Introduction

Scott A. Golder and Michael W. Macy

> [J]ust as the invention of the telescope revolutionized the study of the heavens, so too by rendering the unmeasurable measurable, the technological revolution in mobile, Web, and Internet communications has the potential to revolutionize our understanding of ourselves and how we interact… [T]hree hundred years after Alexander Pope argued that the proper study of mankind should lie not in the heavens but in ourselves, we have finally found our telescope. Let the revolution begin. . . .
>
> *– Duncan Watts (2012, p. 266)*

Online interaction is now a regular part of daily life for a demographically diverse population of billions of people worldwide. Facebook and Twitter are two of the most popular places where these interactions take place. A key difference is that most content on Twitter is publicly accessible via the Twitter API or through data resellers such as GNIP and Datasift, whereas most Facebook content is private. Thus, Twitter has emerged as the single most powerful "socioscope" available to social scientists for collecting fine-grained time-stamped records of human behavior and social interaction at the level of individual events. These data are also global in scale, allowing researchers to address fundamental questions about social identity, status, conflict, cooperation, collective action, and diffusion. This unprecedented opportunity comes with a number of methodological challenges, including generalizing observations to the offline world, protecting individual privacy, and solving the logistical challenges posed by "big data." This introductory chapter reviews current advances in online social research and critically assesses the theoretical and methodological opportunities and limitations.

This chapter is condensed from a larger survey of social media studies published in the *Annual Review of Sociology*. The material is included here with the permission of the publisher.

Opportunities and Challenges for Online Social Research

Scientific disciplines make revolutionary advances not only through new discoveries, theories, and paradigms, but also because of the invention of new tools and methodologies (Kuhn, 1962). The electron microscope, space telescope, particle accelerator, and magnetic resonance imaging (MRI) have allowed scientists to observe the world at greater scale or at finer resolution, revealing previously obscured details and unexpected patterns and experiencing the "eureka moments" of scientific breakthroughs. Newly developed tools for observing online activity are having a similar transformative effect on the social and behavioral sciences. Recent studies show how "digital footprints" collected from online communities and networks enable us to understand human behavior and social interaction in ways we could not do before. While the societal impact of electronic communication is widely recognized, its impact on social and behavioral science is also profound, providing global yet fine-grained observational data and a locus for population-scale experimentation.

Hard Science

Over the past century, there has been no shortage of social theory, but there are severe constraints on access to data. The reason is simple: social life is very hard to observe. For example, it is much easier to ask an isolated individual about their friends than to observe the ongoing interactions and exchanges that are the stuff of friendship. Ethnographic participant-observation studies and surveys of complete networks make it possible to fully document social interactions, but only at costs that can be prohibitively expensive to implement except in very small groups. The need to collect relational data through direct contact has therefore generally limited studies of social interactions to small, bounded groups such as clubs (Zachary, 1977) and villages (Entwisle et al., 2007). Lengthy time-series data on large populations, such as the Framingham Heart Study[1] or the National Longitudinal Study of Adolescent Health (Harris et al., 2009) are enormously expensive logistical challenges and are usually undertaken by multiple cooperating institutions in government and academia. Attempts to measure network structure at the population level by surveying egocentric networks (a randomly chosen person and his or

[1] http://www.framinghamheartstudy.org.

her network neighbors) can be useful for studying the attributes of network nodes (such as degree) and edges (such as tie strength), but this methodology has serious limitations (Marsden, 1990; Flynn, Reagans, & Guillory, 2010), including the inability to measure essential network attributes (e.g., distances, clustering, connectivity, and centrality) or social interactions (e.g., diffusion and polarization).

