Journal of Communication ISSN 0021-9916

ORIGINAL ARTICLE

A Collaborative Way of Knowing: Bridging Computational Communication Research and Grounded Theory Ethnography

Yotam Ophir[1,*], Dror Walter[2,*], & Eleanor R. Marchant[3,*]

1 Department of Communication, University at Buffalo—The State University of New York, Buffalo, NY 14260, USA

2 Department of Communication, Georgia State University, Atlanta, GA, USA

3 Department Programme in Comparative Media Law & Policy, University of Oxford, Oxford, UK

One of the great strengths of the field of communication is its interdisciplinarity. Yet this strength brings challenges, including rifts between diverse subfields. In this study, we illustrate the rich potential of collaborations across subfields. Specifically, we argue that due to often-overlooked epistemological similarities, unsupervised machine learning and grounded theory ethnography subfields are well-suited for an especially enriching collaboration. To demonstrate, a team of computational and ethnographic researchers together applied the analysis of topic model networks approach to ethnographic field notes. We illustrate how the inclusion of the ethnographer in the modeling stages, and of the computational researchers in the analysis stages, led to mutual reflexivity affecting every stage of the study, enabling profound reflections on the technical, conceptual, and theoretical pillars of both subfields. We conclude by discussing the potential future of such collaborative ways of knowing to open doors for cutting-edge interdisciplinary research for the new information era.

Keywords: Epistemology, Ethnography, Computational Methods, Interdisciplinary Research, Computational Ethnography, Reflexivity

doi: 10.1093/joc/jqaa013

One of the great strengths of the field of communication is its interdisciplinarity, drawing from fields as disparate as anthropology, psychology, computer science, and law to answer questions about the nature of media and technologies in society. Yet this same strength has often created divisions within the field, including around how knowledge should be produced. Bridging these divides should be, on paper at least, a productive scholarly endeavor (Delli-Carpini, 2013). However, in reality,

Corresponding author: Dr. Yotam Ophir; e-mail: yotamoph@buffalo.edu
*All authors contributed equally to this study.

the subfields of communication research, arranged around topics (e.g., health and political communication) and methods (e.g., ethnography and computational communication research) seem to have grown apart in recent decades. This widening divide can be seen in a variety of examples, from researchers publishing in different journals and attending separate conferences, to maintaining memberships at different associations and different interest groups and divisions within associations (Anderson & Baym, 2004). For example, the subfield of computational communication research is represented in the International Communication Association in a Computational Method interest group, which also operates its own journal. Similarly, the subfield of ethnography has its own division at the National Communication Association, and its members often publish in specialized journals. It is not an uncommon practice for communication researchers to attend only one of the associations' conferences, and to focus their time in reading articles from the most "relevant" journals. These divides extend into many communication departments, many of which have different labs and subdivisions for different subfields. Furthermore, the methodological training necessary for students to develop expertise in communication subfields often prematurely divides them along methodological lines in ways that discourage the adoption of mixed-methods approaches.

Yet, this narrative of *difference* between the subfields often overshadows important existing *similarities*—in shared knowledge, values, beliefs, goals, and domains of interest—diminishing the potential for cooperation and collaboration that is urgently needed when dealing with modern communication questions. We believe that attending to these similarities, particularly the hidden ones and those that connect particularly disparate subfields, is a necessary first step in helping to overcome differences. More specifically, we propose that two seemingly incongruent subfields—computational communication research and ethnography––share important commonalities that create a unique opening to insightful collaborations.

In this study, we show the ways in which a computational approach to the analysis of texts—the analysis of topic model networks (Walter & Ophir, 2019)—is in fact well-suited for analyzing complex ethnographic data. Importantly, in contrast with studies that applied preexisting computational methods to ethnographic data, our work was fully collaborative. The ethnographer took part in all preprocessing and modeling stages, and the computational researchers actively participated in the analysis and interpretation of the results. Working together, this procedure encouraged greater reflexivity on both sides, allowing the researchers to revisit, reconsider, and modify the assumptions they made in their work. This reflexivity allowed the generation of new insight into the case study analyzed, and into the nature of the computational and ethnographic methods themselves.

Starting from difference: The philosophical and methodological pillars of the subfields

The contemporary academic system is built around historically institutionalized paradigms, with notable divides between broad disciplines like humanities, social science, and natural science (Kuhn, 1962). Communication as a discipline has long resisted pressure to call just one of these disciplines home; many communication scholars work at the boundaries of the first two, and recent developments in the study of brain imagining and physiological reactions to media opened the door for the third as well (Falk et al., 2016).

Tensions between subfields, defined through shared interests in topics (e.g., health and political communication) and methodological approaches (e.g., ethnography and computational methods), have long played a role in the communication discipline (Katz & Katz, 2016). As explained by Anderson and Baym (2004), while some debates have focused on normative and ethical values (axiology) and the nature of the media itself and its impact on society and individuals (ontology), others evolved around questions of what knowledge about communication means (epistemology), and with what methods it should be obtained (praxeology). Researchers often referred to and self-defined as "quantitative" may begin, for the sake of data collection and analysis, from the assumption that the objects of study exist independently of our perceptions of them, representing a unique truth that is discoverable, able to be studied relatively objectively (Anderson & Baym, 2004). On the other hand, researchers often referred to as "qualitative" often follow a social constructivist epistemology, believing research objects are constructed by social contexts in which they are observed and experienced; they believe that the objects of study are interpreted through, and shaped by, the unique subjective lens that every researcher brings with them to the field (Berger & Luckmann, 1966). Of course, those descriptions are theoretical archetypes; in reality, many researchers operate between the extremes.

Within the field of communication, one of the methods closely associated with the social constructivist epistemology is ethnography (Wagner, 1981); while there are certainly ethnographers, like Abramson and Dohan, whose epistemological approach is more akin to "typical" quantitative research (see Abramson, Joslyn, Rendle, Garrett, & Dohan, 2018), most ethnographers believe that their subjectivities will shape what they find in the field. Developed by early anthropologists like Malinowski (1926) to understand the holistic lived experiences of "exotic" cultural communities, ethnography has historically been defined by extended periods of "participant observation" in communities under study, and by the writing of regular field notes, to record the researchers' observations about the community and to help bring the researcher's own subjectivities into relief (Emerson, Fretz, & Shaw, 2011). As we will demonstrate, many modern ethnographers—for example, those who study the Internet—have embraced computational tools, mostly for the collection and analysis of data.

This process of embracing computational tools occurred at a time when the use of computational methods for the analysis of text had grown rapidly among communication scholars who were looking for novel tools that could cope with the sharp increase in the quantity of available digital data. Importantly, computational methods for text analysis tend to be used predominantly in quantitative contexts, often scoring texts using supervised machine learning methods; as a result, it is often perceived to be far removed from the constructivist approach. These perceived differences often resulted in tensions among qualitative researchers who worried that the adoption of computer-assisted quantitative methods could oversimplify the analysis of complex textual phenomena, obscuring the deep human element necessary for ethnographic analysis. Such tensions were not unique to communication research, and were apparent in other fields, including digital humanities (Gottschall, 2008) or sociology (Nelson, 2017).

Finding common ground: The intersection of machine learning and ethnography

While machine learning approaches to text analysis and ethnography might seem like they fall at quite different ends of the disciplinary divide within communication, they have a surprising amount of *epistemological* and *historical* common ground.

Epistemological common ground

Certain kinds of quantitative and qualitative methods subscribe to the *deductive* approach to research, beginning with theoretically driven hypotheses before turning to the data (i.e., category-applying research). However, both unsupervised machine learning and a grounded theory approach to ethnography are *inductive*, category-forming, empirical, and data-driven (Chen, Drouhard, Kocielnik, Suh, & Aragon, 2018; Walter & Ophir, 2019). Unsupervised machine learning uses the construction of algorithms to identify latent structures from observed data (Blei, Ng, & Jordan, 2003). In grounded theory approaches, ethnographers create meticulous open codes of data gathered in the field, with which they build theory to explain their observations (Glaser & Strauss, 1999). While even among grounded theorists there are divergent approaches to performing this coding process in practice (Muller, Guha, Baumer, Mimno, & Shami, 2016), they are united by the belief that coding, in whatever form it takes, is necessary to make sense of data.

