Trace Interviews: An Actor-Centered Approach

ELIZABETH DUBOIS
HEATHER FORD
University of Oxford, England[1]

The current communications environment is characterized by a complex and hybrid system. Individuals use multiple digital platforms in various ways to communicate politically. This presents both theoretical and methodological challenges. As a response, we propose *trace interviewing,* an actor-centric method that employs visualizations of a user's digital traces during the interview process. Trace interviews are useful for enhancing recall, validating trace data-generated results, addressing data joining problems, and responding to ethical concerns that have surfaced in the current era of surveillance and big data. If the challenges of the method are successfully navigated, trace interviewing could allow researchers to respond creatively to new questions about the current, complex political communication environment.

Keywords: research methods, mixed methods, interviewing, social network analysis, data visualization, practice

Introduction

The Internet age is characterized by a hybrid media system (Chadwick, 2013) in which a diversity of authors use and share multiple platforms and modes for political communications. Although the mass media holds significant sway over political narratives, more conversations now happen outside the purview of media gatekeepers (Bennett & Segerburg, 2012). Citizens comment on the conduct of politicians via Twitter, mobilize toward political action using social media platforms like Facebook, and narrate the stories of political events via Wikipedia. Political communications happen across, and not only within, media platforms in what is best described as a "large technological system" of media and ICTs (Schroeder, 2007).

Elizabeth Dubois: elizabeth.dubois@balliol.ox.ac.uk
Heather Ford: heather.ford@oii.ox.ac.uk

This large technological and hybrid media system is challenging for researchers, because current tools cannot accommodate the multiplicity of modalities and the scale of communications that characterize the current environment. In response, methodological approaches tend to focus on particular tools and tend to be located on either side of the qualitative/quantitative divide.

While examining patterns of political communication according to a single tool or platform is useful, investigating the ways in which people practice political communication across different platforms and within the existing large technological system may be even more valuable (Garrett et al., 2012). Purely qualitative approaches such as interviews and participant observation are useful for gaining rich perspectives of users' experiences, but they can be subject to biases from self-reporting and often fail to capture the material context of use. Although big data techniques, surveys, and other quantitative approaches enable large-scale representations of use within particular platforms, they fail to capture the person in the data and tend to focus on single data sources rather than capturing the media system holistically (boyd & Crawford, 2011).

Additionally, ethical issues are involved in the tracking of individuals' communications. Focusing on actors rather than particular media platforms is useful for gaining a holistic perspective of media and ICT systems, but users are concerned about the loss of control when third parties track their movements through a system. Reaction to experimental research on Facebook in 2014 (Goel, 2014) is one in a series of events that draws attention to the need for more user-centered methods that involve those users in the research process rather than merely harvesting their data.

The question for researchers is both practical and ethical. In a data-rich and multimodal environment, how do we ensure the validity and ethical cogency of our results? As Rogers (2009) points out, there is a need to "research *with* the Internet" (p. 5) wherein new methods are developed using digital tools that allow us to study not only online culture but also the interaction between humans and technologies. This article provides a framework for trace interviews. Trace interviews combine qualitative and digital data-driven approaches to understand citizens' political communication in a large technological system. This actor-centric approach fills a gap in political communications research and could provide unique insights into the ways in which individuals and groups navigate their political worlds.

The framework has been developed using two pilot studies. The first employs visualizations of citations added and deleted by Wikipedia editors during the editing of the "2011 Egyptian Revolution" and "2014 Crimean Crisis" English Wikipedia articles to generate discussions about the relationship between mass media and Wikipedia in narrating accounts of political events. The second uses online and off-line social network analysis and visualizations to investigate everyday political chat among Canadians to understand the role of citizens in influencing one anothers' political opinions and behaviors.

We first review traditions of qualitative and digital data-driven approaches in the study of political communications. We then focus on trace data and describe our trace interview method through a description of our two pilot studies. Next we discuss the strengths and weaknesses of our approach and of the potential contribution of trace interviews to the field of political communications research.

Qualitative Approaches in Political Communications Research

Surveys and experiments dominate political communication research, offering opportunities to generate generalizable findings, promote an understanding of communication patterns at the macro level, and quantify human experiences. Karpf, Kreiss, and Nielsen (2013) argue, however, that there is value in qualitative methods and that an important history of qualitative work exists within the field. Ethnographies, case studies, interviews, and focus groups have informed the work of some of political communication studies' most well-known quantitative scholars. They argue that these qualitative approaches offer the opportunity to ask new questions and investigate phenomena from a crucially different perspective.

The questions qualitative approaches allow us to ask and answer are integral to understanding the state of political communications. It was through in-depth interviews, for example, that Katz and Lazarsfeld (1955) were able to establish the two-step flow hypothesis, a model of political communication that has guided thinking in the field for half a century. Through survey research, Katz and Lazarsfeld found most people discovered political information not from news media but from everyday acquaintances. They used interviews to uncover who people were receiving information from, the context of their information seeking, and the reasoning behind those patterns. Today similar questions can be asked about which tools individuals use to engage in politics and which people (or accounts) make up their audience.

Though qualitative approaches in political communication research have been fruitful in the past, and despite calls for increased use of qualitative approaches, such methods have not been adopted much in the field (Bennett & Iyengar, 2008; Karpf et al., 2013).

Qualitative approaches could be of particular importance to political communications researchers as the field transitions from one that focuses on mass media effects to one that engages with patterns and processes of communication and political interaction broadly (Chaffee & Metzger, 2001). The two-step flow hypothesis suggests the mass media have an indirect effect on the general public via individuals called "opinion leaders." This study is one in a long chain dedicated to understanding how political message senders are able to affect their audiences' opinions, attitudes, and behaviors. Scholars are calling for a shift in approaches to understanding political communications by arguing that the dominant "effects studies" paradigm may overlook important issues (Bennett & Iyengar, 2008; Chadwick, 2013; Chaffee & Metzger, 2001).

The blurring of boundaries between different kinds of political actors, the shared tactics and tools of political players (Chadwick, 2013), and the flexibility afforded by technology (Earl & Kimport, 2011) point to a media environment radically different from the one in which effects studies were developed. Qualitative methods allow us to investigate how and why this environment is different and how and why political communication is conducted within that environment.

Integrating Computational Methods and Digital Data

As an alternative to qualitative work, computational social sciences and digital data-driven approaches have enjoyed an uptake to begin to fill in some of the gaps left by more traditional research tactics. Computational social sciences is "the interdisciplinary investigation of the social universe on many scales, ranging from individual actors to the largest groupings, through the medium of computation" (Cioffi-Revilla, 2014, p. 2). Computational approaches offer new opportunities to access data, combine new data sets (to compare over time, for example), and ask new questions. Computational social sciences and digital data-driven approaches are vast and largely unexplored at this point, but as technology advances so do opportunities for the study of large-scale data analysis.

In political communications studies, the adoption of computational methods has been limited. This is related to the resource-heavy nature of data collection, analysis, and presentation given the technical skill required. Notable exceptions include the *Journal of Communication*'s special issue on "Big Data in Communication Research," which provides a range of high-quality papers that make use of computational approaches (Parks, 2014). Language processing and social network analysis are two methods in particular that have enjoyed the most attention within political communication scholarship (Parks, 2014).

These approaches are effective for tracing relationships among actors, gaining a broad picture of which actors matter, comparing patterns over time, and adding complexity. The large-scale nature of computational approaches improves upon the offers of qualitative research methods. Further, attention to individual- and interpersonal-level communication patterns is an improvement on the more traditional quantitative approaches such as the surveys and experiments that are common to political communications research. The value of computational approaches is in their potential to suggest new questions about relationships and processes on a large scale and to enable researchers to answer questions in new ways. For example, digital data-driven approaches allow us to ask what the communicative processes are on a large scale and to point to unexpected intervening variables such as new political players or the affordances of various tools.

