding the overall structure of a particular scientific network, or of scientific progress in general, but network analysis by definition provides a more complex (and hopefully more valid)

representation than case-based models. Thus, only a very limited number of models are capable of simultaneously accounting for such a variety of network types.[1]

The final element to be considered when defining a network is *research context*. In many cases, research context will be readily apparent from the analytical question, especially for historians and other humanists, for whom analyzing sources or events within a defined corpus or timeframe is standard. Network research, however, often requires narrowing the scope or context being considered in order to obtain high-quality data, that can yield insights generalizable to other related contexts.

A pragmatic approach to defining a network is to force oneself to answer the question, "Given my general research goals, what is the most readily accessible type of topic (node), relationship (edge), and context that I could potentially measure or quantify to answer some or all of my research question?" For multiple workshop participants, the most clarifying step in this process came when I asked them to make a sample dataset with a small subset of nodes and edges. This exercise illuminated situations where membership in the set of nodes or edges was poorly defined whether through overly narrow definitions, reducing the quantity of available data, or overly broad definitions, leading to unclear data. For example, many corpuses of text are publicly available through online archives (such as Project Gutenberg or the Internet Archive) and can be used with techniques such as topic modeling (see ch 6 by Cottle) or Epistemic Network Analysis (see ch 8 by Ruis). Likewise, there are standard online sources for many types of scientific networks, such as PubMed or Web of Science. Remember, though, that not all networks need to be large to be effective. Archival data gathered on a single topic can often be conceived of as a network and then productively visualized or analyzed to gain insight that might otherwise have remained hidden if relying on close reading alone (see ch 1 by Runcie, ch 2 by Smith, and ch 7 by Archambeau).

Finding colleagues who are both interested in your topic *and* data-oriented can be a vital step in this process, whether they serve

in a formal role (such as digital humanities specialists or data consultants) or an informal role, say, meeting over lunch to talk about ideas. Only one workshop participant had prior analytic expertise in the method they used for analysis, but with the help of consultation from a small number of analytic specialists and conversation with others in the workshop, each participant was able either to use network analysis to produce insight into their research questions or to determine that it was a poor fit.

Applying Network Analysis

Now that we've reviewed some basic network terminology and considered how to define a network research question, let's identify the typical steps a researcher in the humanities might go through when applying network analysis.

We've already identified the first step, which is to define the network, identify the context, and settle on a research question. Once this has been done, the next step is to make a trial dataset of a few nodes and edges. Network data can be stored in a number of forms, but the most common way is to use two tables, called a *nodelist* and an *edgelist*.[2] As the names suggest, a nodelist is a list of nodes and an edgelist is a list of edges. The nodelist includes columns with a unique identifier for each node, as well as any *node attributes*, such as personal or organizational characteristics, population size, group membership or word frequency. The edgelist minimally contains two columns, representing the two nodes related by each edge. If the data are *directed*, one column is considered a *source* and one a *target*. If edges have an indicator of strength (e.g. a *valued network*), there should be another column for *edge weight*. Any other information about the edges can be included in *edge attribute* columns. Identifiers in the nodelist and edgelist should match exactly. Comma-separated (.csv) or tab-separated (.tsv) text files, which can be created in any spreadsheet program,

are typically interchangeable across software, but some programs may require different formats of input files; search the documentation for your program to find out preferred formats.

In cases where there are multiple relations or affiliations, network data can be quite complex and it may be worth considering if a database (in Access or SQLite, for example) may be more flexible, allowing you to export multiple combinations or structures of the data as networks. Unlike a single table or nodelist-edgelist format, databases can have many different tables, linked by identifiers (see data in ch 1 by Runcie for a relatively simple example).

In the case of relationship data, nodes and edges are fairly straightforward. For affiliation data, however, both types of entities (actors and affiliations) are represented as nodes in a dataset. Each *tie*, then, represents an actor being associated with an event or affiliation. This is also called a *bipartite network*, because there are two sets of distinct types of nodes that can only have direct ties between (but not within) groups. When analyzing affiliation networks, there are procedures for converting the bipartite network into a single mode network in cases where ties are based on how frequently two nodes of the same type are associated with the same nodes. Doing this allows you to focus on one type as the *topic* and the other type as a relationship.

