Chapter 9

Ethnographic and Qualitative Research on Twitter
Alice E. Marwick

Twitter’s success has made it a rich research site for scholars interested in online interaction, information dissemination, activism, and a plethora of other subjects. The sheer volume of users, tweets, and hashtags has made the site a favourite for quantitative data analysis and “big data” number-crunching. For instance, in an early study of Twitter, Krishnamurthy, Gill, and Arlitt (2008) collected information about nearly 100,000 users, including number of accounts followed, number of accounts following them, and frequency of status updates. The authors created a taxonomy of Twitter users, grouping them into broadcasters, acquaintances, miscreants, and evangelists based on the ratio of following-to-follower. Similarly, Java, Song, Finin, and Tseng (2007) used a sample of 1.3 million tweets from 76,177 users to describe why people use Twitter, which they summarized as “information sharing, information seeking, and friendship-wise relationship [sic]” (p. 60). While such studies are valuable, inferences made on the basis of the properties of a large data set are limited in what they can explain. In the latter study, asking people about their motivations for using Twitter would probably reveal an array of interesting motivations that do not neatly map on to these three groups. Because Twitter is such a vast network with so many user groups, simply collecting a great deal of data may not be adequate for describing use beyond simple queries. Qualitative methods, such as interviews, ethnographic observations, and content analysis, provide a rich source of data that allow us to go beyond description. For instance, qualitative methods can help unpack user presumptions about individual technologies, distinguishing general communicative or social media behaviour from behaviour that is specific to a platform.

Qualitative methods can also reveal much about social norms, appropriateness, or larger social concerns about technology. Twitter’s breadth and diversity requires recognizing that different user groups have different social norms and idioms of practice (Gershon, 2010). Generalizations made about one hashtag, meme, or network of users may not apply to another, providing only a small portion of the picture. Qualitative research allows scholars to investigate the practices of a particular user group, as it can go beyond tracking follower counts or hashtag use to include many more sources of input about a specific community or user segment. Moreover, qualitative data can often be useful for triangulating and augmenting quantitative results (see, for example, Honeycutt & Herring, 2009; Naaman, Becker, & Gravano, 2011). This chapter discusses a variety of qualitative research methods, including interviews, ethnographic fieldwork, and textual analysis.

Interviews
Interviews are a basic tool of qualitative methods in a range of disciplines, including sociology, media studies, anthropology, and human-computer interaction (Spradley, 1979; Wengraf, 2001). The content and protocol of the interview will depend on the research questions being asked and type of interview method (semi-structured, ethnographic, narrative, and so forth). While interviews can be conducted via direct conversations on Twitter, this approach produces a very particular and constrained style of interview, due to the 140-character limit. More common is interviewing Twitter users in person, or using a medium like the telephone or Skype to conduct long-distance interviews.

**Interviews on Twitter.**

The simplest way to interview Twitter users is to ask one’s own Twitter followers, or to @reply individual users and ask them quick questions. This approach has several advantages. It is quick and easy, does not cost anything, and allows the researcher to target broad populations in relatively small amounts of time. On the other hand, it is hardly representative (although one could argue that virtually nothing on Twitter would represent “society as a whole”). Besides the obvious bias of using a convenience sample made up of one’s own followers, many Twitter users will not reply to @replies from people they do not know, and getting the attention of specific accounts is easier said than done. The researcher’s earnest question may look like intrusive marketing spam, or simply get lost in the rapidly changing stream of tweets. And, obviously, it is difficult to conduct interviews of any depth using the service. Question-and-answer tweets might more properly be referred to as a very short survey.

