Papers 2 – Shifting the Disciplines

A Mixed Method Approach for Identifying Emerging Fields and Building Collaborative Teams: Leveraging Network Ethnography to Design Experimental Interventions

THERESE KENNELLY OKRAKU *Microsoft*

VALERIO LEONE SCIABOLAZZA University of Florida

RAFFAELE VACCA University of Florida

CHRISTOPHER MCCARTY University of Florida

Rapid innovation in science and technology has led to the development of new fields that transcend traditional disciplinary boundaries. Previous studies have retroactively examined the emergence of these fields. This paper outlines a mixed method approach for using network ethnography to identify emerging fields as they develop, track their evolution over time, and increase collaboration on these topics. This approach allowed us to simultaneously analyze organizational trends and gain an understanding of why these patterns occurred. Collecting ethnographic data throughout the course of the study enabled us to iteratively improve the fit of our models. It also helped us design an experimental method for creating new teams in these fields and test the effectiveness of this intervention. Initially, organizational leaders were wary of using a network intervention to alter these fields. However, by presenting insights from both our network analysis and ethnographic fieldwork, we were able to demonstrate the strategic need and potential impact of this type of intervention. We believe that network ethnography can be applied in many other research contexts to help build strategic partnerships, facilitative organizational change, and track industry trends.

Keywords: Ethnography, Social Network Analysis, Collaboration, Community Detection, Mixed Methods

INTRODUCTION

Scientific collaboration often produces new discoveries, perspectives, and fields of knowledge. However, it is difficult to track the emergence of new fields at an organizational level or evaluate teams using an experimental research design. We address this issue in this paper by outlining how researchers can use a mixed method approach called network ethnography to holistically detect communities and study their evolution over time. Our team used this method to identify emerging research communities within a university's scientific collaboration network, track their evolution over time, and design an intervention to increase collaboration in these fields.

Network ethnography combines methods from social network analysis (SNA) and ethnography to simultaneously analyze the structure and cultural context of communities (Velden & Lagoze, 2013; Berthod et al., 2017). Previous studies have highlighted the need for using network ethnography to better understand the rationale behind network structures and designing interventions (Valente et al., 2015; Berthod et al., 2017). Using network

2017 Ethnographic Praxis in Industry Conference Proceedings, ISSN 1559-8918, epicpeople.org/intelligences

analysis methods allowed our team to visualize what these collaboration trends and organizational structures look like, as well as translate big data into actionable insights. Using ethnographic methods allowed our team to improve the accuracy of our models by helping us to identify more meaningful research teams, as well as gather the thick data necessary to understand why these communities emerged. By combining network and ethnographic methods and data sources, we believe that organizations can cultivate a more nuanced understanding of communities and their structures. In this paper, we provide an overview of our process and discuss ways that other researchers can leverage network ethnography to detect and study communities.

LITERATURE REVIEW

Scientific Collaboration

Scientific collaboration networks are a crucial channel for the diffusion of knowledge and innovation across disciplines and organizations. The number of research collaborations have been steadily growing each year (Dhand et al., 2016; Wuchty et al., 2007; Leahey, 2016). As a result, the field of team science has developed new metrics and methods for measuring team functioning in academic, government, and industry contexts using rigorous scientific methods (Börner et al., 2010, Stokols et al., 2008); Fiore, 2008; Falk-Krzesinski et al., 2010; Falk-Krzesinski et al., 2011; Rozovsky, 2015). These studies have found that research collaborations vary greatly both within and across organizations. This variation occurs primarily based on the number of collaborators, the amount of team members/organizations involved, the group's disciplinary orientation, and the team's end goal (Stokols et al., 2008a).

On one end of the collaboration spectrum, pairs or teams of researchers from the same department and university often work together to address a problem that advances the discipline (unidisciplinary). Many scientific collaborations are unidisciplinary due to the organizational structures, training process, and reward systems of many institutions. Two major advantages of single discipline collaborations are the ability to build consensus on what is at the edge of the discipline and work more quickly to produce results due to a shared disciplinary training and language (Sonnenwald, 2007; Jacobs, 2014). As a result, unidisciplinary collaborations are often prioritized by departments in the tenure and promotion process. However, this disciplinary focus causes silos of knowledge to emerge and fragments scientific research. This insulated process limits knowledge diffusion, innovation transfer, and general awareness of others work on similar topics across disciplines. It also has caused disciplines to develop very different standards for evaluating the impact of research and the productivity of researchers.

Departments also play an important role in compartmentalizing science. These disciplinary units are reflected in federal programs, funding opportunities, hiring practices, money allocation, graduate training, and the division of university resources. As a result, other studies have also found that disciplinary affiliation is a significant predictor of grant and publication network structure (Dhand et al., 2016). Our study seeks to disrupt this trend by identifying cross-disciplinary communities that present unique combinations of knowledge and providing researchers in these fields with incentives to collaborate on grant and publications.

We operationalize cross-disciplinary communities as groups with collaborations between researchers from at least two different disciplines which are often operationalized as departments. We believe that tracking these unique combinations of knowledge can help us to identify the emergence of new scientific fields that often transcend disciplinary boundaries. These emergent research communities also highlight the unique research focuses of a particular university and allow organizations to track the evolution of a specific field at their institution.

This study uses SNA and community detection algorithms to identify cross-disciplinary emerging research fields within a university's scientific collaboration network that could benefit from additional support. SNA provides a method for visualizing the social connections (edges) between individuals or other entities such as organizations (nodes). It is also an interdisciplinary field that illustrates the role that social relationships play in shaping individual and group behaviors (Wasserman & Faust, 1994).

This study builds on Valente (2012), Vacca et al. (2015), and other network scientists' research on the role of network interventions in science, health, management, and other fields. These studies outlined methods for identifying influential individuals within a community, segmenting communities, inducing increased interaction between already connected community members, or altering the whole network to create or remove individuals or connections within an existing community. In our study, we altered the network to increase the number of connections (edges) between group members who had not collaborated before.