Because of the difficulty observing social interactions at population scale, most surveys rely on random samples composed of observations that are selected to be independent and to provide an unbiased representation of the underlying population distribution. However, independent observations preclude the ability to directly measure influence from a respondent's friends. We know that people do not entirely "think for themselves," but when we study opinion formation using random samples, we are left with little choice but to assume that a respondent's opinions are shaped entirely by his or her other traits, such as demographic background, material self-interest, or personal experience. As a result, we cannot rule out the possibility that demographic differences in opinions (e.g., the social liberalism of college graduates) are spuriously generated or exaggerated by the unmeasured effects of peer influence (McPherson 2004; Salganik, Dodds, & Watts, 2006; Della Posta, Shi, & Macy, 2013). Conversely, snowball sampling makes it possible to obtain relational data among network neighbors with which to measure demographic differences in beliefs and behavior net of the similarity between network neighbors, but the path dependence in selecting respondents makes it more difficult to obtain an unbiased representation of the population distribution.

Longstanding limitations on the ability to observe social interaction are rapidly disappearing as people all over the globe are increasingly choosing to interact using devices that provide detailed relational records. Data from online social networks – email archives, phone logs, text messages, and social media postings – allow researchers to relax the atomistic assumptions that are imposed by reliance on random samples. In place of path analytic models of social life as relationships among variables that measure individual traits (Wright, 1934; Duncan, 1966), data from online social networks makes it possible to model social life as relationships among actors (Macy & Willer, 2002).

Among these networks, Twitter stands out as by far the largest and most comprehensive publicly accessible source of online data on human behavior and social interaction. Each day, Twitter users leave billions of time-stamped digital footprints of social interactions, affording unprecedented opportunities for the collection of observational data on a scale that is at once massive and microscopic – massive in the sense that the people under study can number into the hundreds of millions and the data grow into the terabytes, and microscopic in the sense that individual time-stamped microinteractions are recorded. In place of retrospective reports about respondents' behavior and interactions, Twitter data can provide a detailed record of daily activities and the frequency and intensity of social relationships. These methods greatly expand our ability to measure changes in behavior, not just opinion; to measure these changes at the individual level yet on a global scale that spans diverse cultures; to observe the structure of the underlying social network in which these individuals are embedded; to travel back in time to track the lead-up to what later becomes an event of interest; and to find the "dogs that don't bark" (e.g., the failed outcomes that escape the attention of publishers, editors, and authors).

This research strategy is not new. For many decades, social and behavioral scientists have acquired data collected as a byproduct of the administrative or record-keeping processes of governments and organizations. Organizations track their membership lists, firms track the purchases of customers and the performance of employees, and banks collect massive data from credit card transactions. What is new is the macroscopic global scale and microscopic behavioral extensiveness of the data that are becoming available for social and behavioral science. Every tweet resides in a data warehouse waiting to be mined for insights into behavior, and to enable useful functions from spam detection to product recommendations to targeted advertising.

The Social Telescope

The ability to observe hundreds of millions of global tweets makes it possible to measure differences with small effect sizes that might otherwise be swamped by random variability. Just as an enormous antenna such as the Arecibo Observatory is required to detect the low-frequency radiation emitted from neutron stars (Lovelace & Tyler, 2012), Twitter comprises a massive antenna for social science that makes visible both the very large (e.g., global patterns of communication densities between countries; State, Abrahao, & Cook, 2012) and the very small (e.g., hourly changes in emotional affect and microbehaviors such as doing homework, getting drunk, and getting a headache; Golder & Macy 2011).[2]

[2] The associated website timeu.se (http://timeu.se/) provides an interactive tool for plotting the prevalence of keywords over the course of the day and week.

Tweets are recorded in real time rather than retrospectively. In social network studies, when individuals are given "name generators" and surveyed about their communication patterns, they are subject to a variety of potential biases. Question wording and ordering can cause respondents to artificially limit or otherwise vary the individuals they report, leading to underestimates of network size (Fischer, 2009; Pustejovsky & Spillane, 2009) or even measures of some other network (Burt, 1997) when survey questions mistakenly elicit report of a social tie outside the researcher's intended scope. In contrast, every tweet is time-stamped and passively recorded. If the message is retweeted, the data include not only the message content but also from whom the author may have received the message.

