The Impact of Network Structure on Breaking Ties in Online Social Networks: Unfollowing on Twitter

Funda Kivran-Swaine, Priya Govindan and Mor Naaman
Rutgers University
funda@rutgers.edu, priyagn@cs.rutgers.edu, mor@rutgers.edu

ABSTRACT
We investigate the breaking of ties between individuals in the online social network of Twitter, a hugely popular social media service. Building on sociology concepts such as strength of ties, embeddedness, and status, we explore how network structure alone influences *tie breaks* – the common phenomena of an individual ceasing to "follow" another in Twitter's directed social network. We examine these relationships using a dataset of 245,586 Twitter "follow" edges, and the persistence of these edges after nine months. We show that structural properties of individuals and dyads at Time 1 have a significant effect on the existence of edges at Time 2, and connect these findings to the social theories that motivated the study.

Author Keywords
Social networks, tie breaks, social media, Twitter.

ACM Classification Keywords
H.5.0 [Information interfaces and presentation]: General

General Terms
Human Factors

INTRODUCTION
Many of today's popular Web applications and social media services, like Facebook, Twitter, or LinkedIn, are built on sets of contacts articulated by each user of the application. This contact list often drives the majority of user's activity in the system (e.g., users of *Social Awareness Streams* [12] like Twitter and Facebook see the stream of updates from their contacts when they log in). These contacts reflect various types of relationships, including friendship, kinship, common interests, attention, or information exchange. The aggregates of contact lists in each of these services result in immense, articulated *online social networks*. As these services shift the communication and information fabric of our society, the dynamics of the networks they support are important to understand and reason about in depth.

In particular, the articulated online social networks in these services change and evolve as individuals form new ties, or break existing ties to others. These structural changes, observed over time, are far from random, and depend on various factors that affect the relationship between the users. Most work in the Computer Science and HCI fields has focused on the dynamics and models of tie *creation* [5,9,13]. In particular, researchers considered the *structural* aspects of the social network that predict formation of ties [5,10,15]. Here, we also focus on the social network structure, but examine breaking and persistence of existing ties, rather than tie creation.

The topic of breaking and persistence of ties is exceedingly important. Tie breaks impact the dynamics and activity in online services over time, and, as we show below, are common on services like Twitter. Moreover, the act of "de-friending" in online social systems might impact social relationships within and beyond the online world, in part due to the articulated nature of the network, where the ties can be observed, and their nonexistence could be detected.

While a number of sociology studies have examined tie persistence and decay [1,2,11], the topic has not received enough attention even in the sociology field [1]. This gap is perhaps due to the lack of longitudinal data, or the fact that unlike online ties, social ties often decay rather than "formally" break, making measurement of the phenomena harder. The initial findings from sociology research, however, suggest that the structural aspects of individuals' social networks are strong determinants for persistence and decay of ties [1,11]. We now have an opportunity to examine these theories in online social networks, and at a scale that was not available before. The dynamics of these online networks may help us, then, to reason about other types of networks and social ties, and discover patterns that may shed light on social phenomena in different contexts.

We use two snapshots of the Twitter social network to study the breaking of ties. Twitter is founded on an articulated online network, and allows users to read updates from others that they "follow". Thus, users create a *directed* social network that reflects attention and transfer of information [5,7,12,13]. We borrow from sociology theory to frame our investigation of tie breaks and network structures using the concepts of tie strength, embeddedness as well as power and status. Our large-scale analysis uses information from 245,586 dyadic relationships on Twitter at Time 1, and the persistence or break of edges in these dyads by Time 2. The analysis aims to answer the following research question: *What structural social network properties of nodes and dyads predict the breaking of ties*?

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

THEORETICAL BACKGROUND

We build on several key sociology concepts that are likely to influence tie persistence and decay: *tie strength*, *embeddedness within networks*, and *power and status.* These concepts guide our choice of variables and models to use in the investigation. After providing the theoretical background, we explain the measurements and variables available from the Twitter data. We then connect the theoretic constructs to the results of the data analysis. Our examination is conducted in the context of a *dyad*: a relationship between two individuals A and B in the social network. However, note that we are interested in a *directed edge* in this dyad: an edge that indicates that one of the individuals in the dyad follows another.

Tie Strength Measures

Sociologists developed the concept of strength of ties [6] between individuals as a measure of their relationship. Stronger ties, such as family and close friends, offer emotional support and intimacy, while weaker ties may allow for exchange of novel information, or, for example, help in finding jobs. Tie decay is likely to be slower between people who share a strong tie [2]. As there are some indications that tie strength can be measured in social media data [4, 7], we expect the edges in dyads that exhibit greater tie strength to be less likely to break over time.

