

A Structural Approach to Contact Recommendations in Online Social Networks

Scott A. Golder
Cornell University
Ithaca, NY 14850
sag262@cornell.edu

Alice Marwick
New York University
New York, NY 10003
alice.marwick@nyu.edu

Sarita Yardi
Georgia Institute of
Technology
Atlanta, GA 30332
syardi3@gatech.edu

danah boyd
Microsoft Research New
England
Cambridge, MA 02142
dmb@microsoft.com

ABSTRACT

People generally form network ties with those similar to them. However, it is not always easy for users of social media sites to find people to connect with, decreasing the utility of the network for its users. This position paper looks at different ways we can make friend recommendations, or suggest users to follow on the microblogging site Twitter. We examine a number of ways in which similarity can be defined, and the implications of these differences for community-building. People may want different things for friend suggestions, which has implications for both future research and product design.

Categories and Subject Descriptors

J.4 [Social and Behavioral Sciences]: Sociology

General Terms

Human Factors. Measurement

Keywords

Twitter, Social Networks, Homophily

1. INTRODUCTION

The structure of our personal social networks are driven by homophily. The principle of homophily is the idea that “contact between similar people occurs at a higher rate than among dissimilar people” [13]. This principle affects the formation of every kind of network tie, from friendship and emotional support to work and information transfer [13]. Shared demographic characteristics and personal interests have strong effects on which others people choose to form ties with. Moreover, the people to whom we are exposed

as potential new ties is heavily dependent on our existing set of ties – e.g. friends of friends – and on the organizations of which we are part, such as schools, workplaces and community groups [8].

At the same time, since homophily lessens exposure to people who are different in terms of their attitudes, experiences and information, people often actively seek new sources of information, and so they must venture outside of their local networks and create social ties that bridge relatively lengthier social distances to individuals who are different from them, and thus have access to different information [4][10].

Depending on the social and informational goals individuals have, they might choose to form different kinds of ties at different times. People do this quite naturally, and partition their many contacts into different facets of their social life. However, as has been shown with online social network services, all of one’s contacts are lumped together into a single contact list. When examining this collection of individuals and ties as a single network, the reasons for the formation and maintenance of each tie are not clear. Likewise, recommending new people with whom to form ties is challenging because of the decontextualized nature of the ties themselves.

While the theoretical topic of this paper is the structural basis of tie recommendations, the substantive object of study is the microblogging service Twitter. Begun in 2006, Twitter supports users posting very short messages (“tweets”) – up to 140 characters in length. These messages are then visible to those users have chosen to receive that particular user’s messages (“followers”)¹.

Twitter is different from other online social networking services in one significant way, which is that ties are directed, rather than undirected. To illustrate, consider friendship ties in LinkedIn, Facebook or MySpace. In services such as these, when two users share a friendship tie, the tie is bidi-

¹One may choose to make one’s account public or private, in order to make one’s tweets available to everyone, or to only one’s followers

rectional; if A is friends with B, this implies B is friends with A. This is not so in Twitter. For example, celebrities or even people who are not famous but simply popular, about whom many people are interested, can be “followed” by thousands or even millions of people, but they themselves can pay attention to as many or as few individuals they would like, including nobody at all. Twitter’s links bear more resemblance to weblog links or to hyperlinks on the web in general. Just as a webpage may link from itself to some second page and this does not imply any endorsement by that second page, so too can Twitter users follow one another without any implications about the preferences of the user being followed.

Having directed ties presents significant analytical benefits. Though social ties in systems such as Facebook and MySpace do not inherently contain information about the power relations between two users, directed ties necessarily do; if some user A has a directed tie to B (i.e. pays attention to B) and B does not have such a tie to A, then we might say B has a power advantage over A, since B is more important to A than A is to B [6]. We may also distinguish A’s ties into two sets of people – those who A pays attention to, and those who pay attention to A; of course some individuals may fit into both sets. By extension, the former group may be those who satisfy an information acquisition goal for A, and the latter group may satisfy an information dissemination goal.

At large scales, it may be fruitful to study the directed graph of Twitter in the same manner as the World Wide Web, using spectral methods, such as eigenvector centrality to discover nodes that are “authoritative” or of high status [12][2]. However, in this paper we focus on suggestions that are derived strictly from the local structure of nodes and their relations to one another.

2. DISCOVERABILITY

There is a problem of discoverability in Twitter; there is no easy and intuitive way to find other Twitter users to follow. When users join, they are isolates – disconnected nodes – in the social network of users that is constructed by follower relations. The first thing users might do, then, is seek out people to follow. However, people have a variety of motivations for joining Twitter. If they like sports, they could follow @THE_REAL_SHAQ. If they like stand-up comedy, they might follow @jimmycarr, and if they just wanted a good news source, they might recommend @BreakingNews or @BBC. They also might have friends who are already using Twitter, and therefore seek to establish bidirectional followership ties with those friends. Others still might seek to amass large audiences themselves, where they can share their ideas and build a following [3].

The challenge is that after following a few known users, there is no mechanism for locating and following a larger network of users that may be of interest. Armed with a small but growing network, how would they go about finding more Twitter users that they are likely to find interesting?