In their recent piece on machine learning and grounded theory, Muller et al. (2016) make a clear case for the epistemological link between machine learning—especially unsupervised approaches—and grounded theory ethnography: "Each process begins with, and returns to, the data. Each process develops interim components of theory that describe major differences across the range of the data. Each process is iterative. Neither process is complete when the data are analyzed; rather each process requires interpretation and theory-building after the data are complete." (pp. 4–5). However, while the two methods both begin with the data and look for patterns within that data from the ground up, they differ in their approach to how they look for patterns with ethnographers typically adopting a more open exploratory approach. In this project, we build on the important work in this area by the likes of Muller et al. and Nelson (2017) to make the case for the mutual benefit inherent in collaborative work between ethnographers and computational social scientists, particularly through integrating a more open exploratory ethnographic approach to the process of pattern-finding in unsupervised machine learning.

In addition to a common inductive, category-forming, data-driven approach, the two also share a common understanding of how necessarily subjective and context-dependent the process of data interpretation can be. This widely-known characteristic of ethnography was recognized by scholars of unsupervised machine learning (DiMaggio, Nag, & Blei, 2013), who have argued that in most applications, the process of *interpreting* data produced through machine learning is a qualitative rather than a quantitative one, requiring the involvement of a researcher with deep knowledge of the context from which the data has been collected. As Elish and Boyd have demonstrated, “it is impossible for a system to detect whether or not the correlation it has identified is meaningful” (2018, p. 71). Simply put, machines can identify linguistic patterns within data, but a human eye is needed for rendering meaningful what they have produced (DiMaggio et al., 2013). We believe this common epistemological approach means that unsupervised machine learning methods and grounded theory ethnography are uniquely positioned to help bridge disciplinary divides within communication.

Historical common ground

As long as computers have been around, ethnographers have been using them for tasks as simple as writing field notes and transcribing audio recordings (Dohan & Sanchez-Jankowski, 1998). Perhaps most notable in the early integration of computers into ethnographic work was the development of computer-assisted data analysis software intended to help researchers enter, organize, retrieve, and code their data after collection. Even as early as 1998 there were at least 20 such programs, including the still widely used ATLAS.ti and NVivo (Dohan & Sanchez-Jankowski, 1998). More recently, many ethnographers across disciplines (particularly in human–computer interaction studies and communication) have become actively experimented with the integration of new technologies throughout the whole research process. This has ranged from experimenting with new tools for data *collection*, like eye tracking software (Zheng, Hanauer, Wiebel, & Agha, 2016) or deep immersion in online worlds, to new tools for data *analysis*, like the machine learning advancements that have emerged in the last 5 years (Abramson, Joslyn, Rendle, Garrett, & Dohan, 2018), even to new tools for data *distribution*, like the “Ethnoarrays” developed by Abramson and Dohan to visualize findings from ethnographic interviews (Abramson, Joslyn, Rendle, Garrett, & Dohan, 2018). However, while studies like those from Muller et al. (2016) and Nelson (2017) have demonstrated the novel epistemological utility of bringing computational approaches to ethnography, to our knowledge, none have looked at the effect of bringing these two approaches together for ethnographic reflexivity or at what it might bring to computational methods themselves, as we do in this study.

Among those scholars wary of integrating computational methods into ethnographic data analysis, a few arguments seem to dominate. The first is that computational techniques risk stripping qualitative data of the richness that makes it valuable (Jones & Diment, 2010). The second and more prolific perspective is the fear that it could act as a methodological "straight jacket" (Holbrook & Butcher, 1996, p. 60) leading researchers to design projects around the technical parameters of software rather than leveraging the software to support existing project goals. The third is a more practical concern—that computational tools are not worth the investment. Many of the qualitative data analysis programs are proprietary and quite expensive, have a steep learning curve, and as Marathe and Toyama found recently, do not help accelerate the frequently tedious process of qualitative coding (Marathe & Toyama, 2018). Even when more advanced computational methods, like machine learning, have been applied to the task of coding ethnographic data—often the primary focus of a lot of qualitative data analysis software—there continue to be concerns about the low accuracy of machine-generated codes as opposed to the traditional human coding (Chen et al., 2018).

We address these previous discussions by making two arguments. First, we believe not all computational tools should be considered equal; they vary widely in their applicability and affordances. While the worst can feel like they are shoehorning complex ethnographic data into a positivist framework, others, like Abramson and colleagues' "Ethnoarray" aim to remain within the epistemic tradition of ethnography grounded in a "commitment to depth, context, and interpretation" (Abramson et al., 2018). In doing so, they make important contributions to the ability of ethnographers to understand and communicate their findings. However, work that claims to be helping to "solve" ethnography's weaknesses can often further entrench the epistemological divides within interdisciplinary fields like communication. We suggest that approaches like Abramson's—and the approach proposed in this piece—which work *with* ethnography's epistemic traditions, can be particularly productive because they work to enhance ethnography's inherent strengths rather than solve its "weaknesses." In the case of the method we explore here, the right computational tool employed in the right way can even enhance one of ethnography's core strengths: reflexivity. We find the analysis of topic model networks approach (Walter & Ophir, 2019) to be particularly well-suited for enhancing ethnography's inherent strengths, and for helping to improve cross-method collaboration, communication, and empathy within communication.

Second, while other ethnographers have used computational methods to enhance their work, these prior applications tended to take ready-to-use codes and software and apply them as is to the ethnographic text, as Marathe and Toyama (2018) have demonstrated. However, we believe that these prior attempts, while important, were limited by the use of predefined tools, and the inevitable assumptions they brought with them. In this work, the ethnographer worked together with the computational researchers at every stage of the process. As we detail below, the work was iterative in nature (Nelson, 2017). Moreover, at every stage—from the pre-processing and modeling to the analysis and interpretation of the data—the unique character of ethnographic data pushed the researchers to reflect and debate the particular approaches being adopted. Ultimately, working this way is time-consuming, but it created important new space for reflexivity—and an examination of the assumptions inherent in the researchers' arguments and procedures—both for the ethnographer *and* the computational social scientists.

Method: Introducing a tool for enhancing common ground

For the data analysis procedure, we chose a method developed by Walter & Ophir, 2019, combining topic modeling and network analysis techniques for text analysis. The method is based on three steps. *First*, a corpus of texts is analyzed using topic modeling and interpreted qualitatively on the topic level. *Second*, the relationships between topics are calculated based on whether topics occur together in the same documents. *Third*, a community detection technique is employed on the topics network to identify broader themes. While we endeavor to detail our approach below in a way that facilitates other similar collaborations, for readers who are new to the procedures and concepts in topic modeling, we recommend the excellent and accessible piece from Maier et al. (2018).

Topic modeling is an *unsupervised* machine learning approach intended to summarize the thematic content of corpora, meaning that corpora are analyzed inductively, without requiring the use of dictionaries or manual coding (as in *supervised* machine learning). Topic models identify latent "topics" drawn from observed texts. Topics are frequency distribution lists of words that tend to occur together in the same documents (in our study, "documents" refer to paragraphs in the field notes), and are thus assumed to share a common theme (Grimmer & Stewart, 2013). For example, since the words "ball," "stadium," and "bat" tend to appear together often in articles about baseball, the topic model will identify this linguistic phenomenon and suggest a topic characterized by these words. The model will also identify that words such as "politician" or "classroom" have low probability of appearing together with the words associated with the baseball topic. Topic models do not tell the researchers what the meaning of a topic is or what it represents. For example, in context, this group of words could be about cricket instead of baseball. It is up to the researcher to make sense of the topics using their knowledge of the context and their qualitative reading of the topics and the documents that use them.