SNA has also been used to measure and describe the structural patterns of scientific collaboration networks (Newman, 2004). Scientific collaboration network structures are driven by a variety of organizational, disciplinary, geographic, linguistic, and cultural factors. Spatial proximity, homophily, transitivity, past collaboration experiences, shared funding sources, disciplinary training, and organizational divisions like departmental and college affiliation play a major role in shaping the structure of scientific collaboration networks. Previous studies have found that spatial proximity is also a strong predictor for research collaboration, with those closest to one another most likely to collaborate (Katz, 1994; Newman, 2001; Olson & Olson, 2003). Analysis of homophily (the idea that birds of a feather flock together) in collaboration networks has found that researchers are more likely to work with other researchers who are similar to them in discipline, age, gender, and other background characteristics (Powell et al., 2005). Research on the transitivity (the concept that friends of my friend are also my friends) of scientific collaboration networks found that the clustering coefficient was greater than by random chance and occurs due to scientists often introducing their collaborators to their other collaborators (Wasserman & Faust, 1994; Newman, 2001).

Using Network Ethnography to Improve Community Detection

Scientific collaboration networks are often very large. The networks in our study range in size from between 3,000 and 5,000 nodes (individual researchers) and between 10,000 and 20,000 edges (collaborative links between researchers). Other studies have used a variety of different quantitative methods to examine specific research fields or groups (Chen, 2003; Chen, 2004; Braam & Van Raan, 1991a; Braam & Van Raan, 1991b; Massy, 2014). Community detection algorithms allow researchers to identify smaller meaningful units

within a larger network (Newman & Girvan, 2004; Blondel et al., 2008). This helps to compare and contrast communities within the network and identify trends.

The community detection literature has focused on identifying and examining communities through quantitative approaches. Many of these approaches have been adapted from methods developed in computer science, sociology, and statistics (Newman, 2004; Porter et al., 2009). Previous studies have highlighted the difficulty of testing and evaluating algorithms' effectiveness at community detection in real-world networks (Gregory, 2008; Fortunato, 2010). Newman (2008) also acknowledged this issue in his critique that,

The development of methods for finding communities within networks is a thriving sub-area of the field, with an enormous number of different techniques under development. Methods for understanding what the communities mean after you find them are, by contrast, still quite primitive, and much more needs to be done if we are to gain real knowledge from the output of our computer programs...Moreover, it's hard to know whether we are even measuring the right things in many cases.

We hypothesized that applying a network ethnography approach could help to better understand these communities, explain why these fields developed, and address these measurement issues. Therefore, we designed a study where we participated and observed these research communities to examine their cultural patterns and norms. We also interviewed a sample of the researchers identified as community members to gain insights on what these communities meant to them and whether we were measuring the right things.

There is a long history of anthropologists combining ethnographic and network methods to examine personal and small community networks (Radcliffe-Brown, 1940; Mitchell, 1974; Wasserman & Faust, 1994). Ethnography provides a scientific method for analyzing cultures. Our ethnographic study of a university used participant observation and semi-structured interviews with researchers to systematically study scientific collaboration culture by exploring organizational dynamics and disciplinary norms. This paper is part of a small but emerging body of literature applying ethnographic approaches to examine networks and other models created by big data sources (Velden & Lagoze, 2013; Wang, 2013; Nafus, 2014, Berthod et al., 2017). This study builds upon both this long ethnographic tradition and emerging field by adding two new dimensions. First, we introduce the concept of using ethnography to test the fit of community detection algorithms. Second, we describe how ethnography can help researcher gain the necessary buy-in and funding from organizational leaders to design and implement a network intervention.

In order to design a network intervention, researchers must possess a strong understanding of both the underlying network structures of a community and how its cultural context shapes group behavior. Ethnography provides a lens for uncovering these deeper sources of meaning. Therefore, network ethnography combines SNA and ethnographic methods to explore communities more deeply (Velden & Lagoze, 2013; Berthod et al., 2017). Combining qualitative and quantitative approaches enables scientists to cultivate a deeper understanding of both what network structures exist and why these cultural processes developed. This dual analysis adds depth to both sides of the analysis. It helps to develop more realistic models of community behavior, address measurement issues, and analyze decision-making processes at both an individual and group level.

Our study also draws on the vast literature on translational science (Woolf, 2008; Hörig et al., 2005; Sussman et al., 2006; Stokols et al., 2008a). This field has two main focuses. First,

the translation of scientific knowledge to facilitate the adoption of new clinical practices and improve public health outcomes. Second, translating knowledge within teams or transferring knowledge between teams, institutes, or other organizations. The translation process plays a vital role in the development of novel therapies, replication of scientific studies, and dissemination of findings to the public. This literature provided the study with a framework for translating this network intervention to other contexts.

METHODOLOGY

Our study used a multi-stage mixed method iterative approach to identify emerging scientific fields, study these communities' team dynamics and cultural norms, design an experimental intervention, and create new research teams. Combining the methods of participant observation, social network analysis, community detection, ethnographic interviews, web profile analysis, and surveys allowed our team to simultaneously analyze big and thick data. Utilizing this ethnographic network approach helped us gain a more nuanced understanding of both the network structures and cultural norms of communities we were studying. It also allowed us to design and implement our network intervention.