Other social media sites have adopted a variety of approaches to overcome this challenge: Facebook encourages existing users to recommend friends to new users, Amazon recommends books based on other people’s past purchasing pat-

terns, and LinkedIn shows 2nd degree connections. Offline, people rely on social and geographic proximity markers to form and grow ties. Shared local geographies, interests like book clubs, music, or movies, and institutions like churches, schools, and sports groups can help people to find groups of like-minded others. Relationships are formed and networks grow through these existing ties. But do these ties relate to tie formation on Twitter? Do people want to follow the same people who follow them? Do people want to follow people who share the same interests as them? Does homophily guide Twitter network formation the way it guides much social interaction offline [13]?

In this position paper, we examine ways of measuring and detecting kinds of network ties that Twitter users might want (but haven’t yet found themselves). We describe four methods for identifying who users might want to follow. We conclude by questioning some of the assumptions built into recommender systems and whether people want to build diversity or density in their networks.

2.1 Existing Mechanisms for Finding Users

The “Find Friends” tab at the top of a user’s Twitter page provides mechanisms to find pre-existing friends. A user can compare their email contacts to existing Twitter users, search by username, or invite a friend to use Twitter. These mechanisms assume that users will already know whoever they might want to follow. In contrast, the “suggestions” tab shows a pre-defined list of popular Twitter users like Demi Moore, *People Magazine*, and Tim O’Reilly, regardless of the individual user’s existing ties.

Third-party developers have built a number of applications using the Twitter API to compensate for Twitter’s lack of customized friend recommendations. MrTweet.net scans a user’s social network and suggests potentially interesting people based on common friends; the site claims to find “great folks relevant to your objectives”. WhoShouldIFollow.com allows users to input their username and see a list of users similar to themselves and other Twitter users they follow, which can be filtered by location (anywhere or near a certain place) and popularity. Twubble.com looks at who a Twitter user’s friends follow, and ranks potential contacts by number of mutual followers.

This paper examines critically what users might want when looking for friends and followers on Twitter. This is complicated by the open-ended nature of Twitter. The site can be used to communicate with existing, offline friends, to talk with celebrities, to self-promote, for work or marketing, and to find useful and timely information; most users engage in a combination of these activities. Thus, it is unlikely that a single friend similarity metric will work for all users. We discuss four ways in which this metric might be conceptualized.

3. RECOMMENDATION PRINCIPLES

In this section, we consider four structural properties of dyads that might be associated with the members of those dyads having higher propensity to establish follower relationships. With the exception of the first one, these measures are based on different kinds of similarity between users. As described earlier, spectral methods that rely on information

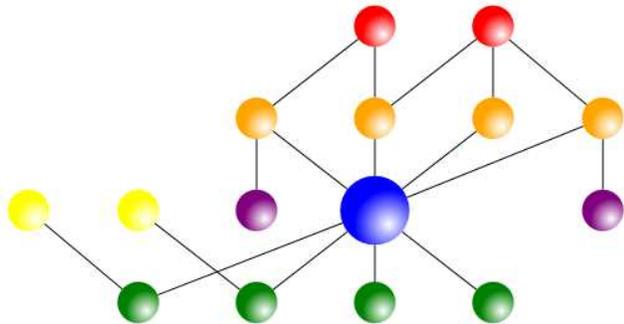


Figure 1: An example graph showing a target user (“ego”) and their local network.

about the structure of the entire social network might identify users who are important in a global sense, but these recommendations might not be useful to particular individuals, whose information needs might be different from what such important nodes can provide. Plus, such methods require information about the entire network and are computationally intensive.

Instead, we focus on people who are in the *second-degree network* for a given target user (“ego”). Ego’s second degree network consists of all of the other people with whom ego shares a tie (first-degree ties, or alters), and all the people with whom those alters share ties. This second-degree network likely consists of people who ego shares a social tie to outside of the online social network, with whom a tie has not yet been established online. We can take advantage of existing homophily among ties and help establish this online tie.

There is good reason to believe that, even among users’ second-degree nodes, there are undiscovered yet desirable ties. Granovetter points out that, despite living in a modern world in which computer systems yield global data about social ties, individuals typically know very little about the ties that exist even in their local network [11]. Further, the key insight of Milgram’s original small world study [16] is that individuals are pleasantly surprised when they learn two of their ties they believe to be strangers turn out to know one another, hence the phrase, “it’s a small world.”

3.1 Reciprocity

Reciprocity is a source of social cohesion [9]. If two individuals mutually attend to one another, then the bond is reinforced in each direction and both people will find the tie rewarding [7]. Reciprocal exchange – exchange relationships in which individuals give something of value, in this case attention, to one another in turns – also leads to stronger affective ties [14]. For this reason, we include reciprocity as a method for recommendation. Furthermore, if A is interested in B, then whatever basis for that interest might also serve to motivate B’s interest in A.

However, we recognize that the cost of feigning attention to another Twitterer by following them but not reading their posts is very low, and so it would be unreasonable to pay back each follower with honest attention. Indeed, it is speculated that spam accounts create follower links to legitimate users (i.e. feigning attention) in order to motivate a reciprocal link and to boost their publicly visible follower count and appear to have a larger audience than they otherwise might.

