

**Social networks that matter:
Twitter under the microscope**

**by Bernardo A. Huberman,
Daniel M. Romero, and Fang Wu**

Abstract

Scholars, advertisers and political activists see massive online social networks as a representation of social interactions that can be used to study the propagation of ideas, social bond dynamics and viral marketing, among others. But the linked structures of social networks do not reveal actual interactions among people. Scarcity of attention and the daily rhythms of life and work makes people default to interacting with those few that matter and that reciprocate their attention. A study of social interactions within Twitter reveals that the driver of usage is a sparse and hidden network of connections underlying the “declared” set of friends and followers.

Contents

[Introduction](#)
[Discussion](#)
[Conclusion](#)

Introduction

Social networks, a very old and pervasive mechanism for mediating distal interactions among people, have become prevalent in the age of the Web. With interfaces that allow people to follow the lives of friends, acquaintances and families, the number of people on social networks has grown exponentially since the turn of this century. Facebook, LinkedIn and MySpace, to give a few examples, contain millions of members who use these networks for keeping track of each other, find experts and engage in commercial transactions when needed (Huberman, et

al., 2008). Furthermore, commercial enterprises try to exploit them for marketing purposes, as they provide a ready made medium for propagating recommendations through people with similar interests (Korgan, et al., 2001).

On the academic side, a large body of knowledge has accumulated on the formation and dynamics of these networks, fueled by the easy availability of data and the regularities found in the statistical distribution of nodes and links within these networks (Golder, et al., 2007; Granovetter, 1973; Kleinberg, 2008; Leskovec, et al., 2007; Wasserman and Faust, 1994).

While the standard definition of a social network embodies the notion of all the people with whom one shares a social relationship, in reality people interact with very few of those “listed” as part of their network. One important reason behind this fact is that attention is the scarce resource in the age of the Web. Users faced with many daily tasks and large number of social links default to interacting with those few that matter and that reciprocate their attention. For example, a recent study of Facebook showed that users only poke and message a small number of people while they have a large number of declared friends (Feld, 1991). And a casual search through recent calls made through any mobile phone usually reveals that a small percentage of the contacts stored in the phone are frequently contacted by the user.

These initial observations suggest a systematic investigation into the nature of the social networks that actually matter to people. By networks that matter we mean those networks that are made out of the pattern of interactions that people have with their friends or acquaintances, rather than constructed from a list of all the contacts they may decide to declare.



Discussion

In order to find out how relevant a list of “friends” is to members of the network, we collected and analyzed a large data set from the Twitter social network. Twitter.com (<http://twitter.com/>) is an online social network used by millions of people around the world to stay connected to their friends, family members and co-workers through their computers and mobile phones. The interface allows users to post short messages (up to 140 characters) that can be read by any other Twitter user. Users declare the people they are interested in following, in which case they get notified when that person has posted a new message. A user who is being followed by another user does not necessarily have to reciprocate by following them back, which makes the links of the Twitter social network directed.

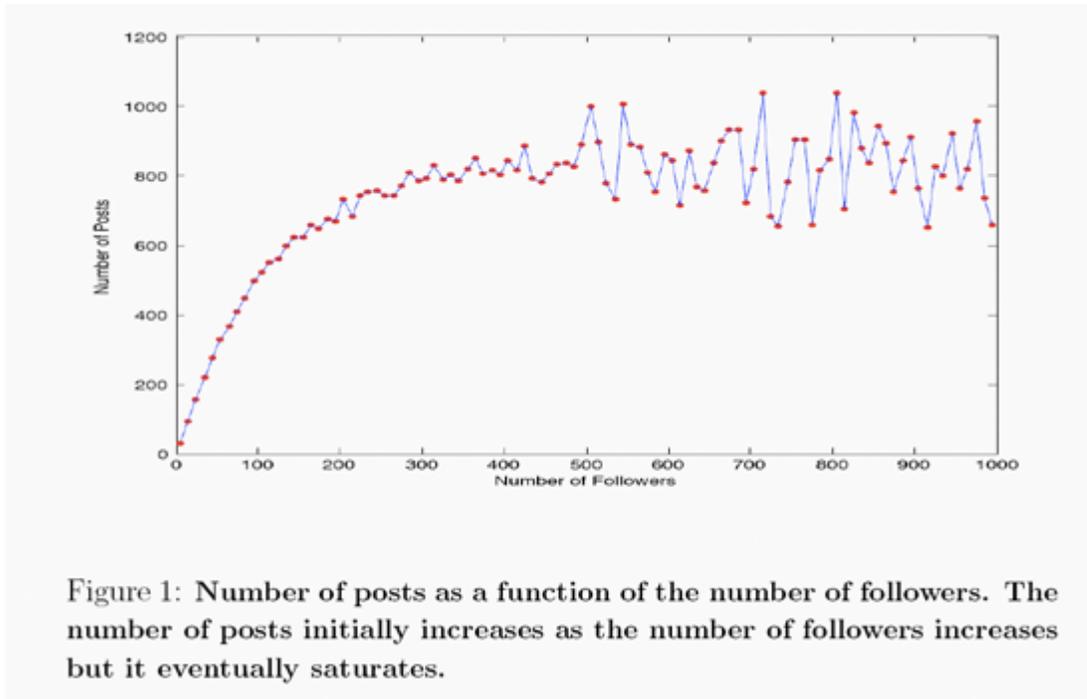
For each user of Twitter in our data set we obtained the number of followers and followees (people followed by a user) the user has declared, along with the content and date stamp of all his posts [1]. Our data set consisted of a total of 309,740 users, who on average posted 255 posts, had 85 followers, and followed 80 other users. Among the 309,740 users only 211,024 posted at least twice. We call them the active users. We also define the active time of an active user by the time that has elapsed between his first and last post. On average, active users were active for 206 days.

