Bernie Hogan

Comment on Elena Pavan/1. Considering Platforms as Actors

(doi: 10.2383/75767)

Sociologica (ISSN 1971-8853)
Fascicolo 3, settembre-dicembre 2013

Ente di afferenza:
()

Essays

Comment on Elena Pavan/1

Considering Platforms as Actors

by Bernie Hogan

doi: 10.2383/75767

1. Introduction

In "Collective Action and Web 2.0," Elena Pavan charts a course for future sociological analyses of political mobilization on Twitter, and the Internet in general. In particular, she charts a middle course between pundits claiming either the utopianism of Shirky's collective mobilization or dystopianism of Morozov's vision of surveillance and slacktivism. She instead seeks to situate internet-mediated communication in a multilayered web of social life. In doing so, she argues that these media allow for the global consolidation of symbolic practices that nurture offline interaction. This is accomplished through the "extended set of social relations established via social media" and the newfound capacity for "creating shared symbolic systems and visions." She explores these claims empirically by looking at the case of "Take Back the Tech," an awareness campaign designed to highlight and ultimately end violence against women.

Despite a great deal of early conciliatory remarks suggesting she is going to chart a middle ground (say between Morozov and Shirky), Pavan still makes a rather strong claim: "all these different 'transformative effects', or impacts of social media stem directly from the augmented network potential that is proper of the Web 2.0 tools." If this is truly the case (as I suspect it is not), then it is worth stepping back from the "potential" of ICTs and consider these media as actors in their own right, with their own agendas and means for facilitating all this transformative discourse. I

argue that these social media do not simply 'transform' collective action. They insert themselves, curate information and shift discourses from being something that can be described as a network, to being something that is a network *a priori*. In doing so, they do not merely substitute for old media but act as "micro-arbiters" of what constitutes legitimate discourse.

Twitter is neither a public utility nor a decentralized community-driven project. Like Facebook, Google and their ilk it is a private corporation. Like Facebook and Google, Twitter is a publicly traded company interested in appeasing shareholders, and it is driven by a general profit motive that needs to be sustained indirectly through participation and advertising. It would be a mischaracterization to make a strong claim that all digital goods corporations are solely and exclusively by profit and profit through increased activity. For example, Craigslist, the highly popular US classified site is resolutely interested in being a useful service more than being a profitable one. However, profit through audiences is a general logic that persists and is institutionalized through public trading and viral features designed to encourage people to get their friends and family to join.

From this basic point stems two key concerns about research on ICTs, and particularly platforms that host and redistribute user-generated content. First, there is no guarantee of getting the right data or the complete data. They do not have an obligation to share such data. In fact, in many cases, accessing and distributing such data might be a violation of one's privacy rights. Second, the way the data is distributed to individuals is as much about the platform's interest in profit and increased activity as it is in being part of the public good. This is not to suggest that such platforms are antithetical to the public good, or even that they are neutral. Rather, I suggest that they intervene in ways both direct and indirect. The bumpy terrain of who can access data, for what purpose and how it is to be represented reinforces one of the core concerns of digital social research: to be publicly accessible is not the same thing as to be in the public domain or the public's interest. To fairly contextualize Pavan's work, I discuss these issues in turn.

2. Data Access

That Twitter is a corporation and not merely a brave new storehouse of public content has become particularly obvious to researchers in 2013. This year Twitter made significant changes to its API to facilitate more access to some kinds of data and less access to others. In doing so, Twitter makes the sort of analysis done by Pavan virtually a thing of the past. For the benefit of the technologically less informed, an

API is an Application Programming Interface. It is the mechanism by which large data providers allow machine-readable versions of slices of its data to be accessed by third parties. APIs allow real time tweets to be displayed in all sorts of places from public billboards at the Olympics to an academic department's homepage. Often such APIs require authentication as a condition of access in order to regulate how much data is distributed and in what form.

APIs do not give complete access to a data store for a host of reasons such as privacy and intellectual property. But these restrictions are also a way to mitigate competitiveness and retain the value of the platform's data. As an example, LinkedIn, the "professional" social networking site charges its users to see a variety of social connections. In this respect, withholding this data from a public API is an integral part of LinkedIn's business model and one they are careful not to undermine through data sharing. It is also the reason that despite the presence of InMaps, a visualization product from LinkedIn, there is otherwise a paucity of work done on LinkedIn ego networks – they simply are not available through the API.

