Bonding vs. Bridging Social Capital:
A Case Study in Twitter

Matthew S. Smith
Department of Computer Science
Brigham Young University
Contact: http://m.smithworx.com

Christophe Giraud-Carrier
Department of Computer Science
Brigham Young University
Email: cgc@cs.byu.edu

Abstract—Online communities are connecting large numbers of individuals and generating rich social network data, opening the way for empirical studies of social behavior. In this paper, we consider the widely-held view of social scientists that bonding interactions are more likely than bridging interactions in social networks, and test it within the context of the large online Twitter community. We find that indeed users who request to follow others having similar profile descriptions (i.e., attempting to bond) increase the number of Twitter users who reciprocate their follow requests. From a practical standpoint, this result also informs how a new user might interact on Twitter to maintain a high follow-back ratio.

I. INTRODUCTION

Online communities are groups of individuals connected by some generally well-defined, explicit relation, such as a shared medical condition in a health community, a trusted contact link in a business network, or an established friend or family relationship in a photo-sharing community. Technology, of course, has been the great enabler for the creation and evolution of such communities. Many of the most visited websites on the Internet, such as YouTube, Facebook, Wikipedia, and Blogger, allow users to connect and maintain ties via a social network. Furthermore, various organizations and initiatives are advocating the creation of standards that support this trend. For example, the Friend of a Friend project describes itself as "a simple technology that makes it easier to share and use information about people and their activities," while the OpenSocial initiative observes that "the web is more interesting when you can build apps that easily interact with your friends and colleagues."

The emergence of global and easily accessible online communities is revolutionizing the way in which individuals, and now even businesses, interact with each other. In turn, the science of building, discovering, understanding and leveraging such communities, or social networks, becomes increasingly important, as the Internet continues to grow into the largest collection of ideas, attitudes, personalities, and cultures in human history.

At the heart of social network analysis is the notion of social capital, aggressively pursued and popularized in the past couple of decades by sociologists and political scientists, such as Coleman [1], Lin [2], and Putnam [3]. Unlike other forms of capital that are centered around the individual, social capital is a property that emerges from the relationships that exist among individuals. While there is no consensual definition of social capital, most definitions focus on the value of social relations in achieving some individual or group benefit based on the resources present in the underlying network. The focus of social capital may be on the relations one specific individual maintains with other individuals, on the structure of the relations within a group of individuals, or on a combination of these [4], [5]. An interesting study of the role of social capital in creating group-level benefits is Paxton's work on the mutually reinforcing effects of social capital and democracy [6]. In this paper, we restrict our attention to a consideration of the relationship of social capital to individual-level benefits or goods.

There is still an active discussion in the social sciences of exactly what social capital is, what forms it may take, or what it may entail. It is clear though that in order to create and leverage social capital, individuals must interact. For the purposes of our study, we consider one of the three forms of social capital identified by Coleman, namely information. Informational social capital arises from relations that provide information that, in turn, facilitates action [1]. Furthermore, we adopt Putnam's high-level dichotomy of social capital into bonding social capital and bridging social capital, where bonding social capital refers to the value assigned to social networks among homogeneous groups of people and bridging social capital refers to the value assigned to social networks among heterogeneous groups of people [3], [7].

The study of social capital requires the availability of sufficiently rich social network data. In the physical world, the acquisition of such data remains one of the biggest challenges. To cite just one example, Haynie's recent work on delinquency is based on the National Longitudinal Survey of Adolescent Health, which draws information from kids at 132 schools, yet the network sample includes kids from only 16 schools [8]. Studies have to be rather large to obtain even adequate network data. The cost of compiling such studies is significant as it involves the design and administration of expensive surveys. By contrast, cyberspace has no such limitations, either of size or cost. Indeed, the ease with which connections can be made online means that rich social network data is becoming available, opening the way for authentic, large-scale analyses of social behavior. We show one such analysis here, focused

on bonding and bridging social capital, in the context of the Twitter community.

Twitter is a fast-growing online social network, which went from 2-4 million users at the beginning of 2009 to about 40 million users by the end of that year. This relatively new community allows users to contribute short free-form status updates, called *tweets*, about themselves, and to follow the updates of others. Individuals are using this service to interact with friends, while businesses are beginning to use it to reach out and respond to customers. Twitter status updates can be a rich source of information about individuals, while the *following* and *follower* relationships provide the backbone of the underlying social network.

The principle of homophily, that contact between similar people occurs at a higher rate than dissimilar people, has been examined extensively [9]. Social capital researchers have also suggested that bonding interactions are more likely to occur than bridging interactions. Lin, for example, points out that interacting homogeneously (i.e., bonding) "should be the expected pervasive pattern of interactions observed," because it requires the least effort, while interacting heterogeneously (i.e., bridging) demands effort due to resource differentials and the lack of shared sentiments [2]. Or, as Burt puts it, "closure is the more obvious force. People advantaged by barriers between insiders and outsiders have no incentive to bring in outsiders. People too long in their closed network have difficulty coordinating with people different than themselves." [10]. In this paper, we use Twitter to verify that bonding interactions are indeed the pervasive pattern.