Topic models work "from the text up," in the sense that they infer the latent linguistic and thematic structure of a corpus based on its observed documents (Blei et al. 2003). In other words, the model "asks" what topics would need to exist in the author's mind in order for her to create the observed set of documents, in this case, paragraphs from ethnographic field notes. An advantage of topic modeling is that it allows each document in the corpus to be composed of a mixture of topics rather than forcing each document to consist of only one topic. The annotation of documents for topic use enables, among other things, the analysis of changes in topics' prevalence over time or by different authors. Additional meta-data about the documents—for example, their source—could allow additional analyses (e.g., whether articles in the finance section of newspapers talk about baseball differently from articles in the sports section).

Finally, while other proprietary data analysis software are often costly, multiple open-source libraries for topic modeling are free to use for popular open-source coding platforms (such as R, Python, and Java) with codes shared freely by researchers. Learning how to use existing packages has become easier, thanks to the rich repositories, tutorials, and accessible books available for free online. However, while we do recommend ethnographers familiarize themselves with the code and the logic of the method, we encourage working within collaborative teams that bring the expertise of ethnographers and computational communication scientists together, as was done in this study.

Ethnographic method and data

The field notes we use in our demonstration were taken from a study that explored the role of narratives and storytelling in shaping the ability of technology entrepreneurs in Kenya to engage in a meaningful way in the production of new technologies. The study concentrated on a community center, known as the iHub, built to support local technologists and entrepreneurs based in Nairobi, Kenya. The iHub had featured quite prominently in international news coverage of a "rising Africa" and had served as a gateway for many international investors and venture capitalists interested in working with African tech start-ups. After two preliminary research trips in 2013 and 2014, the ethnographer (Marchant, 2018) conducted a full year of ethnographic participant observation at the iHub from January to December 2016, becoming embedded in the organization's communication department. In doing so, she was able to focus on the role that the communication ecology—particularly the organization's storytelling and narrative building strategies, as well as existing dominant cultural narratives about the role of technology in society and the role of Africa in the world—had played in shaping the ability of the iHub itself and its entrepreneurs to be taken seriously by those abroad.

Over the course of the year, the ethnographer conducted semistructured and unstructured interviews with every staff member of the iHub, as well as over 80 entrepreneurs, investors, and other figures in Nairobi's tech sector. In addition, she conducted participant observation at iHub events, such as entrepreneurial training sessions, start-up competitions, and a series of talks and presentations. In addition, she participated in many unofficial meetings, after-work drinks, parties, and informal group outings like hikes and trips. She collected media content pertaining to the iHub and Kenya's technology sector, including articles from local and international newspapers, blogs, and Twitter discussions. She recorded details about verbal communication she heard as well as nonverbal elements in the field, such as clothing choice, posture, gestures, and the characteristics of the physical space in which the iHub was situated. This resulted in a complex array of data from visual images and videos to "sensory postcards" to audio recordings, speeches, promotional brochures, and, of course, field notes. The field notes featured reflexive writing about the role of the ethnographer in the field, increasing transparency and reflexivity via documenting "ethnographic moments" (Dolby & Cornbleth, 2001). This produced over 400 pages of field notes. When it came time for the final data analysis when she returned from the field, the ethnographer adopted a version of the Straussian approach to coding and conducted open coding of her field notes (Corbin & Strauss, 2014). She then worked towards identifying links between the codes: Connecting categories and dimensions that define the data (Corbin & Strauss, 2014) and comparing them repeatedly to the data, and later to existing theory. Eventually she found that combining a few existing theories from across disciplines in a novel way—namely integrating narrative theory with community of practice theory (Polletta, 2009)—was the best fit for explaining the data gathered and the categories identified.

Analyzing ethnographic field notes using unsupervised machine learning

All steps taken in this collaborative study differed from how most scholars, including the authors of this work, usually conduct computational and ethnographic research. Specifically, we did not apply a ready-made software or code to ethnographic data. Instead, our work was iterative and collaborative, in the sense that at every stage—from the preprocessing and modeling to the analysis and interpretation—the ethnographer and the computational researchers worked closely together (Nelson, 2017). This decision had far-reaching consequences for the work on both sides, creating new space for reflexivity for both the ethnographer and the computational researchers, leading both to find new ways to incorporate the unique contribution of the other side into their own work. For the ethnographer, that mainly meant using computational tools to access a new perspective on her field notes (detailed in the analysis section); for the computational researchers, it meant incorporating ethnographic reflexivity strategies and having the unique opportunity to work directly with the author of a text they were studying, and with a "raw" text not in its final manuscript form. This ultimately meant that we integrated an ethnographic open exploratory approach to pattern finding into the modeling process itself, impacting the chosen computational approaches on technical, conceptual, and theoretical levels, as we elaborate on in the findings section. In addition, we believe these insights are relevant to other researchers who are applying computational methods to unique contexts and corpora that differ from the news media contexts for which these models were originally developed.

This is different from how computational communication researchers often work; it is almost always the case that the text is seen as a finite product, ready to be processed and analyzed. By doing so, researchers rely on the assumption that all the information needed to understand a particular text lies within the words themselves. But in ethnographic field notes, as in ethnographic research in general, this is often not the case, with important contextual information emanating from beyond the literal words on a page. Embracing Geertz' (1973) views about the ability of ethnography to generate "thick description," many ethnographers strive for a holistic picture of a single field site, incorporating the views and experiences of their research participants (emic) as well as themselves (etic), both of which they believe to be necessary for any understanding of a field site (Agar, 2011). While much of this information ends up in an ethnographer's field notes, other observations are collected in mediums like photographs or videos and omitted from the field notes themselves.

Thus, unlike situations in which computational researchers have teamed up with qualitative researchers to analyze finished products, like domestic and international news articles (Tenenboim-Weinblatt & Baden, 2018), the present study involves the direct participation of the author of the text itself, who brings a deep and often unwritten understanding of the context in which the text is situated.

For the present study, field notes were imported and parsed by date of entry, resulting in 277 items. In our first iteration, we attempted to model the 277 entries, treating each as a document. However, working closely together, we came to realize this resolution did not fit this particular study. From the computational researchers' side, unsupervised machine learning is more effective for larger corpora. From the ethnographer's perspective, treating every date entry as equal masked the diversity within them—as some are much more complex and multi-topical than others—leading her to suggest that the thematic unit should instead be the paragraph. We therefore broke the daily items into paragraphs, resulting in a corpus of 1865 paragraphs. To remain consistent with common language used in machine learning, we refer to these 1865 paragraphs from now on as "documents." Next, the text was preprocessed based on best practices suggested by Maier et al. (2018), including lowercase conversion, and the removal of stopwords (e.g., "of," "the"), punctuation, and numbers. We avoided stemming (a data reduction practice where, for example, the words "team," "teams," and "teaming" are treated as the same word) due to its possible negative influence on model quality (Schofield and Mimno, 2016), particularly in this case. While we believe this is always good practice (see e.g., Schwartz et al., 2013), it is especially important when working with the fluid use of language by ethnographers, which does not adhere to a particular format (like, for example, that of journalists), often fluctuates as the ethnographer explores different writing styles to make sense of their data in new ways, and captures different kinds of language and information, from the ethnographer's own (etic) to that used by informants and participants (emic). Lastly, we removed all words that appear in more than 95% of the documents, and words that appeared in less than two documents (following Maier et al., 2018).

