Ethical and Effective Visualization of Knowledge Networks

Chelsea Canon <canon_at_nevada_dot_unr_dot_edu>, Department of Geography, University of Nevada, Reno
https://orcid.org/0000-0002-0431-343X

Douglas Boyle <douglasb_at_unr_dot_edu>, Department of Geography, University of Nevada, Reno
https://orcid.org/0000-0002-3301-3997

K. J. Hepworth <katherine_dot_hepworth_at_unisa_dot_edu_dot_au>, UniSA Creative, University of South Australia,
Adelaide, Australia https://orcid.org/0000-0003-1059-567X

Abstract

Knowledge mapping combines network analysis and data visualization to summarize research
domains and illustrate their structure. In this paper, we present a framework for ethical and
effective visualization of knowledge networks, which we developed while building a knowledge
map of climate communication research. Using the climate communication knowledge map as
an example, we highlight the practical and ethical challenges encountered in creating such
visualizations and show how they can be navigated in ways that produce more trustworthy and
more useful products. Our recommendations balance tensions between qualitative and
quantitative and objective and subjective aspects of knowledge mapping. They demonstrate the
importance of critical practices in the development of knowledge maps, illustrate the intertwined
nature of analysis and results in such projects, and emphasize the constructedness of the
resulting visualization. We argue that the only way to produce an effective knowledge map is to
produce an ethical one, which requires attention to the ways trust and accountability can be
produced at every step of analysis and production. This extends the literature on ethical
visualization in digital humanities projects by offering a clear example of the utility of a critical
approach for a traditional, science-oriented knowledge mapping project.

1. Introduction

Scholars have long been captivated by the idea that knowledge generation is outpacing their ability to keep up, necessitating creative techniques and solutions to organize, manage, and access information. From scientist Vannevar Bush's memex[1] to futurist Richard Buckminster Fuller's Geoscope,[2] researchers have envisioned high-tech tools for information management and knowledge creation since before the advent of big data. Today, well into the era of hyperlinked text and geographic information systems not unlike the miracle tools envisioned by 20th century inventors, making knowledge on even very specific areas of scholarship tractable remains a complicated challenge. As scholarly knowledge continues to grow, so does interest in quantitative techniques for summary, survey, and synthesis of knowledge domains.

1

Knowledge mapping is one such survey and synthesis technique. It combines data visualization and network analysis to depict large collections of information spatially, as if they were a landscape viewed from above. Knowledge mapping is a general term for an analysis informed by and applied in multiple areas of scholarship, including information science [Börner 2010], literature [Moretti 2007], history [Graham, Milligan, and Weingart 2015], and bibliometrics [Chen 2016]. These maps are at once analytical, managerial, and communicative: analytical because they provide tests of network theories and because bibliometric data was one of the earliest sources of big data for such analyses; managerial because they guide decision-making, funding allocation, and evaluation; and communicative because they aim to support discovery and guide hypothesis formation by depicting knowledge as a collaborative, interlinked whole. Knowledge maps are simultaneously rhetorically powerful (because they make visual arguments) and ethically fraught

2

(because visual arguments presented in such charts are often accepted as objective fact), but knowledge mapping techniques have not yet been brought into conversation with recent work in digital humanities surrounding ethical visualization and data feminism [D'Ignazio 2019]; [D'Ignazio and Klein 2020]; [Hepworth 2017]; [Hepworth 2020].

Ethical visualization is the practice of acknowledging and mitigating the potential for harm that is inherent to particular visualizations [Hepworth and Church 2018]. Ethical visualization and data feminism require consideration of the whole data pipeline, from acquisition to analysis to visualization, and attention to the ways in which some perspectives are marginalized, omitted, or elided [D'Ignazio and Klein 2016]; [D'Ignazio and Klein 2020]. Both traditions require not only acknowledging the potential negative effects of a visualization or drawbacks of a particular dataset, but actively working to mitigate them in a way that increases understanding while minimizing harm [Cairo 2014]. Given nascent discussions in bibliometrics about the potential harm perpetuated by uncritical application of bibliometric techniques for evaluation, it is an important time to bring critical practices into conversations about knowledge mapping [Bornmann 2017]; [Conway 2014]; [Donovan 2019]; [Furner 2014]; [Rolf 2021]; [Zuccala 2016].

In this paper, we demonstrate how principles of ethical visualization and data feminism can be applied in the development of a knowledge map to produce a more trustworthy and more useful product. We do this by providing a detailed account of the path we followed to build a knowledge map of climate communication research, highlighting practical challenges encountered in creating such a visualization and how ideas from ethical visualization literature helped navigate them. Specifically, we: (1) demonstrate the importance of critical practices in the development of a complex type of chart, (2) illustrate the intertwined nature of analysis and results in big data projects, and (3) take a first step toward a feminist bibliometrics informed by critical work in digital humanities. This addresses a previously identified need for clear guidance on the visual communication of knowledge networks [Conway 2014]; [Gavrilova, Kudryavstev, and Grinberg 2019]. It has relevance for digital humanists working with knowledge networks and bibliometricians seeking practical strategies for mitigating harmful externalities of evaluative bibliometrics. It also enriches ongoing conversations in the digital humanities about ethical visualization and data feminism with examples of how frequently ethical choices arise in a data visualization project.

Ultimately, we emphasize the constructedness of these types of visualizations, showing why a purely quantitative, purely objective depiction of a knowledge network is neither possible nor particularly useful. We argue that the best way to produce an effective knowledge map is to produce an ethical one, which requires attention to the ways trust and accountability can be produced at every step of analysis and production.

2. Knowledge mapping

Knowledge maps are powerful tools for planning, collaborating, teaching, and communicating, because they depict metaknowledge about how topics and problems are structured [Börner 2010]; [Evans and Foster 2011]. *Metaknowledge* – literally, knowledge about knowledge – may sound like a dangerously totalizing "god trick," but it can in fact be understood as a concept gesturing at the positionality and situatedness of all knowledge, and the necessity of combining multiple partial perspectives to understand a system [Haraway 1988]; [D'Ignazio 2019]. Sociologists James A. Evans and Jacob G. Foster define metaknowledge as the result of "critical scrutiny of what is known, how, and by whom" and suggest that we might best understand what metaknowledge is by thinking about what we, as individual scholars with particular trainings, career paths, and expertise, think about as we read a journal article in order to "index a broader context of ideas, [scholars], and disciplines" – the kind of knowledge gained from being a participant in a system through time [Evans and Foster 2011, 721]; [Kwan 2002]. Knowledge maps offer an alternative way to gain a similar kind of qualitative understanding via quantitative techniques, affording viewers an intuitive summary of a large collection of information [Lima 2011]; [Moretti 2007]; [van Geenen and Wieringa 2020]. Metaknowledge is important because it sparks abduction, or diagnostic-style logic, where experts may arrive intuitively at a conclusion based on partial information combined with their existing understanding of the system and situation at hand, and where knowledge may be generated instead of simply synthesized [Brooke 2017]; [Douven 2017]; [Graham, Milligan, and Weingart 2015].

While it may be preferable to gain metaknowledge the traditional way, by participating in a knowledge system over a

lifetime, this is not always practical or possible. Knowledge maps act as a tool which – just like regular maps of landscapes – are useful for sharing information, for planning where to go and how to get there, for noticing patterns or pathways that might not have been apparent to boots on the ground, and for coordinating collective effort in the face of complexity. They support switching between general knowledge and specific pieces of information and evidence, an ability which has been identified as a fundamental practice of critical scholarship [Edmond 2018]; [Moretti 2007]; [Kwan 2002]. Their utility comes even though knowledge offered by such a bird's eye view is clearly different from the kind gained through personal experience, immersion, and reflection [Evans and Foster 2011]; [Moretti 2007]. Essentially, it is worth simultaneously being cautious about the algorithmic ways in which knowledge maps synthesize big citation data and present it as knowledge, while also emphasizing the incredible utility of this form for producing, transmitting, and constructing collective knowledge [D'Ignazio and Klein 2016]; [D'Ignazio 2019]. Bibliometrician David A. Pendlebury summed this up effectively when he observed a need for "charting a path between the simple and false and the complex and unusable" [Pendlebury 2019, 549].