Twitter users are able to publicly post direct and indirect updates. Direct public posts are used when a user aims her update to a specific person and are signaled by an “@” symbol next to

the person's username, whereas indirect updates are used when the update is meant for anyone that cares to read it. Even though direct updates are used to communicate directly with a specific person, they are public and anyone can see them. Often times two or more users will have conversations by posting updates directed to each other. Around 25.4 percent of all posts are directed, which shows that this feature is widely used among Twitter users.

We are interested in finding out how many people each user communicates directly with through Twitter. We define a user's friend as a person whom the user has directed at least two posts to. Using this definition we were able to find out how many friends each user has and compare this number with the number of followers and followees they declared.

Based on our previous finding about the role of attention in eliciting productivity within a social network (Huberman, et al., 2008), we conjecture that the users who receive attention from many people will post more often than users who receive little attention. Therefore we expect that users with more followers and friends will be more active at posting than those with a small number of followers and friends. Figures 1 and 2 show that indeed the total number of posts increases with both the number of followers and friends. However, as Figure 1 shows, the number of total posts eventually saturates as a function of the number of followers. This implies that users with a large number of followers are not necessarily those with very large number of total posts. On the other hand, the number of total posts does not saturate as a function of number of friends, as seen on Figure 2. Rather, the number of updates increases until it reaches a maximum point of 3,201. This suggests that in order to predict how active a Twitter user is, the number of friends is a more accurate signal than the number of his followers.



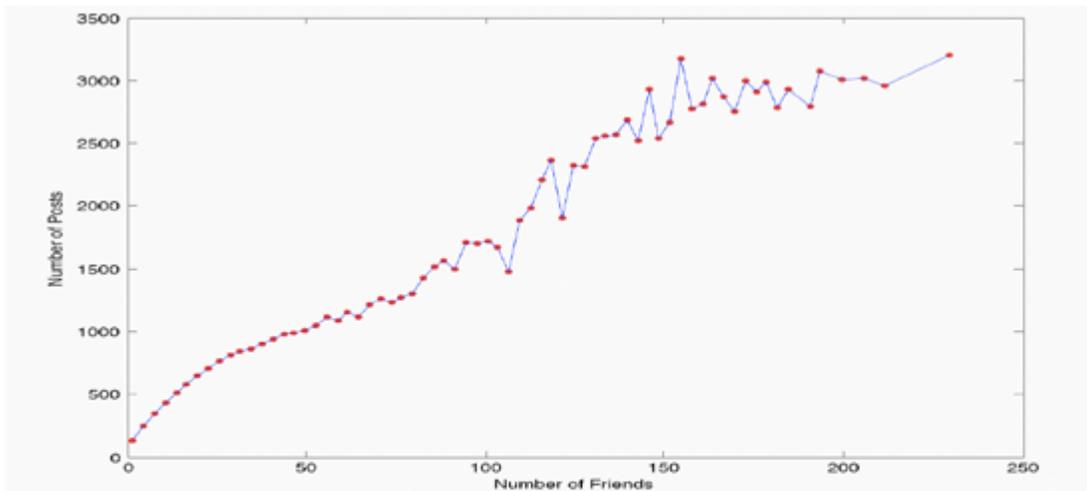


Figure 2: Number of posts as a function of the number of friends. The number of posts increases as the number of friends increases, reaching 3200 without saturating.