We began our project by conducting conducted six months of participant observation at a large university. We participated and observed a variety of activities across the university's sixteen colleges. These activities included participating in multidisciplinary scientific collaborations, observing scientists collaborating, talking with thirty-six researchers about their past and current collaboration experiences, and attending meetings across scientific disciplines both prior to and throughout this process. This contextual inquiry provided us with the necessary context to plan our study and understand disciplinary differences in collaboration norms. This context also helped us understand why this intervention was needed and where to focus our efforts for the greatest impact. This was necessary for pitching our project to organizational leaders in order to get the necessary buy-in, funding, and other resources that were requires to implement our experimental network intervention. We continued our participant observation throughout the duration of the study. This continued participant observation helped us to ask better questions during our interviews, test our models, and adapt to organizational changes.

Our next step was to acquire the scientific collaboration data for all researchers at the university so that we could map their collaborations. We defined scientific collaborators as co-investigators on at least one funded grant or co-authors on at least one academic publication. Therefore, we used grant and publication data to create scientific collaboration networks. We chose this approach because our participant observation and interviews with researchers clearly demonstrated that these were the two most common products of formal research collaborations in academia. This finding is consistent with previous studies on scientific collaboration. We then used this data to map the university's scientific collaboration network using UCINET and R. Network analysis allowed us to visualize university level trends in collaboration (Figure 1).



Figure 1. The university's scientific collaboration network in 2013. Each node (dot) in the visualization (Figure 1) represents a researcher and each tie (line) represents a collaboration between two researchers. Nodes are colored to distinguished between collaborations on a publication, grant, or both (multiplex).

After looking at organizational trends, we used community detection algorithms to identify emerging research communities within this network. We defined an emerging research community as a group of researchers collaborating on a related scientific topic. This definition is consistent with previous studies and our ethnographic interviews with researchers. These communities were identified using the Louvain method of community detection (Blondel et al., 2008). We chose this method because it maximizes modularity. This allowed us to identify very specific communities (Fortunato, 2010). It also allowed us to identify persistent (longitudinal) rather than temporal (single-year) communities. Our participant observation and interviews with researchers revealed clear differences in the scope and type of collaboration that occurs in short and long-term research projects. As a result, we chose to prioritize longitudinal communities because we wanted to map emerging trends across the university rather than individual research projects.

We identified emerging research communities in a few different ways in order to capture the variety of collaborations we observed through our participant observation and learned about through our interviews. First, we used a sum method to identify communities where investigators collaborated at some point between a certain period. Second, we used a comembership method to identify communities where investigators collaborated consecutively for a certain period of time. Figures 2 and 3 highlight the differences between the sum and co-membership methods by looking at the same community of researchers over a two-year period. We then examined how the communities changed when we changed the criteria from between one and five years.



Figure 2. An example of a community using Louvain sum method. The image on the left highlights the community's location within the university's scientific collaboration network. The image on the right represents a community that contains individuals who have been in the community at any time in two of the past five years (not necessarily consecutively). Nodes are colored by academic unit.



Figure 3. An example of a community using Louvain co-membership method. The image on the left highlights the community's location within the university's scientific collaboration network. The image on the right contains investigators who have been part of the same community for two or more consecutive years in the past five years. Nodes are colored by academic unit.

Examining how the communities' network structures changed based on the method and criteria helped us to visualize the different types of collaborations that participants described in interviews. Detecting a variety of different communities also allowed us to test multiple networks with participants during our interviews. This user feedback helped us understand the key differences between these different community visualizations and modify our criteria to create models that better fit reality and their cultural context.

After we identified these communities, we performed a profile analysis of each researcher or clinician identified in an emerging community to determine their research topics. We began this process by looking at each researcher or clinician's university website and performing a web search to identify their scientific expertise. We read through the names of the publications, grants, and collaborators they mentioned on their faculty web page or other online profiles. Then, we identified common keywords for each researcher. We compared these keywords to other researchers identified as part of their community to identify shared topics within research groups. We classified their research field by combining the most commonly shared topics across the community. Comparing keywords across community members was often difficult because their profiles varied greatly in their scope, tone, the level of detail, and the types of information they provided.

Next, we solicited feedback from twenty-three researchers on our network visualizations through ethnographic interviews. We scheduled three distinct stages of interviews in order to iteratively test our visualizations and get participant feedback.

In the first stage, we conducted semi-structured interviews with three researchers to learn about their collaboration experiences and ask for feedback on the visualizations. Our goal for this first round of ethnographic interviews was to evaluate the measurement of our community detection instrument and determine if we should adjust our time frame, inclusion criteria, or exclusion criteria. We selected three researchers from two of the communities (from different departments and colleges) that were working on similar research topics (hypertension) to compare and contrast their visualizations. We then interviewed them for approximately one hour about their research collaborations to get feedback on our network visualizations. The user feedback from this stage let us know that we needed to revisit our criteria.

In the second stage, we solicited additional feedback from six researchers on our revised visualizations via email. Two of these researchers participated in the first round of interviews, one was unable to meet for an interview due to scheduling issues, and three had been previously interviewed about their collaboration experiences for another research project we conducted. Our goal for this stage was to test the new visualizations to determine whether the changes to the time frame helped to identify emergent research communities that reflected researchers' perceptions. The user feedback from this stage verified that the criteria changes we made allowed us to identify emerging research fields that fit users' mental models and were more positively perceived by respondents.

In the third stage, we conducted additional ethnographic data on these emerging research fields through interviewing sixteen scientists and clinicians. We used these interviews to further evaluate the revised visualizations, get feedback on their strengths and weaknesses, and collect narratives on team science and the translational process. Our goal for this stage of the process was to cultivate a better understanding of these communities in order to design a network intervention to propose new research pairs. The feedback from this stage allowed us to move forward with our network intervention.

In the final stage of this project, we identified pairs of researchers from the same emerging research community. In order to be eligible to participate in our network intervention, communities had to contain at least one member from the health sciences and could not have collaborated on a grant or publication in the past three years. We identified both a treatment and control group that had similar network structures and shared similar group characteristics (i.e. level of interdisciplinary). We then selected pairs from the periphery of these communities.