3.2 Shared Interests

Social ties are often established around shared interests, organizations, activities and so on [8]. If two people share an interest, in this case, a person to whom they both pay attention, then those two people may be interested in one another.

Applying this idea to Twitter, if we wanted to find people who share interests with ego, then we might do well to look at the set of users who follow the same users ego does. In Figure 1, ego (blue) follows several other users (orange), and two users (purple) also follow those orange users. Since ego and the purple nodes have a shared interest in the form of the orange nodes, ego and the purple nodes might be interested in following one another.

3.3 Shared Audience

An analogous argument can be made for pairs of users who share audience members in common. In Figure 1, ego has an audience consisting of four (green) users. Two of these green users also pay attention to two yellow users. Since the green nodes find both yellow nodes and ego worthy of paying attention to, it is possible that yellow nodes and ego have something in common such that they might like to follow one another.

Of course it is possible that the basis for the green to yellow tie is different from the basis for the green to ego tie. However, if many such green users follow both yellow nodes and ego, then the likelihood that the basis for following is similar, or else the topics that support the green-yellow and the green-ego ties are similar. This opens up the possibility of treating the Twitter network as a bipartite or two-mode graph of users and topics; see [17].

3.4 Filtered People

Twitter users perform a valuable curatorial service. They follow other users and filter their messages, passing on the ones they see value in, perhaps as retweets [3]. In Figure 1, ego follows orange users, and some of them follow red

users. The only path through which ego might encounter red nodes' tweets is through the filter of orange users. However, if many orange users follow a particular red user, then ego might seek to eliminate the filter and follow the red user directly. Therefore, we might suggest to ego to follow the people who are followed by the people he follows already.

4. DISCUSSION

The methods we describe above act to increase the local density in social networks. Since the contacts that are suggested are already friends of friends, adding a link between ego and the suggested contact creates *triadic closure* or adding the "third leg" onto a "triangle" formed by these three nodes, a phenomenon first noted by Simmel but developed further in the "strength of weak ties" concept [10].

In another sense, suggestions such as the ones we propose create ties between users whose network positions are nearly structurally equivalent [17]. In the case of shared audiences for example, if users A and B are each followed by the same set of actors, then A and B would be highly suggested as a potential tie.

These methods increase information flow within the local network but do not increase information coming from outside the local network. Therefore, they can be considered to build community at the expense of diversity of information access. We know that there is informational value in having ties, especially unique ties, outside one's local network and that people often act consciously to create those ties [4][5].

The principle of homophily describes how people form ties with like-minded others [13], and that people who interact with people like themselves will become more polarized and extreme in their views [1]. What is not clear is to what extent people want to find others like themselves or would want to find those that they might not otherwise interact with. To what extent might we want to encourage homophily or heterophily in Twitter and other online social networks? Though having a dense local network with strong ties encourages continued use of such services, the "echo chamber" phenomenon observed in relatively well-connected and isolated groups of individuals is known to have a negative, polarizing effect [15].

As more people join Twitter, and the ways that it is used for socializing, collective action, sharing news, or debate continue to evolve, understanding what people's follower networks look like and the motivations underlying their growth will be of social and political importance.

5. NEXT STEPS

We have written scripts that collect the data necessary to make the recommendations described above, and have made the recommendations for each of the authors. Next, we will build these recommendations into a web-based tool and examine the extent to which Twitter users in general do or do not benefit from each of the kinds of recommendations we describe.

6. REFERENCES

[1] R. S. Baron, S. I. Hoppe, C. F. Kao, B. Brunzman, B. Linneweh, and D. Rogers. Social corroboration and

opinion extremity. *Journal of Experimental Social Psychology*, 32:537–560, 1996.

[2] P. Bonacich. Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182, 1987.

[3] D. Boyd, S. A. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. Under review.

[4] R. S. Burt. *Structural Holes: The Social Structure of Competition*. Harvard University Press, 1992.

[5] V. Buskens and A. van de Rijdt. Dynamics of networks if everyone strives for structural holes. *American Journal of Sociology*, 117(2):371–407, 2008.

[6] R. M. Emerson. Power-dependence relations. *American Sociological Review*, 27:31–41, 1962.

[7] R. M. Emerson. Exchange theory, part II. In J. Berger, M. Zelditch, and B. Anderson, editors, *Sociological Theories in Progress, Vol. 2*, pages 58–87. Houghton Mifflin, 1972.

[8] S. Feld. The focused organization of social ties. *American Journal of Sociology*, 86(5):1015–1035, 1981.

[9] A. Gouldner. The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25(2):161–178, 1960.

[10] M. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78:1360–1380, 1973.

[11] M. Granovetter. Ignorance, knowledge and outcomes in a small world. *Science*, 301:773–774, 2003.

[12] J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[13] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.

[14] L. D. Molm. Theoretical comparisons of forms of exchange. *Sociological Theory*, 21(1):1–17, 2003.

[15] C. R. Sunstein. *Republic.com*. Princeton University Press, 2001.

[16] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.

[17] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.