

January 2012 Issue • Volume 40 • Issue 1

Social Science with Social Media

Scott Golder and Michael Macy, Cornell University

As a discipline devoted to explaining patterns of human behavior and social interaction, sociologists often have to choose whether to rely on direct real-time observation of very small numbers of non-representative individuals (e.g., in field observation or in the laboratory) or to rely on indirect retrospective accounts obtained through survey responses from large representative samples. Social media offers us the opportunity for the first time to both observe human behavior and interaction in real time *and* on a global scale.

This is possible because many activities of everyday life are now taking place online. We maintain touch with distant friends and relatives using Facebook⁵. We exchange news and opinions with friends and follow politicians¹ and celebrities on Twitter. We find dates and spouses in online dating sites^{4, 11}. We put our professional social capital to use on LinkedIn. We engage in market transactions on eBay⁹ and Amazon. We meet up with friends to battle for treasure in virtual worlds¹². We work together to author Wikipedia—one of the most phenomenal examples of large-scale mass collaboration in human history⁸ (and one that ASA's President is encouraging sociologists to be involved with. See www.asanet.org/about/wiki_Initiative.cfm).

This is not to say that the population of users who interact online perfectly mirror the offline world. Disadvantaged and elderly people continue to be under-represented online in terms of both participation and skills^{2, 7}, despite rapid increases in access to the web, even in the developing world. Although social life online is certainly different in many ways from social life offline (i.e., the absence of face-to-face interaction, the lifting of geographic constraints, and the ability to search and filter our friends), we remain the same people whether we are online or off. We want to find desirable jobs and romantic partners, we need to be able to cooperate and coordinate with others to complete a task successfully, we like to share personal news, argue, commiserate, and celebrate with our friends, we need to quickly mobilize the members of our social movement, and we worry about our status in our social groups.

The web, and the internet on which it operates, sees everything and forgets nothing. Every email you receive, song or movie you stream, and URL you click is digitally recorded on the computer servers that host the web. But the same passively-generated digital traces of activity that make social media services functional to their users, that enable services to detect spam, offer recommendations, and target advertisements, can also be used by social scientists to directly observe human behavior in detail as it unfolds over time on a global scale.

Mood Rhythms and Twitter

Our recent study of mood rhythms⁶ is an example of how social scientists can take advantage of the digital archives of online activity. Twitter messages are real-time, spontaneous reports of what millions of people around the world are seeing, feeling, thinking and doing. We took advantage of this unprecedented research opportunity by collecting over 500 million messages from 2 million users worldwide. Using a prominent text analysis tool¹⁰, we measured the incidence of hundreds of words that express positive and negative affect in users' messages to map how individual mood varies from hour-to-hour, day-to-day, and across the seasons. The results were striking. We found that there are robust rhythms across diverse cultures, from India and Africa to Australia and Brazil. People appear to be in better moods in the morning that deteriorate throughout the day, better moods on the weekend that improve (!) over the work week, and better moods when the days are getting longer.

We also created a public web service (<http://timeu.se/>) where users can track when and how often people mention a given activity, such as celebrating, doing homework, and shopping online. In the first month of the site going live, 14,000 people have queried the site 120,000 times to explore the rhythms of human activity. These patterns can also be useful for social and behavioral scientists, policy makers, politicians, and industry. When do staff meetings usually take place (Tuesdays at 9)? What about accidents (weekdays at 7 and 5 but earlier on Friday), headaches (Sunday morning – no surprise there – and Monday at 5 pm), contractions (Monday at 5 am) or fighting (Saturday night into Sunday morning)?