We then conducted another round of web profile analysis to evaluate whether or not the pairs were viable. Through this analysis, we identified fifteen pairs of thirty investigators and sent them an email outlined the study's goals and incentives. Each member of the pairs who participated was awarded \$1,500 in professional development funds. This funding was contingent upon the pairs attending a group meeting to provide an overview of our network intervention and jointly submitting a letter of intent that described a new collaborative project for potential pilot funding from the institute. Based on the quality of their letter of

intent, three pairs were asked to submit a full grant proposal. These proposals were peerreviewed and the institute we partnered with selected which pair received a pilot award of up to \$25,000 to complete their project.

RESULTS

Our analysis revealed two distinct types of communities: cores and bridges. Core communities are research groups that are concentrated within the same section of the university's scientific collaboration network with dense connections typically between a small (<20) group of investigators (Figure 4). Bridging communities represent groups that span across different sections of the university's scientific collaboration network with sparser connections typically between a large (>20) group of investigators (Figure 5). We also examined the network structure and composition of each community (Leone Sciabolazza et al., 2017).



Figure 4. An example of core community. Each node (dot) represents a researcher. Nodes are colored by the researcher's college affiliation.



Figure 5. An example of bridging community. Each node (dot) represents a researcher. Nodes are colored by the researcher's college affiliation.

First Stage of Interviews

The first stage of interviews revealed five common themes: measurement issues due to criteria limitations, missing data, respondents' poor perceptions of the visualizations, the role of mentorship in these networks, and researchers' motivations for transdisciplinary collaboration. All of the respondents made it clear that these visualizations were too narrow. One scientist responded to the visualization by explaining,

The search criteria are pretty strict [...] I get what you are getting at, it's just you are gonna miss some things because [you know] it can still be a consistent collaboration but some of the people may change over time [...] It's a way to reduce the complexity of this but I just worry that you lose a lot.

They also reported that many of their critical collaborators and more recent projects were missing from these visualizations. This feedback indicated a need to adjust the time frame to better capture their current collaborations and breadth of their research.

The comments from all three interviews suggested that though we were trying to capture emerging communities we were actually identifying stable communities in this first iteration. By identifying only communities where members had collaborated in all five of the past five years, we were capturing small core groups that had a strong history of collaboration. These communities often represented strong relationships between mentors and their mentees. Though these are important relationships to examine, they were not the target of our analysis.

We also found through the interviews that the networks looked simpler and more homogeneous than they actually were in terms of their department/college identification and the strength of the collaboration. The respondents said that though the collaborations highlighted in this group were true, they did not highlight their most innovative or interdisciplinary work on emerging topics. This is a measurement issue that was addressed by adjusting the criteria from five years to a two to three-year period. This shift allowed us to capture much larger, more dynamic communities, working on more recent scientific issues in emerging research fields.

Interviewees also noted that many of their critical collaborations from this period were missing though they believed that they fit the inclusion criteria. It seemed like on a large scale the network data was great at identifying trends, but its accuracy from year to year can be limited. The issues of some missing data points are problematic when you have set strict community criteria. Thus, it is essential to loosen the criteria to improve the accuracy of the models. If a collaborator needs to be in the community all five years, small issues can result in big problems like a lack of appearance within the network. However, if we relax the criteria to two or three years these collaborations will still appear though the strength of the tie between researchers may be underreported.

These initial interviewees also found the network visualizations interesting but perceived them negatively. They were frustrated that these visualizations did not highlight what respondent's self-identified as their most emergent research collaboration/topic. They also believed that they minimized the interdisciplinary work they were often proudest of. These responses highlight the importance of getting feedback on models created through big data sources. Negative perceptions of visualizations can signal a deeper problem with the tool you are building. It is important to create visualizations that fit people's mental models of the phenomenon you are trying to map or they will simply dismiss your results as irrelevant.

The interviews highlighted the critical role that mentorship plays in developing stable research communities. Mentors play a major role in socializing mentees and setting collaboration norms. Mentees introduce mentors to new ideas, challenge their assumptions, and introduce them to other faculty members (typically through co-membership on graduate committees). The mentor-mentee relationship forms a strong bond that does not respond to funding mechanisms the same way that other collaborations do. Mentors and mentees also play a critical role in brokering new collaborations by introducing them to new collaborators outside their network. They also provide a sounding board for questioning assumptions and developing new insights.

Interviewees explain that their primary motivations for collaborations were intellectual curiosity and passion for the project. They also shared the belief that interdisciplinary collaborations can be more difficult but they are also more rewarding long term. They highlighted different expectations, administrative barriers, disciplinary norms, and academic language as the primary challenges when engaging in interdisciplinary team science.

After this first round of interviews, we shared this feedback within our team and redefined our criteria. Based on our user feedback from these initial interviews, we revised our models and adjusted the time frame from five years to both two and three years. We chose to create both two and three-year visualizations because we believed based on the interview feedback that sharing both types of communities with interviewees would allow us to examine different types of collaborative projects.

Second Stage of Interviews

After revisiting our criteria and soliciting additional user feedback to test our changes, we collected additional ethnographic data on these emerging research fields through interviewing sixteen scientists and clinicians. These interviews highlighted the many challenges of capturing emerging research fields. Emergent fields are difficult to capture due to the fast pace of science and slow speed of the peer-review process. The networks highlighted the general trends in the community but were not able to capture the subtle nuances of group dynamics and many informal collaborations between researchers that did not result in a publication or grant. In order to address this issue, interviewees suggested that we integrate data sources from other collaboration products like grant applications, patents, and abstracts from conference presentations to identify more recent research collaborations and emerging topics. However, interviewees agreed that these networks were helpful tools for starting a conversation about their collaborations.