Now it's time to create the dataset. The three main ways to do this are by hand coding, machine coding, and hybrid (or augmented) coding. This first trial dataset is typically made by hand, unless you are importing data from an existing database, such as Web of Science or PubMed (see ch 9 by Phillips), already in a network format. For smaller networks and archival research, the entire network may be hand-coded using the models above, customized to reflect the types of nodes, edges, and attributes included in your data. Machine coding is useful for very large or complex datasets, as well as data that was originally digital such as citation networks, text/topic networks (see ch 6 by Cottle), and web-scraped data. The advantages over hand-coding are time and scale, but it is also easier to miss poor-quality or irrelevant data. Hybrid coding is a relatively

recent development and frequently involves coding a portion of the data by hand, then using either automated tools such as machine learning or crowdsourced workers to create a larger dataset modeled on the initial cases.[3]

The first two steps, definition and data creation, are fairly structured and should be undertaken at specific, definable points in the analytic process. The next two steps, ideally, should be iterative, with the researcher moving back and forth between adjusting visuals and considering the research insight they provide. Don't hesitate to consider multiple approaches to visualization. Visualization early in a project is intended to help discover patterns in the data that might be further investigated. Nicole Archambeau (ch 7) discovered through early visualization that, although there weren't notable gender or age patterns in canonization testimony, her analysis revealed a surprising pattern of people using the first plague mortality as a time marker, rather than a significant event. As you consider your early visualizations be sure to look at some basic network and node characteristics that are calculable in nearly every network package. Each iteration of visualization should reveal important characteristics of the network as well as answers to the research question.

As you move toward a final visualization, be sure to tease out the story your research is telling, in both its layout and design features. Visualizations are, above all else, a form of communication. They should be clearly labeled and free of visual elements that do not represent data (i.e. drop shadows). Often, peripheral elements, such as node labels, isolated nodes or very weak ties, can be removed entirely to improve clarity. Creating effective visualizations, like good writing, requires multiple drafts, critical reading by colleagues, experimentation with formats, and willingness to fail. (Always save backup copies of the data and each version of the visualization.)

Practical Advice

Assisting the cohort of scholars in the Viral Networks Workshop offered me a unique vantage point from which to observe the challenges that traditionally-trained humanists face when attempting for the first time to do network-related research. The following tips come directly from this experience.

First, not all research problems benefit from network thinking and analysis–though many can. To address this challenge, think creatively and critically about what network you are interested in and how it addresses your research question. For humanists, in particular, I would encourage starting by hand-drawing a model of what the visualization product could look like at the end–and consider how this outcome will advance your research agenda. Researchers often invest substantial effort into a project thinking it will fit a particular analytic model, only to discover that they had missed something important that they otherwise would have caught had they followed these preliminary clarifying steps. Nothing is more frustrating than spending hours hand-coding data, only to have to go back and repeat it all because of a simple oversight.

Second, get to know your data and talk about your early thoughts and findings with others outside your discipline. Doing so is vital to developing and communicating network research. A number of the authors in this volume detail the development of their research as they worked in Cytoscape or other software to explore and refine their visualizations. In every case, seeing the possibilities sparked new insight for their project–connections that might never have been made without turning a traditional history project into digital data. Each participant started the workshop with his or her own, distinctive project, but by coming together and talking with each other and a small number of outsiders, they were able to clarify their questions, goals, and processes, ultimately leading to an impressive array of chapters. Collaboration is a vital aspect of creative and scientific growth, even in disciplines where the solo scholarly endeavor is normative.

Third, any researcher who can produce an article or monograph can also succeed in creating a network analysis. The very process of applying digital humanities tools and methods to one's humanities research can be a powerful analytic stimulant. None of the projects here has the broad scope of the most prominent digital humanities projects, yet all benefited from the discipline required to turn their research materials into digital data and the possibility for unexpected discovery that comes from letting others, even computers, participate in the process.

Selecting and Learning Software Tools

Workshop participants worked primarily with two software packages, Cytoscape and Epistemic Network Analysis (ENA). These tools were chosen because of their ease of use and broad range of potential applications. Two additional packages, Gephi and the Python package scikit-learn, were used for their specialized mapping and text analysis capabilities respectively. In this section, I will provide some advice for getting started in Cytoscape and ENA, followed by an overview of other options and when they might be worth considering.

Cytoscape is a free network analysis and visualization package for all operating systems. It is most commonly used in health and biological sciences, although Miriam Posner has created an excellent tutorial,[4] used by many workshop participants, on Cytoscape for humanities applications. Additionally, Cytoscape has a large and growing collection of plug-ins, including ones calculating network and node statistics, downloading citation networks from PubMed, and allowing for easy publishing of interactive visualizations to the web.