Constraints to large-scale computational studies do exist. Concerns related to big data approaches, for example, include the fact that one cannot assume generalizability simply because researchers have access to a large data set (boyd & Crawford, 2011). Computational approaches tend to lack information about the social context in which data were created and do not enable an understanding of the reasoning behind language or actions. It is difficult for computer programs to capture nontextual concepts such as "influence" or "credibility" (Bail, 2014; Freelon, 2014).

Finally, most studies in this genre tend to focus on specific media tools or platforms in isolation rather than considering the larger system. This is the result of technical challenges in combining data from multiple sources. Twitter and Facebook data are not stored in the same way, for example. User IDs for each platform are different, the type of data related to each ID varies across platforms, and the availability of data at different times is not necessarily consistent (Wesler, Smith, Fisher, & Gleave, 2008). Even if data from multiple sources are collected and organized, the interpretation of that data cannot be

similarly interpreted. A tweet is not the same as a Wikipedia edit, which is different still from a Facebook comment. Particularly when it comes to understanding the ways in which political messages are communicated in a large technological and often hybrid system, the actors involved in the communication are important information sources, because they can provide contextual cues and explanations of rationale.

Traces: Digital Data and Interviews

Trace data represent some of the most valuable digital data that researchers are making use of. Trace data are digital records that humans consciously or unconsciously leave behind as they navigate the digital world (Wesler et al., 2008). These data provide opportunities to track patterns of communication and to quantify political interaction on a scale not previously imaginable.

Trace data are used by computer scientists and ethnographers alike to uncover patterns of information sharing and interaction. Although the quantifiable nature of digital data makes it tempting to ignore the constructed nature of information generated from this data, it is important to recognize that uses of trace data are laden with assumptions. Freelon (2014) explains how scholars in different fields define the same traces and types of traces in very different ways. Dubois and Gaffney (2014), in an investigation of how to identify political influentials on Twitter, highlight the importance of theory-driven approaches to deciding which types of traces to collect and how to understand them. For example, when seeking influentials in marketing research, popular individuals are sought. In the context of the two-step flow (Katz & Lazarsfeld, 1955), however, personal connection to others close to the individual is more important. Different types of trace data shed light on these different facets of what influence as a general term is, and theory directs us to which facets to focus on. Trace data are interpreted by those who analyze them; this interpretation is not neutral.

With this caution in mind, trace data allow researchers to generate insight on multiple levels. For example, trace data may be used to gain a broad picture of what is going on at the macro level, as illustrated by Kwak, Lee, Park, and Moon's (2010) investigation of Twitter as a news medium or social networking site. It may also inform meso-level understandings by helping researchers understand actors in a particular system. For example, Wu et al. (2011) reexamined Lasswell's (1948) model of communication, "who says what to whom," in the context of Twitter using a Twitter graph of 42 million users, a collection of 5 billion tweets, and a snowball sample of lists to determine who are the elite users on Twitter and who is being listened to. Finally, at the micro level, trace data can provide details about what specific people do and who they interact with (Wesler et al., 2008).

Qualitative researchers could benefit from integrating trace data in two key ways. Trace data offer the ability to confirm past actions, thus helping to counter memory and self-reporting bias. Trace data also offer the opportunity for researchers to triangulate their data. Instead of relying on an interviewee's account of events or relationships and the researcher's informed interpretation of that account, the researcher can use trace data to consider the traces of digital interaction that the interviewee left behind. In so doing, researchers can begin to produce meaningful accounts of the interactional context as articulated by Geertz (1973) in his work on "thick description," or what Wang (2013) calls "thick data."

Geiger and Ribes (2010) propose "trace ethnography," which involves the analysis of traces left by users of a sociotechnical system to reconstruct their practices. Trace ethnography provides an innovative method for analyzing the ubiquitous and heterogeneous data (including transaction logs, version histories, institutional records, conversation transcripts, and source code) to build "rich narratives of interaction . . . coordination practices, information flows, situated routines, and other social and organizational phenomena across a variety of scales" (Geiger & Ribes, 2010, p. 1).

Other methods, such as those proposed by Anderson, Nafus, Rattenbury, and Aipperspach (2009) on "ethno-mining," show how sensing and behavioral tracking technologies can be used to invite conversations with research participants and within the corporations that develop technologies for use by the public. Ethno-mining merges database mining and field research to combine the semiautomated collection and analysis of behavioral data with the collection and analysis of qualitative data. The method relies on data visualizations as shared artifacts for the collaborative construction of meaning (Anderson et al., 2009).

In contrast, "network ethnography" (Howard, 2002), uses social network analysis to define the most significant informants in an ethnographic study of what he calls "hypermedia organizations." Hypermedia organizations are those whose members use ICTs to "conduct the business of social organization over large areas and disparate time zones, and at all hours of the day" (Howard, 2002, p. 552). Researchers such as C. W. Anderson (2013), for example, employed network ethnography to select informants for his ethnographic study of the news ecosystem in Philadelphia to understand the transformation of U.S. journalism.

Introducing Trace Interviews

Building from these approaches, we propose trace interviews as an effective means of combining the benefits of trace data with those of the qualitative interview. Trace interviews involve the collection, visualization, and discussion of a participant's traces with that participant. Visualized data of a user's interactions are employed in the interview setting so that the participant might reflect on his or her actions as depicted in the data. This process enables participants to interpret data by providing contextual details and clues about their motivations for undertaking particular actions represented in the data as well as to point to missing or inaccurate data. During the interview, the participant is asked to examine and then comment on the visualizations by answering questions to guide interpretation, such as, "Is this [data] surprising?" and "Can you explain X [data point]?"

Trace interviewing complements existing approaches, including digital ethnography, network ethnography, ethno-mining, case studies, and systems approaches. The method would work well in concert with approaches such as trace ethnography (Geiger & Ribes, 2010) and systems approaches (capturing data relating to users' activities on several media platforms) as a means of verifying a researcher's analysis of digital traces and complementing such analysis with the user's own perspective on what happened and why. Trace interviewing also combines the use of visualizations for the co-construction of meaning as explained by Anderson et al. (2009), but with a particular focus on social

media systems. Trace interviewing also can be used as a method for interviewing informants selected through the network ethnography process.

An actor-centric rather than tool-centric approach allows researchers to investigate political communications without embedding assumptions about the value or impact of particular tools. Placing the actor at the center of analysis, we believe, is the appropriate way of answering questions about his or her political communications in the current age. We suggest that participants, when armed with appropriate tools and basic guidance, can provide much-needed insight and that visualizations of personal data brought into the interview setting for joint analysis by the researcher and participant might lead to meaningful insights and new information. By asking participants to comment on visualizations of trace data left by them, we are able to add important contextual meaning to the collected data, which ultimately improves the researcher's ability to make sense of that data.

Cases and Methods

We conducted two pilot studies to evaluate the method and determine appropriate steps. Both studies involved analysis of processes by which political actors share information online. The first concerns sourcing practices of Wikipedians editing articles relating to political events and how this reflects on the perspectives that are generated about those events. The second is a study of Canadian citizens' political talk on social media sites to understand why and how different platforms are used for particular messages and conversations.

Wikipedians editing political events articles and Canadian citizens discussing political opinions both can be seen as nontraditional political actors, distinct from their more traditional counterparts, including journalists and politicians. These actors are ordinary citizens whose political communications are unpaid but whose communicative acts affect the ways in which political events play out and are understood by the larger populace. The ways in which these communications are produced and shared is distinct from professional routines, and therefore little is currently known about how they operate. Communication by ordinary citizens is, however, a critical form of media power that impacts the futures of nation-states and their governments around the world (Bennett & Segerberg, 2012).