The co-membership (consecutive) networks were identified by respondents as the most accurate by interviewees. The two-year consecutive co-membership networks were identified by the most interviewees as the best representation of their emerging research field. However, the three-year consecutive co-membership network was seen as the most meaningful by many respondents, especially those with larger collaboration networks.

We found that interviewees received medium sized communities more positively than small or large communities. Communities of ten to forty researchers were typically most meaningful to participants. When communities were smaller than ten people, respondents often said it was too small and highlighted only their core collaborators. Communities of more than forty researchers were helpful for situating researchers' context within the larger university's network, but they often did not know of all of the community members. This was especially true in communities that had more than one hundred members.

Respondents said that they enjoyed seeing where their work fit into the larger scientific research landscape at their university. However, they were often overwhelmed by looking at the larger visualizations and did not derive a lot of meaning from these types of networks beyond being excited to share them with others. Most respondents identified their middle-sized community as the most accurate representation of their collaboration network and the emerging research field in which they were identified. They explained that this community best balanced the breadth and depth of their research in a meaningful way.

Scientists' perceptions of themselves and their research also shapes the way they interpret these visualizations. Those with a high level of status within the university and collaborators viewed these visualizations as proof of their accomplishments. Whereas, those with a lower level of status within the university and fewer collaborators often saw these visualizations as tools that demonstrated their self-worth, the level of their connections, and the value of their research. One associate professor even said that she planned to use these visualizations as resources for her tenure packet. Other faculty members expressed interests in using the visualizations for other purposes. These included fostering new collaborations at a department level or starting new collaborations within their research group. The act of sharing and starting a conversation about these visualizations with faculty members served as a form of intervention to raise community members' awareness of these emerging research fields and their collaboration patterns. Though these efforts were not part of our initial goals for the project, they demonstrate the potential network effects that sharing these visualizations can have on research communities.

Researchers define their research community in different ways. This definition varies based on their individual conception of collaboration and the disciplinary context of their field. These different conceptions of communities vary greatly by research context. Interviewees provided different narratives about their collaboration in response to each community they were shown. This means that there is no clear best way to identify and visualize a research community. It also means that different communities reflect the team dynamics and focuses of many of the same researchers and clinicians during a distinct time period. As one researcher explained,

All of them actually represent probably components of my research...Each of the little connection charts encompass kind of different components of the collaborations I've done. Ranging from I think the first one tended to be more cell biological and biochemical and I think the last one was more focused on kind of the translational.

This description highlights the potential utility of sharing these network visualizations as a tool for starting conversations about the processual nature of translational and team science.

Most researchers responded positively to the network visualizations. They seemed excited to view the visualizations, provide feedback, and highlight connections between themselves and other researchers throughout the university. They also appreciated seeing the overlap between their different fields. Examining these connections between themselves and other researchers and clinicians seems to give many interviewees a sense of purpose. Most respondents were able to draw connections between their research and others they knew or had heard of before but had not collaborated with. It also allowed them to highlighted the impact of their work and its impacts beyond their discipline or academia.

Though the overall comments were very positive, researchers also pointed out several limitations of using these visualizations to examine emerging research fields. Most researchers were able to identify one or two of their key collaborators who were missing from the visualization. Often, these were collaborations that had not yet resulted in an awarded grant or publication, but that had resulted in a grant proposal or conference presentation. A few times the person they identified was involved but they were missing because either they had been grouped into another community or there was a disambiguation error with the data.

Another limitation was the presence of investigators who had either left the university or died in the past three years. Interviewees did not really mind seeing the names of students who had graduated, post-docs who had received positions at other universities, or faculty members who had recently retired. Rather, they often used this as an opportunity to share stories about their past collaborations or their current position. However, interviewees were often upset to see the name of researchers who had passed away recently. Many interviewees were also sad or frustrated to see the names of researchers who recently left for other positions. They explained that though the networks were technically correct, the presence of investigators in their community who were no longer at the university seemed wrong. We tried to remove investigators who were no longer at the university. However, this process was more difficult than we initially anticipated.

Collaborations within these fields were described by interviewees as a way to both improve the quality of their research and ride out the storms of funding in an increasingly competitive market. Many interviewees described the turn taking process of authorship/investigator status that often occurs when publishing articles and writing grants in in long term collaborations. They also highlighted the role that career level plays in collaboration dynamics in terms of status, power, funding, and priorities. Some respondents said that it was often easier to work with those at the same level so their goals were aligned; whereas, others said it was helpful to work with people at different levels to help them cope with funding lulls within the group. Interviewees' research narratives illustrated the role their mentors played in shaping their collaboration style and choices when choosing collaborators. Mentees often mirror the collaboration patterns demonstrated by their mentor, whether consciously or unconsciously.

LIMITATIONS

Using network metrics to select pairs and introduce investigators to one another provides a new and exciting method for testing and evaluating the effects of a pilot program or network intervention. This method has the potential to increase the impact of the intervention. However, it is also important to highlight some of the potential limitations and ethical challenges of applying these approaches to alter networks and their implications for other studies.

Designing a network intervention that alters community dynamics and structures requires researchers to make a series of decisions that have implications for the people and communities they identify, as well as those who are not selected to participate. Community detection and other algorithmic decisions are often marketed as objective criteria. However, as Gregory (2008) writes,

There is no standard definition of community and no consensus about how a network should be divided into communities.

Also, as this paper demonstrates, these methods require individuals to make subjective choices about what and how to measure. These individual choices are also shaped by the priorities and culture of the organizations they are embedded within. Ethnography can help us better understand these choices and their organizational context.

However, it is important to remember that algorithms are powerful and not neutral. There is a great deal of power that comes from creating these algorithms and making decisions based on their results (Lustig et al., 2016). Using ethnographic methods to engage with the community members you are attempting to map and soliciting user feedback can help to reduce this power imbalance and create more realistic models of big data sources. Yet, it does not negate the power imbalance between researchers and participants that is built into experimental design. Therefore, it is important to also acknowledge the potential limitations of these models. Researchers' subjective decisions can have a significant impact on the groups that they map and those that they ignore.