The paper is organized as follows. We first give a brief overview of some of the most relevant related work. We then proceed to describe our framework to quantify and measure both bonding and bridging social capital in online communities. Finally, we outline our experimental design and present the results.

II. Related Works

The interactions and structure of online social networks is dynamic and complex. Social network analysis assumes that the relationships among interacting units are a critical source of information [11], [12]. Within the social sciences, the study of these interactions has given rise to a number of interesting results. We mention only a few here, that are most relevant to our own analysis. Granovetter introduced the idea of the strength of weak ties, where otherwise dissimilar individuals engage in significant social interactions [13]. While it precedes such work, this idea is captured by the notion of bridging social capital as we use it. Haythornthwaite, in her work on the impact of communication media on social interactions, distinguishes among three types of ties, namely latent ties, weak ties and strong ties [14]. Latent ties correspond to technically possible, but not yet activated communication channels (e.g., belonging to the same email network); weak ties exist once individuals begin to use any medium of communication between them; and strong ties eventually arise as individuals expand their use of existing

and create new media of communication to maintain their interactions. If sharing a communication medium is regarded as type of affinity among individuals, then Haythornthwaite's latent ties are the same as our implicit affinities, and the weights we assign to explicit connections capture the variable strength of ties among individuals. Coleman, in his work on the relationship between social and human capital, discusses the important ideas of obligations, expectations and trust in social networks, where what someone may expect of others depends both on what one has done for them and whether one can safely count on their reciprocating [1]. We capture these ideas through directed, weighted connections.

Most studies have been done in the context of static networks. Recently, however, some researchers have begun to study the actual dynamics of social network formation and evolution, leading to the discovery of several interesting patterns such as degree power laws and shrinking diameters (e.g., see [15], [16], [17], [18], [19]). Other studies have focused on analyzing explicit group formation and evolution [20], [21], [22]. Similarly, our formalism takes into account the inherently dynamic nature of social networks, which, according to Coleman is essential to the formation of social capital [1]. In particular, the notion of implicit affinities is used to further allow the nature of underlying relationships and groupings to vary over time.

In practice, social network analysis has been used to understand an assortment of complex group phenomena, such as terrorist networks [23], [24], [25], [26], animal sociality [27], wasp colonies [28], and spread of diseases and behaviors [29]. Studies with an explicit focus on social capital have been used to explain, for example, how certain individuals obtain more success through using their connections with other people. In an interesting study about CEO compensation, Belliveau and colleagues show that social capital plays a significant role in the level of compensation offered to CEOs [30]. In another study on social capital in the workplace, Erickson concludes that "good networks help people to get good jobs" [31]. Social capital has also been used in computer science to analyze the impact of the number of organizers with whom a potential author is friend on that authors publication records [32], and indirectly to distinguish between factual and relational content in social media communities [33]. Our work continues this tradition of using computational methods to explain social behavior.

III. Social Capital Framework

Social capital within a community is grounded in relationships, individuals' attributes, and available social resources. To exploit this information, we find it useful to distinguish between two types of relationships among individuals, as follows.

- An *explicit* connection links one individual to another based on some purposive action (e.g., sending an email, visiting) or a well-defined relationship (e.g., being a friend of, collaborating with). Individuals thus linked are aware of the explicit connections among them.

- An *implicit* affinity connects individuals together based on loosely defined affinities, or inherent similarities, such as similar hobbies or shared interests. Individuals may not be aware of the similarities in attitudes and behaviors that exist among them.

We call *explicit social networks* (ESNs), social networks built from explicit connections and *implicit affinity networks* (IANs), social networks arising from implicit affinities [34]. A network with both implicit affinities and explicit connections is a *hybrid network*. In social network analysis terminology, a hybrid network is a multigraph having an explicit and an implicit relation among actors.

Implicit affinities are weighted by the amount of similarity estimated between individuals. The similarity metric chosen uses relevant attributes derived from the description and behavior of an individual. We make the important assumption that online personas are accurate. In other words, we assume that "you are what you say you are" online.

Social capital is naturally interested in implicit affinities, since it clearly has some relation to shared affiliations or activities among individuals [30]. On the other hand, social capital can really only accrue when individuals are aware of it, that is, when they establish explicit connections among themselves. Hybrid networks thus play a key role in the definition of social capital, and the kinds of connections that exist among individuals determine whether that capital is realized or not. Note that in a strict sense, social capital is only realized, or accrued, once actions are taken and their result evidences the presence of said social capital. Hence, typical studies of social capital are retrospective. Here, however, we wish to use the notion of social capital to predict how one should leverage one's relations. For example, given that X and I are friends, that X is a head hunter and that I am looking for a job, I would want to ask X to help me find a job. Clearly, I may be misguided in the trust I place in X in this context (i.e., there may not be any social capital for me to leverage here), but it seems most reasonable to assume that I am not and to try to leverage what I perceive as social capital. For simplicity, we say that such capital is realized.

