

Maintaining Ties on Social Media Sites: The Competing Effects of Balance, Exchange, and Betweenness

Daniel M. Romero
Cornell University
dmr239@cornell.edu

Brendan Meeder
Carnegie Mellon
bmeeder@cs.cmu.edu

Vladimir Barash
Cornell University
vdb5@cornell.edu

Jon Kleinberg
Cornell University
kleinber@cs.cornell.edu

Abstract

When users interact with one another on social media sites, the volume and frequency of their communication can shift over time, as their interaction strengthens or weakens. We study the interplay of several competing factors in the maintenance of such links, developing a methodology that can begin to separate out the effects of several distinct social forces. In particular, if two users develop mutual relationships to third parties, this can exert a complex effect on the level of interaction between the two users – it has the potential to strengthen their relationship, through processes related to triadic closure, but it can also weaken their relationship, by drawing their communication away from one another and toward these newly formed connections. We analyze the interplay of these competing forces and relate the underlying issues to classical principles in sociology – specifically, the theories of balance, exchange, and betweenness.

In the course of our analysis, we also provide novel approaches for dealing with a common methodological problem in studying ties on social media sites: the tremendous volatility of these ties over time makes it hard to compare one's results to simple baselines that assume static or stable ties, and hence we must develop a set of more complex baselines that takes this temporal behavior into account.

1 Introduction

In studying the interactions on a social media site, a basic question is to understand what causes relationships among users to be strengthened and what causes them to weaken. This is an issue that is not well understood: there are multiple forces that govern the strengths of social ties and pull in competing directions. It is an important problem to design methods of analysis for these systems that can begin to separate out the effects of these different forces. Existing work in on-line domains has approached this issue by identifying dimensions that characterize the strength of ties (Gilbert and Karahalios 2009), and by incorporating factors such as triadic and focal closure (Kossinets and Watts 2006), similarity among individuals (Anagnostopoulos, Kumar, and Mahdian 2008; Crandall et al. 2008; Aral, Muchnik, and Sundararajan 2009; Kossinets and Watts 2009), and the role of positive and negative relationships (Leskovec, Huttenlocher, and Kleinberg 2010).

Copyright © 2011, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Here we develop a framework for using social media data to begin isolating the effects of three distinct social forces on relationship strengths: *balance*, *exchange*, and *betweenness*. We first describe how these forces operate in a social media context, which will also show how they can produce opposite effects. For this discussion we will focus on undirected links, in which relationships are symmetric.

Balance and Exchange. First, we consider *balance*. Suppose user B is friends with users A and C . The principle of balance argues that if A and C do not have a social tie, this absence introduces latent strain into the $B-A$ and $B-C$ relationships, and this strain can be alleviated if an $A-C$ tie forms (Heider 1958; Rapoport 1953). Hence, balance is a force that causes the formation of an $A-C$ tie to strengthen the $B-A$ tie, when C is also linked to B .¹

Counterbalancing this is an equally natural force, which is the principle of *exchange* (Emerson 1962; Willer 1999). Let's return to the user B who is friends with users A and C . If A were to become friends with C , this provides A with more social interaction options than she had previously. The theory of exchange argues that this makes A less dependent on B for social interaction, thereby weakening the $B-A$ tie.

Figure 1(a) is a schematic illustration of how balance and exchange can act on a set of three nodes. We first study the aggregate effect of these forces on communication among Twitter users. For this, we say that a *tie* between two Twitter users has formed when they have each sent at least 3 @-messages to the other.² We consider scenarios, such as in Figure 1(a), where a user B has ties to users A and C , and look at whether an $A-C$ tie does or does not form.

Decaying Relationships and Outside Opportunities.

We find first that the formation of an $A-C$ tie in our Twitter data makes it significantly more likely that the $A-B$ tie will persist (as measured by the generation of future messages

¹There is a related version of balance in which there is a negative link between A and C , but we do not consider this here.

²@-messages are a basic Twitter mechanism in which one user directs a tweet to another; since they are used between people who know one another as well from users toward celebrities, we require multiple reciprocations before we consider the messaging to constitute evidence of a tie.