This implies that to assess the size of the social network that matters we need to consider those people who actually communicate through direct messages with each other, as opposed to the network created by the declared followers and followees.

Having shown that the number of friends is the actual driver of Twitter user's activity, we compared it with the number of followees the users declare. We define ρ as the number of friends a user has, divided by the number of followees she declared. Since 98.8 percent of the users have fewer friends than followees, almost all the ρ values are less than 1. [Figure 3](#) shows a histogram of the ρ values. As we can see most users have a ρ value less than .1, with the number of users with a ρ close to 1 extremely small. The average of the ρ values is 0.13 and the median is 0.04. This indicates that the number of friends users have is very small compared to the number of people they actually follow. Thus, even though users declare that they follow many people using Twitter, they only keep in touch with a small number of them. Hence, while the social network created by the declared followers and followees appears to be very dense, in reality the more influential network of friends suggests that the social network is sparse.

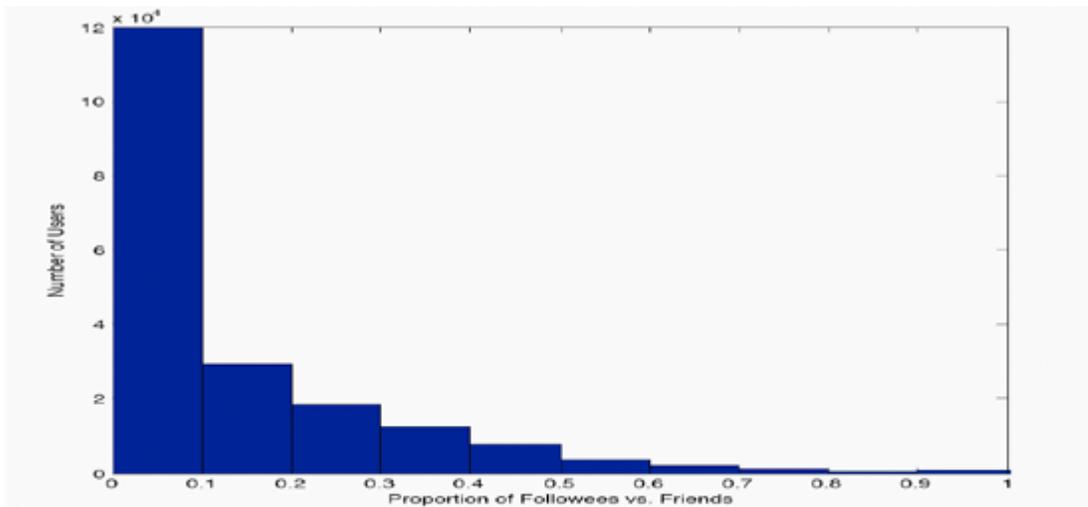


Figure 3: Histogram of contributor's number of friends divided by the number of followees. Most users have a very small number of friends compared to the number of followees they declared.

Another interesting aspect is to consider how the number of friends and the \bullet values change as the number of followees increases. Figures 4 and 5 show that even though the number of friends initially increases as the number of followees increases, after a while the number of friends saturates. This trend can be explained by the fact that the cost of declaring a new followee is very low compared to the cost of maintaining friends (i.e., exchanging directed messages with other users). Hence, the number of people a user actually communicates with eventually stops increasing while the number of followees can continue to grow indefinitely.