Wikipedia references and citations form one of the building blocks of the encyclopedia and are the subject of frequent debates about reliability of content added. Different media sources tend to portray different points of view and narratives about political events, depending on which country they originate from. When an issue or event is based in a country outside the Western English media, editors must decide which point of view to accept as the "neutral" one (according to Wikipedia's "Neutral Point of View" policy). The adding and removal of sources based on particular points of view can be seen as political acts, because they affect the narrative that becomes historicized in the final encyclopedic account. Understanding the processes involved in adding and removing sources is thus important for understanding an important way in which narratives about political events are being imagined and brought into being.

In the Canadian citizens study, everyday political chat is an important catalyst for an informed citizenry. Political messages spread through the personal networks of individuals as they discuss policy

changes over coffee or vent about the inaction of their local government to fix potholes on Twitter. The study of everyday political chat among Canadian citizens explores the ways in which individuals known to be politically influential among their peers navigate the hybrid media system. The goal is to understand why and how they engage with different members of their personal networks using different media tools such as Twitter, Facebook, and face-to-face communication.

The Process

The trace interview process involved three key steps outlined below: case selection, visualizations, and interviews.

Case Selection

The first step in the trace interview process involved identifying relevant actors, content, and platforms within particular networks and then selecting cases. The actors in our cases consist of Wikipedians editing breaking news-related political events, and Canadians involved in everyday political chat. Their networks, in the Wikipedia case, are composed of media sources being added and removed and, in the Canadian case, of social networks in which particular messages are shared (Table 1).

Table 1. Cases and Network Data.

Cases	Platforms	Nodes	Edges
Wikipedia sourcing	Wikipedia	Editors	Sources added and removed
Canadian politics	Twitter, Facebook, and off-line social networks	Followers, friends, "very close" and "somewhat close" ties	Tweets (mentions), Facebook friendships, and social ties

For the Wikipedia case, we chose the "2011 Egyptian Revolution" and "2014 Crimean Crisis" articles on English Wikipedia. We chose these articles because they represent breaking news stories that were accompanied by a strong international media presence but were located in countries outside the headquarters of English mainstream news production. We predicted that these articles would reveal how editors navigate the local/international news source divide, and we were curious about whether any consolidation in news sources that present particular points of view occurred. We chose articles at different stages of their development, because we were interested in whether editors could recall their experience of source editing after a long hiatus. We were also interested in understanding how snapshots taken at different points in an article's history could affect the narrative about it. By choosing articles according to the time of their construction, we were able to compare results to better evaluate the method.

We contacted the editors of both articles who represented the top 10 most frequent editors of (a) the article, in the case of the 2011 Egyptian Revolution article and (b) citations and references, in the case of the 2014 Crimean crisis article. We received responses from five editors, one of whom said that he did not remember editing the Egyptian Revolution article and so would not be much help, and two others whose requirements for anonymity during the interview and work schedules could not be accommodated during the time scheduled for this pilot study.

Two Wikipedians were interviewed for this study. Wikipedia editor Ocaasi was the third most prolific editor of the Egyptian Revolution article at the time of the interview (with 324 edits), and Aleksandr_Grigoryev was the most prolific editor of the Crimean crisis article (with 189 edits). 1,563 edits of the Egyptian Revolution article involved the addition of citations, compared to 773 in the Crimean crisis article. This is probably due to the recentness of the article, since it had only been in existence for five weeks when the data were collated.

In the Canadian political chat case, a month-long sample of the hashtag #CDNpoli was collected using the Twitter streaming API in May 2013, resulting in a data set containing 411,138 tweets by 45,986 users. This hashtag is one of the most enduring hashtags in Canadian politics and is the most inclusive in terms of crossing partisan boundaries, political issues, and types of political players that include journalists, politicians, nongovernmental organizations, and average citizens. Twitter was selected because it is used as both a social network and news media (Kwak et al., 2010) and because we were particularly interested in those individuals who are digitally enabled and make use of social media in their everyday political chat.

Any account that had been mentioned at least twice in the sample was considered "influential."[2] In March 2014, the 7,881 influential accounts were randomized, and 1,000 were selected for manual categorization. All active individuals were invited to respond to a survey and provide their contact details for a follow-up interview.[3]

For this pilot study, three individuals, who will be called Jill, Matt, and Frank, were selected.[4] Jill and Matt are from Halifax, Nova Scotia, a large city on the east coast of Canada. In her mid-20s, Jill is very active on Twitter and on Facebook, using each several times per day, yet her political chat is mostly limited to in-person meetings and phone conversations. Matt is in his early 30s and is active on Twitter, Facebook, and Reddit daily. He also actively uses a wide range of other social media tools, but when it comes to political chat, he mostly limits his communication to Twitter, Facebook, and in-person communication. Frank is from a small town with a population of just over 1,000 in New Brunswick, also on the east coast. In his 60s, Frank is the moderator for multiple politics-focused Facebook groups and spends most of his day curating these groups, producing content, and discussing politics on Facebook. He

---

[2] The threshold of two Twitter mentions, determined by manually coding a random sample of 1,000 tweets, is used to separate those who simply follow the conversation from those who interact by following (Dubois & Gaffney, 2014).

[3] This is part of a larger study. As such, the response rate cannot yet be calculated.

[4] At the time of writing, an additional 20 trace interviews have been conducted as part of a larger project.

rarely talks about politics in an off-line setting, and, although he is semiactive on Twitter, he prefers Facebook.

Visualizations

In preparation for the interviews, we produced a range of visualizations of the users' activities in relation to their networks. Visual images have been shown to be useful stimulants in interview settings (Rose, 2012). Visuals help jog interviewees' memories and focus them on a specific example that helps them think about and reflect on their actions (Gauntlett, 2007). Carefully crafted visual representations of data are also often more accessible than simple tables or lists, because they can provide a quick overview of large amounts of data and information (Tufte & Graves-Morris, 1983). In the Wikipedia case, such preparation involved visualizing sources added by editors in relation to their cumulative addition and removal over time (Figures 2–5); in the Canadian citizens case, it involved visualizing the users' friendship networks on Facebook and Twitter (Figures 6–8).

For the Wikipedia case, we scraped the 2011 Egyptian Revolution and the 2014 Crimean crisis articles' edits over time to gather a list of URLs that were added over time and the editor who added them. We used the Wikipedia API to download a list of the current URLs on the page. We removed links to Wikipedia, to archive.org (because most of these links were copies of previous links that were being archived on archive.org) and to language or regional references in the domain (e.g., http://uk.reuters.com/news/south_africa would become reuters.com, and en.pravda.ru would become pravda.ru).

We then grouped the summarized domains together and organized them according to the editors who added them to find the most frequent editors of citations. We plotted a graph showing which domains were most frequently added by particular editors and added the number of times that a domain on the list was cited in the article at the time of our extraction of links.[5] We then divided the frequency of the domain at the time of extraction by the cumulative frequency of the domain to get a rough reading of stability of the source in question (see Figures 2 and 3). As shown in Figure 2, rt.com (Russia Today) is the least stable domain with only about 30% of citations to its URL remaining after being removed by other Wikipedians. Kyivpost, Bloomberg, Euronews, Slate, Blitz, and National Public Radio are the most stable domains, because citations to their domains had never been removed when we downloaded the data set.

Stability = <u>frequency of domain at the time of data collection</u>
cumulative frequency of domain

Figure 1. Calculating the stability of domains.