1) Implicit affinities only. In this case, the individuals have much in common (e.g., similar occupation or hobbies) but they are unaware of it. If they were to connect explicitly, they would be bonding, but since they have not yet, we say that there is only potential for bonding social capital here.
2) Implicit affinities and explicit connections. In this case, we say that the potential for social capital is now realized as similar individuals connect to one another explicitly.
3) No implicit affinities and no explicit connections. In this case, the individuals have little or nothing in common and they are unaware of each other. If they were to connect explicitly, they would be bridging, but since they have not yet, we say that there is only potential for bridging social capital here.
4) No implicit affinities but explicit connections. In this

case, the mostly dissimilar individuals are now connected to one another (e.g., colleagues collaborating across disciplines or members of a church choir). Hence, we say that there is realized bridging social capital.

Both implicit affinities and explicit connections are therefore necessary to predict the network's social capital. Based on this framework, we have derived an effective mathematical formulation of social capital, as follows. An earlier version of this formulation is in [35].

Let s_{ij}^{IAN} be the strength of the implicit affinity between nodes i and j. s_{ij}^{IAN} ranges over [0,1] and is a measure of the similarity between nodes i and j. Similarly, let s_{ij}^{ESN} be the strength of the explicit connection between nodes i and j. s_{ij}^{ESN} may be as simple as 1 or 0, to reflect the presence or absence of a link, but may also range over [0,1] to capture degrees of connectivity (e.g., best friend vs. casual friend vs. acquaintance). Finally, let N be the set of nodes in the network.

We define the *potential* bonding social capital of an individual i as the sum of the individual's implicit affinity strength to every other individual. That is,

$$pb(i) = \sum_{j \in N, j \neq i} s_{ij}^{IAN}$$

Likewise, we define the *potential* bridging social capital of an individual i as the sum of the individual's implicit dissimilarity strength to every other individual. That is,

$$pbr(i) = \sum_{j \in N, j \neq i} (1 - s_{ij}^{IAN})$$

While it seems appropriate for implicit affinities to be "undirected," since two people either share or do not share a specific affinity, it is less so for explicit edges. It is clear that the value of some (explicit) relationships is not necessarily reciprocal and may vary among participants. For example, one person may consider another person as their best friend, while that other person may look at the first as only a good friend. Thus, our framework recognizes that the amount of social capital an individual i may realize from a relationship with another individual j is not predicated upon the value that i places in the relationship, but rather upon the value that j places in it. While i may think highly of that connection, for example in the context of obtaining a job reference from j, the reference will only be as strong as j thinks of i, and not the other way.

Accordingly, we define the bonding social capital *realized* by a node i, when (explicitly) connecting with node j, as the product of the strength of the implicit affinity between i and j by the strength of the explicit edge connecting j to i: $s_{ij}^{IAN} s_{ji}^{ESN}$. Now, as expected, if j is unaware of i, even when i may be aware of (and possibly even count on) j, there is no social capital available for i from that relationship. The (realized) bonding social capital of an individual i is the sum of its realized bonding social capital with all other individuals. That is,

$$b(i) = \sum_{j \in N, j \neq i} s_{ij}^{IAN} s_{ji}^{ESN}$$

Likewise, the (realized) bridging social capital of an individual i is the sum of realized bridging social capital with all other individuals. That is,

$$br(i) = \sum_{j \in N, j \neq i} (1 - s_{ij}^{IAN}) s_{ji}^{ESN}$$

Finally, as mentioned earlier, social capital is comprised of the two types of social capital. Therefore, the social capital for an individual i is the sum of its bonding capital and bridging capital. That is,

$$sc(i) = b(i) + br(i)$$

IV. Experimental Setup

The following experiment was designed to test the social scientists' hypothesis that bonding is more likely than bridging in social networks. Given our framework, this hypothesis may be recast into the following measurable Twitter hypothesis.

Hypothesis. *Following users with whom the most affinities are shared (i.e., attempting to bond) produces more follow-backs (i.e., bonding) than other following strategies.*

The idea is that individuals who follow-back others who deliberately adopt a following strategy motivated by a desire to bond create bonding social capital. If there is a stronger tendency for individuals to follow-back those of their followers who are like them rather than others, then this, in some sense, establishes that bonding is more pervasive than bridging. Hence, by comparing the relative number of follow-backs and followers (i.e., the bonding social capital accrued) by adequate strategies, we can verify our hypothesis.

For our experiment, we created a set $\mathcal{A}$ of Twitter accounts. Each account was setup to behave approximately the same as all others for everything, except for the individuals it chooses to follow. Specifically, each Twitter account in $\mathcal{A}$ was given a screen name that varied only by the random three-digit number appended to a pre-specifed name (e.g., $jon287$, $jon797$, $jon853$). Each account was also given the same profile information (see Figure 1 for an example), and all accounts were scheduled to tweet at approximately the same times. This rigorous setup allowed us to test the unique following strategies assigned to each account.