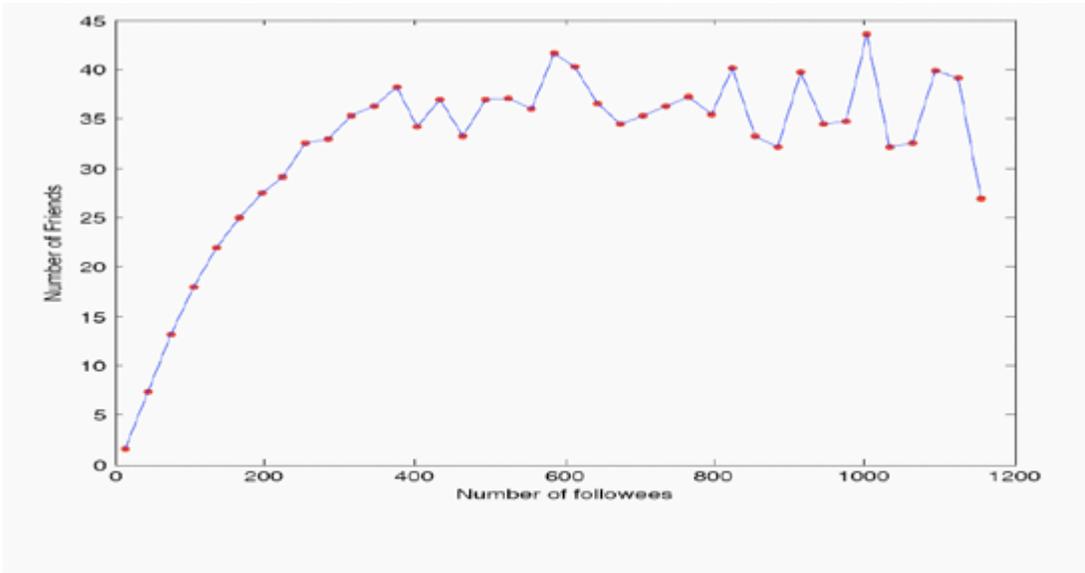


Figure 4: Number of friends as a function of the number of followers. The total number of friends saturates while the number of followers keeps growing due to the minimal effort required to add a follower.

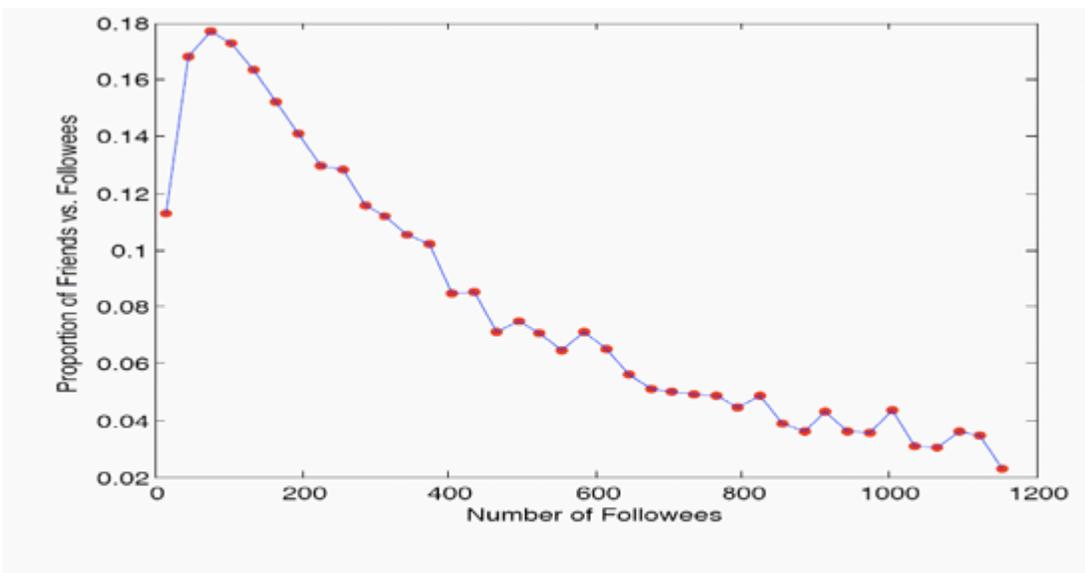


Figure 5: Proportion of friends vs. followers as a function of followers. It initially increases but rapidly approaches zero as the number of followers increases.

Reciprocity plays an important role in many economic and social interactions (Fehr and Gächter, 2000). At the same time, the plenitude of signals that people are flooded with makes attention a scarce commodity and thus a valued private good (Huberman, et al., 2008). In the case of Twitter, we found that the notion of reciprocated attention is present. While our definition of friend allows for a user X to be a friend of user Y while Y is not a friend of X, we found that on average, 90 percent of a user's friends reciprocate attention by being friends of the user as well. This shows that reciprocity of attention plays an important role in defining the "hidden network." [Figure 6](#) shows that reciprocity of attention is a very consistent trend as it holds for both users with many friends as well as for users with very few friends.