In the case of this study, these communities were used to identify pairs of researchers that would be offered exclusive access to apply for pilot award/seed grant funding. This type of award can make a significant impact in a researchers' career. Thus, providing access to some faculty members over others requires critical decision making and hard choices. Controlling this process provides a great deal of power to algorithm creators (Lazer, 2015). We believe that our network intervention can help to reduce some of the biases of traditional funding mechanisms that have been reported in previous studies (Helmer et al., 2017; Marsh et al., 2008; Lee et al., 2008). However, it also introduces its own bias and challenges. These issues would be compounded if the stake of the intervention were higher. Therefore, it is important to keep these ethical challenges in mind when using community detection algorithms to design interventions.

Planning and implementing these types of interventions also presents many logistical challenges. However, we believe that they can also lead to new ways of fostering innovation and knowledge production across an organization. Incorporating ethnography throughout our research design helped us to gain buy-in from stakeholders and test our assumptions. However, we experienced several recruitment challenges when contacting pairs about participating in the network intervention.

Since the pair's participation is contingent upon their partners' desire to participate, it can be difficult to identify viable pairs. Some investigators were eager to participate but we could not identify an eligible collaborator in their community who was interested in our study. One way to address some of the issues we experienced in our study would be to ask a researcher who knows both members of the pair to introduce them and suggest that they collaborate. A request from an existing collaborator could be more impactful than an email from someone they do not know. Our participant observation and interviews revealed that these types of recommendation from current collaborators often play a critical role in fostering new scientific collaborations and developing rapport between new team members. Future studies could also adopt this type of approach when recruiting users for a site or proposing a new collaboration with another team.

CONCLUSION

Identifying emerging research communities provides a useful method of visualizing team science its network effects. As demonstrated throughout this paper, combining network analysis with ethnographic research can help researchers identify trends and opportunities organizational change. Our mixed method approach improved our knowledge and models of scientific communities by helping us combine insights from big data and thick data (Wang, 2013). Applying an ethnographic lens allowed us to connect scientific collaboration networks

with larger narratives about team science and translational research. Other researchers could also use the network ethnography method to gain a more nuanced understanding of both the network structures and cultural norms of communities they are studying and leverage this knowledge to design better experiments. We believe that using this approach can help organizations track emerging research fields in their industry. It can also help organizations better support existing teams, provide funding to encourage new groups to collaborate on these topics, and translate their findings to the public.

Social network analysis and community detection algorithms allowed our team to create compelling visuals that clearly documented organizational trends in a way that was easy for organizational leaders to consume. Ethnographic approaches helped our team gain the necessary cultural context to develop new perspectives on team dynamics and tell compelling stories about why our intervention was needed and what type of organizational impact we expected it to generate. As expected, our analysis revealed a significant overlap between the research communities we detected and traditional department, colleges, and academic units. However, our participant observation and interviews with community members revealed that interdisciplinary collaborations were central to their research aims. Researchers took pride in their interdisciplinary work in these emerging fields even when it was actively discouraged by their peers or dis-incentivized by their departments.

Understanding both what interdisciplinary research collaborations looked like, how they evolved over time, and why they were so important to both scientists and clinicians helped us to demonstrate to organizational leaders the need for funding and support for this type of network intervention. Other researchers could use a similar approach to identify users' pain points and demonstrate to their managers or C-suite executives the need to invest in their solution. Combining the network visualizations that prove that a problem exists with user stories that demonstrate their pain points helps researchers make a stronger case for their proposed changes.

Soliciting feedback from respondents through ethnographic interviews was a critical part of improving our models to better fit the realities of team science. During the first round, all three interviewees agreed that the models we showed them were too simplistic, focused too much on their long-term collaborations, and were not emphasizing their current research agendas. They were also disappointed that the networks did not capture the diversity of their achievements, especially when these were interdisciplinary collaborations. We were able to incorporate their feedback into the models by shortening the length of time individuals needed to be in the same community. After reducing the time period from five to between two and three years, interviewees responded positively to the networks and reported that they saw them as meaningful representations of their emerging research field. Other researchers often adopt this type of approach when testing prototypes or designs. However, our study shows that researchers could also use this method to iteratively test their plans for experimental interventions.

Sharing these visualizations with researchers transformed the way they looked at their research community and their place within the larger university. They also solicited thick data from scientists on the cultural context on these research communities, which allowed us to improve the fit of our models. Respondents enjoyed seeing their position in the community and explaining the narrative behind why that connection existed and how the collaboration started and developed over time. Respondents were also excited to interact with the models and provide feedback on who was missing and share their explanations of why they might be

missing. It was rather intuitive for them to make connections between their work and the work of others in the group even if they have not worked with them directly. In many cases, people were excited to see that a colleague they knew but had not formally collaborated with was also a part of their community.

Typically, there were only one or two people who they did not know in their research community, especially the one they believed most accurately reflects their collaboration. This finding suggests that these research communities are more than just abstract representations of science but represent real communities of scientific inquiry and connection. They highlight both actual collaborations on grants and publications as well as thought communities that researchers use to disseminate and share their ideas with their colleagues working on related but distinct topics. By modeling five different ways that respondents could conceptualize their research communities, we were able to examine different types of communities and collect ethnographic insights on the group's formation and development over time. Some scientists found the community of their closest collaborators on a specific topic as the most meaningful group. Yet, they also drew meaning from the fact that they were connected to other researchers through a training or larger grant that focused on broader issues. Ethnographic interviewing also allowed us to explore how a sense of belonging shaped a team's research focus, norms, and power dynamics.