---

[5] These numbers are different because editors may add Russia Today as a citation 100 times, for example, but other editors may have removed 50 of those citations. In this case, the number of citations in the article when we downloaded the URLs would result in 50 remaining citations.
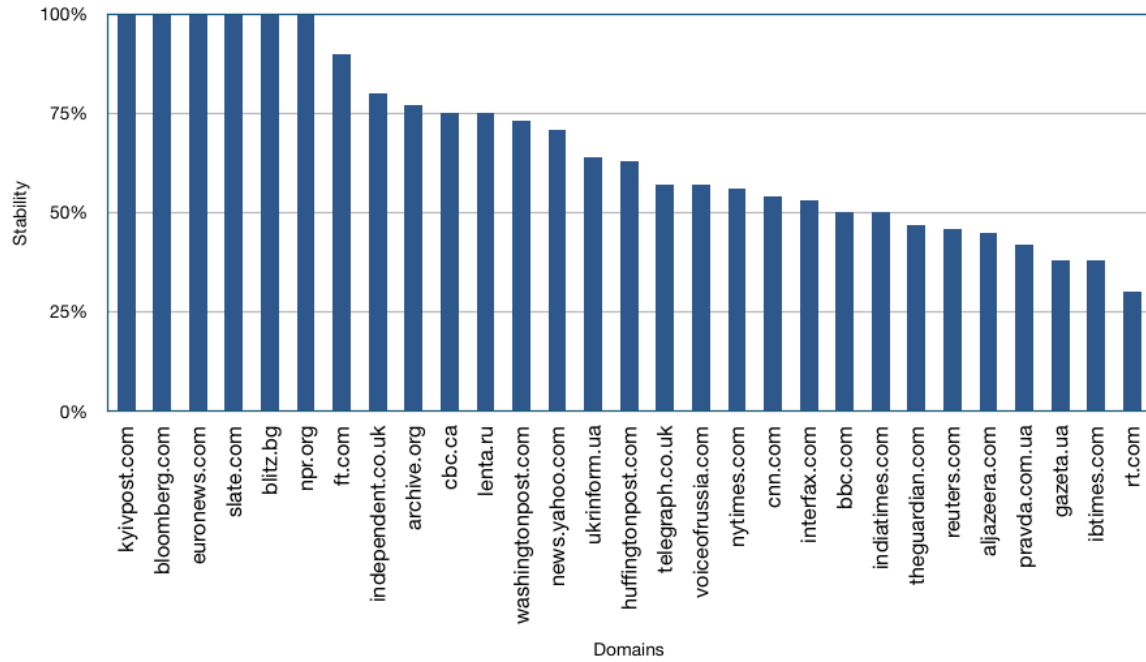
*Figure 2. Stability of the top 30 current domains for the "2014 Crimean Crisis"
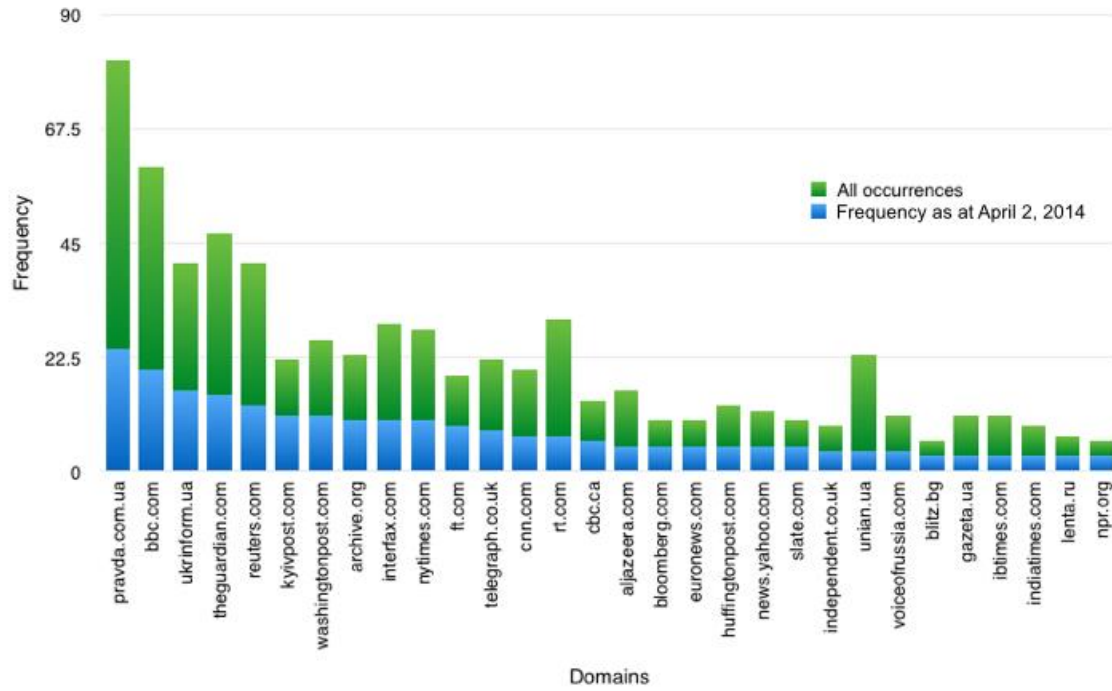Wikipedia article as of April 2, 2014.*