Twitter has been a central part of the digital researcher's toolkit for the last few years. Granted, Twitter is also a significant cultural force. But it is an *easily accessed* cultural force. Papers from the social computing paradigm have engaged in a veritable arms race of "Big Data" consolidating millions of users and billions of tweets [*cf.*, Cha et al., 2010]. However, Twitter does not see any money directly from such work, and the scores of research labs around the world that were constantly querying Twitter's servers have been a significant drain of time, energy and thus expenses. So, despite many cries from the research community, Twitter has restricted access to a number of key API features. In particular, it has removed "whitelisting" and severely limited access to lists of friends and followers.

Whitelisting on Twitter granted some users the ability to make a large number of queries within a given time window. Normal users could make a paltry 350 queries an hour. This was enough for a regular user to keep up on their tweet stream, even under heavy use. However, researchers would often want much larger volumes of data. Whitelisting permitted users up to 20,000 queries an hour. In full disclosure, I had one of these coveted accounts, although I mainly used it for teaching purposes. Now, instead of 20,000 queries per hour, I'm restricted to 180 every 15 minutes. While this still seems like a lot of data throughput, the capacity to capture the follower graph is now significantly limited. What used to take me a matter of 10-20 minutes of querying follower lists now takes days as the calls run out and the computer simply sits idle while I wait for my access to be renewed. Some dimensions of the Twitter graph are now simply intractable for all practical purposes.

In her paper, Pavan notes that only 3 percent of the possible ties in the network were activated. That is to say, there were 287 edges (including retweeting, following and mentioning) between the 100 early tweeters of the TBTT movement. This is a meaningful claim and one that significantly contextualizes participation on Twitter. In many respects it suggests only very modest engagement. Given the way the figure is reported it appears that even less (<1 percent) of the dyads involved had any direct interaction concerning Take Back the Tech, while most of the relationships were more longer term "following" relations.

To determine this figure Pavan had to download the follower graph for all individuals who used the Take Back the Tech hashtag. As it is currently structured through Twitter's API this entails getting the list of every one of these 100 people and for each person in the group a list of their followers. If one person appears in the other's following list, there is an arc from the second person to the first (or an arc the other way around if we think about message flowing from first to second). Only 5000 followers are returned at a time. It only takes a small number of people with very large followings to really slow down this process and sap all the queries offered in a very short period of time.

Reporting these results demonstrates a central concern of research using "live" online data. The researcher is beholden to the wishes of the data controller, and their research can be driven in large part by what the controller permits. Pavan used NodeXL, a third party plug-in for Excel, to capture this data. NodeXL has tried to manage the new limited query issues by providing abilities to stop, start and wait through query downtimes. However, it is beholden to the same rules set out by Twitter that everyone else must consider. Marc A. Smith, founder of NodeXL, says that he no longer captures following information because it is simply too onerous and not likely to finish in time to make a meaningful report.[1]

In the absence of large pots of state money to buy access to this data from third party resellers such as Gnip or DataSift, researchers are thus left trying to shift their research questions to compensate for this difference in access.

3. What Does Data Mean?

Beyond the sociopolitical matter of who owns and accesses data on Twitter is a deeper question about what such data actually means. A full critique of contemporary information society theorists could fill a book or a career, but it is nevertheless worth raising a small number of essential concerns if only superficially.

[1] Personal communication.

These issues tend to orbit the academic device of the "social network." Basically, networks are useful, generally mathematical, representations of relationships between actors. That is to say, in many cases, networks are not real but convenient abstractions that have a great deal of correspondence with reality. Networks are not alone in dealing with issues of correspondence with the external world. In fact, most methodologies have to make some concessions to external validity. Surveys can carry biases, interviews can be loaded, and focus groups can suffer from groupthink. Yet, this issue is not easily resolved, and tends to be an ideological one. Methodological realists would contend that networks are real and we merely discover them. Methodological constructivists would suggest that networks are abstractions from reality, and this abstracting process necessarily distills or alters the reality it seeks to study.