Several approaches to topic modeling are currently available that can serve as efficient and accurate alternatives to traditional methods (see as Guo et al., 2016). In this piece, we used Latent Dirichlet Allocation (LDA) with Gibbs Sampling (Blei et al., 2003), using the *topicmodels* package for R. As is the case in every clustering question, in order for an algorithm to find patterns in data, the researcher should first define the number of clusters/categories (the "k" parameter), as well as a parameter that defines whether documents tend to be multi-topical or tend to focus only on one topic each ("α" hyperparameter). Both k and α were chosen based on statistical examinations of the corpus, using statistical indicators for "goodness of fit" for different combinations of the number of topics and hyperparameters (estimating how well the model fits the data). We used 10-fold cross-validation iterating model specifications over k range of 2–50 and α levels of 0.01–1, and evaluated their fit based on perplexity scores (Maier et al., 2018). We found the level of $\alpha = 0.1$ and $k = 30$ to offer an optimal solution (where additional topics contribute little to the model and lower/higher alpha levels do not improve results). As a sensitivity analysis, to gain additional confidence in our choice of 30 topics, we qualitatively examined two alternative models, with 40 and 20 topics. We found the 20 topics model was lacking and the 40 topics model was redundant. We elaborate on the meaning of the α hyperparameter and the k choice process when discussing computational reflexivity. Finally, for each of the 30 topics, we extracted the 30 top words and 10 top documents that were representative of each. Top words are the words that are most likely to appear in a topic. Top documents are the paragraphs in the field notes that used the language from the specific topic the most. We also extracted 30 top unique words for each topic, balancing word likelihood for each topic with its exclusivity to that topic (Roberts et al., 2014).

At this stage, this new machine-produced data was analyzed again by the ethnographer and the computational researchers. As we have noted earlier, this qualitative step is crucial in any topic modeling work but involving the author of the field notes herself—essentially the "object" of the study—brought into stark relief just how integral this qualitative stage is for rendering machine-produced data analytically useful and how important reflexivity is to the work of computational social scientists. In this case, not only was a human necessary, but one particular human with a uniquely rich understanding of the original data (DiMaggio et al.,). The authors reviewed the 30 top words, 30 top unique words, and the 10 top documents for each. For some topics, the names came easily. For example, one topic contained top words like "uber," "road," "taxi," and "construction"; it was clear to the ethnographer that this topic revolved around transportation around the Nairobi; after consulting the top documents, the name of "transportation" was chosen. For other topics, this process took much longer. For example, another topic contained many quite general top words like "people," "things," "work," and "iHub." Even with the ethnographer's deep knowledge of the context, the theme for this topic was not readily apparent; it took looking more deeply at the relevant documents to realize that much of the topic's content contained stories from iHub staff members about the iHub, leading to the label "narratives from the iHub."

Next, the computations researchers produced charts that illustrate how the prominence of the 30 topics had changed over the course of the year the ethnographer was in the field. In order to do so, they aggregated the document topic loadings (represented in the "theta matrix"—one of the algorithm's outputs) over date. In other words, they calculated the average of every topic salience over every day in the dataset. After consultation between the ethnographer and the computational researchers, a final chart was produced, a map that visualizes the full network of the 30 topics and the connections between (based on their use in similar documents). This step, suggested by Walter and Ophir (2019) allows the analysis to move from specific detailed topics to broader themes, giving researchers different levels of resolution from which to observe the data. Moreover, such transformation allows for the exploration of intertopic connections and topic–communities structure. We elaborate more on the potential advantages of the network representation in our conclusion section. To create the network of topics, we treated each topic as a node in a network. Relationships between nodes were calculated as the rate of co-occurrence between topics in documents (topics that tended to appear in the same field notes paragraphs are represented in the network as being close and strongly connected). In technical terms, the strength of association between topics was calculated using cosine-similarity over the theta matrix columns. Finally, after calculating the relationships between all topics, we used a method called "community detection" to identify clusters of topics (based on the similarity measure). Using the common Louvain community detection algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) we identified two clusters of topics.

Findings

Initially, this collaboration started out as an attempt to examine the ability of a new unsupervised machine learning tool to improve on existing qualitative data analysis software and help an ethnographer code her data in a way that was both accurate and worth the time. However, while we were pleased with the tool's ability to assist data coding, to our delight, the deeply collaborative work necessary due to the nature of ethnographic field notes produced a number of unexpected and perhaps far more valuable results. Specifically, (a) the topic modeling tool provided the ethnographer with an effective and efficient tool with which to navigate her data; and (b) the unique collaborative way of knowing developed through this project created space for more reflexivity about processes and assumptions on the part of *both* the ethnographer and the computational researchers. Specifically, as we will elaborate below, the network analysis, and particularly the community detection step, helped the ethnographer reflect on her own research process, and helped the computational researchers reflect on the nature of topic model networks and the algorithms they employ, as well as their underlying logic, appropriate usage, and linguistic and thematic mechanisms. Ultimately, the process of bringing two computational researchers and an ethnographer together for this project and the close collaboration on all stages of work led both sides to rethink some of their assumptions, procedures, and perspectives. This resulted in important insights that will alter the way all of the authors approach their work going forward. Furthermore, due to the recognition of the

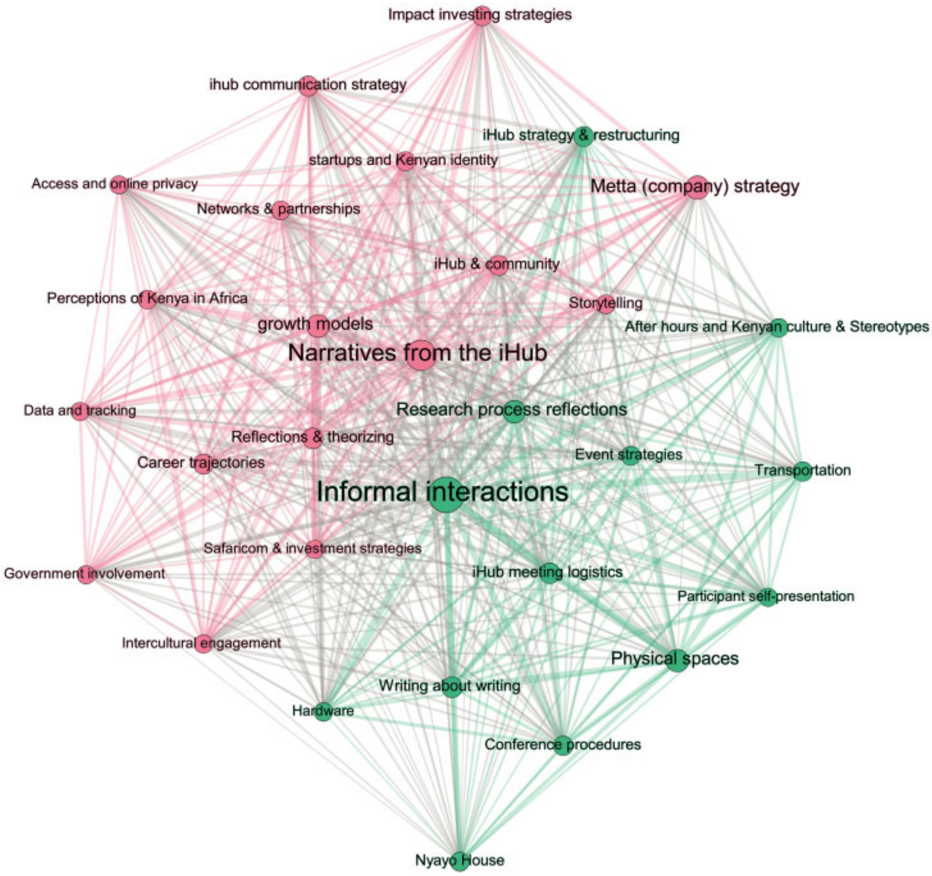

Figure 1 The topic network drawn from the field notes. Nodes represent topics; edges represent the cosine similarity in document co-occurrence between topics; size represents the prominence of each topic in the corpus; node color represents community membership using the Louvain algorithm; edge color represents the color of community for intracommunity edges, and gray for intercommunity edges; layout created using the Force Atlas algorithm.

often-unseen epistemological common ground, it also increased mutual methodological appreciation. The modeled network of topics can be seen in Figure 1.