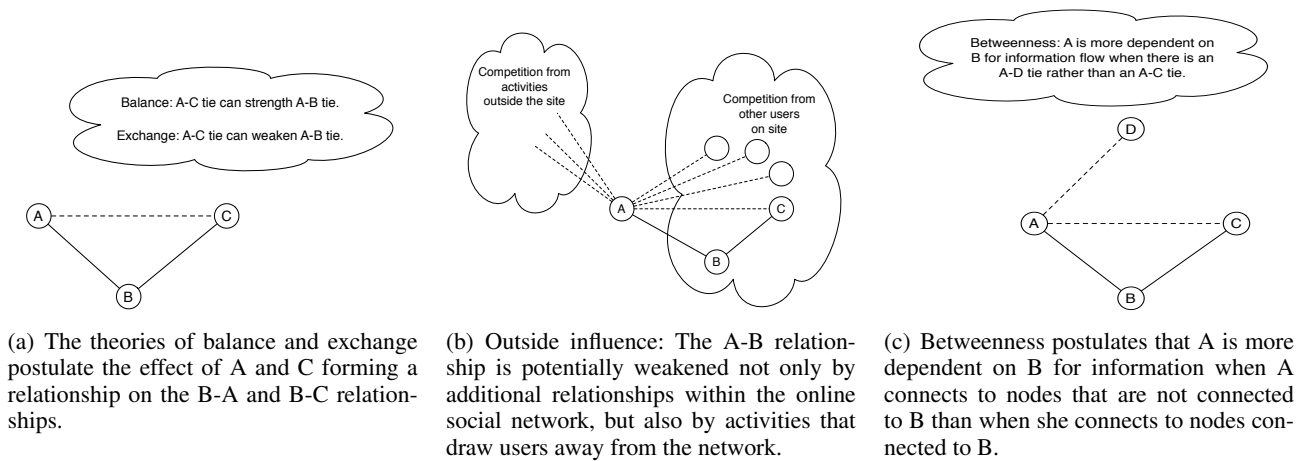


Figure 1: Explanations of related sociological theories

from A to B). At one level, this points to the dominance of balance over exchange in this particular scenario; however, as we investigate the effect of tie formation on tie persistence further, a more subtle picture emerges. Going back to users A , B , and C , suppose that we consider the effect on the A - B tie of A 's sending k messages to arbitrary users other than B , for some relatively large value of k — potentially even requiring these messages to go to users not linked to B . Even in this case, these messages from A to others lead to an increase in the persistence of the A - B tie.

This observation underscores the need to be careful in reasoning about how the persistence of ties operates on a social media site. One might suppose, via the principle of exchange, that the k messages from A to others would divert A 's attention from B , to the detriment of the A - B tie. But consider the full set of activities that might draw A away from B . Interaction with other users on Twitter is one source of such activities. However, there are many activities completely outside Twitter that might also draw A 's attention away from B . Thus the picture from Figure 1(a) should be expanded to look more like the larger picture in Figure 1(b).

The principle of exchange is operating in Figure 1(b); the point is that we are applying it too narrowly if we view other Twitter users as the only sources of outside opportunities for A in the A - B relationship. And the point, then, is that k messages from A to many users other than B still provide strong evidence that A is actively involved in Twitter, rather than in other activities. This increased involvement makes it easier for A 's Twitter activity to “spill over” to the A - B tie.

In Section 4, we consider ways of capturing this spillover effect, and propose a reconceptualization of exchange theory in the particular context of social media to integrate the outside opportunities of a user A at both the “micro” level (to other users on the site) and the “macro” level (to potentially unobserved activities off the site).

This framework also suggests an important methodological consideration underscored by our analyses. Social media sites are domains in which the typical relationship exists in a state of rapid decay, since either user involved in the rela-

tionship may begin to rapidly reduce their involvement in the site, or leave it altogether and never return. Such issues are much less of a constraint (even if they are present at lower levels) in analyses in the off-line world — but in on-line settings, they need to be carefully controlled for.