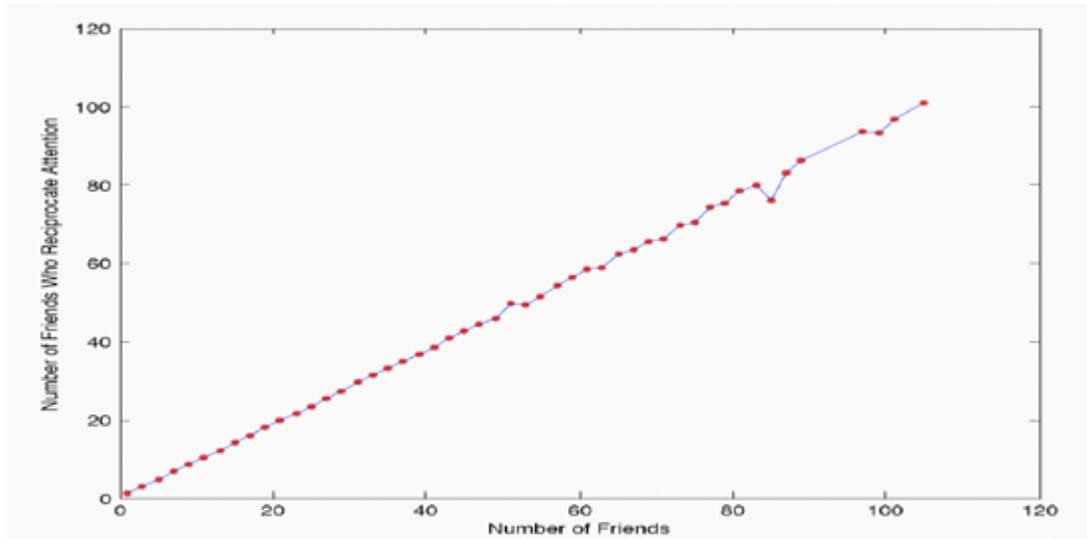


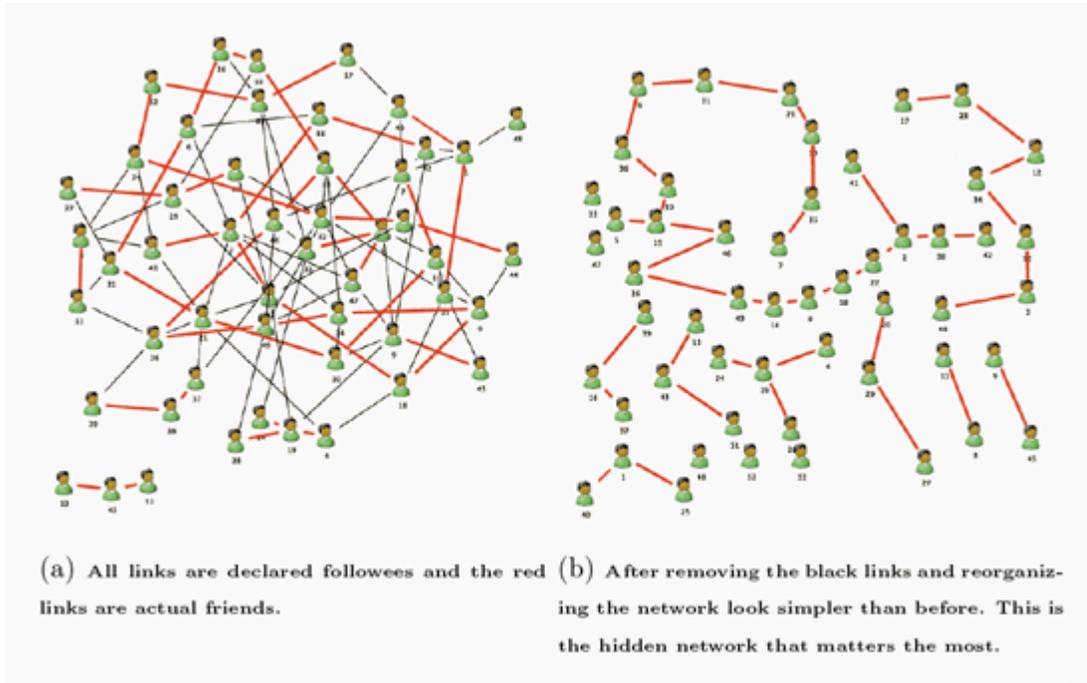
Figure 6: Number of users who reciprocate attention as a function of the total number of friends.



Conclusion

In conclusion, even when using a very weak definition of "friend" (i.e., anyone who a user has directed a post to at least twice) we find that Twitter users have a very small number of friends compared to the number of followers and followees they declare. This implies the existence of two different networks: a very dense one made up of followers and followees, and a sparser and simpler network of actual friends. The latter proves to be a more influential

network in driving Twitter usage since users with many actual friends tend to post more updates than users with few actual friends. On the other hand, users with many followers or followees post updates more infrequently than those with few followers or followees.