When constructed in partnership with residents from the mapped research domain, knowledge maps can prompt identification and reassessment of taken-for granted assumptions and highlight forces shaping problem definition in a field. In fact, examples of this work in climate domains are what initially inspired our attempt to map the climate communication research landscape [Canon, Boyle, and Hepworth 2022]. For example, informationist Christopher Belter and climate scientist Dian Seidel [Belter and Seidel 2013] created network visualizations to reveal (dis)connections in geoengineering knowledge, using sparse network charts to show how some planetary systems receive more focus than others, and that lack of information exchange between areas of study means there may be inadequate consideration of potential interactions between different geoengineering strategies. Another example is the work of psychologist Laura Kati Corlew and her team [Corlew et al. 2015], who successfully used social network analysis to build a knowledge sharing and resource discovery tool for climate professionals in the Pacific Islands.

8

These projects are notable because they leverage knowledge maps as communication and collaboration tools while being sensitive to audience needs and effects on those depicted in the visually powerful network charts. Rather than viewing analysis as an endpoint, these studies distill the information spaces into clear visualizations for interested audiences. Knowledge maps should be built from this perspective, to be both useful and usable and to assist viewers in extracting meaning from information spaces and analyses that may otherwise be esoterically complex.

9

3. Designing the climate communication knowledge map

The climate communication knowledge mapping project was initially situated in the tradition of science mapping, a kind of knowledge mapping undertaken by computer scientists and network analysts. Science mapping uses bibliometric data to reveal the structure of a knowledge domain by creating networks of papers or authors linked by citations or collaborations. However, despite having domain expertise in climate communication and fluency in network analysis techniques (often considered the ideal for this type of project), we immediately ran into ethical and practical challenges that were not well addressed in the science mapping literature, perhaps because too much emphasis is placed on the objectivity of network maps, the assumed impartiality of algorithmic curation, and the potential of data speaking for itself [Conway 2014]; [D'Ignazio 2019]; [D'Ignazio and Klein 2020]; [Hepworth 2020]; [White and McCain 1998]; [Yang et al. 2016]. In our project, limited technological resources necessitated static depictions of the knowledge network, data quality precluded purely algorithmic curation, and it was frequently unclear how fundamental tasks like disambiguation and layout should be approached when the goal was to create a useful knowledge map (instead of just to arrive at an analytical result). In this section, we describe in detail the challenges encountered in building the climate communication knowledge map and point out how attention to conversations in ethical visualization and data feminism provided needed guidance on navigating them.

10

Perhaps because of the impact of scientific approaches to visualization on knowledge mapping, a significant amount of data in knowledge mapping projects is not pre-processed before it is visualized. It is thought that network analysis itself controls for any errors or ambiguities in the underlying data [Chen 2016]. If ambiguities do exist in the network map, conventional wisdom holds that resulting irregularities cancel each other out, and overall conclusions are not affected [Fegley and Torvik 2013]; [White and McCain 1998]. We did not find this to be the case in our knowledge map,

11

especially when imagining how it might be useful and trustworthy for climate communicators. If the first thing a user notices when looking at a knowledge map is a data error (such as a single leading scholar being split into multiple author nodes), it cannot be expected that the user will trust other elements of the knowledge map. Given that the point of a knowledge map is to support reasonable inferences about the structure and function of the research domain, it is a serious problem if data issues affect interpretation. Engineer Jinseok Kim's research reveals the nature and extent of factors that distort knowledge maps [Kim, Kim, and Diesner 2014]; [Kim et al. 2014]; [Kim and Diesner 2016]; [Kim 2019]. For example, datasets rife with author name ambiguities appear *scale-free*,[3] a feature long thought to characterize citation networks [de Solla Price 1965], when in fact they aren't. Consequences to both the mapped research community and to theorists studying network processes can be profound when data visualizations uncritically display conclusions that may be based on irregularities in the data, and when funding decisions are made based on this data. As a result, the bulk of our paper focuses on data processing, and the ethical considerations that may be glossed over when the network analysis itself is the jumping off point for a knowledge mapping study.

In this section, we undertake the important data feminist approach of showing our work [D'Ignazio and Klein 2020]. We provide a complete description of how we produced the climate communication knowledge map, attempting to be transparent about elements of data preparation and visualization design that are not generally discussed in published literature, but which are of vital importance in following an ethical and accountable process of visualization design [Callaway et al. 2020]; [D'Ignazio and Klein 2020]; [Hepworth and Church 2018]. All described data management and network analysis was conducted in Python with metaknowledge [McLevey and McIlroy-Young 2017] and NetworkX [Hagberg, Schult, and Swart 2008]. Visualization was performed with Gephi [Bastian, Heymann, and Jacomy 2009].

3.1 Data search and selection

Data for the knowledge map came from Clarivate's Web of Science database, selected for its accessibility despite known gaps in its coverage and because previous work suggests that it entails about 50% as much cleaning and curation time as Scopus and about 5% as much as Google Scholar [Andres 2009]; [Harzing 2015]; [Meho and Yang 2007]. Though it would have been preferable to combine multiple datasets to get the broadest possible bibliometric survey of climate communication literature, there are significant practical challenges to combining relational citation data from different databases. We used a search query (TS = (climat* NEAR chang* AND communicat*)) from a previous systematic review of climate communication, conducted by an expert in the field [Moser 2016]. This was intended to ensure that this knowledge map complemented previously accepted delineations of the climate communication field and would therefore hopefully be more trustworthy to members of the mapped research domain. Query-based data selection for any knowledge mapping study faces a trade-off between *recall* (getting every single relevant item) and *relevance* (making sure no single irrelevant item is included), and it is always a concern that excluded materials are not evenly distributed across the field in question. The Web of Science search yielded 5,934 results on July 31, 2020.

One of the reasons data is not generally cleaned before analysis is the expectation that selecting and extracting the *giant component* (the largest set of connected nodes in a network) will filter papers that don't belong (because they'll be in disconnected communities). In the climate communication knowledge network, some combination of one-time participation by many researchers, the nebulous border between climate science, environmental science, and climate communication, and the broad application of some permutation of the terms "climate," "communication," and "change" in nearly every academic discipline meant that the giant component included many irrelevant items when constructed with the complete Web of Science data pull (Figure 1).

We explored programmatic filtering strategies to prune the irrelevant data, but these all failed, to the point that it seemed much more time would be spent fine-tuning an algorithm specifically suited to disambiguating this one dataset than it would take to disambiguate it manually. Manual (human) review resulted in the removal of ~50% of the returned records (2,997 retained; 2,937 removed). Though in some ways it seemed alarming to remove half of the returned data, it also seemed misleading not to do so, because clearly irrelevant items (for example, papers about inter-model communication of climate signals, effects of climate change on insect pheromonal communication, and changes in workplace communication climates) distorted the visualized network (Figure 1). This highlights the impossibility of

12

13

14

15

achieving an objective delineation of a particular research area from an unprocessed dataset and the need to rely on human judgment in data visualizations even at these early stages. This process of data preparation opens a knowledge mapping analysis up to perpetuating harm in two ways: first, that whether an algorithm or a human applies inclusion/exclusion criteria, some error is likely to occur, and second, that the underlying data does not adequately or equally represent the breadth of scholarship in a particular knowledge domain. To mitigate this, we added annotations to our final product (Figure 6) intended to guard against interpretation of the knowledge map as a complete picture of every possible research output in the climate communication space.
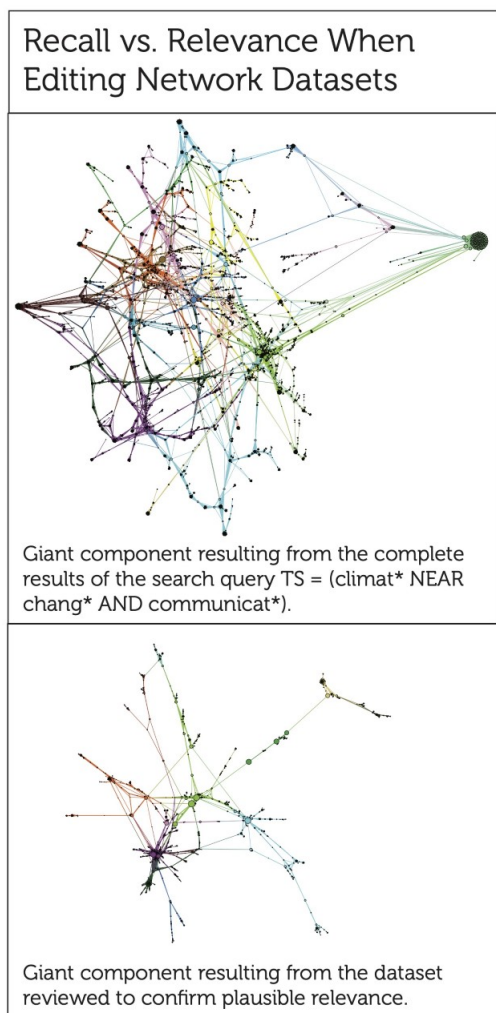

Figure 1. Recall vs. relevance when editing network datasets. An illustration of the trade-off between recall and relevance when including or excluding data for network analysis. The upper panel shows what the giant component looks like when all returns from a Web of Science query are included; the lower panel shows the same network's giant component after manual review was conducted to ascertain relevance of the returned data.

