

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

Bernardo A. Huberman¹, Daniel M. Romero^{1,2} and Fang Wu¹

¹Social Computing Lab, HP Laboratories, Palo Alto, CA 94304

²Cornell University, Ithaca, NY 14850

December 5, 2008

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.

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 [6]. 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 [8].

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 [1, 3, 4, 7, 9, 10].

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 [2]. 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.

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** is a online social network used by millions of people

around the world to stay connected to their friends, family members and coworkers 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 datestamp 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% 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.

¹**Twitter** only displays up to 3201 updates per user so we only have the complete set of updates for users who have posted 3200 or less updates. A very small set of users showed 3201 updates so we have the complete set for about 99.6% of all the users.

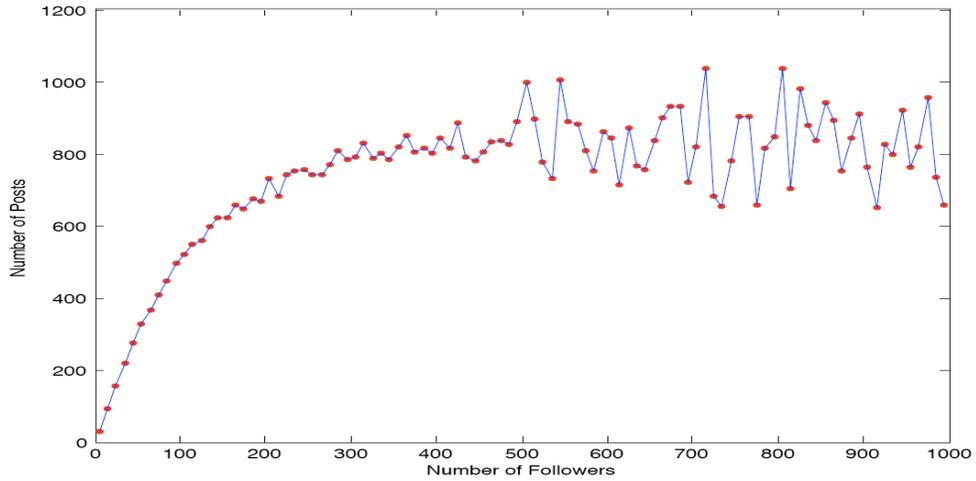


Figure 1: Number of posts as a function of the number of followers. The number of posts initially increases as the number of followers increases but it eventually saturates.

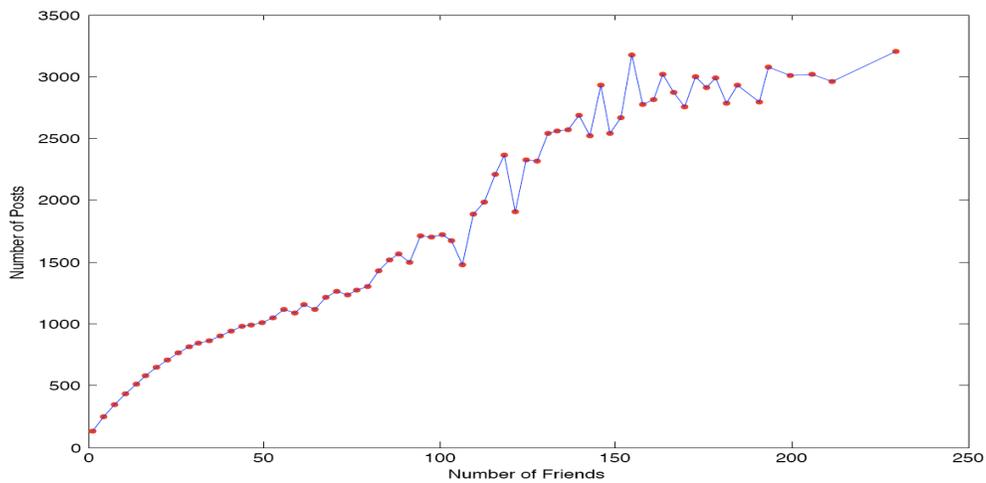


Figure 2: Number of posts as a function of the number of friends. The number of posts increases as the number of friends increases, reaching 3200 without saturating.

Based on our previous finding about the role of attention in eliciting productivity within a social network [5], 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 3201. 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.