Figure 3. Domains added by Aleksandr Grigoryev (Crimean Crisis example) as of April 2, 2014.
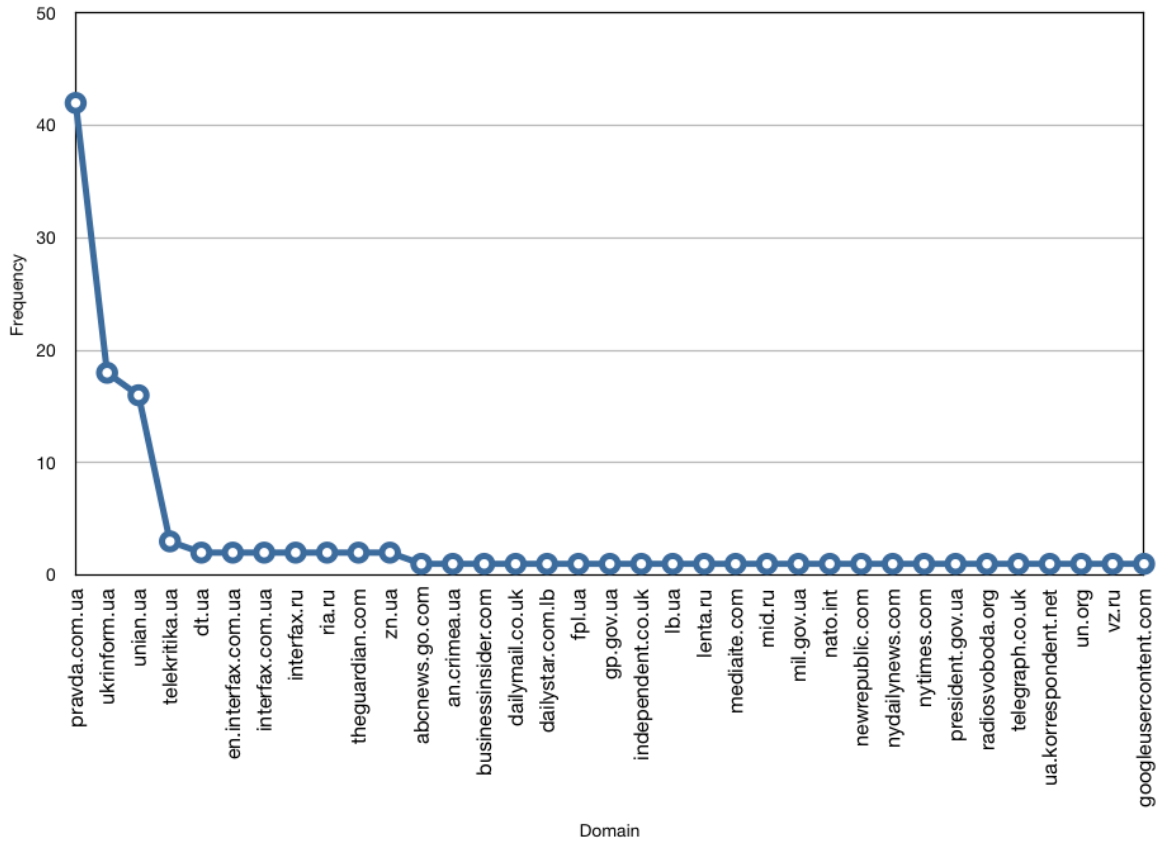
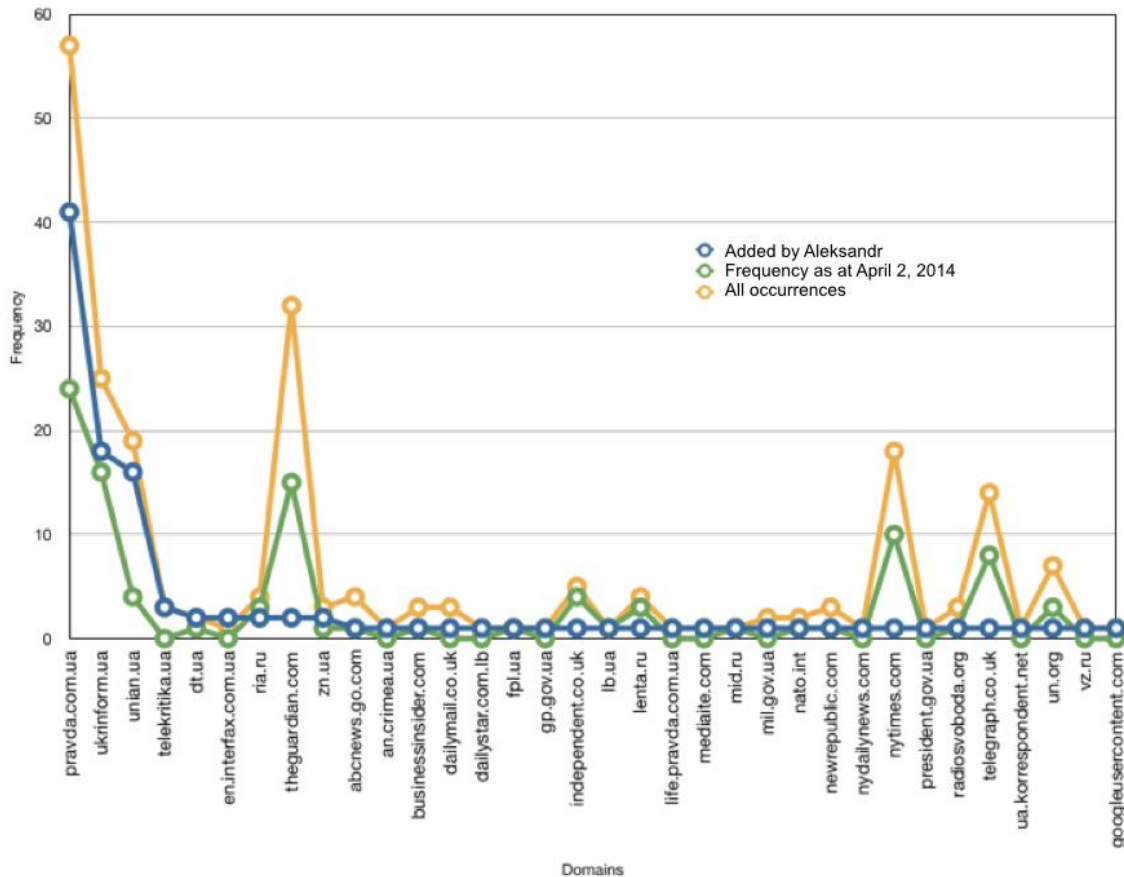*Figure 4. Domains added by Aleksandr Grigoryev.*

Figure 5. Domains added by Aleksandr Grigoryev in relation to domains added at the time of data collection as well as all occurrences ("2014 Crimean Crisis" example).

For the Canadian case, interviewees' Twitter mentions and Facebook friend networks were presented to them during the interview. The Twitter mention networks were collected and visualized using Netlytic software, and the Facebook friend networks were collected and visualized using NameGenWeb.[6] Netlytic and NameGenWeb are free data collection and visualization programs created for academic use by researchers. Figures 6 and 7 provide examples.

---

[6] Participants logged on to their Facebook accounts during the interview, ran the application, and downloaded the graph file, which they provided to the researchers.
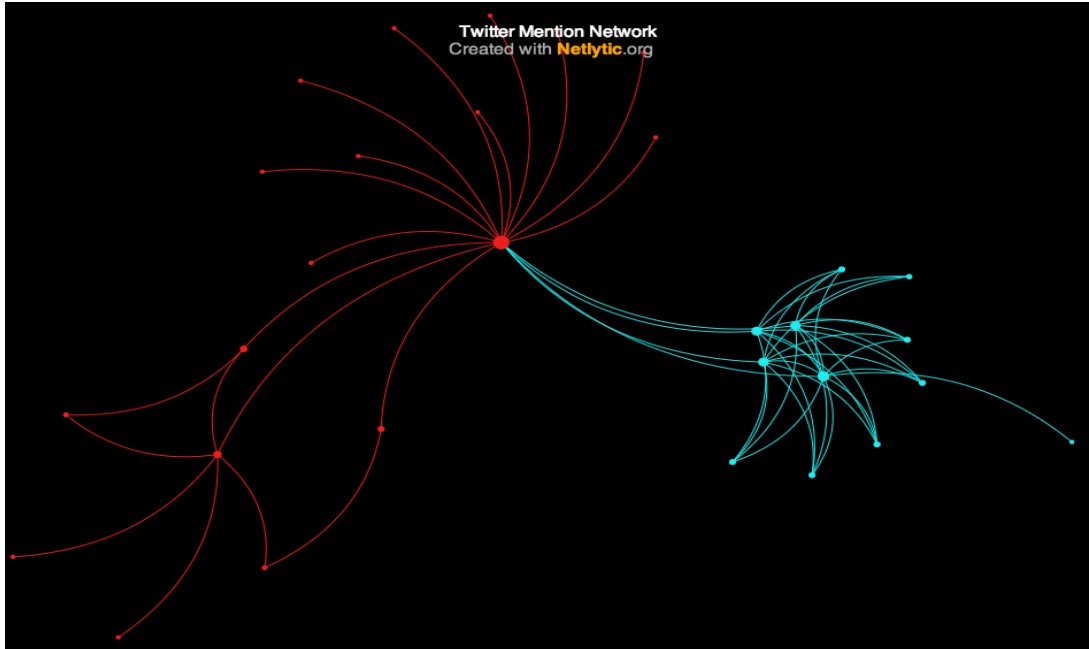
Figure 6. Netlytic: Twitter mention network example (Matt).
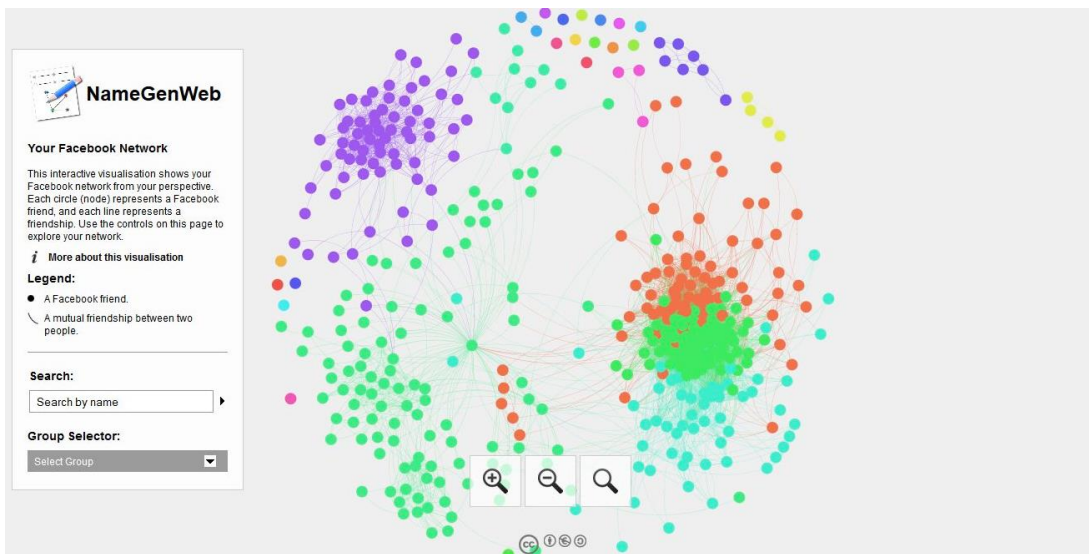