3.2 Author name disambiguation

Because we wanted to produce a knowledge map of collaboration patterns within climate communication scholarship, we projected the bibliometric data from Web of Science into a co-authorship network, where nodes are individual scholars and links are formed between them when they write papers together. To present a clear co-authorship network, it is necessary to disambiguate author names [Fegley and Torvik 2013]; [Harzing 2015]; [Milojevic 2013]; [Strotmann and Zhao 2012]. Co-authorship network analysis rests on the assumption that there is one node for everyone in the network, but in reality, names are frequently split (identifying a single author as multiple authors) or merged (collapsing multiple individuals into a single mega-researcher) [Harzing 2015]; [Milojevic 2013]. This can be due to inconsistent name spelling (for example the occasional omission of a middle initial), name change (for example due to a change in

marital status), or database limitations and errors. Merged names are such a problem in bibliometrics that many databases contain entries for mega-authors who appear to publish multiple papers every day for years on end [Harzing 2015]. Clearly, the process of author name disambiguation has enormous ramifications for whose contributions are legible or visible in a knowledge network, and whose are not. Recent efforts to use unique author identifiers such as ORCIDs may eventually solve this problem, but they are not yet used widely enough to offer a practical author disambiguation solution.

Author network studies wishing to disambiguate name data face a choice: implement algorithmic disambiguation strategies of varying complexity or undertake the time-consuming process of manually disambiguating the data. Simple algorithmic disambiguation is the norm, partly due to research suggesting that more computationally or time intensive approaches don't improve results [Milojevic 2013]. Manual disambiguation is rare, though not unheard of; some find the time investment comparable to designing an algorithm, at least for up to about five thousand authors [Burckhardt 2017]; [Fegley and Torvik 2013]; [Strotmann, Zhao, and Bubela 2009]. ¹⁷

The simplest algorithmic approaches are "first initial" disambiguation and "all initial" disambiguation. A pattern of initials is selected and all names are converted to match this pattern (for example, a full author name being converted to first initial and full last name). First initial disambiguation is the most common but also the most fraught: split individuals change network statistics less overall than do merged authors, and first initial disambiguation's weakness is that it merges many authors [Fegley and Torvik 2013]; [Strotmann and Zhao 2012]. When authors are merged, their actual networks combine with others, sometimes resulting in the appearance of bridges connecting areas of the network that are actually distinct and making the mega-author seem like an important synthesizer of knowledge at a level that may not reflect reality [Kim and Diesner 2016]. Networks containing many merged individuals become smaller, less transitive, denser, more productive, and more connected (or more frequently connected) than they actually are, and will have larger giant components [Kim et al. 2014]; [Kim and Diesner 2016]; [Fegley and Torvik 2013]. Initials-based disambiguation has been shown to introduce error into the entire gamut of statistics describing network structure and function, including those assessing author productivity, collaboration patterns, and cohesion [Kim and Diesner 2016]. Given the potential for harm inherent in miscalculating network statistics, both for evaluated individuals and for their broader knowledge community, we explored and visualized networks resulting from each of these disambiguation approaches so that we could make a clear determination of which strategy to use. ¹⁸
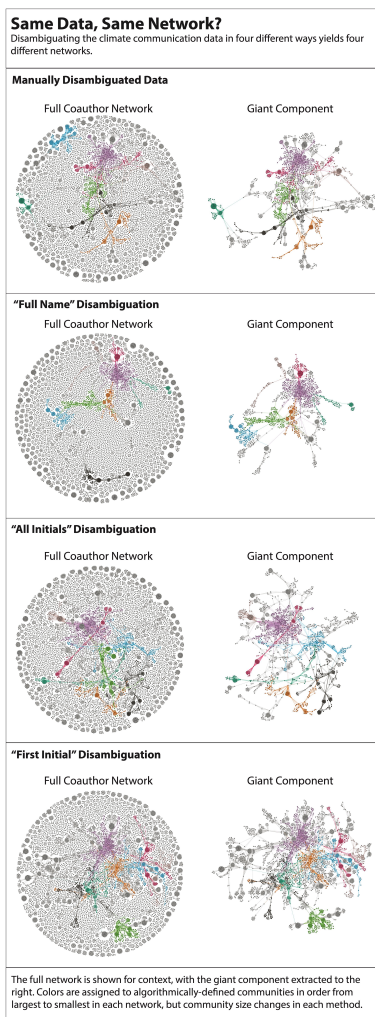
Same Data, Same Network?
Disambiguating the climate communication data in four different ways yields four different networks.

Manually Disambiguated Data

Full Coauthor Network   Giant Component

"Full Name" Disambiguation

Full Coauthor Network   Giant Component

"All Initials" Disambiguation

Full Coauthor Network   Giant Component

"First Initial" Disambiguation

Full Coauthor Network   Giant Component

The full network is shown for context, with the giant component extracted to the right. Colors are assigned to algorithmically-defined communities in order from largest to smallest in each network, but community size changes in each method.

Figure 2. Same data, same network? An illustration of how distortions to the network introduced by different disambiguation strategies look when visualized.

Very different pictures emerged when each of the four different disambiguation approaches was applied (Figure 2). Though the overall network is smaller, many more nodes are included in the "First Initial" giant component than in "Full Name," with "All Initials" falling somewhere in the middle. Table 1 provides quantitative support for these visual impressions. The number of authors contained in the giant component and the number of links connecting them changes significantly, by hundreds of individuals in the overall network and by over a thousand individuals in the associated giant components. Most notable in Figure 2 is the fact that a salient feature of the manually disambiguated network, the bridge that spans from the left to the right of the giant component without traveling through the core, is not as clear in any of the networks generated from algorithmic disambiguation. But each of the networks appears individually reasonable, both from a network standpoint and a climate communication one. This makes it clear how researchers could (wittingly or unintentionally) select a disambiguation strategy that conveys a particular message. In this case, the "First Initial" network suggests thriving exchange and dense connection, "All Initials" offers better-defined clusters, and "Full Name" presents neat divisions between communities with just a few bridges between. It's therefore important to be attentive to how interpretations may be influenced by visualization choices.

Entire Network	Manually Cleaned	All Initials	Full String	First Initial
Collection Length	2995	2995	2995	2995
Nodes (Authors)	7255	7402	7694	7061
Links (Collaborations)	23232	23503	23638	23272
Isolated Nodes	438	448	480	416
Giant Component	**Manually Cleaned**	**All Initials**	**Full String**	**First Initial**
Nodes (Authors)	1676	1604	1182	2730
Links (Collaborations)	7888	6667	4882	13659
Density	0.006	0.005	0.007	0.004
Transitivity	0.829	0.787	0.781	0.842
Assortativity	0.687	0.661	0.661	0.728
Average Clustering	0.823	0.818	0.808	0.829
Modularity	0.898	0.892	0.844	0.932

Table 1. Metrics describing the networks produced by four different disambiguations of the climate communication author network.