The best way to learn Cytoscape is, frankly, to try it out. Original projects for all of the Cytoscape visualizations in this volume are available in the online supplements, and can give you a good feel for the software. When you first open Cytoscape, the splash screen

will present you with options for accessing an existing project or creating a new one. In Cytoscape each project is a single file corresponding to related analysis or networks to which you can add multiple datasets, layouts, and style sets. The main window has three panels. When you open a project, the first one you will probably want to look at is the visualization at the top right. You can drag or use standard zoom gestures to get a better feel for different parts of the network, and many options are available by right-clicking on nodes or edges. You can also drag nodes with your mouse or change the layout using the Layout menu. On the bottom right is the Table Panel, where you can view or edit the source data Cytoscape used to produce the visualization. The Control Panel on the left is the heart of customization for the visuals, and allows you to select from multiple networks, adjust the appearance nodes and edges, and select subsets of the data. The Style tab in particular allows you to use colors, size, shapes, labels, or even images to represent node and edge attributes and help tell your network's story.

To import your own data into Cytoscape, start a new project and choose File-Import-Network-File from the menus and import the edgelist; then repeat the process with File-Import-Table-File and the nodelist. To add extra features like auto-imports of web data or network statistics, use the Apps menu. Once your data is imported, think about your research questions and how they might be elucidated visually and then play around with options. When you're satisfied with the product, you can save the project for use in Cytoscape, save the diagram as a picture, and save the project as a web page.

ENA is a relatively new software package, available for free, both through a web interface and the rENA package for R statistical software. Unlike Cytoscape, ENA's design is based on a specific methodology and not useful for more general exploration of networks. ENA answers variations on a single question: "How do a set of concepts co-occur throughout a corpus of coded material."

For instance, it is excellent for answering questions a researcher might have about a particular word or phrase, such as its usage and meaning vary over time or between contexts.

The original intent of ENA was to analyze and compare different stages and adaptations of educational activities, but it can be applied to any collection of sources that can be coded in terms of a small number of key ideas. Source data for ENA must match a specific format, and codes must already be created and applied prior to importing. The sample data provides helpful examples of the different elements of an ENA project, and the web interface can help guide new users through selecting variables. There are a number of good resources and tutorials to help you determine if ENA is right for your project, and to get you started with the web tool.[5] The web interface provides a user-friendly way to experiment with data and produce attractive and useful visualizations. The R package, while still in a preliminary form at the time of writing, is useful in documenting your work and making it available and replicable to others, as well as providing simple data transfer for current R users and a way to share datasets exported from the web tool.

Other widely-used, standalone network packages include Gephi, Pajek, and UCINET. Gephi and Pajek are free and cross-platform; UCINET is Windows-only and is free to try with full functionality for 90 days. Gephi is similar to Cytoscape in many ways, although the controls are less intuitive for new users. It is focused on visual design of networks, provides a great deal of customizability and multi-format exports, and has some key features that Cytoscape lacks, such as geographic network visualization with map overlays. Pajek provides a mix of both visualization and statistics features, and is particularly good for working with very large networks. UCINET has a larger variety of statistics and is among the best-documented, but its visualization tool NETDRAW is less refined and works best with smaller networks. While Cytoscape and Gephi work by importing all data into a single project that stays open and

accessible throughout the session, Pajek and UCINET are more modular and require combining input and output files for each step of the process, adding flexibility but increasing the learning curve.

Network modules are available for a number of more general software packages, as well. The most user-friendly of these are NodeXL, a plug-in for Microsoft Excel for working with small to medium networks, and Tableau, an interactive data visualization tool. NodeXL's greatest advantage is its integration with Excel; editing data and moving between worksheets will be familiar to many users, and there is no need to export data to another package. Tableau, available free to students and educators, is excellent at rapidly producing clear visuals without the need to code, although its drag and drop interface can limit its flexibility. R and Python both have extensive network analysis packages, although R's are more full-featured and include many statistical procedures for simulation and modeling that are not available in other packages.