This tension between realists and constructivists pervades the network analysis community, and creeps into introductory books. For example, Newman, a realist and physicist, notes that "the main disadvantages of network studies based on direct questioning of participants are that they are first laborious and second inaccurate" and "data based on direct questioning are also plagued by uncontrolled biases. Answers given by respondents are always, to some extent, subjective" [Newman 2010, 43]. It is known that respondents have biases. From the earliest work on respondent inaccuracy in networks (generally considered Bernard, Kilworth and Sailer's studies, 1979) towards more recent work, this bias has shifted from being a bug to being a feature. That is, it seems that individuals have well defined biases, rather than random ones. For example, when a respondent lists people emotionally closest to her, if she mentions one family member, she is likely to then mention another less close family member rather than the person who would be 'rank-ordered' as immediately less close than the first [Marin 2004]. Individuals might be bad at remembering who was in the most recent class, but they are good at remembering who attends class regularly [Freeman, Freeman, and Romney 1987]. These biases make sense and help individuals separate social noise from important facts.

From the social network community, Hennig *et al.*, begin their recent introductory text in no uncertain constructivist terms: "Social network studies entail the use of network representations to understand social phenomena. Social networks do not exist as such but only as concepts" [Hennig *et al.* 2013, 1]. Social scientists, often sensitized to the many minute decisions made in creating a boundary for what constitutes a tie between two people, tend to take a more constructivist perspective (although not always).

The use of social media networks for social network analysis can be a way to defer concerns about a constructivist network: one either is a Twitter follower or not; one is either a Facebook friend or not; one retweets or does not. Thus, the nature of

the data that is imported permits us to be realists by studying the data traces from well-structured sociotechnical systems.

Those who study the meaning of these links within these systems tell a more complicated picture that unfortunately pulls analyses back to a constructivist mode. In particular, boyd [2006] has noted the problems with the ontological status of friends on social network sites. From interviews with teenagers on MySpace she suggests one might be friend for a dozen different reasons such as status signaling, social pressure, and the fact that it is easier to say yes than no. Thus Twitter will give us a graph based on who befriended whom. Yet, the meaning of this befriending process might, to the chagrin of realists such as Newman, be even more biased than links drawn from personal recall.

When visualizing networks, as was done at several points in Pavan's paper, it is easy to consider the visualized network as a self-contained and coherent entity. However, many decisions must be taken in order to select that network in the first place: what time window, who is a member, should the links represent followers or mentions? When links represent multiple potential relations, such as mentions, retweets and following, how are we to interpret the topological features? What do higher order features such as clusters really mean at the microsocial level?

4. The Network: From Technique to Technology

Even if we resolve these issues to our satisfaction – namely that we have graphs that are based on stable, well sampled and well bounded connections between individuals – we have to contend with a deeper issue. As mentioned above, a study of Twitter networks merely defer a discussion of realist versus constructivist positions, but do not resolve such a discussion even if the data is easily accessible. This is because Twitter, by its nature exists as a network *a priori*.

Social network analysis used to be a one-way process moving from relationships to a network. People would communicate, interact, meet, etc., and academics would distill these relationships into networks as data structures. On the Internet, the network comes first in a deliberate technologically contrived way. As the Internet became more complex this process started to go the other way, as developers with little or no experience with social network analysis sought to find ways to facilitate sociality online. Many forms of sociality were tried, such as MOOs, chat rooms, online journals, email, instant message, random chat partners a la Chatroulette, etc. Aside from the now ubiquitous email, one technology above all has been successful: the social network site. These sites have captured the public consciousness from California

to Cairo to China and everywhere in between. Statistics on their use are constantly changing, but at the moment, Facebook is second only to Google in traffic (and in some countries such as Indonesia even outpaces Google; *cf.*, Alexa 2013). Twitter's participation is significantly more skewed but it is still among the top social network sites globally alongside Twitter, LinkedIn, Tumblr, Pinterest, Vkontakte, and Sina Weibo at present.

Social network sites took their inspiration from social networks from very early on in their development. One of the first sites was an explicit homage, called "Six Degrees" [boyd and Ellison 2007]. This is based on the notion that a third party manages two key features: the generated content of the users and the distribution of that content based on a logic drawn from explicitly denoted connections between users. Content typically would be distributed to those denoted as friends or inside a specific social circle, but in almost all cases these routes of distribution would be clearly denoted person-to-person connections. Thus social network sites represent a world made in the image of a social network, rather than created as networks after the fact.