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 δ as the number of friends a user has, divided by the number of followees she declared. Since 98.8% of the users have fewer friends than followees, almost all the δ values are less than 1. Figure 3 shows a histogram of the δ values. As we can see most users have a δ value less than .1, with the number of users with a δ close to 1 extremely small. The average of the δ 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.

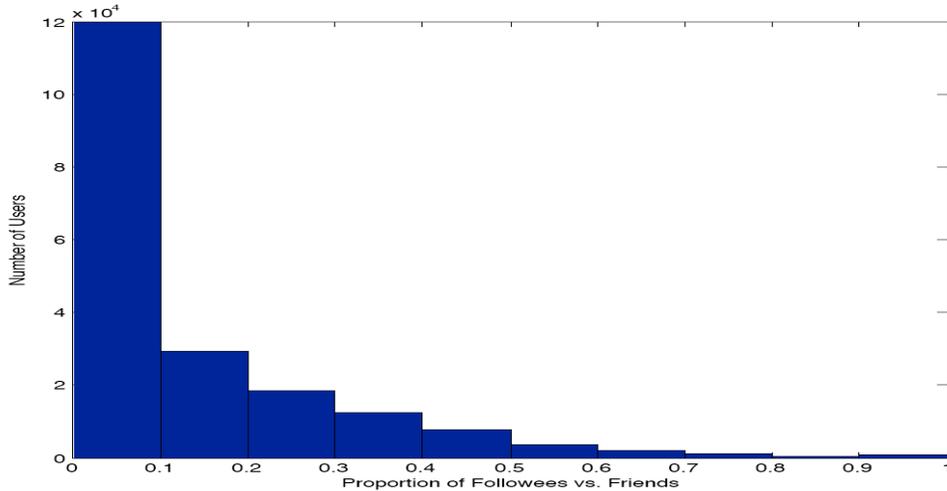


Figure 3: **Histogram of contributor’s number of friends divided by the number of followees. Most users have a very small number of friends compared to the number of followees they declared.**

Another interesting aspect is to consider how the number of friends and the δ 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 starts to saturate and stays nearly constant. 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 a 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.

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 influ-

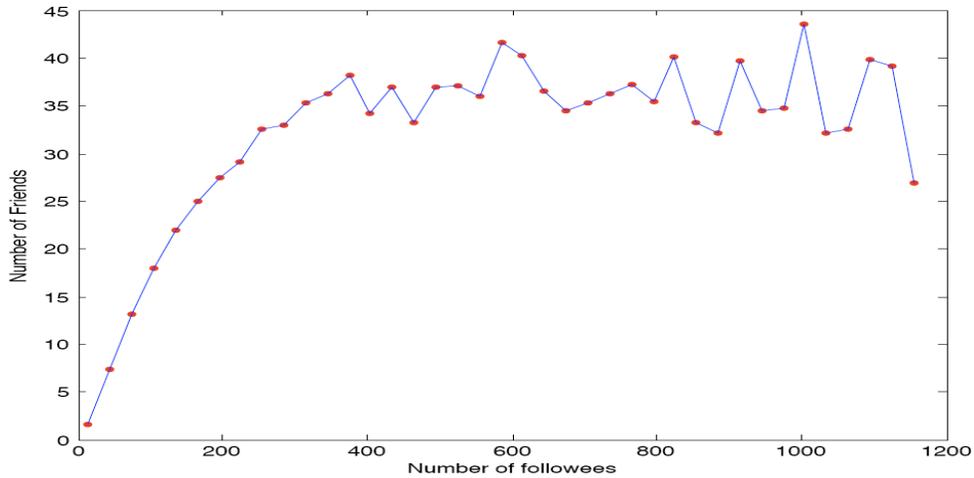


Figure 4: Number of friends as a function of the number of followees. The total number of friends saturates while the number of followees keeps growing due to the minimal effort required to add a followee.