Usefulness as a coding tool

As discussed above, the vast majority of efforts to integrate computational tools—including machine learning—into the analysis stage of the ethnographic process have been aimed at improving the laborious process of data coding. This was also the original intention of the collaboration that led to this article. In this regard, the tool proved helpful. Specifically, it (a) supported many (though certainly not all) of the codes the ethnographer had originally identified; (b) found one topic the ethnographer did not include in her original analysis; (c) identified a number of topics the ethnographer had paid attention to but chose not to include in her final analysis; and (d) identified a number of topics about particular categories of data, including the research process itself.

Firstly, after the algorithm identified the 30 topics and the ethnographer (with the computational researchers' assistance) interpreted and named (labeled) them, it supported many of the codes the ethnographer had found herself in the original coding process. In particular, the following topics were rough replications of substantive codes the ethnographer had originally produced herself: "perceptions of Kenya in Africa," "iHub community," "iHub communication strategy," "narratives from the iHub," "start-ups and Kenyan identity," "networks and partnerships," "intercultural engagement," "impact investment strategies," and "storytelling." Crucially, these topics featured prominently in the ethnographer's development of theory.

Secondly, the tool identified at least one topic that the ethnographer had not identified, namely "data and tracking." This topic included field notes on the collection and use of data by start-ups, particularly among fintech start-ups, and the increasing role of "data science" in the Kenyan tech community. Once the tool identified this topic, the ethnographer clearly recalled discussing these issues with participants, but it had not been a topic she had coded, nor was it a topic that really featured in her final written analysis (Marchant, 2018).

Thirdly, a number of the topics the tool identified were things the researcher was interested in but had not included in her final analysis. Transportation is a particularly good example of this. In the first few months, the ethnographer took particular note of the ways in which technology start-ups were changing the transportation landscape in Kenya, including the role of Uber, but with time she shifted away from this and began to focus more on the role of narratives. This whittling way of the topic of focus is common in lengthy ethnographic fieldwork as the researcher begins by looking holistically at the field site and, over time, homes in on the specific elements that become the primary object of study.

Finally, the tool identified topics about particular *categories* of data that were not substantive, but were instead about the *research process* itself. We elaborate more on this finding below.

For ethnographers, one of the most fundamental elements of their methodological approach is reflexivity (Emerson et al., 2011). Over time, as the ethnographic communities in anthropology and other disciplines like communication have become aware of the often painful colonially embedded legacy of the discipline, they began to confront the limits and subjectivities inherent in any process of knowledge production (Ferguson, 2006). In order to do this, methodological reflexivity became a key part of the ethnographer's approach, helping the researcher to confront things like their own positionality in the field, how their presence might influence findings, and how their own experiences might influence what they find "important." In many ways, reflexivity can be described as a, if not the, defining characteristic of modern ethnography. In the field notes, ethnographers write up their reflections while surrounded by their data and their field site. And then when they often (but not always) leave the field site and begin the next phase of data analysis, that separation creates more distance in both space and time between the researcher and the field allowing for further reflection. Reflexivity at any stage can be a challenging and sometimes even uncomfortable process, one that forces researchers to confront deeply embedded aspects of their identity or unwanted and previously unacknowledged biases. As a result, being able to see such biases and assumptions can take time, emerging in different ways at different stages of the research process from the writing of field notes to the analysis of the data.

Given the centrality of reflexivity to the ethnographic approach and the far more limited role it often plays in other methods—including computational approaches to communication—it came as a bit of a surprise when looking at the data produced by this computational tool, revealed new insights that allowed the ethnographer to reflect further on her data, her process, and her own assumptions. This realization became most apparent when the salience of the different topics was examined over time, as can be seen in Figure 2 below.

One topic, named "research process reflections" illustrates this especially well. It is clear the ethnographer wrote quite a bit on this topic in the early months of her fieldwork, but then gradually dropped off over time. In the end, this produced quite a vivid visualization of the uncertainty that defined the early stages of her (and many ethnographers') fieldwork. The inductive nature of ethnography means that the first few months are often necessarily quite broad, when the researcher tries to observe the widest variety of events and trends in the field before narrowing down her focus and gaining greater certainty about her project and her approach. It is no surprise then, yet still insightful, that she wrote a great deal more "research process reflections" in the early months as she used writing to work through her uncertainty. Seeing it visualized in such a clear way put this tendency in stark relief.

Both the process of writing field notes and of removing oneself from the field to write up the final analysis can be seen as tools ethnographers use to better "see" their data. By applying an unsupervised machine learning tool to her field notes, a new way of "seeing" the data emerged, one that somewhat de-personalized the data, forcing the ethnographer to see herself almost as a research participant, at times creating a feeling of being "objectified" and of vulnerability of having quite personal information examined by researchers (in this case the computational ones), a kind of vulnerability that would be encouraged by anthropologists, like Behar (2014). In the end, this forced new, important, and difficult to resolve reflections on the role that "objectification" of research participants often plays in the research process, even for ethnographers who work hard to minimize it and to see their participants holistically.

In the end, this new tool did provide the ethnographer with a different way to code her data in the future. But it was in the novel data visualization afforded by the tool—both over time and as a network map showing the importance of different

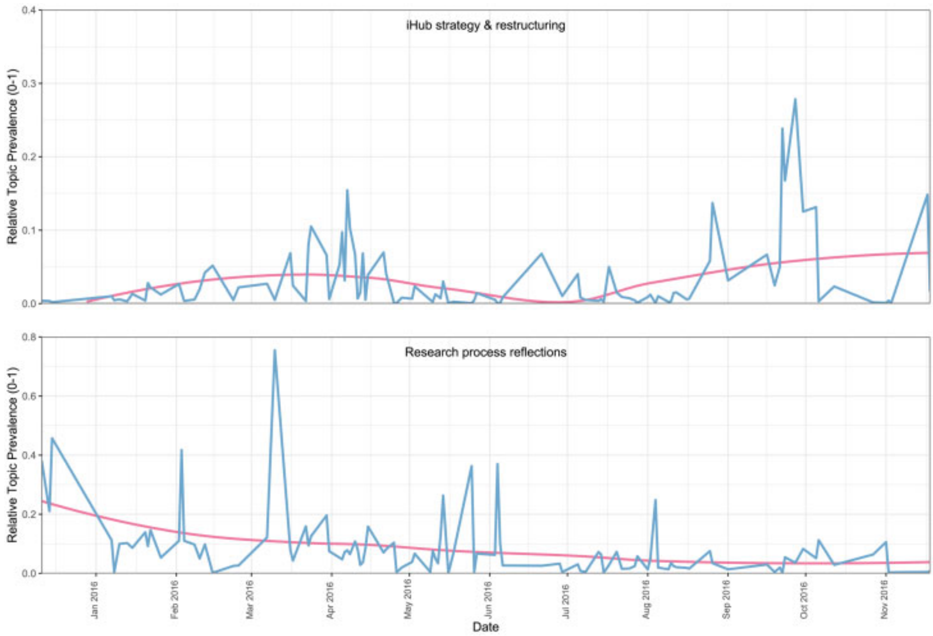

Figure 2 Change in topic salience over the research period for topic 1 (iHub strategy and restructuring) and topic 6 (research process and reflections). The blue line represents the share of topic out of total text on a daily scale. The pink line represents the loess smoothed curve (span = 0.75) of topic share out of total text on a daily scale with 95% confidence intervals in gray. Note: as we are interested in trends in topics over time and not between topics, scales on the y-axes are different between the two timelines.

topics and the connections between them—where new insights were gained. By combining her existing deep knowledge of the field site with the new distancing from her data this tool created, the ethnographer began thinking about her research process differently, particularly how she related to participants, how her approach changed over time, and the kind of knowledge she produced.