By explicitly managing both the structured data constituting a user's profile and the structured data representing links, a data curator can effectively push personalized content to its users. This is a marvel not just of social network analysis, but also of highly scalable computing and database engineering. Consequently, the methodological realists are winning – not by demonstrating their position as more accurate – but creating and analyzing a world coded and deployed as network. It is this world in the network's image that Pavan is explicitly speaking to, as she notes an "explicit focus on the networked structures of participation that derive from the use of these communication tools." Reading her paper one might almost presume that such networks merely emerge from the interactions on the site. By contrast, these networks are the *sine qua non* of social network sites.

This newfound structuring has a number of consequences for social mobilization. In Pavan's case, these are often presented in a particularly positive light. Networks here are seen to "enhance communication potentials […] through a common framework of reference." She also refers to Twitter as "enriching" and "exploiting the potentialities." However, Twitter also leads to a number of disciplining features as well. As she notes in one footnote, those who received no attention for their tweets had the same intent but did not use the correct or most attractive symbolic resources such as the correct hashtags or practices. Sadly for these individuals, the lack of a shared presence does not allow such individuals to easily learn from their failure to attract attention – they simply recede into the background.

When using an ostensibly virtuous example of women's rights and technology, this structuring of relationships can be seen as a positive phenomenon. However,

the very same tools can be used to exclude and marginalize individuals. Far-right extremist groups are also on Twitter. Those interested in persecuting queer and LBGT groups now have new ways to harass, name and shame, and the newfound configuration of technology allows such like-minded individuals to also find each other as easily. In which case we ought to pay careful attention to what exactly is being enhanced and enriched on Twitter – does the site enable greater forms of social justice to percolate or does it merely elevate many local contestations to a global stage? If it is the latter then the "networked" nature of the medium may actually encourage global homogeneity in discourse as individuals feel the need to focus their diverse issue sets into large global agendas (and practically, be found among the hashtags and search terms that facilitate much of this networked communication).

5. The Personalization Agenda

Twitter's logic is one of individualized consumption and selectivity. Each person gets their own feed and no two feeds are alike (except in the highly unlikely case that two people follow exactly the same set of individuals and are so similar that promoted tweets come in the same). But despite the fact that each account is as unique as a snowflake, give or take an order of magnitude, small world clustering still persists.

Pavan here uses time zones to suggest that for the Take Back the Tech campaign people from around the world were tweeting and participating. It is an argument for the global and diverse nature of this campaign. Taking aside the methodological concerns regarding the association between time zone and actual locality, there is an argument to be made that spatial diversity does not entail relational diversity. While Pavan noted this is not a concern about "cyberbalkanization" where the Internet is literally filtered by nation states into separate networks, there is still the concern about "Filter Bubbles" where content is so personalized as to become ideologically homogenous [Pariser 2011]. Filter bubbles do not necessarily require people to be close in space, but rather nearness in terms of network structure and content. The filters, hashtags, friend recommenders and promoted content of these sites foster different forms of similarity than mere propinquity. Thus, while the Take Back the Tech campaign was not local in the traditional geographic sense, it could still have been primarily concentrated among a like-minded group of highly connected individuals. Certainly the networks shown in the paper suggest a relatively high level of closure among the early actors.

One of the consequences of such clustering and content-based similarity is that it can reinforce ideological feedback loops and filter out disconfirming information.

Even with hundreds of friends it is possible to skew almost entirely left or right wing within an online network. Despite the fact that Justin Bieber is followed by almost one tenth of the global Twitter population, I cannot recall the last time I saw one of his tweets in my stream. More seriously, despite only limited work on political polarization (rather than mobilization) on Twitter (e.g., Conover et al. 2011), if it follows past work on blogs, such polarization is likely to be a mainstay of the Twitter experience.

In such a case, I argue that Twitter does not enhance or enrich political mobilization, *per se*. Instead, it reconfigures our political attention in often highly globally concentrated ways. Some of this reconfiguration is a function of the medium itself (140 character limits, icons next to tweets, etc.). In other cases, this reconfiguration is tweakable, such as recommending friends, promoting tweets, and sorting tweets (or any other personalized 'feed'). In these latter cases, the data curation decisions made at Twitter can reinforce local content or ignore it. Twitter can recommend friends that close triads (as was formerly the case at Facebook) or do so using some A/B-tested secret sauce, as is reportedly done at Google for most forms of personalization. But in almost all cases there is some guiding imperative behind such curatorial decisions that is about the self-sustaining (often profit generating) motives of the company itself. It is often indirect, such as clicks on advertisements, likes per hour or time spent on the site. But even if it is indirect, the motive is still there: we mediate (and thereby curate) your global communication and in return you help us by paying attention in ways that contribute to our long-term viability.