When activities are recorded via mobile devices, real-time geotagged mosaic accounts of collective behavior become possible that otherwise could not be reconstructed. As smartphone use increases in prevalence, the offline context of online behavior becomes available, such as common participation in a public event. For example, sampling a corpus of tweets that occurred during a certain time range and within a limited radius of a given event, we can reconstruct how online activity complemented a parade or demonstration or add a geographic variable back into an analysis that is otherwise blind to spatial location.

Relatedly, scientists can observe tweets unobtrusively, limiting the potential for Hawthorne-type effects in which researcher-induced desirability bias makes it difficult to observe normatively inappropriate behaviors (e.g., expressions of racial and ethnic prejudice), which participants may self-censor in surveys and in laboratory studies (Zizzo, 2010). Observing behavior unobtrusively ensures that the social pressures and normative constraints on individuals are exerted by their peers rather than by the researchers. Moreover, Twitter messages can be targeted to specific users with "@mentions," creating threaded conversations that can be observed in real time.

The task for the researcher is to see Twitter behavior as social behavior, the kind that might occur in any field site, be it a remote village, a law office, or a high school cafeteria. Some researchers explicitly conceptualize online sites as field sites in the ethnographic sense (Lyman & Wakeford, 1999). Relatedly, Twitter behavior represents social action in the Weberian sense – action that is oriented toward others (Weber, 1922), involving what Weber called "*verstehen*" – the subjective meaning for the actors involved. Paccagnella (1997) noted the multiple ways one might interpret the purpose, use, and limitations of technology, hence the need not to conflate the meaning to the researcher with the meaning for users (Pinch & Bijker, 1984).

Research Applications

Social Networks, Contagion, and Diffusion

Social network analyses have been among the earliest studies to use Twitter data. Twitter allows users to view indirectly the content received by those they follow only if the user also follows those same people.[3] In contrast to Facebook, which requires symmetric social ties (two friends must each indicate friendship with the other), Twitter (like most blogging platforms) allows asymmetric ties, leading to an extremely long-tailed degree distribution (e.g., celebrities often have many thousands of followers). Twitter also differs from Facebook by not demanding a clear tie to one's offline identity. Thus, the social network among Twitter users cannot be equated with an offline network based primarily on face-to-face interaction, as might exist in a school or workplace. Nevertheless, a recent study compared the volume and direction of messages, retweets, and @mentions among Twitter followers with the same users' offline friends and discovered a close correspondence (Xie et al., 2012).

Romero, Meeder, and Kleinberg (2011) found evidence to support the theory of complex contagions (Centola & Macy, 2007) by examining the spread of the use of Twitter hashtags. Hashtags for controversial topics such as politics were more likely to be adopted following exposure to multiple adopting neighbors, compared to topics such as music or sports. More recently, Weng, Menczer, and Ahn (2013) used Twitter hashtags to confirm a key implication of the theory of complex contagions – that the spread of complex contagions depends on network structure, a result that is consistent with the experimental findings reported by Centola (2010). Other studies have used online data to test longstanding theories about information diffusion, including the existence of well-connected "influentials" who initiate cascades. Popularized by Gladwell (2000) in *The Tipping Point*, the theory of these high-degree network nodes (or "hubs") was earlier proposed by Katz and Lazarsfeld (1955), who referred to them as "opinion leaders" in a two-step model of the flow of influence. Billions of advertising dollars are targeted at so-called influentials based on this theory, but a growing number of studies cast serious doubts. Cha et al.'s (2010) study of 1.7 billion tweets found that hubs "are not necessarily influential in terms of spawning retweets or mentions," a result consistent with Kwak et al. (2010), which also casts doubt on the influence of widely followed users on Twitter.

[3] That is, if A follows B, then A can see all of B's messages, but if B and C engage in a conversation, this is visible to A only if he follows both B and C. The purpose of this is more to prevent cluttering A's message stream with irrelevant conversations than to protect the privacy of B and C's conversation.

Using Twitter to measure the effectiveness of advertising is only one of several ways that these data have been used to study economic behavior. Chapter 3 of this volume shows how Twitter data have been used to measure unemployment, consumer confidence, social mood, investor sentiment, and the direction of financial markets. The chapter also refers to other sources for mining data for tracking and predicting economic behavior, including search queries and mobile phone data. The authors conclude by addressing the challenges that researchers face and identifying strategies for addressing these in future work.

