

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 vused 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 with those several rarities of bookes and Manuscript papers for which I am much obliged and returne you many thanks. I sent you back in my former letter according to your desire a catalogue [see 32/1/12] of every particular 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 from 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 from my brother mr John Richards from 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 that 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 that 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, which 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 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, & its rare also among the Indians who vse it constantly: mr Harwood or any [*H underlines*] New England man will or woman can direct the making of & dressing of that sampe or direct to some

New England woman 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 & ingagements: 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 parts of England for such cures, & will heare of many experimentall cases in that kind it may be of some yet living: & will know which may be the fittest for your particular case: & whether they may be transported with their intire virtue from 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 vtilis ad expellendos calculos renum, I have not the bookes at present but find this in some papers which I overlooked lately in reference to your trouble as a [word deleted] memorandum I had taken, I suppose out of one of those authors my note also speakesmentions De fonte Bollensi ex Fallopiæ de aquis medicatis In & I thinke Bauhicuss hath something of the same In Regiense agro aput castellum vocatumBrondale 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 parts where he liveth which is much used for other distempers. 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 attained 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, exopt in some few. that friend of whose talents you desired to be informed, hath an other very reall way *which* may be probably attainable, 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 settis before vs is greate sed vita brevis: this way *which* he intends hath some concomitants *which* would greatly advance commerce & other publique concernments for the benefitt of poore & rich in great Britaine & the good of these plantations would easily be involved therein [*word deleted*] but it cannot 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 *which*> the vndertakers of such a banke would be invested, be conveniently intrusted in a letter but if he could by any oportunity speake with 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: *which* 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 from what you seeme to hint in your letter to be surmised, he is contented with a wilderness condition & I beleive can truly say Fælix cui deus obtulit Parca quod satis est [*manu?*] yet I know when he can have such a visible stock, is not without thought of one voyage more into Europe: I know it is his iudgement *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* cannot but spend much, *which* I think hath made him speak of a visible stock as I have mentioned from his owne expressions: though he might have supply by what traffique 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 hartly desires to that great phisitian to give you perfect recovery, and my most reall respects presented, I shall take leave to subscribe myselfe

Honored Sir

Hartford Jan: 7: 1660 Youre cordiall friend
in New England John Winthrop
Sir If you can receive pay for them according to this inclosed
letter I desire you to procure me these few bookes: viz:
Selenographia
Systema Saturnium
All Glaubers bookes exe in duch or latine
except his Fur booke of New Furnaces with
appendices & .. de auro potabili
& his thre books operum mineralium.
and his Miraculum mundi: 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 Weinstein
printed I think at Hamburg
[Kesslerus?] Fur auserlegene
process the last edition
I think it is funff Hundred
auserlegene processen

Acknowledgments

This work was supported in part by the National Endowment for the Humanities, the National Library of Medicine, the National Science Foundation (DRL-0946372, DRL-1247262, DRL-1661036), and the Wisconsin Center for Education Research. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

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). DOI: 10.1111/lic3.12388.

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).