One might argue that Twitter or social network sites are not new in this regard – all media have a self-interested imperative and in some ways redirect our attention. Certainly. This argument can date back to Innis [2007; first published in 1950] if not before. Then at the very least, the discussions about Twitter's potential ought to consider how its curation patterns reshape our communication practices (including who receives what messages), not merely enrich them. Yet it is still possible to look for that something new among social network sites. It is not global reach or instantaneousness. It is the structuring of social relations as a social network *a priori* and then using this structure to deliver highly personalized content whether such content is political, entertainment, current affairs or personal performance; and to do this personalization in real time. In doing so, this makes Twitter (or Facebook, etc.) much more than a medium for new forms of mobilization, but a political actor contributing to the types of mobilization and the types of research done on this mobilization.

6. Conclusion

Digital data presents exciting opportunities for the testing and assessment of all forms of social life that heretofore might have been far more difficult to collect and analyze. However, this comes with a whole host of caveats and pitfalls, including those in the rhetorical guise of wisdom. Below I point out three such rhetorics and offer caution based on the previous discussion.

1. The online-offline false dichotomy: This is a rhetorical strategy of suggesting that because individuals contribute to online spaces with others they tend to know offline, that we need not think of the online world as a separate sphere of activity. While it is fair to think of the Internet as not arriving *ex nihilo* and to not draw the line at offline-online, it is fair to still recognize that the affordances and constraints of online life regulate the patterns of behavior therein much the same way that affordances and constraints of houses, malls and roads do offline. We are constrained in different ways in different places. But in the rush to legitimate online spaces, it is possible to forget to scope these spaces as inherently reductive and simplifying.

Much can be done with such simplification; network analysis as a field is based on the promise that by simplifying relations to a graph the outcome is far more useful than the constraints. Yet, there are also losses. When we think of 'friends' online, we are measuring a database-oriented technological construct, not an internally felt social construct. As such, challenges to the online-offline dichotomy are often a way to allay concerns about external validity. In cases where we are interested in opinion changes or alterations of behavior, we ought to still remain vigilant as to what constitutes a good research design. Pavan alludes to these issues by pointing to next steps involved in measuring a community both on and offline. I heartily agree. In fact, *a fortiori*, studies merely demonstrating that collective action happens on any given medium run the risk of being data journalism rather than social science. If that particular medium shuts down, alters its algorithms or users flock to a new service, social scientists must be left with theoretical tools that do more than merely legitimize popular technologies.

2. Speakers are more important than audiences and non-speakers: This is a strategy that masks the fact that for most online analysis it is terribly difficult to get a count of exposure. Not everyone who reads a tweet is going to retweet it. Not everyone who has a person in their tweet stream is going to be attending Twitter and noticing said tweet. And such variation is certainly not homogenous with different communities being more or less attentive to Twitter. In the absence of coherent metrics for audience reach tweets we focus on speakers instead of listeners. For an awareness campaign this is a particularly onerous constraint.

This constraint is further reinforced by shifts in the twitter API towards streaming volumes of content and forward facing tweets. The search API only gives data going back six days but will give upwards of 18,000 tweets within this window. As such, Twitter is steering academics towards analyses that are speaker-focused, not follower-focused.

Pavan's analysis follows this general trend of focusing primarily on the speakers. It is a reasonable approach to describe the events in a specific community, but it is insufficient to get a cohesive analysis of an event as an unfolding series of symbolic acts. Work on protests offline not only looks at the posters and speeches of the leaders, but the scale of those low-commitment people that constitute the masses also involved. Are these masses invisible now? What if they have nothing to say apart from what the organizational leaders say – is there now an imperative to retweet in order to be seen?

3. We have more data than we need: This quantity over quality strategy makes an appeal to the wealth of online data is a means for compensating for something – to be frank, in most cases, grants. This is one issue where social scientists could potentially shoot themselves in the foot. As more governments across the world shift from costly long-form censuses to trace data, sociologists could end up being complicit in the extinction of some of the best "gold standard" population-level estimates available while simultaneously complicit in the gross privacy intrusions required to compensate for such extinctions. No single study is necessarily responsible for this shift, but a discourse that suggests corporate sites like Twitter and Facebook enable easy and cheap access to data both suggest that our research can be cheap and that it is okay to be beholden to incorporated (often American) third parties for our analyses.