Balance and Betweenness. Given these considerations, we explore a further set of questions about social forces and relationships in which we control for A 's overall level of involvement in the site. Specifically, consider again a user B who has ties with users A and C . Now, let a fixed amount of time pass, and consider two possible scenarios: (i) A forms a tie with C , or (ii) A forms a tie with a user D who is not connected with B . In which scenario is the A - B tie more persistent? (See Figure 1(c).) Both (i) and (ii) provide evidence of comparable involvement by A in the site, and so we must look to the finer structure of the interaction pattern to decide which has a more positive effect on the A - B tie.

As before, the principle of balance argues that the A - B tie should be more strengthened in scenario (i). But competing forces provide arguments suggesting that scenario (ii) could be better for the A - B link. In particular, access to information is a crucial aspect of Twitter, and when there is no A - C link, user B plays an important brokerage role in her relationship with A : B provides A with access to information from C . If a direct A - C tie forms, this brokerage role is sharply diminished; on the other hand, the role is not as strongly diminished if A forms a tie with D . This argument is based on the principle of *betweenness*, with connections to brokerage and the theory of *structural holes* (Burt 1992).

In Section 3, we carry out a careful analysis of the trade-off between balance and betweenness, finding significant evidence that the balance argument is operating more strongly than the betweenness argument in the setting of Twitter: the closing of the A - B - C triangle (as in scenario (i)) has a more positive effect on the A - B relationship than the formation of ties by A that leave it open (as in scenario (ii)).

2 Data Set and Network Extraction

From August 2009 until January 2010, we crawled Twitter using their publicly available API, collecting over 3 billion messages from more than 60 million users. From this, we extract all @-messages to build an “attention network” that evolves over time: there is a directed edge from user A to user B if A sends at least k @-messages to B (we use $k = 3$), and this edge is created at time $t_D(A, B)$, the point at which the k th message is sent. There are multiple ways of defining a network, and our definition is one way of defining a proxy for the attention that a user A pays to other users. The resulting network contains 8,509,140 non-isolated nodes and 50,814,366 links.

From this directed network we construct an undirected one: we define an undirected edge between A and B when A has sent at least 3 @-messages to B and B has sent at least 3 @-messages to A . The edge $e = (A, B)$ has time-stamp $t(A, B) = \max\{t_D(A, B), t_D(B, A)\}$, the later of the times when the two directed edges were formed. This tie network contains 20,492,393 ties between 3,701,860 users, and although fewer than half of the users remain in the tie network, over 80% of relationships contribute to a tie.

Finally, we define an *open triad* O as a graph of three nodes A, B , and C containing the ties (A, B) and (B, C) . The time-stamp of the open triad is $O_t = \max\{t(A, B), t(B, C)\}$, the time at which the last of the two ties forms. Open triads $O = (A, B, C)$ in which the undirected (A, C) edge eventually forms are said to *close*. We define an open triad that closes d days after O_t ($t(A, C)$ is d days after O_t) to be a *d -closed triad*.

3 Balance Vs. Betweenness

We begin with the contrast between balance and betweenness. We take an open triad (A, B, C) , and as in Figure 1(c), we compare the amount of interaction from A to B after one of the following two events takes place: (i) the $A-C$ tie forms, or (ii) A forms a tie with a user D who is not connected to B . Because we have the times of edge formation, we can control for factors such as the delay between triad formation and the creation of the additional tie. Additionally, we will control for A being ‘active’; we make sure that A was communicating when the triad formed, when the new tie formed, and some time after the new tie formed.

Representing the competing scenarios. In particular, we consider the percentage of messages that A directs to B in two comparison sets of triads designed to represent scenarios (i) and (ii). First, we choose a value for d and consider all d -closed triads; to guarantee A had a certain minimum level of activity overall, we require that A sent between 200 and 1000 messages in total after the open triad (A, B, C) was formed, and moreover that A sent at least one message 1, d , and $2d$ days after the open triad was created.