The implicit affinity network among Twitter users was derived from the profile description, labeled *Bio* in Figure 1. Alternatively, the implicit affinity network could naturally be derived from tweets made by users, thus creating a more dynamic but also more computationally intensive network. We chose to be conservative by utilizing just the profile description for this study. Collecting status updates (i.e., tweets) requires additional calls to the Twitter API and could require a significant amount of text mining, which could slow down the pace of the experiment and possibly limit the applicability of the results for regular Twitter users. Profile descriptions are relatively easy to obtain, are less dynamic, and can be sufficiently descriptive of a particular user. Implicit affinities were calculated by counting the number of matching unigrams and bigrams within profile descriptions after removing common

stop words (e.g., a, by, on, the, with). Thus, users having more profile affinities (with the accounts in $\mathcal{A}$) offer opportunities for bonding. Those with few or no affinities offer opportunities for bridging.
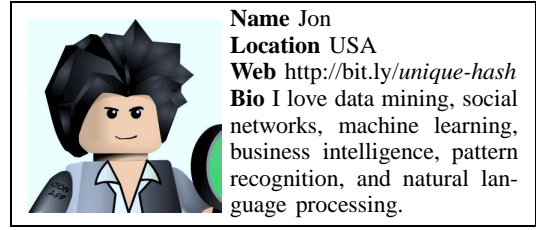

Fig. 1. Twitter profile information for each account in $\mathcal{A}$.

A selection of Twitter users, $\mathcal{U}$, was sampled from the Twitter public timeline. Rather than including anyone that recently tweeted something, we decided to restrict our sampling to those that had recently tweeted either "data mining" or "social networks". Sampling from the Twitter public timeline without restriction opens the possibility that all of the accounts sampled in $\mathcal{U}$ could share very little or even nothing with the niche accounts in $\mathcal{A}$. We chose to avoid this possibility, and effectively narrowed the candidate pool to those that could possibly share an interest with the targeted focus of the accounts in $\mathcal{A}$.

Next, each Twitter account in $\mathcal{A}$ was assigned a following strategy as follows.

A *Bonding* - this strategy attempts to bond by following users, in $\mathcal{U}$, with whom the most affinities are shared. In other words, at the time of selection, the user with the largest s_{ij}^{IAN} is followed.

B *Bridging* - this strategy attempts to bridge by following users, in $\mathcal{U}$, with whom the least affinities are shared. In other words, at the time of selection, the user with the smallest s_{ij}^{IAN} is followed.

C *Median affinities* - this strategy follows the users in $\mathcal{U}$ having the median number of affinities shared at the time of selection.

D *Randomly* - this strategy randomly follows a user in $\mathcal{U}$ at the time of selection. Every remaining user in $\mathcal{U}$ has the same probability of being followed.

E *Minimum absolute following/ers difference* - This strategy follows the user in $\mathcal{U}$ having the smallest difference between following and followers (i.e., $|following_{count} - followers_{count}|$) at the time of selection. For example, the absolute following difference is 400 for a user following 100 and followed by 500.

F *Maximum absolute following/ers difference* - This strategy follows the users in $\mathcal{U}$ having the largest absolute difference between following and followers (i.e., $|following_{count} - followers_{count}|$). The absolute following difference is calculated as described in the previous strategy.

G *Median number of followers* - This strategy follows the user in $\mathcal{U}$ having the median number of followers at the time of selection.

rank$_f$	strategy	following	follow-backs	↓ followers	rejects	churn	follow$_{total}$	follower$_{total}$
1	*bonding* (A)	500	158 (32%)	202 (40%)	12	127	512	329
2	*max. following/ers diff.* (F)	500	84 (17%)	172 (34%)	12	324	512	496
3	*random* (D)	500	118 (24%)	154 (31%)	20	103	520	257
4	*median affinities* (C)	500	99 (20%)	123 (25%)	25	93	525	216
5	*bridging* (B)	500	99 (20%)	120 (24%)	25	91	525	211
6	*min. following/ers diff.* (E)	500	87 (17%)	99 (20%)	50	55	550	154
7	*median num. followers* (G)	500	63 (13%)	86 (17%)	31	51	531	137
8	*min. num. followers* (H)	500	33 (07%)	42 (08%)	79	29	579	71
9	*follow nobody* (I)	0	0 (—%)	3 (—%)	0	24	0	27

TABLE I

FOLLOWER STATISTICS: EACH OF THE NINE ACCOUNTS ARE LISTED BY *strategy* AND RANKED BY THE NUMBER OF *followers* OBTAINED DURING THE EXPERIMENT, DENOTED *rank$_f$*. THE *following* COLUMN IS THE NUMBER OF USERS THAT THE ACCOUNT WAS FOLLOWING AT THE END OF THE EXPERIMENT. THE *follow-backs* COLUMN REPORTS THE NUMBER OF USERS FOLLOWED THAT WERE FOLLOWING THE ACCOUNT BACK AT THE END OF THE STUDY, THE PERCENT OF FOLLOWING IS SUPPLIED FOR REFERENCE (I.E., *follow-backs/following*). THE *followers* COLUMN IS THE NUMBER OF USERS FOLLOWING THE ACCOUNT AT THE END OF THE EXPERIMENT (INCLUDING THOSE THAT WERE NEVER FOLLOWED BY THE ACCOUNT). THE PERCENT OF FOLLOWING IS ALSO SUPPLIED FOR REFERENCE (I.E., *followers/following*). THE *rejects* COLUMN REPORTS THE NUMBER OF USERS THAT COULD NOT BE FOLLOWED ON TWITTER AT THE TIME (E.G., ACCOUNT WAS PROTECTED, USER ATTEMPTING TO FOLLOW WAS BLOCKED, OR USER WAS SUSPENDED). THE *churn* STATISTIC REPORTS THE NUMBER OF USERS THAT FOLLOWED THE ACCOUNT FOR A TIME, BUT WERE NO LONGER FOLLOWING THE ACCOUNT AT THE END OF THE EXPERIMENT. THE *follow$_{total}$* IS THE TOTAL NUMBER OF USERS THAT WERE FOLLOWED BY THE ACCOUNT, I.E., THE SUM OF *following* AND *rejects*. THE *follower$_{total}$* IS THE TOTAL NUMBER OF USERS THAT FOLLOWED THE ACCOUNT DURING THE EXPERIMENT, I.E., THE SUM OF *followers* AND *churn*.