Usefulness as a mapping tool

Another finding of this collaboration is that while other computational data analysis tools exist that help ethnographers code their data, this one produces a uniquely rich visualization, a network map that provided a new way to navigate data. To our knowledge, this map is novel among the current computational tools being used with ethnography, and, we believe, can provide ethnographers with a birds' eye view of their data and the interrelationships within it, much as a geographic map of the world might do. In this map, both the proximity of the different topics (nodes) to one another as well as the thickness of the lines (edges) that connect them give some indication of how frequently these topics tend to appear together in the original field notes. Similarly, the size of the nodes themselves denote how frequently the ethnographer wrote about a particular topic. The larger the node, the more it featured in her field notes. Finally, the map also breaks all the topics down into two macro frames (topic clusters as identified using community detection), illustrated here in green and pink, which we will discuss shortly.

There are a few interesting observations that can be gained from this map. The first is the role of the largest topics, "informal conversations" and "narratives from the iHub," both of which appear quite central to the map as a whole; both of which were also quite central to the ethnographer's own original analysis. This visualization of the first topic, "informal interactions" helped to illustrate how important those kinds of less structured discussions were to the ethnographer's data collection process. While she employed other processes, like conducting interviews or observing staff meetings and community events, this visualization makes clear how much her field notes focused on observations made during more "informal" interactions. The second key topic, "narratives from the iHub," was also very reflective of the ethnographer's own substantive interpretation of her data. In her original analysis, not only did she code a similar topic, but she found that topic to be so central to her analysis that it played a key role in the theoretical framework she developed (Marchant, 2018).

These two key topics also point to another interesting observation that the map makes possible: the distinction between the two macro categories (communities), green and red. On the map, "informal conversations" is green while "narratives from the iHub" is red. Not coincidentally, these two topics also seem to represent two different kinds of writings that could be found in her field note data: writing about the research *process* (characterized by the "informal conversations" topic) versus writing about more *substantive* observations (characterized by the "narratives from the iHub" topic). Many (though not all) of the other topics coded in green—like "iHub meeting logistics," "research process reflections," or "writing about writing"—seem to represent the ethnographer's writing about the research process. Likewise, many of the topics coded in red—like "perceptions of Kenya in Africa," iHub and community," or "start-ups and Kenyan identity"—seem to represent more substantive observations, many of which eventually featured in some way in the ethnographer's original analysis and theoretical framework.

Usefulness as a tool for computational reflexivity

While our concluding remarks have focused thus far on insights gained by the ethnographer using the computational tools, the benefits ran both ways. The computational researchers were affected by the collaboration in ways that we believe can and should influence current and future studies using such methods. Just as the ethnographer gained a new perspective on her data from which to reflect on herself and her approach to the research, the computational researchers also gained a new perspective from which to reflect on their own methodological assumptions, procedures, and method use cases, reevaluating and modifying them based on the fully collaborative work with the ethnographer. Below, we elaborate in more detail how this collaboration led to the reassessment of the researchers' work on multiple levels, including technical, conceptual, and theoretical.

First, on the technical level, working with the ethnographer enabled the computational researchers to better prepare the data for analysis. For computational researchers who are used to treating a text as a finished product, it was illuminating to work with an ethnographer—the author of the text being analyzed—who brought deep additional insights about the data beyond what could be found in the text alone. Computational communication researchers, who usually lack access to the original authors of the texts they study, have to assume that every data point exists in the words on the page. In contrast, in our study, the access to the original author shaped every step of the modeling procedure, with choices about the number of topics, level of documents, preprocessing, hyperparameter optimization, and interpretation relying on both statistical indicators of model fit and the ethnographer's qualitative assessment of the quality of the model.

Most importantly, having an ethnographer work directly with the computational scientists enriched the process of model specification. Using 10-fold cross-validation, perplexity was found to be optimized at the level of $\alpha = 0.1$and $k = 30$. While the computational researchers expected, based on their experience, that different types of data might be optimized at different levels of α, the ethnographer's knowledge was necessary to make sense of it. A lower α value indicates that the documents (paragraphs in the field notes in this case) were inclined to be more monothematic in nature (that is, a single document tended to focus on fewer topics). The ethnographer's interpretation was that this resulted from the nature of ethnographic writing, or at least her own writing style. As opposed to the concise and simplistic social media posts or news articles for which topic modeling is often used, ethnographic writing is expansive, lengthier, and far more varied, from person to person and even from entry to entry. The choice of k (number of topics) also benefitted from this dual perspective. While the computational scientists found the model to be statistically optimized at the level of $k = 30$, the ethnographer qualitatively assessed two additional models ($k = 20$and $k = 40$), and corroborated that in her own reading, the 30 topics model made most sense.

After completing the main analysis, we reflected once again on the method chosen. We came to the conclusion that the analysis of topic model networks using LDA was the right choice in this case. However, the LDA algorithm is one of many possible algorithms for topic modeling. We chose it based on its popularity and availability in the field of communication research (Maier et al., 2018), and its superior performance compared to alternatives, such as Probabilistic Latent Semantic Analysis (Blei et al., 2003; Lu, Mei, & Zhai, 2011). Other algorithms may be more suitable for specific research questions. For example, researchers interested in incorporating covariates such as authorship or time of publication into the modeling stage may choose Structural Topic Modeling (Roberts et al., 2014).

Due to the nature and richness of ethnographic data and the complexity challenges involved in its analysis, the collaborative approach to this project shaped the data interpretation process as well as the modeling process. Most notably, the inclusion of the author enabled the incorporation of knowledge that did not make it into the final field notes. This includes knowledge about the subject, the field site, the approach, and the way of knowing, as well as knowledge about the different kinds of data collected, like videos, images, brochures, and other sensory data difficult to capture in words or that were otherwise omitted from the field notes. This led the team to abandon the often-used practice of working on the document level (in this case the individual field notes entry for a specific day) in favor of a unit that was more suitable for the specific writing style and decisions of the creator of the text, the ethnographer. While contemporary quantitative researchers tend to acknowledge the role of the research process in shaping their findings, these insights tend not to make it to the final write up in most quantitative publications, where writing conventions allow less space for methodological reflection than in more qualitative writing. This conclusion is consistent with Elish and Boyd's recent argument in favor of "the development of reflexive practices in machine learning" (2018, p. 58).

Second, in addition to reflecting on the processing, modeling, and interpretation of the data, this collaborative work influenced the researchers on the conceptual level, leading them to reevaluate the purpose, logic, and usage of their chosen method, in this case the analysis of topic model network approach. The team of researchers made use of this approach to examine broader themes in the corpus rather than just individual topics. As stated by the authors of this computational method, it was originally designed with the purpose of identifying journalistic frames. In that framework, topics represent frame elements, and their clustering creates more coherent and parsimonious frame packages. Therefore, when applying the method to ethnographic data, the researchers expected to uncover various substantial frames. However, the process yielded surprising results, showing two clusters, one of substantial topics and one of reflexive topics. This seems to indicate that the original designation of the method may have been too limited, and that the method is able to identify other thematic categorizations of data that are not limited to framing.

As communication researchers hoping to continue and develop new tools for understanding communication behavior and cultural objects, we see this experience as a crucial lesson, shedding light on the logic, application, and possible usage of this, as well as other, methods and algorithms. We encourage future researchers who employ computational methods in novel contexts to pay attention to the possibility of methods performing in qualitatively different, and often unexpected, ways.

For example, topic modeling is often referred to as a *generative* algorithm, in the sense that the model tries to mimic a hypothetical process of an author generating text. However, researchers using topic modeling often do not pay attention to the unique author or context characteristics that originally shaped the way in which the text was created. In addition, topic modeling researchers often treat the method as off-the-shelf, rarely inquire whether an existing method and its operationalizations can or should be transferred into the new context as is. In our case, in analyzing ethnographic field notes, a topic modeling approach resulted in a number of substantive topics as well as topics that were clearly a linguistic artifact of stylistic choices, information which is critical to the understanding of the ethnographic work. These results should inform the expectations future researchers bring when employing topic models for the analysis of ethnographic data; it suggests a more cautious and open-minded approach when handling new types of corpora with existing computational methods.