A final software class to mention here is interactive html and JavaScript visualization tools. Gephi, Cytoscape, R (via plotly), and Tableau all export visualizations that web users can visit and explore themselves, changing display options or even the network itself. However, a new collection of tools, such as d3.js and node.js, have emerged in the last few years to allow embedding network data natively in web pages with extreme customizability and interactive flexibility. Their application is limited by the need for fluency in their coding language, but they remain an option for high-impact visualization for code-savvy researchers or those collaborating with programmers. These tools are not limited to network data; they are designed as full-featured data visualization tools. To get a sense of what is possible with these packages, you can browse visualization galleries such as those at d3js.org or FiveThirtyEight.[6]

Getting Started on Your Own

At this point, some people considering network analysis for the first time may feel overwhelmed by the variety of options available. So how should one get started? The best options, depending on your access to support, are, first, to take a hands-on, instructor-led workshop or course in network analysis, and, second, to find a colleague who uses network techniques. In addition to departmental colleagues, many universities have statistics or research data consultation available through the library, statistics department, or social and demographic research centers. Tapping into experience in this way can save a great deal of frustration both on learning the language and processes involved and finding the right tools.

If in-person help is not practical or available, the next-best option is to start with a user-friendly tutorial or textbook. The Cytoscape tutorial by Miriam Posner (discussed above) combines an introduction to network concepts and data with application to real data. At present, the most accessible textbook on applied network analysis is *Analyzing Social Networks*, by Borgatti et al.[7] It provides both a readable introduction to a wide variety of network concepts and a good overview of the elements of network visualization, all using UCINET software. NodeXL, Pajek and Gephi all have hands-on books to help you get the most out of your chosen software.

I wouldn't recommend starting with software documentation for the simple reason that all of the major packages assume existing familiarity with network analysis. Once you have experience with a single tool, you may choose to stick with it or you may discover it doesn't meet your needs and try something else. Either way, just getting started, creating and working with network data, will be invaluable regardless of the tool you choose in the end.

Another option is to find a paper that employs methods or a particular visual approach you would consider adapting for your own research. The greatest advantage here is that you can more quickly discern whether your research question is a tractable

network question and what tools or techniques may be most relevant. The challenge, however, is that many network papers are written by and for people who live and breathe network analysis or statistical programming. Often the techniques they use would be difficult if not impossible for a novice, even if they are an expert in the same subject area. Still, if you see something that makes sense to integrate in your research, you can try to learn a little more about the methods or ask a colleague if they are feasible for you. Understandably, this approach is best used in combination with the others; start with an idea of where you want to end, read carefully to find out how previous researchers got there, and then use that information to help select the tools or approaches you'll need to learn to pursue your research question.

Conclusion

My goal in this chapter has been to convince readers that, if they are successful researchers in their own substantive fields, more than likely they will be able to productively use network analysis provided that they take a few basic steps. First, they need to learn to think in terms of networks and network data. Second, their research questions and data sources must be appropriate for network analysis. And third, they must be prepared to match their goals to the appropriate tools and learning resources.

Based on my experience of the Viral Networks Workshop in the capacity of data consultant, I would encourage all humanities scholars to keep talking to colleagues, keep coming back to your research question and sources materials, and keep playing. With these conditions and exhortations in mind, you should have the tools to embark in a new direction toward network research, whether your networks consist of friends, enemies, letters, places, patients, doctors, ideas, or anything else. Whether you intend to become a network or digital humanities specialist or you simply

want to enhance and complement other approaches, network thinking, tools, and visualizations are useful additions to your toolbox.

Endnotes

1. Tom A. B. Snijders, "Statistical Models for Social Networks," *Annual Review of Sociology* 37 (2011): 131-153.

2. Examples can be found in the supplemental data for this chapter.

3. Matthew J. Salganik, *Bit by Bit: Social Research in the Digital Age* (Princeton University Press, 2017).

4. Miriam Posner, *Creating Network Graphs with Cytoscape* (web page), https://github.com/miriamposner/cytoscape_tutorials.

5. David Williamson Shaffer, Wesley Collier and A.R. Ruis, "A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social and Interaction Data," *Journal of Learning Analytics* 3 (2016): 9-45.