Even with these limitations, I found this method quite useful as part of a larger project. I worked on one study that examined how highly followed individuals conceptualised their audience (Marwick & boyd, 2011a). My co-author and I were interested in “context collapse”, the phenomenon where large social-network sites like Facebook and Twitter “collapse” acquaintances from different social contexts into the single word, friend, or follower. We wondered if Twitter users recognized the coexistence of these multiple audiences, or had only a subgroup of followers in mind when they tweeted. Using the site, Twitterholic.com, which ranks Twitter accounts by number of users, we generated a list of the top 300 most-followed individual users on Twitter, removing media and business accounts. I created a research Twitter account separate from my personal account, which clearly identified my affiliation and purpose. I then sent individual tweets to each person, asking them who they thought of when they tweeted. My response rate was very low, but a number of people did respond. I then created a similar list of 300 accounts with 10,000–100,000 followers and repeated the process. The response rate was higher, and I followed up with each responder via Twitter. Two agreed to be interviewed, one via email and one over the phone. I then tweeted my own followers and asked for responses. The response rate was still higher. My co-author, danah boyd, had approximately 15,000 followers at the time (a very high number for 2009), and she retweeted my inquiry, garnering still more responses.

At this point we still did not have anything resembling a ‘representative’ sample, but we had several hundred responses and could group them into a rough taxonomy of “how people thought about their audiences.” We noticed that these categories remained constant regardless of the number of followers; in other words, many of the accounts with only a few hundred followers carefully curated their tweets in the same way that people with hundreds of thousands of followers did, and several of the most highly followed accounts claimed that they tweeted...
only for themselves. We also found several categories we had not considered while formulating our research questions. We could use this information, combined with what we had gleaned from our literature review of previous studies of the audience, to draw some rough conclusions about conceptions of the audience on Twitter. We used full-length interviews to test these assumptions (Marwick & boyd, 2011a).

The goal of the second study was to understand how teenagers use Twitter, and whether there are significant differences between teenage and adult Twitter use. We collected a large sample of tweets (400,000) that contained the hashtag, “#IGoToASchoolWhere”. This topic involved young people complaining or making funny observations about their high schools (the most popular tweet was “#IGoToASchoolWhere the kids are higher than the grades!”). An intern and I spent many hours going through the corpus, determining the most frequently retweeted tweets, the most prolific authors, and the highest-followed accounts that participated. We used quantitative methods to determine these three factors, but I also spent a lot of time reading through the tweets to get a “feel” for the sample. I searched for various college-related terms, and randomly sampled accounts to feel confident enough to make the assumption that most of the participants were teenagers, not adults.

However, in order to test this assumption, we needed to talk to the people participating in the #IGoToASchool hashtag. I again used my research account to send inquiries to the 300 most-frequent tweeters in our #IGoToASchool sample. I created a webpage with a URL-shortened link (e.g., bit.ly/teentwitter) which I included in the tweets, so users could verify that the study was legitimate. I got a single response, and it was of the “Uh, what?” variety. Unfortunately, the methods I had used in the audience study did not work. Teenagers are less likely than highly followed adult accounts to @reply strangers, and they change their usernames more frequently than the average Twitter user. I had waited too long after data collection to talk to the participants; I should have tweeted participants while the hashtag was trending. In general, when studying a particular hashtag or event, it is best to act quickly and try to get requests out while the topic is still trending or current. We had to abandon Twitter interviews and rely primarily on quantitative data and content analysis of the sample, along with a close reading of the tweets themselves.

After this experience, I think it is best to use the Twitter interviews as a supplement to triangulate results gleaned through other methods such as in-person interviews, content analysis, or quantitative analysis. Designing a research project so that it required interviews with specific Twitter users (as opposed to “Twitter users”) was a mistake, given the low response rate.