This leads us to the third, namely theoretical, aspect of computational reflexivity. As we have seen, this collaboration resulted in the identification of topics that were both substantive and reflections of the ethnographer's research process. One interpretation of this result is that the method did not identify frames as originally intended, but instead identified a different kind of construct. However, this interpretation is worth more consideration. A different interpretation might be that the method actually did identify frames in the data, but that frame analysis in the context of ethnographic writing is a more complex phenomenon differing from its definition in the journalistic context. As communication research originated primarily from an interest in mass media, most conceptualizations of frames have focused on journalistic work (e.g., Gamson & Modigliani, 1989). However, ethnographic writing is its own form of communication that differs in many ways from other professions, including journalists. In that regard, we could interpret the finding of substantive and reflexive topics in the ethnographic field notes as in fact their own kinds of frames, perhaps common across different kinds of ethnographic field notes, similar to the repeated appearance of episodic and thematic frames in journalistic writing. Much in the same way that journalistic frames are rooted in the socialization of journalists into their profession (van Gorp, 2010) and the conventions of their practice (Tuchman, 1973), these frames may be the result of ethnographers' academic socialization. One frame focuses on the object of interrogation for the ethnographer, while the other situates the research itself within context, and sheds light on the methodological considerations of the work. This interpretation ought to be carefully considered by computational social scientists as they seek out new kinds of texts to analyze. As the volume and variety of these texts continue to increase, researchers will need to pay greater attention to the professional, social, and cultural contexts in which texts have been written and embedded when deciding what theoretical approach to take to the interpretation of their findings.

Conclusion and implications: Towards a collaborative way of knowing

The epistemological intersection of ethnography and unsupervised computational communication research is an exciting opportunity to reconsider the boundaries in our field. To many, no two subdisciplines will look more different. Indeed, collaborations between ethnographers and computational communication researchers are still relatively scarce, while the adoption of computational tools into the ethnographic work often faces resistance within the subfield (Abramson et al., 2018). Even when computational tools are used in ethnographic research, it is often done in a way that assumes that either the computational method or the ethnographic data, or both, is finite and ready for use by the other side.

In this article, we have sought to shed light on the similarities and the potential for collaboration between these two diverse subfields. Through our example we hope to encourage other ethnographers and computational communication researchers to experiment with collaborative ways of knowing that can benefit both sides and help them develop their research in unexpected, yet fruitful, ways. Our collaboration demonstrated that combining topic modeling and network analysis (Walter & Ophir, 2019) with ethnographic data and an ethnographic mindset could be mutually beneficial by creating space for reflexivity about assumptions, procedures, and perspectives. While reflexivity is already a well-known and integral element of ethnographic research, this collaboration demonstrates the ways in which computational research also stands to benefit from the integration of more purposeful reflexivity into its process.

Specifically, this study harnessed recent developments in the process of automated text analysis to revisit field notes written during an ethnographic study in Nairobi, Kenya during 2016 (Marchant, 2018). The collaborative employment of a computational method for the analysis of topic model networks (Walter & Ophir, 2019) resulted (after multiple iterations) in a 30-topic model that was then clustered into two major network communities. Interestingly, only one community consisted of topics focusing on substantive observations. While we entered this project expecting to look only at substantive topics, we were surprised and delighted to find a rich and informative second community identified by the automated method consisting of reflexive topics about the research process itself. An intriguing characteristic of this second community was that its topics included a strong component of self-reflection on the ethnographer's role in the field. This not only aided the ethnographer in her goal of analyzing her data, but also the computational researchers in reassessing the applications, weakness, and boundaries of their chosen method.

Reflexivity is a cornerstone of the ethnographic approach and a crucial part of any ethnographic writing. Grounded in social constructivism, ethnographers use their field notes to bring front and center the ways in which their own subjectivities shape how they approach their research and what they find. But as Olive (2014) has demonstrated, it can be a particular challenge for ethnographers to balance the emic and etic information in their final analysis because of how intertwined they can be. By highlighting the relationship between different topics within her field notes, the community detection algorithm method provided a map that helped the ethnographer think more reflexively about how she had gathered her data. Importantly, because we worked with field notes, almost every text has some element of etic in it, and it is rarely possible to pull the etic and emic components apart entirely. However, in the end, the novel use of community detection in a topics network helped the ethnographer reflect more on the prevalence of etic ways of knowing in her research and to identify some of the interactions in her data between emic and etic findings.

Similarly, the analysis of topic model networks was useful for the ethnographer as a unique mapping tool, visualizing the research process and results in insightful ways, providing a bird's eye view of the data and the interrelationships between its components. Not only do these data maps help the ethnographer cope with large amounts of data, they also serve as efficient and effective tools for collaborating with other researchers, including sharing data in an anonymous form and communicating findings across subfields, including to computational researchers.

In this demonstration we have focused only on the analysis of edges between nodes (topics) and their clustering into coherent communities, but the calculation and visualization of ethnographic data in the form of a network opens up the door for further explorations. For example, once constructed as a network, the topic model could be analyzed using the rich and constantly developing tools for social networks analysis, such as the calculation of centrality or the identification of brokers. By adding extended metadata into the model, ethnographers could also examine the development of the network over the timeline of the ethnographic study. Considering the impact of reflexivity for both the ethnographers and computational scientists, our project serves as an illustration of the need to move beyond merely borrowing methods from across communication subfields, and towards more collaborative approaches to research. For the computational researchers, the collaboration resulted in a newfound appreciation for ethnographic reflexivity and the role it could play in their own research. For other scholars in the computational communication subfield, particularly those willing to embrace reflexivity, this could lead to new research into how various corpora alter the procedures and assumptions buried in computational research, even, perhaps, incorporating more reflexive writing about their data and methodologies into their final publications. For the ethnographic researcher, the collaboration resulted in new appreciation for the inherently qualitative process of interpreting computational data and the important role of domain expertise in that process. Other computational social scientists have long argued that automated tools should not be used to replicate experts' work, but rather to extend and compliment it (DiMaggio, Nag, & Blei, 2013; Grimmer & Stewart, 2013; Levy & Franklin, 2014). For other scholars in the ethnographic communication subfield, such a deep collaborative and iterative approach to combining computational and ethnographic insights should begin to address some of the concerns of those worried that computational approaches seek to "replace" human researchers (Gottschall, 2008). The present analysis would not have been possible without the involvement of the ethnographer, the domain expert, at every stage of the analysis.

To conclude, in this study we demonstrated the potential and importance of transcending academic silos within the broad discipline of communication through creating new collaborative ways of building knowledge about the role of communication in society. We acknowledge that "research that transcends conventional academic boundaries is harder to fund, do, review, and publish" (Nature, 2015). However, we also strongly agree with Theodore Brown, former vice-chancellor for research at the University of Illinois at Urbana-Champaign who in the early 1980s said that "the problems challenging us today, the ones really worth working on, are complex, require sophisticated equipment and intellectual tools, and just do not yield to a narrow approach" (cited in Ledford, 2015, p. 308). Our collaboration across the at-times-fragmented subfields of communication ultimately yielded an insightful, informative, and even educational experience that created more space for mutual understand and new ways of thinking about seemingly-established approaches to knowledge-building. It led to new and unexpected reflections on existing methods and data sets in ways that changed how the computational researchers prepared and modeled the data and how they perceived the applied method in its possible future applications, as well as how the ethnographer analyzed her data and reflected on how her subjectivities and approaches shaped the data she produced in the first place. This was made possible, in part, by the willingness of those involved to go beyond narrow categories of communication research and instead attend to the hidden epistemological similarities between grounded theory approaches to ethnography and unsupervised machine learning techniques. We are eager to see such future collaborations across communication subfields using other methods for unsupervised machine learning, or supervised machine learning (from dictionary methods to neural networks) where applicable. We conclude with the words of Brown (p. 311) that these days "people focus on big problems, and if you go for a big problem you need to be interdisciplinary."