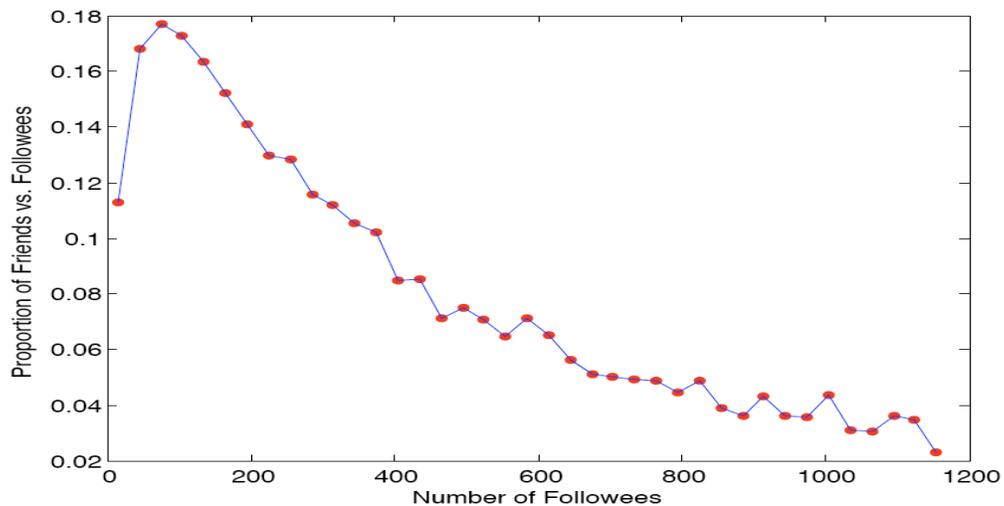
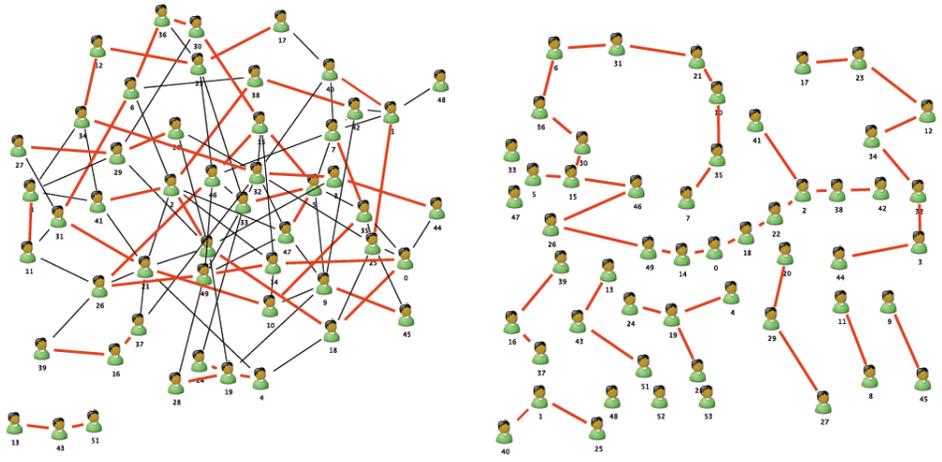


Figure 5: Proportion of friends vs. followees as a function of followers. It initially increases but rapidly approaches zero as the number of followees increases.



(a) All links are declared followees and the red links are actual friends.
 (b) After removing the black links and reorganizing the network look simpler than before. This is the hidden network that matters the most.

ential 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.

References

- [1] S. Feld. Why your friends have more friends than you do. *American Journal of Sociology*, 96(6), pp. 146–1477, 1991.
- [2] S. A. Golder, D. Wilkinson and B. A. Huberman. Rhythms of Social Interaction: Messaging within a Massive Online Network. *3rd International Conference on Communities and Technologies*, 2007.
- [3] M. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6), pp. 136–1380, 1973.
- [4] R. Grinter and L. Palen. Im in teen life. *Proceedings of the ACM Conference on Computer-Supported Work*, 2002
- [5] B. A. Huberman, D. M. Romero and F. Wu. Crowdsourcing, Attention and Productivity. *2009 World Wide Web Conference*, 2008 (Submitted)
- [6] J. Kleinberg. The convergence of social and technological networks. *Communications of the ACM*, 51(11), pp. 66–72, 2008.
- [7] K. Korgan, P. Odell and P. Schumacher. Internet use among college students: Are there differences by race/ethnicity? *Electronic Journal of Sociology*, 5(3), 2001.
- [8] J. Leskovec , L. A. Adamic and B. A. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1), 2007.
- [9] S. Wasserman and K. Faust. Social Network Analysis: Methods and Applications. *Cambridge University Press*, 1993
- [10] B. Wellman and N. Hampton. Living networked in a wired world. *Contemporary Sociology*, 28(6), pp. 64–654, 1999.