Caveat

Of course there are many things that analyzing Tweets cannot tell us. Twitter tells us when people write about their activities, which may not be when they actually occur. We know when people are in traffic but not when they're in the bathroom, and when they're having breakfast but not when they're having sex. We know about only those things that users like to share publicly with their friends. Privacy concerns limit investigations to the information that users are willing to make public and which private companies like Facebook or Google are willing to make available to outsiders. Data from online networks and communities typically include little demographic information about the users, without which many sociological investigations are not possible or useful. Online data should therefore be viewed as a complement to, and not substitute for, data collected by traditional methods. Indeed, in many cases, the value of online data may depend on opportunities to integrate with data obtained from surveys. For example, a recent study of social network structure and social inequality matched a massive collection of UK telephone logs with census data about the exchange areas³. But the detailed, time-stamped, cross-cultural observations that social media makes possible are too valuable to not include as an important part of our methodological toolkit.

Disciplines are revolutionized by the development of novel tools: the telescope for astronomers, the microscope for biologists, the particle accelerator for physicists, and brain imaging for cognitive psychologists. Social media provide a high-powered lens into the details of human behavior and social interaction that may prove to be equally transformative.

However, this development is so new that many graduate programs have not had time to catch up in providing the necessary training. As a consequence, most of the social and behavioral science using online data is coming from computer and information scientists who do not always have the training required to ask the right questions, or to recognize unfounded assumptions and socially unjust ramifications. The digital records of online behavior and social interaction hold the promise of opening up a new era in the social and behavioral sciences, but when and whether this opportunity is realized may depend on the involvement and leadership of sociologists with the necessary technical and computational skills.

References

1. Boutyline, Andrei and Robb Willer. 2011. "The social structure of political echo chambers: Ideology leads to asymmetries in online political communication networks." Working Paper, Political Networks Paper Archive.
2. DiMaggio, Paul, Eszter Hargittai, W. Russell Neuman, and John P. Robinson. 2001. "Social implications of the internet." *Annual Review of Sociology* 27:307–336.
3. Eagle, Nathan, Michael Macy, and Rob Claxton. 2010. "Network diversity and economic development." *Science*, 328:1029–1031.
4. Feliciano, Cynthia, Belinda Robnett, and Golnax Komaie. 2009. "Gendered racial exclusion among white internet daters." *Social Science Research*, 38:39–54.
5. Golder, Scott A., Dennis Wilkinson, and Bernardo A. Huberman. 2007. "Rhythms of social interaction: Messaging in a massive online network." in *Communities and Technologies*, edited by C. Steinfield, B. T. Pentland, M. Ackerman, and N. Contractor. Springer.
6. Golder, Scott A. and Michael W. Macy. 2011. Diurnal and seasonal mood vary with work, sleep and daylength across diverse cultures." *Science*, 333:1878–1881.
7. Hargittai, Eszter. "Digital na(t)ives? variation in internet skills and uses among members of the 'net generation'." *Sociological Inquiry*, 80:92–113.
8. Keegan, Brian, Darren Gergle, and Noshir Contractor. 2012. "Do editors or articles drive collaboration? multilevel statistical network analysis of Wikipedia coauthorship." in *ACM Conference on Computer-Supported Cooperative Work*.
9. Kollock, Peter. 1999. "The production of trust in online markets." in *Advances in Group Processes*, Vol. 16, edited by E. J. Lawler, M. Macy, S. Thye, and H. A. Walker. Greenwich, CT: JAI Press.
10. Pennebaker, J.W., M.E. Francis, and R.J. Booth. 2001. *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Lawrence Erlbaum Associates, Mahwah.
11. Taylor, Lindsay Shaw, G.A. Mendelsohn, Andrew T. Fiore, and Coye Cheshire. 2011. "Out of my league: A real-world test of the matching hypothesis." *Personality and Social Psychology Bulletin*, 37:942–954.
12. Wang, Jing, David A. Huffaker, Jeffrey W. Treem, Lindsay Fullerton, Muhammad A. Ahmad, Dmitri Williams, Marshall Scott Poole, and Noshir Contractor. 2011. "Focused on the prize: Characteristics of experts in massive multiplayer online games." *First Monday*, 16(8).

[Back to Top of Page](#)