6. https://github.com/d3/d3/wiki/Gallery, https://fivethirtyeight.com/tag/data-visualization/

7. Stephen P. Borgatti, Martin G. Everett, and Jeffrey C. Johnson, *Analyzing Social Networks* (London: Sage, 2018).

Contributors

Viral Networks Workshop, January 2018:

Keynote Speaker, Consulting Scholars, and Advisory Board Members:
Ryan Cordell, Northeastern University
E. Thomas Ewing, Virginia Tech
Theresa MacPhail, Stevens Institute of Technology
Amy Nelson, Virginia Tech
Nathaniel Porter, Virginia Tech
Peter Potter, Virginia Tech
Katherine Randall, Virginia Tech
Jeffrey Reznick, National Library of Medicine
Samarth Swarup, Virginia Tech

Contributing Scholars:
Nicole Archambeau, Colorado State University
Katherine Cottle, Goucher College
Michelle DiMeo, Science History Institute
Lukas Engelmann, University of Edinburgh
Melissa Grafe, Yale University
Anna Lacy, University of Delaware
Christopher J. Phillips, Carnegie Mellon University
Andrew Ruis, University of Wisconsin
Sarah Runcie, Columbia University
Kylie Smith, Emory University
Katherine Sorrels, University of Cincinnati

Workshop Participants:
Gabrielle Barr, National Library of Medicine
Ben Busby, National Library of Medicine
Seth Denbo, American Historical Association

Delia Golden, National Library of Medicine
Atalanta Grant-Suttie, National Library of Medicine
Ken Koyle, National Library of Medicine
Christie Moffatt, National Library of Medicine
Jennifer Serventi, National Endowment for the Humanities
Susan Speaker, National Library of Medicine
Elizabeth Tran, National Endowment for the Humanities

Contributing Authors to this Volume:

Nicole Archambeau is Assistant Professor of History at Colorado State University. Her research projects explore plague, war, and diverse healing methods in 14th-century Europe.

Katherine Cottle is Assistant Professor of Writing at Goucher College. Her current research explores anatomical epistolary analysis as a method of mapping historical intimacy.

Michelle DiMeo is Associate Library Director for Collections Development at Hagley Museum and Library. Her research focuses on early modern science and medicine, particularly household experimentation conducted by women and lay practitioners.

Lukas Engelmann is Chancellor's Fellow at the University of Edinburgh. His current research explores the twentieth-century history of epidemiology, with particular focus on the relation of narrative and formal methods in the field.

E. Thomas Ewing is Professor of History at Virginia Tech. His current research explores the intersection of medical history and digital humanities in the context of the Russian influenza (1889-1890).

Christopher J. Phillips is Assistant Professor of History at Carnegie Mellon University. He is researching a history of statistical analysis in mid-century medicine.

Nathaniel Porter is Social Science Data Consultant and Data Education Coordinator in the University Libraries at Virginia Tech.

His current research uses online experiments, big data, and network analysis to study contemporary U.S. Christianity and social psychology.

Katherine Randall is a doctoral candidate in rhetoric and writing at Virginia Tech. Her research focuses on ethical practice in clinical and public health communication, specifically regarding resettled populations in the United States.

Jeffrey S. Reznick is Chief of the History of Medicine Division of the U.S. National Library of Medicine of the National Institutes of Health. His research focuses on the social and cultural history of World War I and its aftermath, particularly in the contexts of health and humanitarianism, memorialization, and material culture.

Andrew Ruis is a researcher at the Wisconsin Center for Education Research and a fellow of the Department of Medical History and Bioethics at the University of Wisconsin-Madison. His current historical research, which focuses primarily on the the history of food, nutrition, and health, explores methods for analyzing historical sources using a combination of qualitative and quantitative approaches.

Sarah Runcie is Assistant Professor of History at the University of Louisiana at Lafayette. Her current research focuses on the intersections of decolonization in Africa and histories of global health.

Kylie Smith is the author of "Talking Therapy: Knowledge and power in American psychiatric nursing," and her current research examines the impact of the Civil Rights Act on psychiatric institutions in the South.

Katherine Sorrels is Associate Professor of History at the University of Cincinnati. Her current research is on intellectual disability and alternative medicine in twentieth-century Europe and the U.S.

Glossary of Network Terminology

Actor In an *affiliation network*, the people or other entities tied by *events*

Asymmetrical tie An *edge* or relationship in a *directed network* that is not *reciprocated*; for example Bob cites Jane but Jane does not cite Bob

Affiliation network A network where the *edges* consisted of a shared characteristic, such as attending a class together, rather than a direct relationship, such as friendship, and the *nodes* are the *actors* and *events*; actors cannot be directly tied to other actors, nor events to other events

Attribute A characteristic of a *node* or *edge*; can be used to select nodes and edges or as an analytic variable; can also be represented visually through size, color, etc.