**Interviews about Twitter.**

A preferred interview method is to conduct long-form in-person, email, phone, or Skype interviews with Twitter users. These have the advantage of providing more information and background than can be garnered in 140 characters. In their study of “unfollowing”, Kwak, Chun, and Moon (2011) interviewed 22 Korean users about why they unfollowed people on Twitter, both in person and on Skype. The researchers compared this interview data with quantitative analysis of the following behaviour of 1.2 million Korean Twitter users. While some of the interview data confirmed their quantitative findings, other findings were surprising—such as people following others reciprocally, even if they did not know the person who had followed them (Kwak et al., 2011). Thus, on the one hand, as in this case, even a small number of interviews may help to augment the quantitative findings.
On the other hand, long-form interviews require more time and dedication, which may be difficult, depending on the population under investigation. The logistics of interview coordination are often difficult. Participants can be recruited over Twitter, but many researchers find that, out of necessity, they must use email or Facebook to reach out to a broader group of individuals, as the response rate on Twitter may be low. For instance, in a study of fans of the Brazilian band, Restart, who use Twitter, Recuero, Amaral, and Monteiro (2012) recruited 43 fans at Restart concerts and another 23 through social media. However, it may be difficult to recruit a very specific sample (e.g., “#IGoToASchoolWhere hashtag users”) or a representative sample, as the only people who will respond are those willing to talk to researchers. In this case, interviews may be used as part of a multi-methodological study to confirm or complicate previous findings. For example, Letierce, Passant, Breslin, and Decker (2010), in their study of how Twitter is used to spread scientific methods, surveyed scientists, collected tweets, and interviewed 10 researchers to clarify points in the data analysis.

As stated in the introduction, interviews can be an effective way to investigate normative assumptions about technology. When I interview people about individual social media technologies (like Twitter or Facebook), I ask a lot of basic questions (e.g., “What is a hashtag?”), and pay attention to how people explain their actions. When I first began interviewing technologists about Twitter (Marwick, 2010), I was sometimes tempted to show off my technical knowledge, but I found it more effective to feign ignorance and ask users to explain principles of the technology to me, which can reveal a lot about implicit norms and social practice. Depending on the study, I have also found it useful to ask interview participants to show me their Twitter accounts and walk me through individual tweets. This can reveal a lot of rich information about content strategies and presumptions that the user makes (as well as a gap between self-reported data and practice!). I have also found that it is necessary to understand Twitter as part of a multiplex of communication options (Haythornthwaite, 2001). Studies show that virtually all Twitter users use another social network, usually Facebook, in addition to Twitter (Brenner, 2012). Thus, it is important to distinguish social media behaviour in general from social media behaviour on Twitter.

**Ethnographic Research on Twitter**

For the purposes of this article, I will differentiate ethnographic interviews (which involve understanding participants’ meaning-making processes) from ethnographic fieldwork, which involves in-person observation and participation, ideally over a lengthy period of time, either online or in a particular physical location (Fetterman, 2009; Madden, 2010).

**“In-person” fieldwork.**

For my doctoral dissertation, I conducted more than a year’s worth of ethnographic fieldwork in San Francisco among members of the “Web 2.0 scene” (Marwick, 2010). My participants were avid users of Twitter and were happy to discuss it in interviews, but I also observed their use of technology in the field. While I was not always able to see people tweeting in social situations, the technologies constantly came up in conversation. I tried to keep records of even small mentions of the technology. I paid close attention to discussions and conversations about the “right” or “wrong” ways to use technology, which revealed many normative assumptions about the “best” way to use Twitter. I tried to track when participants chose to use
When informants mentioned that they read something on Twitter, or explained how they learned to use Twitter, this information was quite useful.

Comparisons of a person’s discussions of Twitter with their Twitter stream can reveal an added layer of useful information. For example, the information gathered by researchers in face-to-face settings may be consistent, or divergent, from the uses demonstrated by collected tweets or the type of information considered proper to share. Moreover, Twitter provides an articulated social graph in the form of the lists of following/followers that appears on every Twitter profile. Examining who chooses to follow—or not follow—who can enable greater understanding of a particular social scene in which ethnographic fieldwork is being conducted. This also applies to tweets about events, such as parties or conferences. Reviewing tweets about an event where ethnographic data was gathered can help flesh out participants’ meaning-making practices about their activities.

Twitter exists as part of an ecosystem of communicative options for users, and often what is posted on Twitter is not limited to that medium. Participants may discuss specific tweets or accounts on Tumblr or blogs; repost certain tweets to Facebook; use Twitter to post Instagram pictures or Foursquare check-ins; or take part in a variety of other social media interactions. Thus, Twitter must be understood as part of a mediascape which includes other forms of social media, as well as texting, phone calls, emails, and in-person discussions. Contextualizing tweets within this rich social web is important.