H *Minimum number of followers* - This strategy follows the user in $\mathcal{U}$ having the fewest number of followers at the time of selection.

I *Follow nobody* - this strategy chooses not to follow any users. It may naturally be viewed as a control group.

A *following round* consisted of each Twitter account in $\mathcal{A}$ selecting users from $\mathcal{U}$ one at a time according to its assigned strategy. Users were removed from $\mathcal{U}$ as soon as they were selected. Thus, users from the pool could only be followed by a single account in $\mathcal{A}$. Each following round began by randomizing the order in which accounts selected users. On the days that following rounds occurred, accounts selected 50 or less users to follow. Following rounds were planned to occur sporadically until every account in $\mathcal{A}$ was following 500 users (an arbitrary, yet substantial number of individuals). Following rounds occurred on 22 of the 105 days in which the experiment was conducted.

For the duration of the study, each of the Twitter accounts in $\mathcal{A}$ published identical status updates to their respective Twitter stream at approximately the same time. There were 117 status updates made across 19 different days during the experiment. Over 90% (106) of the status updates published included a link that tracked the number of times it was clicked. Each link was shortened (using bit.ly) and associated to specific account in $\mathcal{A}$. After all of the users in $\mathcal{U}$ had been selected by the accounts in $\mathcal{A}$ and all status updates had been published the experiment concluded. The following statistics were analyzed for each account at the conclusion of the experiment:

- number of followers
- number of click-thrus (tweets and profile click-thrus)
- individual bonding capital
- individual bridging capital

V. RESULTS

The final follower statistics for each account after the experiment are shown in Table I. Each account in $\mathcal{A}$ is listed by the assigned strategy and ranked by the number of *followers*. The number of *follow-backs* is the subset of *following* users that reciprocated follow requests made by the account. The *followers* column reports the number of followers that the account had at the end of the experiment. Unlike the statistic reported in *follow-backs*, this statistic includes followers that discovered the account through alternative methods. Although we do not know all of the ways that accounts can get noticed through the numerous Twitter apps, a few alternative methods for being noticed on the Twitter website include being discovered through Twitter search or by traversing the explicit social network (e.g., being discovered through a "friend of a friend").

The *rejects* column reports the number of users that could not be followed on Twitter, at the time the request was made, due to some reason, such as the account was protected, the user attempting to follow was blocked, or the user was suspended. Twitter is a constantly evolving community where users can block other users on a whim and where users are regularly suspended for "strange activity." For instance, attempting to follow a user that has been suspended produces the following error message: "Could not follow user: This account is currently suspended and is being investigated due to strange activity." Rejection errors occurred most often for strategies H and E, perhaps suggesting something about users that fall into these groupings (i.e., users in in group H likely block the accounts in $\mathcal{A}$ and users in group E tend to get suspended more often.)

The *churn* statistic represents the number of users that followed the account for some time during the study, but no longer followed the account at the end of the study. The current guidelines on Twitter's website state "if you decide to
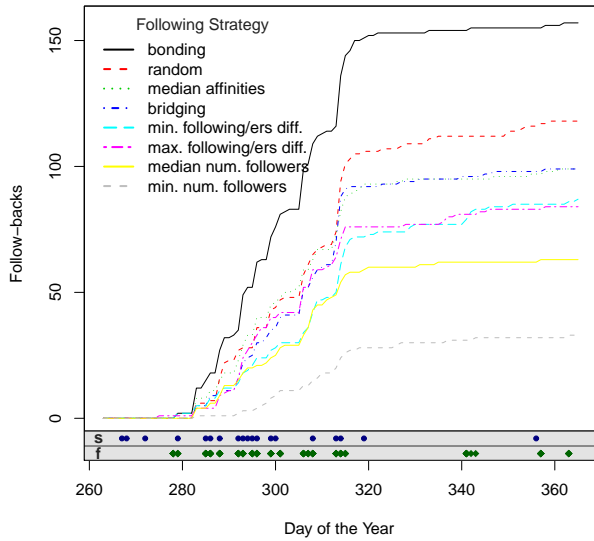
Fig. 2. **Follow-backs Over Time**: Follow-backs obtained by accounts in $\mathcal{A}$ throughout the duration of the study. Days in which following rounds occurred (i.e., accounts in $\mathcal{A}$ followed users in $\mathcal{U}$) are marked in the row labeled **f**. Days that new status updates were posted to the accounts in $\mathcal{A}$ are marked in the row labeled **s**.