For scenario (ii), we want an open triad (A, B, C) where A sends a message to a node not connected to B . Thus, for each triad $O' = (A, B, C)$ that never closes, we look at all of the nodes D that are not connected to B , and with which A forms a tie after O'_t . We pick such a node D at random and

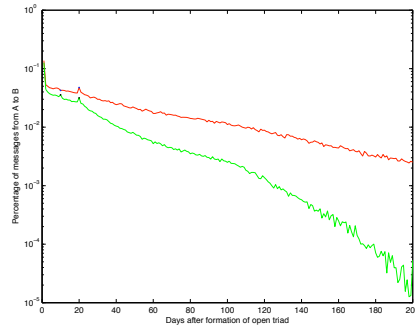


Figure 2: Percentage of message from A to B vs. the number of day after creation of open triad. The green curve is based on the d -open triads and the red curve is based on the d -closed triads. A must have sent from 200 to 1000 messages in total after day = 0 and A must have sent at least one messages on days 1, d , and $2d$. Graph for $d = 10$

say that O' is d -open, where d is the number of days after O'_t that the $A-D$ tie formed. As before, we also require that A sent between 200 and 1000 messages after the open triad (A, B, C) was formed, and that A sent at least one message 1, d , and $2d$ days after the open triad was created.

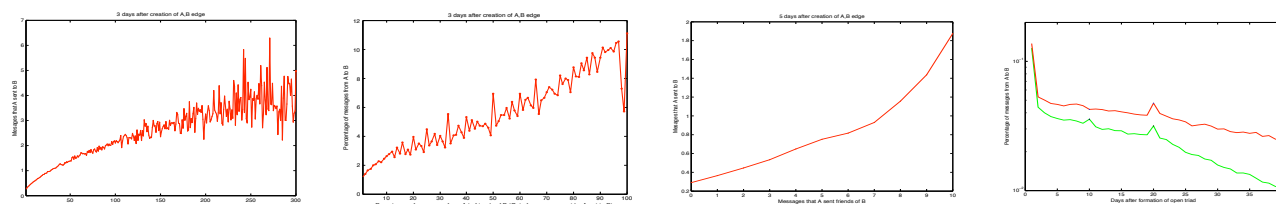
For each population, we measure the percentage of A 's communication that goes toward B , as a function of the time since the formation of the open triad. As noted in the introduction, relationships on social media sites have a default tendency to decay, but by observing which scenario provides a slower aggregate decay rate for the $A-B$ tie, we can begin to learn about the different effects of balance (scenario (i)) and betweenness (scenario (ii)).

Results. Figure 2 shows the results of this test with $d = 10$: the average percentage of messages A sent to node B as a function of the number of days after O_t . The red curve is based on the d -closed triads, while the green curve is based on the d -open triads.

We observe first that for all choices of d , the red curve decreases at a slower rate than the green curve. This indicates that the $A-B$ tie decays more slowly in the population corresponding to scenario (i). But beyond this, the gap between the two curves is widening. After day 100, the communication percentage for the open triads decreases much faster. This suggests that closing the triad benefits communication from A to B by slowing the inevitably decreasing amount of online interaction. In interpreting these results as evidence for the effect of balance, it is important to understand that the formation of the $A-C$ tie is not causing the extent of $A-B$ interaction to increase in an absolute sense, but rather for its rate of decay to be slowed.

4 Exchange Theory and Spill-Over Effects

In the previous section, we observed that the communication between A and B benefits in the long run from the closing of a triad (A, B, C) . More generally, we now ask what can be predicted about the $A-B$ interaction from knowledge of A 's activity level with users other than B .



(a) Number of messages A sends to everyone but B vs. number of messages A sends to B , 3 days after the creation of the A - B edge.

(b) Percentage of messages A sends to B as a function of the percentage of A 's non- B messages that go to friends of B . These messages take place 3 days after the A - B edge forms.

(c) Number of messages A sends to friends of B vs. number of messages A sends to B , five days after the A - B edge forms. Node A sent exactly 10 messages to users other than B .

(d) Zoom-in of figure 2. We observe jumps on the green curve at days d and $2d$ and on the red curve at day $2d$ but not on day d .