[20] Studies do not agree about which algorithmic disambiguation method is most effective at approximating the true network, though it's been shown that between 8-39% of individuals in a dataset can be affected by merging or splitting [Kim and Diesner 2016]; [Milojevic 2013]. Estimates to quantify the amount of error often use "First Initial" disambiguation as the lower limit of the range of error on the actual number of authors in the database, and , "All Initials" as the upper limit, but they do not usually visualize the resulting differences to see how more qualitative structural conclusions may be affected. Clearly, the success of a disambiguation strategy depends on meta-characteristics of the dataset itself. The most common example of this is the near-impossibility of using initials-based disambiguation on datasets with high participation from individuals with Chinese or Korean last names [Harzing 2015]; [Kim et al. 2014]; [Kim and Diesner 2016]; [Milojevic 2013]; [Strotmann and Zhao 2012].

[21] Even if it were possible to estimate the average distortion effects from different approaches to algorithmic disambiguation, it would be hard to predict which authors would be the most affected, meaning that analysis choices may have differential effects on different communities of scholars. Even if only a small percentage of error is introduced overall, that error may be distributed unevenly across the network, raising real concerns about whose and what types of contributions may be emphasized or obscured. Figure 3 shows how different authors' network positions are affected by different disambiguation strategies: node size changes (e.g., the node for S. Dessai), as does community membership (e.g., the node for S. Lewandowsky) and the visibility of underlying data errors (e.g., two nodes for E. Maibach linked together). The relative importance of nodes acting as brokers (large nodes here have high betweenness, one measure of a brokering role) changes too, and lumped or split authors open or close network pathways and affect community delineation. Table 2 shows how overall rankings of authors change order as nodes are merged or split.

[22] If the goal of the network analysis is to be useful to the people represented, this sort of error can't be waved away as acceptable distortion, and the only way to understand how such error affects an analysis is to visualize it and see. Though author disambiguation is rarely framed as an ethical issue, it clearly is one, as disambiguation strategies affect not only author rankings but the structural roles of authors in a network and the apparent structural function of the knowledge domain as a whole. In this case, we mitigated potential harm caused by inaccurate evaluations by building our final visualization on the manually disambiguated author dataset (though it's important to note, there are almost certainly errors even in this carefully curated dataset). However, most knowledge mapping projects will not have the benefit of a manually disambiguated dataset to compare to. In those cases, the best thing to do is understand the potential consequences of choosing different disambiguation strategies and avoid over-interpreting the results.

Manually Cleaned	All Initials	Full String	First Initial
Maibach, E	Leiserowitz, A	Leiserowitz, Anthony	Maibach, E
Leiserowitz, A	Maibach, E	Maibach, Edward	Leiserowitz, A
Pidgeon, N	Maibach, EW	Maibach, Edward W.	Pidgeon, N
Hart, PS	Hart, PS	Pidgeon, Nick	Hart, P
Moser, SC	Pidgeon, N	Nerlich, Brigitte	Moser, S
Schafer, MS	Moser, SC	Moser, Susanne C.	Nerlich, B
Nerlich, B	Nerlich, B	Hart, P. Sol	Myers, T
Myers, TA	Schafer, MS	Lewandowsky, Stephan	Lewandowsky, S
Lewandowsky, S	Lewandowsky, S	Roser-Renouf, Connie	Roser-Renouf, C

Table 2. Different rankings of top authors from four different disambiguations of the climate communication author network.
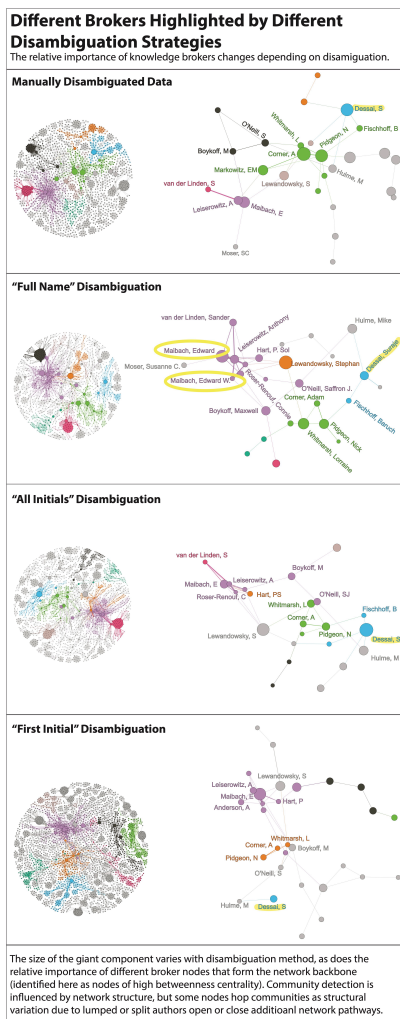
Figure 3. Different brokers highlighted by different disambiguation strategies. Here, nodes of high betweenness are extracted from the giant component produced by each disambiguation strategy applied to the same dataset. Node size indicates betweenness, or how many paths in the network pass through a single node. Larger nodes would be presumed to have brokering or gatekeeping functions.

3.3 Network layout

Several strategies have been developed for transforming networks into 2D visualizations [Leydesdorff 2014]; [Petrovich

2020]. As with geographic maps, all strategies simplify and therefore distort certain aspects of the underlying data, meaning they should be undertaken intentionally and transparently to mitigate potential harm [Monmonier 1991]; [Welles and Meirelles 2015]. In this study, we applied graph theory informed transformations, which use relationship strength to infer the distance between nodes [Leydesdorff and Rafols 2011]. Layout algorithms use physics calculations to position nodes optimally, by first distributing them randomly in space and then balancing the repulsion between them, the attraction caused by links, and the gravity of the center of the graph, until an equilibrium is reached [de Nooy, Mrvar, and Batagelj 2011]; [Fruchterman and Reingold 1991]. Because the resulting network maps have similar intuitive visual force to a geographic map, they are subject to being produced and accepted uncritically.

Both researchers and viewers can be misled by network layouts that suggest patterns where there are none [de Nooy, Mrvar, and Batagelj 2011]. For example, the *Force Atlas layout* allows longer edges between nodes than does the *Fruchterman Reingold layout*, which emphasizes cohesion [Leydesdorff and Rafols 2011]. Figure 4 illustrates how this changes the visual impression of the climate communication research network. Further complicating this is the possibility to mix and match layouts to achieve results that are different again from either layout method alone. Ultimately, it may be necessary to manually edit data to avoid misleading patterns such as criss-crossing links or overlapping nodes that could suggest incorrect inferences about network topology [Welles and Meirelles 2015]. However, manual edits may distort embedded visual conventions which are based on math, such as distance representing dissimilarity, and so again these should be undertaken cautiously [Stephens and Applen 2016].

24

For the climate communication knowledge map, we selected a layout produced by the Force Atlas algorithm and performed a small amount of editing to space out overlapping nodes and make the community size more interpretable to the human eye. In this case, our decision was less about mitigating potential harm and more about increasing understanding for viewers of the knowledge map while portraying the communities as legible groups of individuals. However, Figure 5 makes it clear that some layouts will emphasize connectivity and cohesiveness, and others will emphasize distance, despite that they are views of the same set of nodes and relationships. This illustrates that there is no option to choose an objective view of a network dataset, so the onus is on the researcher to make layout choices intentionally, as these arguments can have consequences for individuals and communities depicted. It also shows that just like datasets, tools for visualizing data can make implicit arguments which must be considered carefully.
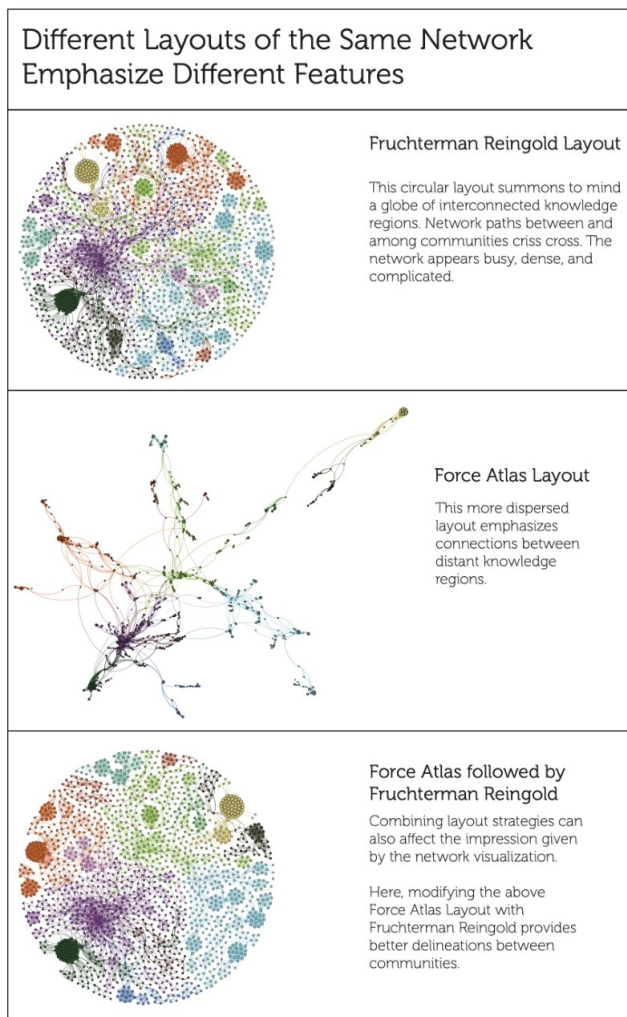
25