Attention is also a valued resource in social exchange. Podolny (2001) suggests that attention is a prism or lens through which one is judged by others; having the attention of powerful others can, in turn, redound to one's financial benefit and is a signal to others about who is worth the investment of attention. Twitter users have been shown, for example, to rate others as more interesting to the extent that their own neighbors expressed interest in those others (Golder & Yardi, 2010).

In addition to social contagions, Twitter has also been used to track, study, and intervene in the spread of disease. Chapter 5 of this volume documents Twitter's strength as a public health medium for two-way communication, both as a health information source for users and also as a central hub for the collection and dissemination of health information that can improve early-warning and preparedness, aid disease prevalence mapping, and provide personally targeted health advice. Chapter 4 extends the research application from physical to emotional health, focusing on the use of Twitter data, in conjunction with census tract data, to study the ecological relationship between language use (e.g., sentiment analysis) and psychological experience.

Collective Action and Social Movements

Twitter data have also been used to study collective action and social movement mobilization. For example, data from Twitter have been used to provide digital traces of the spread of protest information and public sentiment in the Arab Spring (González-Bailón et al., 2011). Because information about protests reaches people through numerous channels besides social media, it is impossible to isolate the effects of social media net of other channels. However, users' messages can be used to measure the rate and extent of mobilization by tracking topic changes in user-generated content at a very fine-grained temporal level, and these changes can in turn be correlated with changes in the users' social and spatial environment as reflected in news accounts as well as the content of other users. For example, Weber, Garimella, and Batayneh (2013) track secular versus Islamist postings by Egyptian Twitter users over the course of the Arab Spring.

Twitter is not only a new channel for social movement organizers, but also for emergency managers to mobilize resources during disaster response and recovery activities. Chapter 6 of this volume extends research on mobilization to the responses of authorities to large-scale disasters. The chapter shows how Twitter has been used successfully to identify emergency events, obtain crowd-sourced information as the event unfolds, and provide up-to-date information to the affected community from authoritative agencies and for resource planning purposes.

Researchers have also used changes in the distribution of user-generated content not only to explain political outcomes but to try to predict them. For example, Digrazia et al. (2013) showed that local U.S. election outcomes were positively correlated with the number of times that Republicans had been mentioned in tweets. Nevertheless, a review of recent papers (Gayo-Avello, 2012) concluded that predictive claims are exaggerated. One important limitation on predictive power is that users of social media are not randomly selected in the way that is possible with survey research. Users preferentially choose to follow sources that conform to their existing worldviews (Sunstein, 2001) and preferentially rebroadcast (“retweet”) conforming messages (Conover et al., 2011). Boutyline and Willer (2011) showed that there is a valence effect to the formation of so-called echo chambers – those further to the political right exhibited more ideological homophily in who they chose to follow on Twitter.

Chapter 2 of this volume uses representative case studies to show how Twitter data can be used to track public opinion through its expression in political discussions. The chapter also identifies challenging problems in measuring opinion and how these might be addressed in future research. These and other studies show that the use of social media to study opinion dynamics provides a potentially important complement to – not substitute for – traditional survey methods. Each can be used to obtain information that is missing in the other. Surveys provide more reliable estimates of the distribution of opinion in the underlying population but typically provide only retrospective responses and lack network data with which to study the flow, diffusion, and clustering of opinion.