Figure 7. NameGenWeb: Facebook friends network example (Jill).

The Twitter mention networks were based on a two-week sample, and the Facebook friend graph was based on the friends list on the day of the interview. Node size represents total degree (a measure of how central an account is in a network, here equaling the number of others connected to that node either by mentions for Twitter and friends for Facebook), and color represents modularity class (an indication of social clusters. Notably, for Twitter total degree includes both times that the account in question mentions another account (out-degree) and when other accounts mention it (in-degree). This is because we are interested in interaction generally. Out-degree and in-degree could be used in future studies to add nuance or for comparison.

Following Hogan, Carrasco, & Wellman (2007), participants in the Canadian case were also instructed to create their own visualization of their "somewhat" and "very close" personal ties (also known as a sociogram) during the interview. This became a visual tool used to discuss the interviewees' personal social network regardless of digital platform (Figure 8).
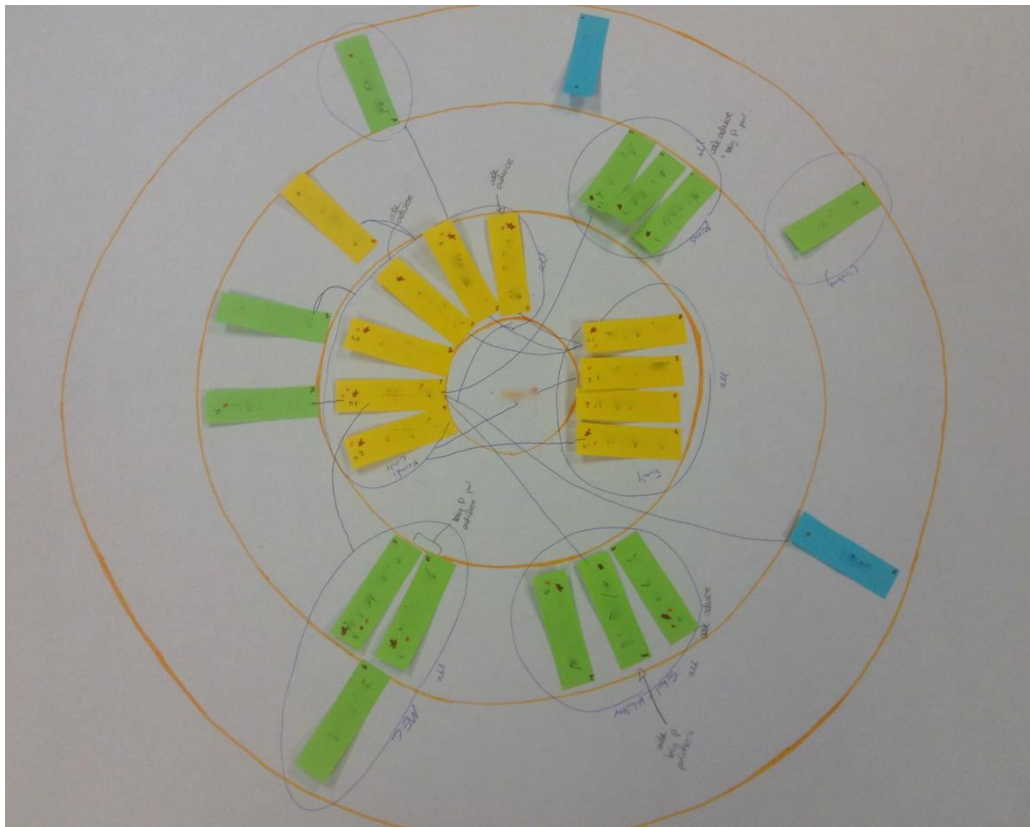

Figure 8. Sociogram created during interview (Matt).

Interviews

Interviews with the Canadian users took place face-to-face, whereas those of the Wikipedia editors took place over Skype. This enabled us to examine how the method could be used in multiple settings.

For the Wikipedia study, we sent the graphs to the editors before the interview and asked them to open particular files as we discussed each one. The interview consisted of nine general introductory and practice-based questions followed by an exploration of each of the graphs.

In the Canadian user case, the interview consisted of two components: The first part involved eliciting the participant's personal social network; the second part involved discussing visualizations of online networks. First, participants were asked to draw a model of their very close and somewhat close ties using the technique outlined by Hogan et al. (2007). This included, but was not limited to, online ties. Participants were then asked about who they communicated with regarding political issues and which media they used for that communication.

Second, visualizations of online networks were discussed. For each graph, the researcher described the parameters of data collection, the implication of certain design elements such as layout, and the meaning behind color and size of nodes. Participants were invited to mouse over nodes to see the names associated with each. They were asked to identify each cluster and to discuss individuals of interest in the network in more detail.

The Benefits of Trace Interviews

There are four key reasons trace interviews could be a useful method in a hybrid and complex media system. Trace interviewing provides opportunities to (1) enhance recall and generate an understanding of an actor's decision-making process, (2) validate trace data-generated results, (3) address the data joining problem as interviewees explain traces on multiple platforms, and (4) respond to ethical problems relating to the use of personal data without consent or involvement by users.

Recall and Reasoning

Visual aids have proven useful in studies that use interviews to investigate the experiences of individuals by enabling them to jog their memory or initiate a new line of inquiry (Rose, 2012). We found that the use of visualizations of the trace data individuals left behind during their political communications to be beneficial for enabling recall and reasoning by the interviewee.

In the Wikipedia case, visualizations sparked interviewees' memories about the context in which they made decisions about sources. During his interview, Ocaasi reflected on a visualization showing that he had added only a single reference to Al Jazeera when this became the most popular citation in the 2011 Egyptian Revolution article. This observation prompted Ocaasi to discuss how he came to employ Al Jazeera as a key source after initial skepticism. He explained how he was "glued to Al Jazeera's live feed

from Tahrir Square" and that when he heard something new, he would look for results in the Associated Press or the New York Times. He also explained how, before the Egyptian Revolution, he had thought that Al Jazeera was a "propaganda network for terrorists" and that this experience changed his attitude toward the network, because he came to rely on Al Jazeera for regular, on-the-ground coverage and for what he saw as its neutrality. This important contextual information was gained by using the visualizations to help with recall, which then elicited important new information about the context in which decisions were made. In some cases, interviewees simply reiterated existing knowledge; in others, the ability of interviewees to engage with a visualization of their data prompted new insights.