While it is a cliché to affirm the importance of field notes, they are the most important source of information a researcher will have once fieldwork is complete. I carried a small notebook in my purse and frequently left events to scribble down notes about what was happening. I wish I had not assumed that I would remember certain things that happened. While I have found that only the most disciplined researchers write up their fieldnotes every night, there is a reason that this is consistently recommended (Emerson, Fretz, & Shaw, 1995).

Digital or virtual ethnography.

Digital, or virtual, ethnography refers to the practice of observing and/or participating in a particular online group or community over a period of time (Hine, 2000; Miller & Slater, 2000). Given the traditional definition of a field site as a space, “the stage on which the social processes under study take place” (Burrell, 2009), many such ethnographies have investigated bounded online “places” such as bulletin boards, forums, or multi-player games like World of Warcraft (Boellstorff, 2008; Kendall, 2002; Nardi, 2010). Twitter challenges this model because it is a large, public site, making it difficult to bound, or even determine, exactly who or what one is studying. Jenna Burrell’s (2009) networked field site approach may be more appropriate, reframing Twitter as one part of a “network composed of fixed and moving points including spaces, people, and objects” (p. 189). In other words, Twitter may be one node on a network of field sites which include other social media sites, in-person locations, and material objects. (This was the case in my own dissertation project.) Twitter can be used as the primary place to observe interactions between people over a period of time, but these may be transient, ephemeral, and difficult to pin down.

Several approaches can be taken in determining the boundaries of Twitter as a field site. For example, a project could “follow” a set number of subjects who have been identified based
on other research, such as “feminist bloggers” or members of a specific gaming guild. When tracking interactions between subjects, and indeed any Twitter users, conversations must be persistently rebuilt “by way of exploring several previous messages that form the conversation threads” (Bougie, Starke, Storey, & German, 2011, p. 5). This can be difficult, as Twitter’s tools for such things are limited. Even when expanding an individual tweet to “conversation view”, items are often missing, such as contributions by other users and messages sent as new tweets rather than as replies. The search function on Twitter is notoriously problematic. The only way to see all messages tweeted by a particular account is from the individual profile page, where all @replies are aggregated, or by collecting tweets through the API. While such tools can aid in tracking down components of conversations, they can also be painfully slow.

Another, albeit incomplete, way to bound a group is to track the use of hashtags. For example, I worked on a collaborative study in which the authors were interested in fan practices around the television show, Glee (Marwick, Gray, & Ananny, 2013). We collected tweets that contained one of three hashtags: #glee, #klaine, and #brittania (the last two are portmanteaus for names of queer couples on the show). However, it is difficult to call people who use a particular hashtag a “community” by any strict definition of the term. Some hashtags do function as spaces of expression with recurring actors (Bruns & Burgess, 2011), but in other hashtags the participants do not interact with each other. Moreover, hashtags can be used for a wide variety of purposes besides identification. And the majority of Twitter users do not use hashtags, as they only appear in between 5–11% of tweets (boyd, Golder, & Lotan, 2010; Suh, Hong, Pirolli, & Chi, 2010). While this can be a convenient method, it is also an inadequate one.

Textual Interpretation

Because Twitter is partially a giant corpus of text, many textual analysis methods are appropriate for analysing Twitter interaction, from qualitative coding of individual tweets to close readings of particular accounts.

Textual and discourse analysis.

Qualitative research on Twitter also includes textual analysis and discourse analysis of individual tweets. Typically, these tweets are collected using an automated tool such as HootSuite Archives (formerly TwapperKeeper) or The Archivist, creating a fairly large corpus (discussed in detail in Gaffney & Puschmann, Chapter 5 in this volume). A subset is then selected for analysis and individually coded using textual analysis software such as Atlas.ti, NVivo, or Dedoose. For example, Zizi Papacharissi (2012) used textual analysis in her study of performative self-presentation in Twitter trending topics. Working with a sample of 1,798 tweets, the research team manually coded for descriptive features such as @replies and hashtags, as well as specific performative strategies which were operationalized based on concepts drawn from performance theory. Papacharissi also undertook discourse analysis on the same sample, identifying patterns and repetition in the text. She concluded that play is a primary performative strategy on Twitter, suggesting that “individuals confronted with a restricted stage for self-presentation seek to overcome expressive restrictions through imaginative strategies that include play” (Papacharissi, 2012, p. 1998). In other words, play provides a measure of deniability when voicing possibly controversial statements in a public forum rife with context collapse. In both
these studies, qualitative textual analysis was used to unearth subtleties of interaction on Twitter which may have been missed using more quantitative methods.