strategy	significantly different
(A) *bonding*	B, C, E, G, H
(B) *bridging*	A, H
(C) *median affinities*	A, H
(D) *random*	E, G, H
(E) *min. following/ers diff.*	A, D, F, H
(F) *max. following/ers diff.*	E, G, H
(G) *median num. followers*	A, D, F, H
(H) *min. num. followers*	A, B, C, D, E, F, G

TABLE II
FOLLOWERS-TO-FOLLOWING: PAIRWISE PROPORTION TEST RESULTS. ($\alpha = 0.01$, BONFERRONI CORRECTED p-VALUES)

strategy	significantly different
(A) *bonding*	B, C, E, F, G, H
(B) *bridging*	A, H
(C) *median affinities*	A, G, H
(D) *random*	G, H
(E) *min. following/ers diff.*	A, H
(F) *max. following/ers diff.*	A, H
(G) *median num. followers*	A, D, C
(H) *min. num. followers*	A, B, C, D, E, F

TABLE III
FOLLOWBACKS-TO-FOLLOWING: PAIRWISE PROPORTION TEST RESULTS. ($\alpha = 0.01$, BONFERRONI CORRECTED p-VALUES)

follow someone and then change your mind later, thats fine!" However, they discourage *aggressive follow churn*, which they define as "when an account repeatedly follows and un-follows large numbers of users."[1] Churn was observed most often for strategy F (more than double strategy A, the next highest) — perhaps more users selected by this group are using automated tools to aggressively follow and un-follow.

The second to last column, $follow_{total}$, is the total number of users that were followed by the account, or the sum of *following* and *rejects*. The last column, $follower_{total}$, is the total number of users that followed the account during the experiment, or the sum of *followers* and *churn*.

Figure 2 shows a plot of the number of follow-backs that each account had during the experiment. Following rounds occurred on the days marked in the row labeled **f**. Status updates occurred on the days marked in the row labeled **s**. The follow-backs plotted is the cumulative sum of followers obtained on the day indicated and that remained at the conclusion of the experiment. Users that followed back but were no longer following at the end of the study are not included. Following rounds are accompanied by noticeable increases in follow-backs for all following strategies, with significantly larger such increases for strategy A, a first indication that bonding may indeed be easier.

We formally tested our hypothesis using proportion tests. Strategy I is left out as it does not follow anyone (i.e., *following*=0) and thus proportions would be undefined (division by 0). All other strategies are included in the results, but our

[1] *Following Limits and Best Practices* available at: http://help.twitter.com/forums/10711/entries/68916 (*Jan. 06, 2010*).

focus is on strategies A and B. Users who follow A, especially when reciprocating (i.e., follow-backs), are clearly bonding since they were first picked by A because they were similar to A. Users who follow B, again especially when reciprocating, cannot be bonding, and must be bridging, since B explicitly chose them for their dissimilarity with itself.

A test comparing the *followers*-to-*following* proportions showed that strategies A and B were significantly different having a p-value < 0.001. Upon performing a pairwise proportion test across all of the strategies, we observe that many of the strategies were significantly different, as shown in Table II. Note that the p-values were Bonferroni adjusted and considered significant only if they were less than alpha ($\alpha = 0.01, p < \alpha$). While it appears that A's bonding strategy is not significantly different from D's random following strategy, this is probably due to the fact that we pre-selected the set $\mathcal{U}$ of users to follow based on their affinities with A. Hence, if our hypothesis holds, a random strategy would exhibit a fair amount of bonding.

As an additional check, a pairwise proportion test was performed on the *follow-backs*-to-*following* proportion, as shown in Table III. Again, this test shows that strategies A and B are significantly different. As above, the p-values were Bonferroni adjusted and considered significant only if they were less than alpha ($\alpha = 0.01, p < \alpha$). These results are similar to the above.

Next, we investigate the click statistics. Table IV shows the clicks obtained through each account (and the number of mentions). Each of the nine accounts are listed by *strategy* and ranked by the number of *total clicks* received, denoted $rank_c$. Each account made approximately 117 status updates

(i.e., tweets), of which 106 included a clickable tracking link. The $clicks_t$ column shows the number of clicks that came through links posted in status updates for each account. The next column, $clicks_p$, shows the number of times that the profile link (i.e., unique tracking link immediately after *Web* in Figure 1) was clicked for each account. The next column, *total clicks*, is the sum of the previous two columns and is the total number of clicks obtained for each account. Lastly, the number of *mentions*, or any Twitter update that contained @*username* in the body of the status update, is listed for each account. Note that due to a small configuration error, not all of the click data was recorded for strategy G. It is therefore not included in the ranking (only 78 of the 106 links posted were tracked).
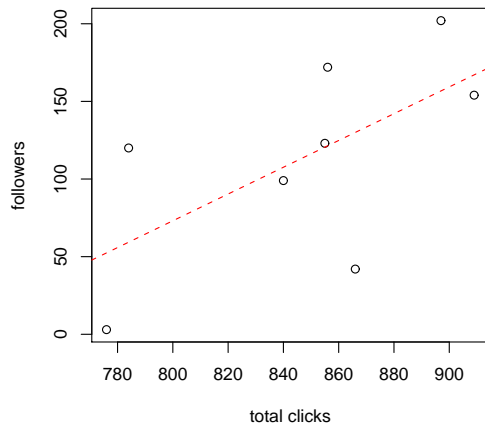
$rank_b$	strategy	↓ bonding
1	*bonding* (A)	10%
2	*max. following/ers diff.* (F)	3%
3	*follow nobody* (I)	3%
4	*random* (D)	2%
5	*bridging* (B)	2%
6	*min. following/ers diff.* (E)	2%
7	*median num. followers* (G)	2%
8	*median affinities* (C)	1%
9	*min. num. followers* (H)	1%