These findings have applications for the way that researchers examine online communities. They could be applied to look at the connections between real and online communities and examine how social networks evolve longitudinally. Future studies could share similar networks (based on their interactions with other users or previous tasks that they have completed on the platform) with users and ask them explain what these networks mean to them and how they would change them. These types of interactions with users are critical to understanding their mental models, motivations, and reason why they engage in a particular activity.

Researchers could also conduct a network ethnography within their own organization to improve communication or collaboration across their team or department. It can also help to empower users or team members to share their stories, understand why the current patterns exist, and identify opportunities for growth. Sharing network visualizations with team members can also be a great way of eliciting conversations about collaborations between a team and its leaders.

Ethnographers are always attempting to map new fields and shift the discipline through their research. This paper highlighted the role that network ethnography played in helping our team identify emerging fields and build new teams at a university. We believe this same method could also be applied in almost any field to answer the questions: what are a community's existing structures, why did they develop in this way, and how could we disrupt this structure. Network ethnography can help researchers understand organizational or user patterns and promote organizational or product changes. We believe that network ethnography can be an important tool for helping ethnographers uncover strategic research insights and gain the necessary buy-in from organizational leaders to test their assumptions or scale their work. The process of conducting and sharing data from a network ethnography can also help researchers build strategic partnerships across their organization and make their research findings more tangible. We believe that building these strategic partnerships can be a first step in making the case for giving ethnographic research a seat at the table. **Therese Kennelly Okraku** is a design researcher at Microsoft who conducts research on how teams of IT professionals use technological support and services. She received her PhD in Anthropology and a certificate in Clinical Translational Science from University of Florida. *thereseokraku@microsoft.com*

Valerio Leone Sciabolazza is a Postdoc Associate at University of Florida who conducts research on spatial econometrics, the econometrics of networks, the new science of networks, team science, and migrant networks. He received a PhD in International Economics and Finance from Sapienza University. *sciabolazza@ufl.edu*

Raffaele Vacca is a Research Assistant Professor at the University of Florida Department of Sociology and Criminology & Law. He is also affiliated with the Clinical and Translational Science Institute (CTSI) and BEBR. He is a sociologist and social network analyst, and his main research interests are migration and ethnicity, health disparities, and scientific and knowledge networks. *r.vacca@ufl.edu*

Christopher McCarty is director of the University of Florida Survey Research Center, the director of the Bureau of Economic and Business Research, and chair of the Anthropology department. He has worked on the adaptation of traditional network methods to large-scale telephone and field surveys and the estimation of hard-to-count population, such as the homeless and those who are HIV positive. His most recent work is in the area of personal network structure. He also developed a program called EgoNet for the collection and analysis of personal network data. *ufchris@ufl.edu*

Acknowledgments – The research reported in this paper was supported by the University of Florida Clinical and Translational Science Institute, which is supported in part by the NIH National Center for Advancing Translational Sciences under award number UL1TR001427. Thank you to the EPIC reviewers and curators, especially Julia Haines for your insightful comments and suggestions which played an important role in shaping the development of this paper. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, University of Florida, or Microsoft.

REFERENCES CITED

Adams, Jimi and Ryan Light.

2014 Mapping interdisciplinary fields: Efficiencies, gaps and redundancies in HIV/AIDS research. *PloS one* 9.12 (2014): e115092.

Berthod, Olivier, Michael Grothe-Hammer, and Jörg Sydow.

2017 Network ethnography: A mixed-method approach for the study of practices in interorganizational settings. *Organizational Research Methods* 20.2 (2017): 299-323.

Börner, Katy, Michael Conlon, Jon Corson-Rikert, and Ying Ding.

2012 VIVO: A semantic approach to scholarly networking and discovery. Synthesis Lectures on the Semantic Web: Theory and Technology 7.1 (2012): 1-178.

Braam, Robert R., Henk F. Moed, and Anthony FJ Van Raan.

1991a Mapping of science by combined co-citation and word analysis I. Structural aspects. *Journal of the American Society for Information Science* 42.4 (1991): 233.

Braam, Robert R., Henk F. Moed, and Anthony FJ Van Raan.

1991b Mapping of science by combined co-citation and word analysis II. Dynamical aspects. *Journal of the American Society for Information Science* 42.4 (1991): 252. Burrell, Jenna.

2009 The field site as a network: A strategy for locating ethnographic research. *Field Methods* 21.2 (2009): 181-199.

Cambrosio, Alberto, Peter Keating, and Andrei Mogoutov.

2004 Mapping collaborative work and innovation in biomedicine: A computer-assisted analysis of antibody reagent workshops. *Social Studies of Science* 34.3 (2004): 325-364.

Chen, Chaomei.

2003 Visualizing scientific paradigms: An introduction. Journal of the Association for Information Science and Technology 54.5 (2003): 392-393.

Chen, Chaomei.

2004 Searching for intellectual turning points: Progressive knowledge domain visualization. Proceedings of the National Academy of Sciences 101.suppl 1 (2004): 5303-5310.

Chen, Chaomei, and Jasna Kuljis.

- 2003 The rising landscape: A visual exploration of superstring revolutions in physics. Journal of the Association for Information Science and Technology 54.5 (2003): 435-446.
- Crabtree, Andy, David M. Nichols, Jon O'Brien, Mark Rouncefield, and Michael B. Twidale.
- 2000 Ethnomethodologically informed ethnography and information system design. Journal of the Association for Information Science and Technology 51.7 (2000): 666-682.

Dunn, Adam G., and Johanna I. Westbrook.

2011 Interpreting social network metrics in healthcare organisations: A review and guide to validating small networks. Social Science & Medicine 72.7 (2011): 1064-1068.

Fortunato, Santo.

2010 Community detection in graphs. *Physics Reports* 486.3 (2010): 75-174.

Geertz, Clifford.