Different Layouts of the Same Network Emphasize Different Features

Fruchterman Reingold Layout

This circular layout summons to mind a globe of interconnected knowledge regions. Network paths between and among communities criss cross. The network appears busy, dense, and complicated.

Force Atlas Layout

This more dispersed layout emphasizes connections between distant knowledge regions.

Force Atlas followed by Fruchterman Reingold

Combining layout strategies can also affect the impression given by the network visualization.

Here, modifying the above Force Atlas Layout with Fruchterman Reingold provides better delineations between communities.

Figure 4. Different layouts of the same network emphasize different features. Three layouts of similar network data. The top is a Fruchterman Reingold Layout, the second is a Force Atlas Layout, and the third represents both layouts carried out in succession.

3.4 Community detection

After the network layout is complete, it is customary to divide the nodes into communities using algorithms such as Louvain community detection, a modularity maximization algorithm that identifies communities by finding groups of nodes sharing frequent connections within their group, but as few as possible connections to other communities [Blondel et al. 2008]. Even scientifically minded knowledge mappers tend to acknowledge that interpreting the results of a community detection algorithm is an iterative process, and as much an art as a science. Interpretation generally draws on non-network attributes (for example, researcher discipline) as well as network metrics like those in Table 1, to explicate the identified community structure. In the climate communication dataset, traditional strategies like summarizing purely by author keywords were minimally successful, as single occurrences of keywords were the norm (e.g., one primary community had 511 author-provided keywords describing 240 published works, but 388 of the keywords were used only once).

This illustrates the conundrum of categorization that is well-recognized in ethical visualization and data feminism: by placing nodes into groups, the knowledge map makes a strong argument that there is some inherent quality of those nodes that justifies their grouping, despite that the underlying data is not always so decisive. This is especially so when the network has been filtered to represent connections above a certain strength, such as authors that have collaborated at least twice, or papers that have been cited together at least ten times. This common practice eases the computational task of creating and analyzing the networks, but it also makes it easier to identify communities that seem much less interconnected than they actually are (or, conversely, give too much weight to connections which are strongly present

26

27

based on the visualized logic, but which may be essentially meaningless in the real world). Figure 5 demonstrates how clear communities become at different levels of filtering, offering another example of how seemingly technical choices made in knowledge network design make implicit arguments. Another way of saying this is that determinations of exactly what is signal and what is noise in a network dataset are subjective considerations.

In the climate communication knowledge map, we attempted to moderate the visual force of the communities by providing extensive annotations on their composition, and by resisting giving the three "core" communities particular names that were simply not justified by their heterogeneous compositions.
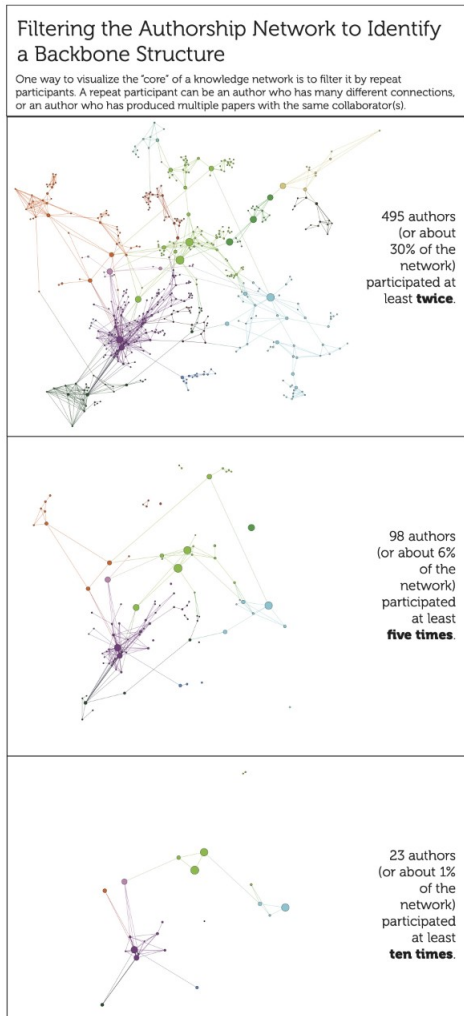
Figure 5. Filtering the authorship network to identify a backbone structure. Nodes here have been included in the network only if they participated a certain number of times.

3.5 Ground truth or trust?

Each of the previous sections has emphasized the constructedness of knowledge maps, demonstrating that every step in the production of a knowledge map entails subjective and value-laden judgments which foreground or obscure elements of the depicted knowledge domain and therefore run the risk of perpetuating different types of harm. Issues of data selection, disambiguation, layout, and community detection processes combine to make it difficult to validate or ground truth a network map. But it's worth asking, given the constructedness and subjectivity of a knowledge map, whether a ground truth would even be meaningful.

The traditional sense of a ground truth – going to a particular spot on a landscape and confirming that reality matches modeled values – is probably going to fail for knowledge maps. Though they should absolutely be intelligible to members of the depicted community, it is unlikely that every scholar would agree with the version of reality presented in

the chart. Knowledge maps approximate a consensus about knowledge structure based on authorship and citation data, but this is knowledge on average. Individual experts may or may not agree, and the knowledge map may not be a good representation of an individual's subjective experience of the knowledge domain, or even of the specific connections depicted based on their existing co-author relationships, since a network represents a diversity of connections as a single type of link. In the absence of a ground truth, perhaps a more important question to ask is how a network visualization can become trustworthy.

While building the climate communication knowledge map, we learned that the only way to produce trustworthy maps is to have attention to ethical dimensions and potentials for harm at every step of the design process. This can be accomplished with vigilance for implicit arguments made by customary analysis choices and tools, with the inclusion of annotations and caveats in the visualization, and by remaining constantly aware that any purely data-driven view of a knowledge domain will be partial. The key utility of ethical visualization perspectives in the development of the climate communication knowledge map was to push back against our worries that we were somehow biasing our results in an inappropriate way by making these human-driven choices surrounding data curation and presentation. In the next section, we offer suggestions for how to apply these perspectives in future knowledge mapping projects. `31`

4. Strategies for ethical and effective visualization of knowledge networks

Ethical and effective visualizations of knowledge networks must inspire trust in two key groups: people depicted in the knowledge map, and people using the map to acquire knowledge or make decisions. We draw on our experience in the climate communication knowledge mapping project and on previous work on ethical visualization in the digital humanities to offer ethical visualization strategies for network data and knowledge maps specifically. This represents an important step in " downscaling " recommendations from ethical visualization and data feminism to demonstrate their application and utility in specific projects. We encourage readers to consult other literature on ethical visualization (especially [Lima 2011] and [Gavrilova, Kudryavstev, and Grinberg 2019] for discussions of additional visual conventions specific to the design of network charts. `32`

Our recommendations are intended to assist researchers in balancing tensions between quantitative and qualitative, and objective and subjective, aspects of knowledge mapping. We advocate a humanistic approach to visualization design, recognizing that working with this type of data successfully requires coupling quantitative fluency with the more humanities-oriented practice of contextualizing findings, and tolerance for and transparency around the trial-and-error approach necessary to creating a useful product [Conway 2014]; [Dragga and Voss 2011]; [Drucker 2011]; [Hepworth and Church 2018]; [Stephens and Applen 2016]. As the climate communication knowledge mapping project illustrates, a purely quantitative, purely objective depiction of a knowledge network is neither possible nor particularly useful. `33`