TABLE V

SOCIAL CAPITAL RESULTS: EACH OF THE NINE ACCOUNTS ARE LISTED BY *strategy* AND RANKED BY THE PROPORTION OF BONDING SOCIAL CAPITAL THEY ACCRUED (I.E., $b(i)/sc(i)$), DENOTED $rank_b$. STRATEGY A HAS SIGNIFICANTLY MORE BONDING SOCIAL CAPITAL THAN ANY OF THE OTHER STRATEGIES.

Fig. 3. **Clicks vs. Followers**. The linear model shown by the regression line (dashed) poorly fits the data having an R^2 value of 0.28. There is, however, a positive Pearson correlation of 0.62, yet it is not as high as might be expected.

The click results were somewhat unexpected. First, the number of clicks obtained for strategy I (i.e., followed nobody) is surprisingly high and similar to the number of clicks obtained for the other strategies. We think that this may be due to how rigorously Twitter data was being consumed by automated tools and web crawlers during the study. Secondly, we expected that the number of clicks for each account would be linearly proportional to the number of followers. This was not the case. Figure 3 plots the number of clicks versus the number of followers. The adjusted R^2 value of 0.28 for the linear model confirms that it poorly fits the data. The number of clicks did not appear to be proportional to the number of followers, nor did the number of clicks vary significantly among strategies with a standard deviation of 46. These results suggest that, in terms of obtaining clicks, tweeting is more important than obtaining more followers.

Finally, we consider the social capital accrued by each strategy. Using the formulas defined in our framework, we compute the social capital realized by each strategy at the end of the experiment. Table V shows the proportion of bonding social capital (to total social capital) for each strategy,

in descending order. Again, these results set A's bonding strategy as a clear winner over all strategies, and in particular significantly higher than B's bridging strategy. Note that the seeming rise of strategy I is due to the fact that social capital is accrued based on followers rather than following, as discussed above. Hence, while I did not follow anybody, it did garner three followers as shown in Table I. The 3% proportion of bonding social capital is, however, artificially inflated by I's small number of followers.

In passing, we note that the above results also seem to confirm the intuition that utilizing a random following strategy produces more follow-backs than following nobody at all. We do have to be a little careful here since, as mentioned above, the random strategy here may be confounded by our "bonding-friendly" pre-selection of users.

VI. CONCLUSION

Social media is becoming an important channel for sharing news and information. For many individuals and businesses, the very dynamic Twitter community is a particularly attractive social network to participate in. We have used a novel computational framework for social capital, together with a well-defined experiment, to verify the widely-held view that bonding interactions are more likely than bridging interactions in social networks.

Our experiments involved analyzing the behavior of a group of Twitter users in reaction to a number of artificial users with pre-defined strategies. The results considered such quantities as ratio of follow-backs to followings as well as accrued social capital. In particular, they show that users who request to follow others having similar profile descriptions (i.e., attempting to bond) increase the number of Twitter users that reciprocate their follow requests, thus generating significantly more bonding social capital. Indirectly, this highlights a strategy that a new user could employ to maintain a high follow-back ratio when interacting with people on Twitter.

rank$_c$	rank$_f$	strategy	clicks$_t$	clicks$_p$	↓ total clicks	mentions
1	3	*random* (D)	900	9	909	2
2	1	*bonding* (A)	882	15	897	3
3	8	*min. num. followers* (H)	850	16	866	1
4	2	*max. following/ers diff.* (F)	849	7	856	1
5	4	*median affinities* (C)	846	9	855	1
6	6	*min. following/ers diff.* (E)	821	19	840	4
7	5	*bridging* (B)	773	11	784	2
8	9	*follow nobody* (I)	775	1	776	1

TABLE IV

CLICK STATISTICS: EACH OF THE NINE ACCOUNTS ARE LISTED BY *strategy* AND RANKED BY THE NUMBER OF *total clicks* RECEIVED, DENOTED rank$_c$. EACH ACCOUNT MADE APPROXIMATELY 117 STATUS UPDATES (I.E., TWEETS), OF WHICH 106 INCLUDED A CLICKABLE TRACKING LINK.

REFERENCES

[1] J. S. Coleman, "Social capital in the creation of human capital," *American Journal of Sociology*, vol. 94, pp. S95–S120, 1988.

[2] N. Lin, *Social Capital: A Theory of Social Structure and Action*. NY: Cambridge University Press, 2001.

[3] R. D. Putnam, *Bowling Alone: the Collapse and Revival of American Community*. Simon & Schuster, 2000.