1994 Thick description: Toward an interpretive theory of culture. *Readings in the Philosophy of Social Science* (1994): 213-231.

Geiger, R. Stuart, and David Ribes.

2011 Trace ethnography: Following coordination through documentary practices. System Sciences (HICSS), 2011 44th Hawaii International Conference, 2011.

Girvan, Michelle and Mark EJ Newman.

2002 Community structure in social and biological networks. Proceedings of the National Academy of Sciences 99.12 (2002): 7821-7826.

Gläser, Jochen, and Grit Laudel.

2001 Integrating scientometric indicators into sociological studies: methodical and methodological problems. *Scientometrics* 52.3 (2001): 411-434.

Gläser, Jochen, and Grit Laudel.

2007 The social construction of bibliometric evaluations. The Changing Governance of the Sciences (2007): 101-123.

Gregory, Steve.

2008 A fast algorithm to find overlapping communities in networks. *Machine learning and knowledge discovery in databases* (2008): 408-423.

Helmer, M., Schottdorf, M., Neef, A., and Battaglia, D.

2017 Gender bias in scholarly peer review. *eLife*, 6 (2017): e21718.

Howard, Philip N.

2002 Network ethnography and the hypermedia organization: New media, new organizations, new methods. *New media & society* 4.4 (2002): 550-574.

Jetson, Judith A., Mary E. Evans, and Wendy Hathaway.

2009 Evaluating the Impact of Seed Money Grants in Stimulating Growth of Community-Based Research and Service-Learning at a Major Public Research University. *Journal of Community Engagement and Higher Education* 1.1 (2009).

Lee, C. J., Sugimoto, C. R., Zhang, G., and Cronin, B.

2013 Bias in peer review. Journal of the Association for Information Science and Technology, 64.1(2013): 2-17.

Marsh, Herbert W., Upali W. Jayasinghe, and Nigel W. Bond.

2008 Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability. *American psychologist* 63.3 (2008): 160.

Massey, Tilmann.

- 2014 Structuralism and Quantitative Science Studies: Exploring First Links. *Erkenntnis* 79.8 (2014): 1493-1503.
- Meltzer, David, Jeanette Chung, Parham Khalili, Elizabeth Marlow, Vineet Arora, Glen Schumock, and Ron Burt.
- 2010 Exploring the use of social network methods in designing healthcare quality improvement teams. Social science & medicine, 71.6 (2010), 1119-1130.

Newman, Mark.

2008 The physics of networks. *Physics today* 61.11 (2008): 33-38.

Newman, Mark.

2004a Detecting community structure in networks. The European Physical Journal B-Condensed Matter and Complex Systems 38.2 (2004), 321-330.

Newman, Mark.

2004b Fast algorithm for detecting community structure in networks. *Physical Review E*, 69.6 (2004), 066133.

Pilerot, Ola.

- 2014 Making design researchers' information sharing visible through material objects. *Journal of the Association for Information Science and Technology* 65.10 (2014): 2006-2016.
- Porter, Mason A., Jukka-Pekka Onnela, and Peter J. Mucha.

2009 Communities in networks. Notices of the AMS 56.9 (2009): 1082-1097.

Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., and Parisi, D.

2004 Defining and identifying communities in networks. Proceedings of the National Academy of Sciences of the United States of America, 101.9 (2004), 2658-2663.

Rohrbach, L. A., Grana, R., Sussman, S., and Valente, T. W.

- 2006 Type II translation: transporting prevention interventions from research to real-world settings. Evaluation & the Health Professions 29.3 (2006): 302-333.
- Sciabolazza, Valerio Leone, Raffaele Vacca, Therese Kennelly Okraku, and Christopher McCarty.
- 2017 Detecting and analyzing research communities in longitudinal scientific networks. *PloS one* 12, no. 8 (2017): e0182516.

2017 EPIC Proceedings

Scott, John.

2017 Social network analysis. Sage.

Small, Henry.

2003 Paradigms, citations, and maps of science: A personal history. *Journal of the Association for Information Science and Technology* 54.5 (2003): 394-399.

Smith, Stephanie S.

- 2016 A three-step approach to exploring ambiguous networks. Journal of Mixed Methods Research 10.4 (2016): 367-383.
- Vacca, Raffaele, Christopher McCarty, Michael Conlon, and David R. Nelson.
- 2015 Designing a CTSA Based Social Network Intervention to Foster Cross Disciplinary Team Science. *Clinical and translational science* 8, no. 4 (2015): 281-289.
- Valente, Thomas W., Lawrence A. Palinkas, Sara Czaja, Kar-Hai Chu, and C. Hendricks Brown.
- 2015 Social network analysis for program implementation. *PloS one* 10, no. 6 (2015): e0131712.

Valente, Thomas W.

2012 Network interventions. Science 337.6090 (2012): 49-53.

Van den Besselaar, Peter.

2000 Communication between science and technology studies journals: A case study in differentiation and integration in scientific fields. *Scientometrics* 47.2 (2000): 169-193.

Velden, Theresa and Carl Lagoze.

2013 The extraction of community structures from publication networks to support ethnographic observations of field differences in scientific communication. *Journal of the Association for Information Science and Technology* 64.12 (2013): 2405-2427.

Wasserman, Stanley and Katherine Faust.

1994 Social network analysis: Methods and applications. Vol. 8. Cambridge University Press.

Wang, Tricia.

2013 Big data needs thick data. *Ethnography Matters* 13 (2013).

Watts, Duncan J.

2004 The "new" science of networks. Annual Review Sociology 30 (2004): 243-270.

Wray, K. Brad.

2006 Scientific authorship in the age of collaborative research. *Studies in History and Philosophy of Science Part A* 37.3 (2006): 505-514.

Zuccala, Alesia.

2006 Modeling the invisible college. *Journal of the Association for Information Science and Technology* 57.2 (2006): 152-168.