4.1 Understand knowledge maps as both process and product

Knowledge maps have a dual purpose: they are a tool used in studying knowledge domains, and a visual artifact representing those domains. The first focuses primarily on creating knowledge and insight for the immediate researchers involved and can be best understood as a process. The second focuses primarily on assisting others in gaining insight into the mapped domain, and can be best understood as a product. However, as the climate communication knowledge map makes clear, these two types of knowledge mapping are generally impossible to separate: knowledge mapping proceeds in an iterative and exploratory fashion, where impressions gained in experimenting with different analyses and layouts inform the final version of the map, and where the layout of the map informs a researcher's conclusions. Franco Moretti described this as a "heterogeneity of problem and solution" in distant reading; we argue that the same applies to knowledge mapping [Moretti 2007]. In the case of the climate communication knowledge map, the desire to create a particular kind of product (a useful and trustworthy knowledge map to support collaboration and discovery in a particular knowledge community) influenced the process of data preparation and analysis, for example leading to the decision to manually disambiguate data, which in turn affected analytical results. `34`

The intertwining of product and process does not guarantee that a good product will follow from a sound process (or `35`

that an apparently good product was generated with a sound process). For example, we often observe knowledge maps developed primarily to facilitate analysis being included in research reports. In these cases, researchers have generally extracted meaning from the visualization to support their interpretations. However, without adjustment, these visualizations often remain incomprehensible to those outside the research team. We suggest that ethical and effective knowledge mappers should remain aware of how process and product intertwine in a knowledge mapping project, and how this can both enhance and trouble the process of producing a knowledge map for use beyond the research team.

4.2 Translate and annotate knowledge maps

Anecdotally, we learned while developing and sharing the climate communication knowledge maps that despite their visual and rhetorical force, viewers are often unsure how to understand these charts. We also learned that, once viewers felt they understood what the nodes and connections and color scheme meant, they quickly formed and held onto inaccurate impressions about what the chart conveyed. While it is not possible to moderate the rhetorical force of a visualization entirely, it is possible to embed guidelines and caveats for interpretation that guide viewers towards the interpretations intended by the visualization designer. This does not preclude the viewer reaching their own conclusions; rather, it fulfills the promise of a knowledge map as a communication tool. Ethical and effective knowledge mappers should take the opportunity to translate and annotate a knowledge map, asking themselves which insights have the most rhetorical force, and which may need to be moderated with caveats. This is why the knowledge maps presented in this paper include textual information about their genesis and interpretation.

4.3 Account for implicit arguments made by network theory, tools, and data sources

A common practice in data feminism is to disclose the positionality of the researchers, so that viewers of the resulting data visualizations might understand how researchers' perspectives and life experiences may influence their understandings of the phenomena portrayed. We encourage knowledge mappers to consider not just their own positionality, but the positionality of network theory, tools, and data sources when drawing conclusions from a knowledge mapping project.

First, it is necessary to consider the implicit arguments made by networks themselves. A network chart necessarily conveys a message of connection, especially since unconnected elements are generally omitted from analysis. Network thinking is fundamentally structuralist, meaning that it privileges relationships over attributes and affords the system a level of agency that it may or may not have [Conway 2014]; [Gochenour 2011]; [Moretti 2007]. A consequence of this structuralist lens is that disconnection is seen as undesirable, and distinctions between quality and quantity are often blurred [Gochenour 2011]. The focus on quantitatively (but possibly not qualitatively) strong connections (after all, there are many ways to become connected in a coauthor network that do not necessarily represent close collaboration) extends to the way networks are generally depicted, showing nodes connected only above a certain link strength threshold (as in Figure 5). The assumption that weak connections are just noise to be removed, and that strong connections are the defining structural feature of a network, means this type of analysis may miss cutting-edge or non-standard knowledge patterns.

Subsequently, it is important to disclose the tools used for analysis and, especially, for visualization. Because the visualization tools actively shape the conclusions of the analysis (as discussed in 4.1), it is important to be transparent about how this may have occurred even if the effects are not apparent to the research team [van Geenen and Wieringa 2020]. None of this threatens the utility of network analysis as a tool for understanding a system of related data, but it should serve as a reminder of network thinking's assumptions, so they are not incorporated uncritically into a knowledge map. An ethical and effective knowledge mapper should be aware of inherent implications stemming from choices about the theory, tools, and data sources which feed into resulting visualizations.

4.4 Consider impacts on people depicted

Knowledge networks are often assumed to be exempt from privacy concerns because they are constructed from published works. It's important to remember that a network (especially the co-author networks presented in this paper) is made up of individual people. Certain nodes may assume standout roles that wouldn't be apparent in a list of search

results or even in tabular data, and certain patterns of connection may occur that do not represent current relationships and collaborations. Because there is no objective way to characterize node roles or performance, or even which nodes truly "belong" in a network, caution is warranted in "calling out" specific people depicted, or associating merit with network position alone [Furner 2014]. It may sometimes be preferable to display anonymized data (for example by grouping nodes into communities) to guard against incorrect or harmful interpretations about individuals.

On the other hand, it's precisely the transparency and associated lack of anonymity that can make knowledge network tools useful. Laura Kati Corlew and her team relied on this in building a network mapping and knowledge discovery tool for climate workers in the Pacific Islands [Corlew et al. 2015]. The correct balance of privacy and publicity will need to be determined on a project-by-project basis. Asking questions about who benefits from the network depiction of participants can help navigate these tensions [Conway 2014]. In the case of the climate communication knowledge map, we felt that attention to data selection and disambiguation mitigated some concerns about privacy for network participants and so we labeled key nodes with names. Climate communicators are likely familiar with these scholars and can use them as landmarks to support orientation and sense-making while they interact with the knowledge map. 41

As knowledge maps proliferate, they may be used increasingly for evaluation and analysis, and so an ethical and effective knowledge mapper should consider the potential ramifications of such a map having a life of its own after publication, and the possible agendas users might bring to interpretation of these visualizations, especially when choosing to use personally identifying data in a knowledge map. 42

4.5 Put your cards on the table

The final strategy for ethical and effective visualization of knowledge networks is inspired by data feminism's maxim of showing one's work [D'Ignazio and Klein 2020]. We encourage ethical and effective knowledge mappers to put their cards on the table alongside their produced visualizations. This means being completely transparent about what you did, why you did it, and how you did it, but also about what insights you identify or arguments you wish to make with your produced visualization. Work in ethical visualization and data feminism has shown real drawbacks to obscuring the work done to produce a data visualization, to attempting to let data speak for itself. Ethical and effective knowledge mappers must resist the temptation to keep analysis and interpretation behind the curtain, recognizing that doing so may hinder precisely the type of knowledge creation and discovery they are likely trying to support. 43

The first way to put your cards on the table is to make analysis decisions transparent, as we have done in this article. This may be done in scholarly publications, or in metadata and annotations accompanying published visualizations. Like disclosing positionalities, articulating the strategies used to create visualization products reveals implicit impressions the designer may not have been aware of, but which could be important to viewers. Sharing this information safeguards transparency by aiding other researchers in creating similar maps, should they wish to. 44

The second way to put your cards on the table is to tell the story you as the researcher see in the visualization. What insight were you seeking that led you to build the knowledge map? Did you find evidence for it, or against it? Narratives explicate network structure and concepts in an intuitive way, guiding viewers to reasonable scales of interpretation and guarding against out-of-context presentations of the knowledge network. They can also shed light on what the system diagram on its own might not reveal clearly, such as a possible function of or reason for absent links between authors or communities. 45

There is a fine line to walk here, given that the point of visual communication is simplicity. Over-annotating a visualization might make it harder to engage with. We attempted to address this in our final knowledge map (Figure 6) by providing a below-the-fold presentation of narrative context for the network. Other network projects, especially interactive ones, may discover additional ways to support the juxtaposition of narrative and network. 46

Figure 6. A fully annotated, final version of the climate communication knowledge map. The key content is provided "above the fold," with additional metadata about the analysis and the depicted knowledge communities included below to support construction of alternate interpretations of the knowledge map.