[4] S. P. Borgatti, C. Jones, and M. G. Everett, "Network measures of social capital," *Connections*, vol. 21, no. 2, pp. 27–36, 2 1998.

[5] P. S. Adler and S.-W. Kwon, "Social Capital: Prospects for a New Concept," *The Academy of Management Review*, vol. 27, no. 1, p. 17, January 2002.

[6] P. Paxton, "Social capital and democracy: An interdependent relationship," *American Sociological Review*, vol. 67, no. 2, pp. 254–277, 2002.

[7] R. D. Putnam and L. M. Feldstein, *Better Together: Restoring the American Community*. Simon & Schuster, 2003.

[8] D. L. Haynie and D. W. Osgood, "Reconsidering peers and delinquency: How do peers matter?" *Social Forces*, vol. 84, no. 2, pp. 1109–1130, 2005.

[9] M. McPhearson, L. Smith-Lovin, and J. Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, vol. 27, pp. 415–444, 2001.

[10] R. S. Burt, *Brokerage and Closure*. Oxford University Press, 2005.

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

[12] J. P. Scott, *Social Network Analysis: A Handbook*. Thousand Oaks, CA: Sage Publications Ltd; 2nd edition, 2000.

[13] M. Granovetter, "The strength of weak ties," *American Journal of Sociology*, vol. LXXVIII, 1973.

[14] C. Haythornthwaite, "Strong, weak, and latent ties and the impact of new media," *The Information Society*, vol. 18, no. 5, pp. 385–401, 2002.

[15] J. Katz, "Scale independent bibliometric indicators," *Measurement: Interdisciplinary Research and Perspectives*, vol. 3, pp. 24–28, 2005.

[16] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006, pp. 611–617.

[17] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: Densification laws, shrinking diameters and possible explanations," in *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2005, pp. 177–187.

[18] S. Redner, "Citation statistics from 110 years of *Physical Review*," *Physics Today*, vol. 58, pp. 49–54, 2005.

[19] C. Tantipathananandh, T. Berger-Wolf, and D. Kempe, "A framework for community identification in dynamic social networks," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2007, pp. 717–726.

[20] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: membership, growth, and evolution," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006, pp. 44–54.

[21] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, "Microscopic evolution of social networks," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008, pp. 462–470.

[22] E. Zheleva, H. Sharara, and L. Getoor, "Co-evolution of social and affiliation networks," in *Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2009.

[23] N. Memon, D. L. Hicks, H. L. Larsen, and M. A. Uqaili, "Understanding the structure of terrorist networks," *International Journal of Business Intelligence and Data Mining*, vol. 2, no. 4, pp. 401–425, 2007.

[24] J. Xu and H. Chen, "The topology of dark networks," *Communications of the ACM*, vol. 51, no. 10, pp. 58–65, 2008.

[25] M. A. Shaikh and W. Jiaxin, "Network structure mining: locating and isolating core members in covert terrorist networks," *WSEAS Transactions on Information Science and Applications*, vol. 5, no. 6, pp. 1011–1020, 2008.

[26] S. P. Borgatti, "Identifying sets of key players in a social network," *Computational & Mathematical Organization Theory*, vol. 12, no. 1, pp. 21–34, 2006.

[27] T. Wey, D. T. Blumstein, W. Shen, and F. Jordán, "Social network analysis of animal behaviour: a promising tool for the study of sociality," *Animal Behaviour*, vol. 75, no. 2, pp. 333–344, 2008.

[28] A. Bhadra, F. Jordán, A. Sumana, S. A. Deshpande, and R. Gadagkar, "A comparative social network analysis of wasp colonies and classrooms: Linking network structure to functioning," *Ecological Complexity*, vol. 6, no. 1, pp. 48–55, 2009.

[29] N. A. Christakis and J. H. Fowler, "The collective dynamics of smoking in a large social network," *New England Journal of Medicine*, vol. 358, no. 21, pp. 2249–2258, 2008.

[30] M. Belliveau, C. I. O'Reilly, and J. Wade, "Social capital at the top: Effects of social similarity and status on CEO compensation," *Academy of Management Journal*, vol. 39, no. 6, pp. 1568–1593, 1996.

[31] B. H. Erickson, "Good networks and good jobs: The value of social capital to employers and employees," in *Social Capital: Theory and Research*, N. Lin, K. S. Cook, and R. S. Burt, Eds. Aldine Transaction, 2004, ch. 6, pp. 127–158.

[32] L. Licamele and L. Getoor, "Social capital in friendship-event networks," in *Proceedings of the IEEE International Conference on Data Mining*, 2006, pp. 959–964.

[33] V. Barash, M. Smith, L. Getoor, and H. Welser, "Distinguishing knowledge vs social capital in social media with roles and context," in *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media*, 2009.

[34] M. Smith, C. Giraud-Carrier, and B. Judkins, "Implicit Affinity Networks," in *Proceedings of the 17th Annual Workshop on Information Technologies and Systems*, 2007, pp. 1–6.

[35] M. Smith, C. Giraud-Carrier, and N. Purser, "Implicit affinity networks and social capital," *Information Technology and Management*, vol. 10, no. 2–3, pp. 123–134, 2009.