Challenges

The Privacy Paradox

Twitter data confront researchers with imposing hurdles, ranging from validity of both the data and how they are sampled to the ethical issues regarding their use. Online data present a paradox in the protection of privacy: data are at once too revealing and not revealing enough. Twitter data lack the detailed demographic profile information that is standard in survey research. For example, while most Twitter data are public, profiles can also be private, and direct messages (which can only be sent to followers) are not world readable even for public profiles. Many users provide sparse, invented, incomplete, or ambiguous profile information, making it difficult for researchers to associate the content of tweets or the attributes of network nodes with basic demographic measures such as age, gender, ethnicity, or location. Identity is slippery and poorly defined in some online communities, where participants are known only by a self-chosen username that they may change at any time. In some cases, it is difficult to tell who is a human; the growing incidence of "spam accounts" is worrisome, and despite progress in spam detection methods (Yardi et al., 2010), spammers manage to keep the arms race going. As spammers become more sophisticated, it becomes harder for social scientists to clean the data they collect without specialized technical training, a problem we explore in more detail later.

Nevertheless, rapid progress is being made to address these limitations. For example, Compton et al. (2013) showed how label-propagation algorithms can be adapted to potentially geotag the vast majority of Twitter users to within a few kilometers, and Jernigan and Mistree (2009) (using Facebook data, although the method can also be applied to tweets) showed how social media content can be used to infer a wide range of user attributes, including age, gender, sexual preference, and political party affiliation. These advances illustrate the other side of the dilemma – that online data may not be private enough. These new sources of data raise challenging procedural, legal, and ethical questions about how to protect individual privacy that are beyond the scope of this review, but there is a growing body of research showing that anonymizing or encrypting data is not sufficient for protecting privacy, as this can sometimes be reverse-engineered (Backstrom, Dwork, & Kleinberg, 2007; de Montjoye et al., 2013) using the unique attributes of individuals' egocentric networks or physical mobility patterns.

Access to Twitter data can be a significant challenge. These data are owned by a private company that restricts access largely to protect the privacy of their subscribers. For example, Twitter removes users who decide to close their accounts, but Twitter has no way to make sure that these data have also been removed from all the copies made via the Twitter API. These restrictions have raised concerns about reproducibility of results, corporate influence, and stratification in the research community between a small elite that is well connected to social media companies and everyone else (Boyd & Crawford 2011; Huberman, 2012). New protocols and institutional arrangements are needed to align the goals and needs of industry and the academic community. In addition, advanced programming and other technical skills are required to access and process large semistructured datasets such as Twitter.

Measurement Issues

Although advances in identifying sentiment and opinion from text are proceeding rapidly (Pang & Lee 2008), we can only measure inner states indirectly, through their behavioral expression. For example, psychological lexicons (Pennebaker, Francis, & Booth, 2001) can be used to measure the expression of affective rhythms on Twitter on a global scale (Golder & Macy, 2011), but these methods cannot account for temporal lags between expression and experience.

An important limitation in all observational studies of network contagion, whether online or offline, is the difficulty distinguishing between homophily and contagion. Homophily refers to a variety of selection mechanisms by which a social tie is more likely between individuals with similar attributes and environmental exposures (McPherson, Smith-Lovin, & Cook, 2001). Contagion refers to influence mechanisms (e.g., imitation or peer pressure) by which traits diffuse along network edges. Homophily and contagion offer competing explanations for network autocorrelation, which refers to the greater similarity in the attributes of closely connected nodes. Based on simulated networks, Shalizi and Thomas (2011) conclude that "there is just no way to separate selection from influence observationally" (see also Manski, 1993). This does not mean that observational studies using online networks are useless, but researchers need to refrain from assuming that the observed network autocorrelation reflects contagion effects and to acknowledge that the similarity between adjacent nodes may reflect the mutually reinforcing effects of influence and selection whose separate contributions may be impossible to tease apart. For example, although Ugander et al. (2012) controlled for demographic similarity (sex, age, and nationality), there are countless other ways in which shared environments, affiliations, interests, and personality traits might cause two friends to join Facebook independently but not on the same day, making it look like the "early adopter" influenced the friend they invited who would have joined anyway.

One solution is to conduct controlled experiments that manipulate exposure to a possible contagion, as in the Facebook experiment by Bond et al. (2012). Where experimental methods are not feasible and the only data are observational, researchers can tease apart influence and selection by using an instrumental variable that is correlated with actors' neighbors' exposure to the contagion but not to actors' own exposure and then comparing the presence of