5. Conclusion

We began this paper with the observation that scholars have long been captivated by the quest to develop high-tech tools for information management and knowledge synthesis. From the vantage point of the 21st century, we recognize that the success of these tools is not simply a matter of technological progress, but one that requires ethical thought. Just as a network mapping platform like Gephi provides the technical tools for working with big network data, an ethical visualization framework provides the ethical tools for reducing the harm and increasing the impact of knowledge maps – essentially, for making them deliver on their promise of augmenting our perceptual abilities when faced with large and complex information spaces. In this paper, we have drawn on lessons learned in the climate communication knowledge mapping project to offer strategies for ethical and effective visualizations of knowledge networks. This approach acknowledges the potential for harm that comes with such visualizations and works actively to mitigate it. Taking an ethical approach to knowledge mapping is vital because the choices made in such an analysis determine whose and which contributions are ultimately legible.

We have attempted to provide guidelines for knowledge mappers seeking to support users in synthesizing and selecting meaningful insights from the mapped domain. Our recommendations balance tensions between qualitative and quantitative and objective and subjective aspects of knowledge mapping. They demonstrate the importance of critical practices in the development of knowledge maps, illustrate the intertwined nature of analysis and results in such projects, and emphasize the constructedness of the visualization. We show the only way to produce an effective knowledge map is to produce an ethical one, which requires attention to the ways trust and accountability can be

produced at every step of analysis and production.

This framework for ethical visualization of knowledge networks contributes to early conversations about a feminist bibliometrics and demonstrates the direct relevance of digital humanities tools to work carried out across diverse other fields. Connecting bibliometrics literature with digital humanities discourse invites more qualitative and interpretive perspectives into bibliometric knowledge mapping, demonstrating that the data-centric and science-focused techniques frequently applied to knowledge network visualization may fail to produce a useful and impactful portrait of a knowledge domain without attending to these ethical concerns. A central purpose of this paper is to offer a clear example of the utility of a critical approach to knowledge mapping in a traditional knowledge mapping project.

49

The idea of separating wheat from chaff is a common trope in knowledge mapping. While knowledge mapping can absolutely facilitate discovery of relevant or applicable knowledge, assuming this is its key strength overlooks its real potential. Essentially, a system-level depiction may provoke a reassessment of what is wheat and what is chaff, and recognition that these designations are situation-dependent. As knowledge production increases and we continue to rely on quantitative and visual tools to help us navigate information spaces, it may be best to abandon the threshing trope and instead adopt a more kaleidoscopic model of knowledge mapping's goals. Knowledge maps provide meaningful and multipurpose ways for individuals to explore a knowledge domain, and different individuals may need to turn the kaleidoscope in unique ways to find useful or compelling information and explanations which gel with, and usefully complement, their own experiences of a knowledge network.

50

Notes

[1] An external brain for forming and tracking relational knowledge [Bush 1945].

[2] A macroscope for harnessing disparate data and visualizing it on a global surface, to support problem definition and problem solving [Fuller 1981].

[3] *Scale-free* means that the structural patterns evident in the network persist regardless of the size of the network. This is often explained by a process called preferential attachment, where newcomers to the network tend to connect to existing network members who already have a high number of connections (an example of a "rich-get-richer" phenomenon). In the example of a citation network, certain papers accumulate high numbers of citations while others are never cited.

Works Cited

Andres 2009 Andres, A. *Measuring Academic Research: How to Undertake a Bibliometric Study.* Chandos Publishing, Oxford(2009).

Bastian, Heymann, and Jacomy 2009 Bastian, M., Heymann, S., and Jacomy, M. "Gephi: An Open Source Software for Exploring and Manipulating Networks ". *International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media* , San Jose, California, May 2009.

Belter and Seidel 2013 Belter, C. W. and Seidel, D. J. "A Bibliometric Analysis of Climate Engineering Research ", WIREs Climate Change, 4.5 (2013): 417–427.

Blondel et al. 2008 Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. "Fast Unfolding of Communities in Large Networks " , *Journal of Statistical Mechanics: Theory and Experiment* , 10 (2008): P10008.

Bornmann 2017 Bornmann, L., "Measuring impact in research evaluations: a thorough discussion of methods for, effects of and problems with impact measurements," *Higher Education*, 73 (2017): 775-787.

Brooke 2017 Brooke, C. "Abduction, Writing, Digital Humanities." In S. I. Dobrin and K. Jensen (eds), *Abduction Writing Studies*, Carbondale, IL (2017), pp. 163–180.

Burckhardt 2017 Burckhardt, D. "Comparing Disciplinary Patterns: Exploring the Humanities through the Lens of Scholarly Communication."*DHQ: Digital Humanities Quarterly*, 11.2 (2017).

Bush 1945 Bush, V. "As We May Think", The Atlantic Monthly, 176.1 (1945): 101–108.

Börner 2010 Börner, K. *Atlas of Science: Visualizing What We Know* . MIT Press, Boston (2010).

Börner 2011 Börner, K., "Plug-and-Play Macroscopes ", Communications of the ACM, 54.3 (2011): 60-69.

Cairo 2014 Cairo, A., "Ethical infographics: In data visualization, journalism meets engineering," *The IRE Journal* (Spring 2014): 25–27.

Callaway et al. 2020 Callaway, E., Turner, J., Stone, H., and Halstrom, A. "The Push and Pull of Digital Humanities: Topic Modeling the 'What is digital humanities?' Genre", *DHQ: Digital Humanities Quarterly*, 14.1 (2020).

Canon, Boyle, and Hepworth 2022 Canon, C. R., Boyle, D. P., and Hepworth, K. J. "Mapping pathways to public understanding of climate science," *Public Understanding of Science* March 2022. doi:10.1177/09636625221079149.

Chen 2016 Chen, C. *CiteSpace: A Practical Guide for Mapping Scientific Literature*. Nova Science Publishers, New York (2016).

Conway 2014 Conway, S. "A Cautionary Note on Data Inputs and Visual Outputs in Social Network Analysis", *British Journal of Management*, 25.1 (2014): 102–117.

Corlew et al. 2015 Corlew, L. K., Keener, V., Finucane, M., Brewington, L., and Nunn-Chrichton, R. "Using Social Network Analysis to Assess Communications and Develop Networking Tools Among Climate Change Professionals Across the Pacific Islands Region", *Psychosocial Intervention*, 24.3 (2015): 133–146.

de Nooy, Mrvar, and Batagelj 2011 de Nooy, W., Mrvar, A. and Batagelj, V. *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, Cambridge (2011).

de Solla Price 1965 de Solla Price, D. J. "Networks of Scientific Papers", Science, 149.3683 (1965): 510–515.

D'Ignazio and Klein 2016 D'Ignazio, C. and Klein, F. "Feminist data visualization," Workshop on Visualization for the Digital Humanities (VIS4DH), Baltimore, MD, IEEE (2016).

D'Ignazio and Klein 2020 D'Ignazio, C. and Klein, F. *Data Feminism*. The MIT Press, Cambridge (2020).

Donovan 2019 Donovan, C. "Do we need a feminist bibliometrics?" PowerPoint presentation. Available at: https://www.researchcghe.org/events/cghe-seminar/do-we-need-a-feminist-bibliometrics/. Accessed 28 February 2022. (2019).

Douven 2017 Douven, I. "Peirce on Abduction." In E.N. Zalta (ed), *Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/entries/abduction/peirce.html .

Dragga and Voss 2011 Dragga, S. and Voss, D. "Cruel Pies: The Inhumanity of Technical Illustrations", *Technical Communication*, 48.3 (2001): 265- 274.

Drucker 2011 Drucker, J. "Humanities Approaches to Graphical Display", *DHQ: Digital Humanities Quarterly*, 5.1 (2011).

D'Ignazio 2019 D'Ignazio, C. "Data Visualization." In R. Hobbs and P. Mihailidis (eds), *The International Encyclopedia of Media Literacy*, (2019), pp. 1–10.

Edmond 2018 Edmond, J. "How Scholars Read Now: When the Signal Is the Noise", *DHQ: Digital Humanities Quarterly*, 12.1 (2018).