References

Abramson, C. M., Joslyn, J., Rendle, K. A., Garrett, S. B., & Dohan, D. (2018). The promises of computational ethnography: Improving transparency, replicability, and validity for realist approaches to ethnographic analysis. *Ethnography*, *19*(2), 254–284.

Agar, M. (2011). Making sense of one other for another: Ethnography as translation. *Language & Communication*, *31*, 38–47.

Anderson, J. A., & Baym, G. (2004). Philosophies and philosophic issues in communication, 1995-2004. *Journal of Communication*, *54*, 589–615.

Behar, R. (2014). *The vulnerable observer: Anthropology that breaks your heart*. Boston, MA: Beacon Press.

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality: A treatise on the sociology of knowledge*. Garden City, NY: Doubleday.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocaiton. *Journal of Machine Learning Research*, *3*, 993–1022.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *10*, 1–12. doi:10.1088/1742-5468/2008/10/P10008

Chen, N., Drouhard, M., Kocielnik, R., Suh, J., & Aragon, C. R. (2018). Using machine learning to support qualitative coding in social science: Shifting the focus to ambiguity. *ACM Transactions on Interactive Intelligent Systems*, *8*(2), 1–20.

Corbin, J., & Strauss, A. L. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Los Angeles, CA: Sage.

Delli-Carpini, M. X. (2013). Breaking boundaries: Can we bridge the quantitative versus qualitative divide through the study of entertainment and politics? *International Journal of Communication*, *7*, 531–551.

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, *41*, 570–606.

Dohan, D., & Sanchez-Jankowski, M. (1998). Using computers to analyze ethnographic field data: Theoretical and practical considerations. *Annual Review of Sociology*, *24*, 477–498.

Dolby, N., & Cornbleth, C. (2001). Social identities in transnational times. *Discourse: Studies in the Cultural Politics of Education*, *22*, 293–296.

Elish, M. C., & Boyd, D. (2018). Situating methods in the magic of big data and AI. *Communication Monographs*, *85*(1), 57–80.

Emerson, R., Fretz, R., & Shaw, L. (2011). *Writing ethnographic fieldnotes* (2nd ed.). Chicago, IL: University of Chicago Press.

Falk, E. B., O'Donnell, M. B., Thompson, S., Gonzalez, R., Dal Cin, S., Strecher, V., . . . An, L. (2016). Functional brain imaging predicts public health campaign success. *Social Cognitive and Affective Neuroscience*, *11*, 204–214. doi:10.1093/scan/nsv108

Felix, Dasgupta A., & Bertini, E. (2018). The exploratory labeling assistant: Mixed-initiative label curation with large document collections. In *The 31st Annual ACM Symposium on User Interface Software and Technology* (pp. 153–164). New York, NY: ACM.

Ferguson, J. (2006). *Global shadows: Africa in the neoliberal world order*. Durham: Duke University Press.

Gamson, W. A., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, *95*, 1–37. doi: 10.2307/2780405

Geertz, C. (1973). *The interpretation of cultures*. New York, NY: Basic Books.

Glaser, B. G., & Strauss, A. L. (1999). *Discovery of grounded theory: Strategies for qualitative research*. New York, NY: Routledge.

Gottschall, J. (2008). *Literature, science, and a new humanities*. New York, NY: Palgrave MacMillan.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*, 267–297.

Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism & Mass Communication Quarterly*, *93*(2), 332–359. https://doi.org/10.1177/1077699016639231

Holbrook, A., & Butcher, L. (1996). Software in educational research: The literature, the hard questions and some specific research applications. *Australian Educational Researcher*, *23*, 55–80.

Jones, M., & Diment, K. (2010). The CAQDA paradox: A divergence between research method and analytical tool. In *Proceedings of the international workshop on* *computer-aided qualitative research Asia (CAQRA2010)* (pp. 82–86). The Netherlands: Merlin Institute.

Katz, E., & Katz, R. (2016). Revisiting the origin of the administrative versus critical research debate. *Journal of Information Policy*, *6*, 4–12.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Ledford, H. (2015). How to solve the world's biggest problems. *Nature*, *525*, 308.

Levy, K. E. C., & Franklin, M. (2014). Driving regulation: Using topic models to examine political contention in the U.S trucking industry. *Social Science Computer Review*, *32*, 182–194.

Lu, Y., Mei, Q., & Zhai, C. (2011). Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA. *Information Retrieval*, *14*, 178–203.https://doi.org/10.1007/s10791-010-9141-9

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., . . . Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, *12*, 93–118.

Malinowski, B. (1926). *Crime and custom in savage society*. London: Kegan Paul.

Marathe, M., & Toyama, K. (2018). Semi-automated coding for qualitative research: A user-centered inquiry and initial prototype. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 348). New York, NY: ACM.

Marchant, R. E. (2018). *Anyone anywhere: Narrating African innovation in a global community of practice*. (Publication No. 10842445) Doctoral dissertation, University of Pennsylvania.

Muller, M, Guha, S, Baumer, E. P. S, Mimno, D., & Shami, N. S. (2016) Machine learning and grounded theory method: convergence, divergence, and combination. In *Proceedings of the 19th International Conference on Supporting Group Work* (pp. 3–8). New York, NY: ACM.

Nature. (2015). Why interdisciplinary research matters. *Nature*, *525*, 305. Retrieved from https://www.nature.com/news/why-interdisciplinary-research-matters-1.18370. 10.1038/525305a

Nelson, L. K. (2017). Computational grounded theory: A methodological framework. *Sociological Methods & Research*. https://doi.org/10.1177/0049124117729703

Olive, J. L. (2014). Reflecting on the tensions between emic and etic perspectives in life history research: Lessons learned. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, [S.l.], 15(2)*. http://dx.doi.org/10.17169/fqs-15.2.2072.

Polletta, F. (2009). *It was like a fever: Storytelling in protest and politics*. Chicago, IL: University of Chicago Press.

Roberts, M. E., Steward, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., . . . Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*, 1064–1082. doi: 10.1111/ajps.12103

Schofield, A., & Mimno, D. (2016). Comparing apples to apple: The effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics*, *4*, 287–300. https://doi.org/10.1162/tacl_a_00099

Tenenboim-Weinblatt, K., & Baden, C. (2018). Gendered communication styles in the news: An algorithmic comparative study of conflict coverage. Communication Research. Advance online publication. https://doi.org/10.1177%2F0093650218815383.

Tuchman, G. (1973). Making news by doing work: Routinizing the unexpected. *American Journal of Sociology*, *79*, 110–131. http://dx.doi.org/10.1086/225510 van Gorp, B. (2010). Strategies to take subjectivity out of framing analysis. In P. D'Angelo and J. A. Kuypers (Eds.), *Doing news framing analysis: Empirical and theoretical perspectives* (pp. 100–125). New York, NY: Routledge.

Wagner, R. (1981). *Invention of culture* (revised and expanded). Chicago, IL: University of Chicago Press.

Walter, D., & Ophir, Y.(2019). News Frame Analysis: An Inductive Mixed-method Computational Approach. Communication Methods and Measures, *13*(4), 248–266. doi: 10.1080/19312458.2019.1639145

Zheng, K., Hanauer, D. A., Weibel, N., & Agha, Z. (2016). Computational ethnography: Automated and unobtrusive means for collecting data in situ for human-computer interaction evaluation studies. In Patel. (Ed.), *Cognitive informatics for biomedicine* (pp. 111–140). Switzerland: Springer International Publishing.