Figure 3: Exchange Theory and Spill-Over Effects Figures

Figure 3(a) shows the number of messages A sends to everyone but B vs. the number of messages A sends to B for various points in time after the A - B edge forms. (In the figure, we look three days after the the edge forms; the plots for other time periods are similar.) The plot's monotonic increase is at odds with the basic prediction of exchange theory, which posits that A 's messages to others should reduce the time A has for communicating with B . It is consistent, however, with the idea underlying Figure 1(b), where messages from A to others indicate A is spending more time on Twitter overall, and hence has more time for B as well.

The Role of Balance in Spill-Over Effects. Thus A 's activity toward users other than B “spills over” in a positive way toward B . We now show that the principle of balance can enhance this spill-over effect.

In particular, we consider the messages sent by A to users other than B , and ask what fraction of these messages go to users C with whom B also has a tie. As Figure 3(b) shows, the fraction of messages from A to B increases as this fraction of messages from A to B 's friends increases: that is, the spill-over in A 's activity toward B is accentuated when A 's activity toward users other than B occurs with friends of B . We also consider a version where A 's activity level is fixed: in Figure 3(c) we consider only users A who sent exactly 10 messages to users other than B , and we ask how many messages A sends to B as a function of the number of these non- B messages that go to friends of B . The increase of the curve again shows how the spill-over is strongly enhanced when A 's non- B activities include many friends of B .

Finally, we discuss one intriguing situation with an apparent lack of spill-over. Figure 3(d) zooms in around the days d and $2d$ (in this case 10 and 20) on the curves from Figure 2. The upward “jumps” on days d and $2d$ correspond to increased probability of an A - B message on days when we stipulate that A must have sent at least one message. But the lack of a jump on day d in the red curve — the day when A messaged a neighbor of B — points to a possible case in which A 's actions toward others are reducing the level of activity on the A - B link. Understanding this effect and the mechanism behind it is an intriguing open question.

5 Conclusions and Future Work

We have developed methods for isolating the effect of three distinct social forces on the strength and longevity of ties in social media contexts: balance, in which ties are strengthened when they close triads; exchange, in which ties are weakened when one end of the tie has other opportunities; and betweenness, in which ties are strengthened when they serve as conduits for information.

Our analysis shows the power of balance in the domain we study, Twitter. It also suggests a broadening of exchange theory to include off-site opportunities for participants in a tie, reflecting the rapid rate at which ties decay. We believe the framework here can be applied to social media settings quite broadly, and suggests ways of comparing sites by the different extents to which these diverse social forces operate.

References

- Anagnostopoulos, A.; Kumar, R.; and Mahdian, M. 2008. Influence and correlation in social networks. *Proc. ACM KDD Conf.*
- Aral, S.; Muchnik, L.; and Sundararajan, A. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci.* 106(51):21544–21549.
- Burt, R. S. 1992. *Structural Holes*. Harvard University Press.
- Crandall, D.; Cosley, D.; Huttenlocher, D.; Kleinberg, J.; and Suri, S. 2008. Feedback effects between similarity and social influence in online communities. *Proc. ACM KDD Conf.*
- Emerson, R. 1962. Power-dependence relations. *American Sociological Review* 27:31–40.
- Gilbert, E., and Karahalios, K. 2009. Predicting tie strength with social media. *Proc. 27th ACM CHI Conf.*, 211–220.
- Heider, F. 1958. *The Psychology of Interpersonal Relations*. Wiley.
- Kossinets, G., and Watts, D. 2006. Empirical analysis of an evolving social network. *Science* 311:88–90.
- Kossinets, G., and Watts, D. 2009. Origins of homophily in an evolving social network. *American J. Sociology* 115(2):405–50.
- Leskovec, J.; Huttenlocher, D.; and Kleinberg, J. 2010. Signed networks in social media. *Proc. 28th ACM CHI Conf.*, 1361–1370.
- Rapoport, A. 1953. Spread of information through a population with socio-structural bias I: Assumption of transitivity. *Bulletin of Mathematical Biophysics* 15(4):523–533.
- Willer, D. 1999. *Network Exchange Theory*. Praeger.