Evans and Foster 2011 Evans, J. A. and Foster, J. G. "Metaknowledge", *Science*, 331.6018 (2011): 721–725.

Fegley and Torvik 2013 Fegley, B. D. and Torvik, V. I. "Has Large-Scale Named-Entity Network Analysis Been Resting on a Flawed Assumption?", *PLOS ONE*, 8.7 (2013): e70299.

Fruchterman and Reingold 1991 Fruchterman, T. M. J. and Reingold, E. M. "Graph Drawing by Force-Directed Placement", *Software: Practice and Experience*, 21.11 (1991): 1129–1164.

Fuller 1981 Fuller, R. B. (1981) *Critical Path* (2nd ed). St Martin's Griffin Press, New York (1981)

Furner 2014 Furner, J. "Ethics of Evaluative Bibliometrics." In B. Cronin and C. R. Sugimoto (eds), Beyond Bibliometrics, Boston (2014), 85–107.

Gavrilova, Kudryavstev, and Grinberg 2019 Gavrilova, T., Kudryavtsev, D. and Grinberg, E. "Aesthetic Knowledge Diagrams: Bridging Understanding and Communication." In M. Handzic and D. Carlucci (eds), *Knowledge Management, Arts, and Humanities: Interdisciplinary Approaches and the Benefits of Collaboration*, Cham (2019), pp. 97–117.

Gochenour 2011 Gochenour, P. "Nodalism", *DHQ: Digital Humanities Quarterly*, 5.3 (2011).

Graham, Milligan, and Weingart 2015 Graham, S., Milligan, I., and Weingart, S. *Exploring Big Historical Data: The Historian's Macroscope*. Imperial College Press, London (2015).

Hagberg, Schult, and Swart 2008 Hagberg, A. A., Schult, D. A. and Swart, P. J. "Exploring Network Structure, Dynamics, and Function Using NetworkX." In G. Varoquaux, T. Vaught, and J. Millman (eds), *Proceedings of the 7th Python in Science Conference*, Pasadena (2008): pp. 11-15.

Haraway 1988 Haraway, D. "Situated knowledges: the science question in feminism and the privilege of partial perspective", *Feminist Studies*, 14.3 (1988): 575-599.

Harzing 2015 Harzing, A. W. "Health Warning: Might Contain Multiple Personalities — the Problem of Homonyms in Thomson Reuters Essential Science Indicators", *Scientometrics*, 105.3 (2015): 2259–2270.

Hepworth 2017 Hepworth, K.J. "Big data visualization," *Communication Design Quarterly*, 4.4 (2017): 7–19.

Hepworth 2020 Hepworth, K.J. "Make Me Care: Ethical Visualization for Impact in the Sciences and Data Sciences." In A. Marcus and E. Rosenzweig (eds), *Design, User Experience, and Usability. Interaction Design*, Cham (2020), pp. 385-404.

Hepworth and Church 2018 Hepworth, K. and Church, C. "Racism in the Machine: Visualization Ethics in Digital Humanities Projects", *DHQ: Digital Humanities Quarterly*, 12.4 (2018).

Kim 2019 Kim, J. "Scale-free Collaboration Networks: An Author Name Disambiguation Perspective", *Journal of the Association for Information Science and Technology*, 70.7 (2019): 685–700.

Kim and Diesner 2016 Kim, J. and Diesner, J. "Distortive Effects of Initial-based Name Disambiguation on Measurements of Large-scale Coauthorship Networks", *Journal of the Association for Information Science and Technology*, 67.6 (2016): 1446–1461.

Kim et al. 2014 Kim, J., Diesner, J., Aleyasen, A., Kim, H., and Kim, H.-M. "Why Name Ambiguity Resolution Matters for Scholarly Big Data Research" IEEE International Conference on Big Data, Washington, DC, October 2014.

Kim, Kim, and Diesner 2014 Kim, J., Kim, H. and Diesner, J. "The Impact of Name Ambiguity on Properties of Coauthorship Networks", *Journal of Information Science Theory and Practice*, 2.2 (2014): 6–15.

Kwan 2002 Kwan, M. "Feminist visualization: re-envisioning GIS as a method in feminist geographic research", *Annals of the Association of American Geographers*, 92.4 (2002): 645-661.

Leydesdorff 2014 Leydesdorff, L. "Science Visualization and Discursive Knowledge" In B. Cronin and C. R. Sugimoto (eds), *Beyond Bibliometrics*, Boston (2014), pp. 167–185.

Leydesdorff and Rafols 2011 Leydesdorff, L. and Rafols, I. "Indicators of the Interdisciplinarity of Journals: Diversity, Centrality, and Citations", *Journal of Informetrics*, 5.1 (2011): 87–100.

Lima 2011 Lima, M. *Visual Complexity* Princeton Architectural Press, New York (2011).

McLevey and McIlroy-Young 2017 McLevey, J. and McIlro. Princeton Archy-Young, R. "Introducing metaknowledge: Software for Computational Research in Information Science, Network Analysis, and Science of Science", *Journal of Informetrics*, 11.1 (2017): 176–197.

Meho and Yang 2007 Meho, L. I. and Yang, K. "Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science Versus Scopus and Google Scholar", *Journal of the American Society for Information Science and Technology*, 58.13 (2007): 2105–2125.

Milojevic 2013 Milojevic, S. "Accuracy of Simple, Initials-based Methods for Author Name Disambiguation", *Journal of Informetrics*, 7.4 (2013): 767–773.

Monmonier 1991 Monmonier, *M. How to Lie with Maps*. University of Chicago Press, Chicago (1991).

Moretti 2007 Moretti, F. *Graphs, Maps, Trees: Abstract Models for Literary History*. Verso, New York (2007).

Moser 2016 Moser, S. C. "Reflections on Climate Change Communication Research and Practice in the Second Decade of the 21st Century: What More is There to Say?", *WIREs Climate Change*, 7.3 (2016): 345–369.

Pendlebury 2019 Pendlebury, D. "Charting a path between the simple and false and the complex and unusable: review of Henk F. Moed, Applied Evaluative Informetrics", *Scientometrics*, 119.1 (2019): 549-560.

Petrovich 2020 Petrovich, E. " Science Mapping", *ISKO Encyclopedia of Knowledge Organization*, (2020).

Rolf 2021 Rolf, H. "Navigating power in doctoral publishing: a data feminist approach", *Teaching in Higher Education*, 26.3 (2021): 488-507.

Stephens and Applen 2016 Stephens, S. and Applen, J. D. "Rhetorical Dimensions of Social Network Analysis Visualization for Public Health." IEEE International Professional Communication Conference, Austin, Texas, October 2016.

Strotmann and Zhao 2012 Strotmann, A. and Zhao, D. "Author Name Disambiguation: What Difference Does it Make in Author-based Citation Analysis?", *Journal of the American Society for Information Science & Technology*, 63.9 (2012): 1820–1833.

Strotmann, Zhao, and Bubela 2009 Strotmann, A., Zhao, D. and Bubela, T. "Author Name Disambiguation for Collaboration Network Analysis and Visualization", Proceedings of the American Society for Information Science and Technology, 46.1 (2009): 1–20.

van Geenen and Wieringa 2020 van Geenen, D. and Wieringa, M. In M. Engerbretsen and H. Kennedy (eds), *Data Visualization in Society*, Amsterdam (2020), pp. 141–156.

Welles and Meirelles 2015 Welles, B. F. and Meirelles, I. "Visualizing Computational Social Science: The Multiple Lives of a Complex Image", *Science Communication*, 37.1 (2015): 34–58.

Wenger 1998 Wenger, E. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, New York (1998).

White and McCain 1998 White, H. D. and McCain, K. W. "Visualizing a Discipline: An Author Co-citation Analysis of Information Science, 1972 - 1995", *Journal of the American Society for Information Science & Technology*, 49.4 (1998): 327–355.

Yang et al. 2016 Yang, S., Han, R., Wolfram, D., and Zhao, Y. "Visualizing the Intellectual Structure of Information Science (2006–2015): Introducing Author Keyword Coupling Analysis", *Journal of Informetrics*, 10.1 (2016): 132–150.

Zuccala 2016 Zuccala, A. "Inciting the metric oriented humanist: teaching bibliometrics in a faculty of humanities", *Education for Information*, 32.2 (2016): 149–164.