Bipartite network A network where ties occur only between (and not within) two distinct *subgroups*; *affiliation networks* are a type of bipartite networks

Betweenness centrality A type of *node centrality* measuring the importance of each node in *geodesic paths* between other nodes

Centralization A *network statistic* measuring how unevenly spread the edges in a network are; a network with high centralization has relatively few key *nodes* connecting a large number of other nodes

Clique A subgroup of *nodes* where each node shares an edge with every other node; the most restrictive subgroup definition

Closeness centrality A type of *node centrality* determined by the *geodesic distance* to all other *nodes* in a *component*; high closeness indicates that most other nodes can be reached in relatively few steps

Clustering A *network statistic* measuring how strongly *nodes* are grouped; high clustering indicates that most nodes are part of distinctive subgroups that are more highly connected to each other than to other nodes in the network

Component A *set of nodes* that are all reachable from each other tracing *edges*; a *network* with only one component is called a connected network

Cutpoint A *node* whose removal from the network would cause two subgroups to become disconnected *components*

Degree The total number of *nodes* a node is directly connected to via all *edges*; for example, if Fred, George and Martha each claim Julie as a friend and Julie claims Fred and Jane as a friend, Julie's degree is 4

Density The proportion of possible *edges* that exist

Directed network A *network* with *asymmetric ties*

Disconnected subgroup Group of nodes with no edges reaching past the subgroup

Edges The relationships or shared characteristics that connect *nodes* in a *network*

Edge attribute Additional characteristics of an *edge*, such as the type or frequency of the tie

Events In an *affiliation network*, the shared characteristics or associations that form the ties between *actors*

Geodesic distance The number of *edges* in the *geodesic path* between two nodes

Geodesic path The shortest path (least number of edges) connecting two *nodes* in a *network component*; in a *directed network*, geodesic path must follow direction of ties

Graph An entire *network*; does not refer to visualization but to the network itself

In-degree For *directed networks*, the total number of *nodes* selecting a given node; for example, if Fred, George, and Martha each claim Julie is their friend, Julie's in-degree is 3

Isolate A *node* that has no *edges* connecting it to other nodes in the *network*

Multirelational network A network that include more than one type of edge or tie between nodes; for example including both co-authorship and citation relationships

Network A set of *nodes* (entities) and *edges* (relationships); can be further differentiated into empirical and observed networks

Network connectivity Whether all *nodes* in a *network* are reachable from all others; a fully connected network has only 1 *component*; can also refer to measure of the number of nodes that would need to be removed to split the network into multiple components

Network statistics Measures that summarize characteristics of an entire *network*

Nodes The entities in a *network that* are connected to each other through *edges*; can be any individual, collective, or in the case of affiliation networks, shared characteristics or activities

Node attribute Additional characteristic of a *node*, such as name, type, or quantity

Node centrality A large family of measures of how important a *node* is within a network based on the number and/or characteristics of *edges* connecting it to other nodes

Out-degree For *directed networks*, the total number of *nodes* selected by a given node; for example, if Julie claims that Fred and Jane are her friends, her out-degree is 2

Power-law (exponential) distribution A distribution where most *nodes* have low *degree* and the proportion of nodes with degree of at least X shrinks rapidly as X increases; corresponds to *scale-free networks* and frequently fits well with rank-order distributions (such as sales rankings)

Prestige centrality Measures that summarize the prominence or prestige of a *node* based on *in-degree* and the prestige of the nodes selecting the node

Random network A *network* where every possible *edge* (e.g. pair of nodes) has equal probability of existing; empirically rare but often used for simulations and baseline models; also called an Erdos-Renyi random graph

Reciprocated tie A *directed edge* where both nodes select the other, such as two classmates that identify each other as close friends

Scale-free network A *network* where degree is distributed roughly according to a *power-law*

Sign An *edge attribute* denoting whether a tie is positive (such as friendship) or negative (such as dislike); many types of edges can only take a positive sign

Sparse graph A *network* where only a small proportion of possible *edges* exist; most large empirical networks are sparse

Subgroup A group of *nodes* which are more closely connected via *edges* to each other than to nodes outside the subgroup

Symmetrical tie An *edge* that is either non-*directed*, such as belonging to the same group, or are *reciprocated*

Tie More general term for an *edge* or relationship

Topic More general term for a *node* or *actor*

Weight An *edge attribute* indicating the strength, volume, frequency or recency of the tie; *networks* with edge strength are call *valued networks*; color intensity or line width of edges often represent edge strength in visualizations

Valued network A network where edges are assigned different *weights*