Many people, including scholars, advertisers and political activists, see online social networks as an opportunity to study the propagation of ideas, the formation of social bonds and viral marketing, among others. This view should be tempered by our findings that a link between any two people does not necessarily imply an interaction between them. As we showed in the case of Twitter, most of the links declared within Twitter were meaningless from an interaction point of view. Thus the need to find the hidden social network; the one that matters when trying to rely on word of mouth to spread an idea, a belief, or a trend. 

About the authors

Bernardo A. Huberman is a Senior HP Fellow and Director of the Social Computing Lab at Hewlett-Packard Laboratories, Palo Alto, Calif.

Web: <http://www.hpl.hp.com/research/idl/people/huberman/>

E-mail: bernardo [dot] huberman [at] hp [dot] com

Daniel M. Romero is a graduate student at the Center for Applied Mathematics of Cornell University (Ithaca, N.Y.) and also a researcher in the Social Computing Lab of HP Laboratories.

E-mail: dmr239 [at] cornell [dot] edu

Fang Wu is a researcher in the Social Computing Lab of HP Laboratories.
E-mail: fang [dot] wu [at] hp [dot] com

Acknowledgment

One of us (BAH) thanks Dr. Josef Falkinger for useful discussions.

Note

1. Twitter only displays up to 3,201 updates per user so we only have the complete set of updates for users who have posted 3,200 or less updates. A very small set of users showed 3,201 updates so we have the complete set for about 99.6 percent of all the users.

References

E.Fehr and S. Gächter, 2000. "Fairness and retaliation: The economics of reciprocity," *Journal of Economic Perspectives*, volume 14, number 3, pp. 159–181.

S.L. Feld, 1991. "Why your friends have more friends than you do," *American Journal of Sociology*, volume 96, number 6, pp. 1,464–1,477.

S.A. Golder, D. Wilkinson and B.A. Huberman, 2007. "Rhythms of social interaction: Messaging within a massive online network," *Third International Conference on Communities and Technologies*, at <http://www.hpl.hp.com/research/idl/papers/facebook/facebook.pdf>, accessed 21 December 2008.

M. Granovetter, 1973. "The strength of weak ties," *American Journal of Sociology*, volume 78, number 6, pp. 1,360–1,380.

R.E. Grinter and L. Palen, 2002. "Instant messaging in teen life," *Proceedings of the ACM Conference on Computer-Supported Work*, pp. 21–30; version at <http://www.cs.colorado.edu/~palen/Papers/grinter-palen-IM.pdf>, accessed 21 December 2008.

B.A. Huberman, D.M. Romero and F. Wu, 2008. "Crowdsourcing, attention and productivity," version of paper submitted for the 2009 World Wide Web Conference (Madrid); version at <http://arxiv.org/abs/0809.3030>, accessed 21 December 2008.

J. Kleinberg, 2008. "The convergence of social and technological networks," *Communications of the ACM*, volume 51, number 11, pp. 66–72; version at

<http://www.cs.cornell.edu/home/kleinber/cacm08.pdf>, accessed 21 December 2008.

K. Korgan, P. Odell and P. Schumacher, 2001. "Internet use among college students: Are there differences by race/ethnicity?" *Electronic Journal of Sociology*, volume 5, number 3, at <http://www.sociology.org/content/vol005.003/korgen.html>, accessed 21 December 2008.

J. Leskovec, L.A. Adamic and B.A. Huberman, 2007. "The dynamics of viral marketing," *ACM Transactions on the Web*, volume 1, number 1, article number 5; version at <http://www-personal.umich.edu/~ladamic/papers/viral/viralTWeb.pdf>, accessed 21 December 2008.

S. Wasserman and K. Faust, 1994. *Social network analysis: Methods and applications*. New York: Cambridge University Press.

B. Wellman and N. Hampton, 1999. "Living networked in a wired world," *Contemporary Sociology*, volume 28, number 6, pp. 648–654.

Editorial history

Paper received 9 December 2008; accepted 20 December 2008; revised version received 1 January 2009.

Copyright © 2009, First Monday.

Copyright © 2009, Bernardo A. Huberman, Daniel M. Romero, and Fang Wu.

Social networks that matter: Twitter under the microscope
by Bernardo A. Huberman, Daniel M. Romero, and Fang Wu
First Monday, Volume 14, Number 1 - 5 January 2009
<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/rt/printerFriendly/2317/2063>