While we have a great deal of data on the number of tweets for the Occupy protest, there is no public data set, to this author's knowledge, of the number of people in each of the most popular occupy camps, their duration and a chronicle of events related therein, such as the lending of generators, toilets, tents, etc. Turning on the Twitter tap is easy, but going to sites that are ostensibly facilitated by these events is not. This leads to a tennis match of ideologies – was the Arab Spring due to Facebook or not? We can get thorough quantitative evidence from public facing Facebook pages, but only anecdotal evidence of what happened in Tahrir Square, such as photographs and journalist accounts. Colleagues from Cairo suggest that Facebook provided significant public visibility to private grievances. Colleagues in the U.S. remind me that many social movements occurred without such media. Without knowing who talked to whom in the "ur-graph" of communication (*i.e.* all communication between individuals regardless of media), scholars can endlessly debate the merits of one specific medium without resolution.

Pavan is right to point out first, the many possibilities for novel social science research in Twitter and the fact that social media such as Twitter are an integral part in many contemporary forms of political action. In this response I have spoken to the fact that social media platforms are not merely media, but structuring devices in their own right. Twitter is not merely the conduit for discussion, but an actor in these discussions. Twitter's actions, such as tweaking the API, sorting discussions, filtering communication, and deciding what is trending have an effect on the agency of political actors and the agency of researchers intending to study such actions. Approaches to this work must embed this sort of reflexivity in its approach. While the data-driven connections may be self-evident (being a follower or not), the meaning behind these data still require interpretive care, triangulation and attention to the structure that gives rise to such communication in the first place.

References

Alexa

2013 *Daily Traffic Rankings*. Available online: http://www.alexa.com/topsites/countries Date Accessed: August 10, 2013.

Bernard, H.R., Killworth, P.D., Sailer, L.

1979 "Informant Accuracy in Social Network Data IV: A Comparison of Clique-Level Structure in Behavioral and Cognitive Network Data." *Social Networks 2* (3): 191-218.

Boyd, D.

2006 "Friends, Friendsters and Top 8: Writing Community into Being on Social Network Sites." *First Monday 11* (12).

Boyd, D., Ellison, N.

2007 "Social Network Sites: Definition, History, and Scholarship." *Journal of Computer Mediated Communication 13* (1): 210-230.

Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.P.

2010 "Measuring User Influence in Twitter: The Million Follower Fallacy." *4th International AAAI Conference on Weblogs and Social Media (ICWSM).*

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., Flammini, A.

2011 "Political Polarization on Twitter." *5th International AAAI Conference on Weblogs and Social Media (ICWSM).*

Freeman, L.C., Romney, A.K., Freeman, S.C.

1987 "Cognitive Structure and Informant Accuracy." *American Anthropologist 89* (2): 310-325.

Hennig, M., Brandes, U., Pfeiffer, J., Mergel, I.

2013 *Studying Social Networks*. Berlin: Springer-Verlag.

Innis, H.A.

2007 *Empire and Communications*. Lanham: Rowman & Littlefield. [First Edition 1950]

Marin, A.

2004 "Are Respondents More Likely to List Alters with Certain Characteristics? Implications for Name Generator Data." *Social Networks 26* (4): 289-307.

Newman, M.E.J.

2010 *Networks: An Introduction*. Oxford, UK: Oxford University Press.

Pariser, E.

2011 *The Filter Bubble: What the Internet Is Hiding from You*. New York, NY: The Penguin Press HC.

Comment on Elena Pavan/1

Considering Platforms as Actors

Abstract: Twitter tends to represent an unfounded resurgence of methodological realism, yet the medium is not merely a conduit for messages that reflect the true sentiment of all actors involved. Rather it is a structuring device for communication based on the idea of a "network" *a priori*. It is also a business. In this article, I discuss the technical and practical issues that simultaneously make Twitter feel like a profound window to the social world while masking many absences and inequalities. I critique the absence of lurkers, the focus on streaming data and the emphasis on personalization. Greater care must be made to triangulate Twitter data with traditional social science theories and methods.

Keywords: Social media, social networks, constructivism, API, personalisation.

Bernie Hogan is a Research Fellow at the University of Oxford's Internet Institute. His theoretical focus is on online identity expression and how individuals understand their audiences as an amalgam of publics and personal networks. His methodological work focuses on network visualisation and interactivity. He received his PhD in Toronto in 2009.