Validation and Meaning

Trace interviewing allows researchers to add to the meaning of digital data, because the researcher has the opportunity to directly query the creator of the data. By interviewing the person whose traces have been stored for analysis, the researcher can ask about missing data or other anomalies. Further, discussion about the perceived meaning of various traces contributes to more contextually relevant interpretations of the data.

Examining his Twitter follower network graph, for example, Matt noticed he had three major clusters of followers, all of whom he discussed political matters with but each concerning different issues. He explained that he posts different types of content with the intention of attracting the attention of those different types of users represented by each cluster. In other words, not all tweets are equal for Matt. Others explained how retweets from accounts that they find credible are more valuable than retweets from others and that Facebook "likes" from a family member may not be as valuable as a retweet from a journalist. The combination of trace data and in-depth interviews made it possible to discern the reasoning behind Matt's tweeting patterns and enable a nuanced understanding of the meaning of various traces in context.

Data Joining

A major constraint on the utility of trace data is that it is often difficult to combine data sets (Wesler et al., 2008)—both in terms of the technical challenge of piecing together data that have been organized differently and the sociological challenge of understanding what each kind of trace is and what it means.

In the study of political chat in Canada, the sociogram (the personal network created off-line at the beginning of the interview) helped identify places where various online networks overlapped. For both Jill and Matt, for example, friends from university clusters were all people they communicated with on Facebook, but only some friends from the university cluster were also Twitter followers. Most of Frank's very close Facebook friends also followed him on Twitter, and those who were less close did not. This observation led to discussions about what it means to be a Facebook friend or to be followed on Twitter. It was possible not only to uncover which off-line names, Facebook screen names, and Twitter handles matched but to understand the relationships between those various profiles using the trace interview process.

People and Their Data

When users create content or leave a trace, they are in some sense leaving something of themselves behind. In the same way that an artist's work is forever connected to his or her person, or a picture forever depicts a moment in a person's life, users' traces reflect an important part of their identity at the time of that trace making. If users produce data that will later be used by researchers to describe them and make inferences about who they are, do those researchers have an ethical obligation to ask for permission from those individuals?

In an age of what Bruns (2008) calls "produsage," methods for eliciting data about user practice need to treat users as creators and authors in the research process. Bruckman's (2002) conception of Internet users as "amateur artists" rather than "human subjects" is appropriate here, as is Berry's (2004) conception of an "open source approach to ethics." Using trace data as a way to mutually explore research questions with users represents a significant advancement in the traditional power relations between the observers and the observed. It necessarily involves a dialogue between the researcher, the participant, and the data rather than between the researcher and the data or between the researcher and the participant. This mode of research can be seen as its own method of triangulation in which the entire process of research is a collaborative one, rather than only at the final stage of publication when participants' comments are sought. Conducting conversations with participants about the use of their data for research purposes also can help us understand user concerns during larger, big data studies, in which asking each individual would be prohibitive. Interviewing a small number of individuals might alert researchers to potential participant concerns that may be broadly shared.

In the Wikipedia case, the data that editors produce in the conduct of their activities on the Wikipedia platform are automatically licensed under a Creative Commons Attribution license or a public domain dedication, which allows others (including researchers) to freely copy and share it. This does not mean that no harm is caused by researchers taking individual editors' data out of specific contexts (Nissenbaum, 2010). Providing opportunities for the researched to speak back to the data collected about them is therefore not only fruitful from a data elicitation perspective, but it also constitutes an affirmative response to the culture and context of open source production and contextual integrity (Nissenbaum, 2010). A similar argument could be made for other social media platforms such as Twitter and Facebook, which each have their own norms regarding privacy of content and intended use (Madden et al., 2013).

Navigating Challenges

Despite the benefits of trace interviews, two key challenges must be navigated to make the most successful use of the method.

Visual Literacy

Visualizations as abstract, graphic representations of taken-for-granted behaviors can be challenging for participants to navigate, because they can require a type of digital literacy not consistent across all age, educational, or socioeconomic groups. The understanding of traces requires a complex

understanding of how user behavior relates to the traces that users leave in the sociotechnical space, which can be complex for relative newcomers to navigate.

The Wikipedia case interview with Ocaasi yielded very different information from that of Aleksandr_Grigoryev, for example, bringing into sharp relief the different experiences, perspectives, and worldviews of editors. Ocaasi used the graphs to explain his sourcing actions in relation to other editors', explaining how they were replacing Al Jazeera citations with the BBC or New York Times when available, but Aleksandr_Grigoryev, as a newer user, was less aware of the dynamics of his own sourcing practices in relation to others. Despite this, Aleksandr_Grigoryev was able to provide invaluable contextual information about the positioning of each of the Russian and Ukrainian sources used in the Crimean crisis article. This experience revealed a huge contrast in the ways that editors reflect on their work and the extent to which they are aware of others in the editing pool of particular articles.

For Frank in the Canadian case, the Twitter mention network was not initially meaningful to him because he did not place value in the message sharing that he was part of on Twitter. Once conversations about the Facebook groups he moderates were introduced, however, he was able to better describe the role of different individuals. He went on to explain why certain users were more or less prominent in the group and what role they played. He even connected their Facebook persona to his memory of them on Twitter and off-line. This highlights the importance of carefully introducing and situating the digital data and visualizations used during the interview and exemplifies the importance of continual reflection on the part of the researcher.

The Bias of Data Collection and Visualization

Visualizations are constructed. The data are not "raw," the representation is not free of assumptions, and the researcher makes various analytical and editorial decisions when generating useful visualizations. As Bowker (2005) writes, "Raw data is both an oxymoron and a bad idea; to the contrary, data should be cooked with care" (p. 184). Researchers must remain aware of the impact of their decisions and continuously reflect on their role in the data interpretation process.

The sheer volume of digital data available can create new obstacles and may be overwhelming for researchers (Wesler et al., 2008). Significant decisions regarding scope must be made; each decision has an impact on the kinds of analysis that will result from it. For example, specific online social networking platforms present multiple options for data collection and analysis. Considering Twitter in the Canadian politics case, we could have conducted social network analysis of followers, friends, mentions, mutual mention, co-listing, and so on, but only mentions were analyzed. Content of tweets, which might be analyzed for a short or long period of time, in segments, or randomized, was not examined in this case. Each of these networks and approaches to content analysis depicts a different image of what constitutes everyday political chat for that individual. Our decisions about what data to collect and include guided the interview.

The express goal of using visualizations in the interview setting is to enable researchers to direct interviewees to answer particular questions. To make best use of trace interviews, the researcher must

carefully think about the timing of introducing the visualizations. In the Wikipedia case, for example, interviewees had the opportunity to reflect more generally on their practice in the first section (without use of the visualizations) and hone in on a particular area guided by the visualization in the following section of the interview. With careful time management, adequate preparation, and careful reflection on the consequences of decisions being made, these challenges can be easily overcome.

Conclusion

We have explained the benefits and challenges of trace interviewing at a practical level, but the ultimate goal of the trace interviewing method is in advancing political communication theories in the context of the multimodal environment that characterizes the current Internet. Returning to the example of Katz and Lazarsfeld's two-step flow (1955), the introduction of in-depth interviews into the repertoire of tools that researchers could use for understanding the impact of mass media and political message senders on the general public enabled scholars to better understand the role of social influence in the political communication process. In a hybrid media system, more channels of communication are available, so understanding an individual's communication practices becomes a much more complex task. Trace interviews allow researchers to use trace data to situate interviewees in the context of their digital practice, which then allows us to ask new kinds of questions. For example, instead of considering only the linear flow of communication from sender to receiver, we can now consider how the participant influences his or her social group across channels and in different contexts. Researchers are also better equipped to deal with the large amount of trace data available, because they can work with the creator of that data to understand what is theoretically relevant and what is not.