Coding itself is a complex process which can be approached in a variety of ways. In Papacharissi’s study, variables were strictly operationalized; for instance, a tweet was coded for “play” if it contained reordering, exaggeration, repetition, fragmentation, exaggeration and repetition, or (in)completion. Each of these strategies was carefully defined so as to make coding easier (for example, “reordering” was defined as “playing around with syntactical or grammatical rules, rearranging conventional sequencing of words to form sentences, and generally going against the norm of presenting thoughts into a written sentence” (Papacharissi, 2012, p. 1996). Other approaches include coding for the presence of a particular word (e.g., “drama” if the tweet contains the word “drama”), coding for particular names or hashtags, and so on; the right coding method will primarily depend on your research questions. For more on coding, see Charmaz, 2006; Corbin & Strauss, 2007; and Patton, 2002. When manually coding, I have found it easier to create a codebook based on pilot coding a subset of tweets, rather than rely entirely on grounded theory methods where categories come up during coding. This is primarily because I tend to create superfluous, repetitive codes without some sort of reference to draw from (for example, “celebrity”, “celebrities”, and “micro-celebrity” as three separate codes). Even though it is inevitable that the codebook will change throughout the coding process, having a fixed reference is invaluable and usually saves time in the end.

Close reading and critical discourse analysis.

In addition to social science methodology, humanities scholarship has provided methods that can be useful when considering Twitter. Close reading is a primary method in literary criticism, in which texts are read paying rigorous attention to individual words, syntax, and diction. Critical discourse analysis is a similar close reading strategy in which the researcher focuses on power relationships and links between texts and ideology (Fairclough, 2003). In both instances, the researcher will need to choose a relatively small sample of tweets to analyse. This may be tweets from top users; all tweets from certain users; tweets containing a particular hashtag; tweets to a particular user, and so forth. In a study of celebrity interaction on Twitter, my co-author and I chose three case studies—Mariah Carey, Miley Cyrus, and Perez Hilton—to demonstrate particular aspects of power relationships inherent in fan-audience interactions. I conducted a close reading of three months of tweets from each celebrity, paying close attention to their interactions with other Twitter users, particularly @replies (Marwick & boyd, 2011b). In providing thick description of specific tweets and interactions, we were able to illuminate specific patterns of use that would have been difficult, if not impossible, to ascertain with a more automated method.

Conclusion

Twitter is an immensely rich site for analysis, with a diverse array of users, multiple language communities, and a variety of subcultures who have taken to it. While, as we have seen, virtually any qualitative method can be better used to understand Twitter, the majority of studies on Twitter to date have been quantitative. While the “big data” approach has advantages, it also has limitations (boyd & Crawford, 2011). Identifying large-scale patterns can be useful, but it can also overlook how people do things with Twitter, why they do them, and how they
understand them. Quantitative studies often determine connections and networks, and interpret them “objectively” *ex post facto*, based on statistics and numbers. Instead, qualitative research seeks to understand meaning-making, placing technology use into specific social contexts, places, and times. Moreover, the claims to “truth” often made by “big data” methods frequently ignore the difficulty in finding any representative sample of Twitter, Twitter users, users, or people in general. Tweets gathered are often incomplete, even from APIs or the public “firehose”. The search function is imprecise. Twitter is used by a relatively small number of people to begin with, and leaves out entirely those who do not use the Internet. Rather than taking statistics for granted, the methods outlined in this chapter, and demonstrated throughout this book, show alternate ways to make sense of user practices, social norms, and power relations as they play out on Twitter, and throughout the digital world.