Another example of how trace interviews might contribute to the advancement of political communication theories comes from the realm of gatekeeping theory. A distributed, networked communication environment has resulted in the need for fundamental shifts in the way that we understand political theory. Barzilai-Nahon (2008), for example, notes that there has been a lack of agreement on what gatekeeping is, as well as who both the gatekeepers and gated are in the context of the complex networks that characterize our current communication environment (Barzilai-Nahon, 2008,). Barzilai-Nahon's theory of networked gatekeeping addresses the gated as well as gatekeepers in the network and extends the methods of gatekeepers beyond issues of selection to a broader conceptualization of information control. Trace interviewing offers promising opportunities to refine the dynamics of gatekeeper and gated activities within a particular community such as Wikipedia rather than maintaining a singular focus on Wikipedia editors as a bloc.

We suggest pairing computational techniques with traditional in-depth interviews allows researchers to understand the political communication process by taking an actor-centered approach. By collecting and visualizing traces left by participants and then discussing those data and visualizations in the interview setting, researchers can improve the quality and utility of that data as well as generate new insights during the interview process. Specifically, our two pilot studies point to four benefits: (1) to enhance recall and generate an understanding of decision-making processes, (2) to validate trace data, (3) to address data joining problems, and (4) to respond to ethical concerns.

Trace interviewing requires high levels of technical skills and reflexivity to navigate challenges related to pairing computational and qualitative interview approaches. If these characteristics are present, trace interviews can be a powerful tool for ethnographers, other qualitative researchers, and computational social scientists alike as the current media environment forces researchers to consider what digital traces mean in context and how political actors contribute to a complex and hybrid media system. This underlines the potential for researchers to collaborate to pool efforts in refining the method, defining best case practice for interviews and visualizations, and sharing code and skills.

Notably, small pilot studies like this one present necessarily limited findings. We selected divergent cases and used different kinds of trace data and visualizations to illustrate the potential wider applicability of trace interviews. More extensive studies and studies across different contexts are needed to build a better understanding of what trace interviewing can be most useful for and for establishing best practices.

Trace interviews present an opportunity for developing more holistic research that addresses political communications in today's large technological and hybrid system. By mixing trace data collection, visualizations, and interviews, researchers can ask questions about the meaning of traces in context and generate new insights about the decision-making process and the roles played by an increasing variety of political players.

References

Anderson, C. W. (2013). *Rebuilding the news: Metropolitan journalism in the digital age.* Philadelphia, PA: Temple University Press.

Anderson, K., Nafus, D., Rattenbury, T., & Aipperspach, R. (2009). Numbers have qualities too: Experiences with ethno-mining. *Ethnographic Praxis in Industry Conference Proceedings*, 2009(1), 123–140. doi:10.1111/j.1559-8918.2009.tb00133.x

Bail, C. (2014). The cultural environment: Measuring culture with big data. *Theory and Society, 48,* 465–482.

Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society for Information Science and Technology, 59*(9), 1493–1512.

Bennett, W. L., & Iyengar, S. (2008). A new era of minimal effects? The changing foundations of political communication. *Journal of Communication, 58*(4), 707–731.

Bennett, W. L., & Segerberg, A. (2012). The logic of connective action. *Information, Communication and Society*, *15*(5), 739–768. doi:10.1080/1369118X.2012.670661

Berry, D. M. (2004). Internet research: Privacy, ethics and alienation: An open source approach. *Internet Research*, *14*(4), 323–332.

Bowker, G. C. (2005). *Memory practices in the sciences*. Cambridge, MA: MIT Press.

boyd, d., & Crawford, K. (2011). *Six provocations for big data* (SSRN Scholarly Paper No. ID 1926431). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1926431

Bruckman, A. (2002). Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology*, *4*(3), 217–231.

Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and beyond: From production to produsage* (Vol. 45). New York, NY: Peter Lang.

Chadwick, A. (2013). *The hybrid media system: Politics and power*. Oxford, UK: Oxford University Press.

Chaffee, S. H., & Metzger, M. J. (2001, November). The end of mass communication? *Mass Communication and Society*, *4*(4), 365–379.

Cioffl-Revilla, C. (2014). Introduction. In *Introduction to computational social science* (pp. 1– 20). London, UK: Springer.

Dubois, E., & Gaffney, D. (2014). The multiple facets of influence: Identifying political influentials and opinion leaders on Twitter. *American Behavioral Scientist*, *58*(10), 1260–1277.

Earl, J., & Kimport, K. (2011). *Digitally enabled social change: Activism in the Internet age*. Cambridge, MA: MIT Press.

Freelon, D. (2014). On the interpretation of trace data in communication and social computing research. *Journal of Broadcasting and Electronic Media, 58*(1), 59–75.

Garrett, R. K., Bimber, B., de Zuniga, H. G., Heinderyckx, F., Kelly, J., & Smith, M. (2012). New ICTs and the study of political communication. *International Journal of Communication*, *6*, 214–231.

Gauntlett, D. (2007). *Creative explorations: New approaches to identities and audiences.* London, UK: Routledge.

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

Geiger, R. S., & Ribes, D. (2011, January). Trace ethnography: Following coordination through documentary practices. In *2011 44th Hawaii international conference on System sciences (HICSS)* (pp. 1–10). Kauai, HI: IEEE.

Goel, V. (2014, June 29). Facebook tinkers with users' emotions in news feed experiment, stirring outcry. *The New York Times*. Retrieved from http://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html

Hogan, B., Carrasco, J. A., & Wellman, B. (2007). Visualizing personal networks: Working with participant-aided sociograms. *Field Methods*, *19*(2), 116–144.

Howard, P. N. (2002). Network ethnography and the hypermedia organization: New media, new organizations, new methods. *New Media and Society*, *4*(4), 550–574. doi:10.1177/146144402321466813

Karpf, D., Kreiss, D., & Nielsen, R. (2013). A new era of field research in political communication? Presented at *The annual meeting of the International Communications Association* .London, UK.

Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications*. Glencoe, IL: Free Press.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World Wide Web* (pp. 591–600). Raleigh, NC, USA: ACM.

Lasswell, H. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 117–130). Urbana, IL: University of Illinois Press.

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., & Beaton, M. (2013, May). *Teens, social media, and privacy.* Washington, DC: Pew Research Center. Retrieved from http://www.pewinternet.org/files/2013/05/PIP_TeensSocialMediaandPrivacy_PDF.pdf

Nissenbaum, H. F. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford Law Books.

Parks, M. R. (2014). Big data in communication research: Its contents and discontents. *Journal of Communication*, *64*(2), 355–360. doi:10.1111/jcom.12090

Rogers, R. (2009). *The end of the virtual: Digital methods.* Amsterdam, The Netherlands: Amsterdam University Press.

Rose, G. (2012). *Visual methodologies: An introduction to researching with visual materials*. London, UK: SAGE Publications.

Schroeder, R. (2007). *Rethinking science, technology, and social change.* Stanford, CA: Stanford University Press.

Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2). Cheshire, CT: Graphics Press.

Wang, T. (2013, May 13). *Big data needs thick data: Ethnography matters*. Retrieved from http://ethnographymatters.net/2013/05/13/big-data-needs-thick-data

Wesler, H. T., Smith, M., Fisher, D., & Gleave, E. (2008). Distilling digital traces: Computational social science approaches to studying the Internet. *The SAGE handbook of online research methods* (pp. 116–140). London, UK: SAGE Publications.

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011, March). Who says what to whom on Twitter. In *Proceedings of the 20th international conference on World Wide Web* (pp. 705–714). Hyderabad, India: ACM.