Embeddedness, Equivalence and Transitivity

In sociology, the *embeddedness* of a tie can be defined by the set of relationships that exists between the individuals in a dyad, through third parties (i.e., common friends). Embeddedness was shown to be a strong predictor of a tie's persistence over time in offline settings [11]. In addition, homophily plays an important role in retention of ties in networks [1], and homophily is connected to increased embeddedness as it is likely to manifest itself in dense, more transitive networks [13]. Related concepts to embeddedness for directed networks (like Twitter's) are inlink and outlink equivalence [14]. Measures of such equivalence are the common inlinks, people with direct ties to both individuals in the dyad, and common outlinks, people to whom both individuals in the dyad have directed ties. The common inlinks or outlinks suggest mutual interests [5], and support tie preservation.

Finally, the concept of transitivity captures another type of relationship individuals in the dyad have through third parties. Transitivity suggests that when individual A is connected to X, and X connects to B, it is likely that A and B will also be connected. The *triadic closure* is the process that drives transitivity. In the case of directed networks, A→X (e.g., A follows X on the Twitter network) and X→B, leads via the triadic closure process to A→B. Transitivity and triadic closure have been shown [5,10,13] to impact *creation* of ties on Twitter, and are likely to play a role in tie persistence as well. In summary, we expect the edges in dyads exhibiting greater embeddedness, inlink and outlink equivalence, or transitivity through third parties, to be less likely to break.

Power and Status

Dyadic relationships may exhibit different power relations [3], for example, due to different status or relative importance of the individuals (e.g., boss and employee). Status difference has been shown to effect tie persistence in work settings [1]. Online social networks may include both extrinsic and intrinsic measures of status and power relationships. For example, a boss and employee may be connected in an online social network (extrinsic); but two individuals may also be compared according to *prestige*, a status intrinsic to the online network and the activities within it. We anticipate the edges in dyads that exhibit greater difference in status or prestige to be more likely to break. We also expect for an edge to be less likely to break if the target individual in the edge holds greater prestige.

DATA

The key elements of our Twitter data are a set of randomly-selected seed nodes S, and, for each node $s \in S$, the nodes $f \in F_s$ that follow s on the Twitter social network. The dyads we examine are therefore *(s, f)* pairs of seeds and followers, where we know that an edge $s \leftarrow f$ (f follows s) existed in the Twitter network.

We consider two snapshots of the Twitter social network, from July 2009 (*Time 1*), and April 2010 (*Time 2*). The Time 1 snapshot is derived from complete Twitter social network as collected by Kwak et al. [8]. This snapshot was used to extract the ego-centric network data for the seed nodes in S, and the followers F_s for each node. The Time 2 snapshot was collected directly using the Twitter API, and was used to identify whether $s \leftarrow f$ follow relationships from Time 1 still exist, or whether f "unfollowed" s by Time 2.

The set S of seed nodes is described in [12] and includes 911 Twitter users that were randomly selected in April 2009, and who were identified as active, personal users of Twitter, not representing commercial entities or celebrities. We retrieved Time 1 and Time 2 data for 746 of these users whose Twitter profile was public at Time 2, and further removed users who had more than 5,000 followers or followees, reflecting our interest in the set of individuals that use Twitter for personal reasons. The final sample included 715 seeds, with a total of N=245,586 followers (343 per seed, on average). Out of these edges, 30.6% were dropped over the 9 months between Time 1 and 2. Seed nodes, on average, lost 39% of their original connections (SD=15%) demonstrating that the tie break phenomenon is significant in Twitter's online social network.

Computed Properties

As mentioned above, we focused our analysis on dyads where a seed s is followed by follower f ($s \leftarrow f$) in Time 1. We calculated the network properties of the seed, of the follower, and of the dyad as detailed next, and as listed in Table 1. We then examined how these properties affect the existence of the edge $s \leftarrow f$ at Time 2.

The *seed properties* include features such as the number of followers, the ratio of followers to followees, and the *reciprocity rate*: the ratio of connections the seed reciprocated to their followers. We also computed the seed's *follow-back* rate: the ratio of nodes followed by the seed that reciprocate a connection to them. The *follower properties* include the number of followers of *f*, and *f*'s ratio of followers to followees. In both seed and follower properties, we did not include the number of followees in the analysis as it is highly correlated with the number of followers. We calculated a number of *dyadic properties* for *s* and *f*. Reciprocity is a binary indicator of whether the connection between *s* and *f* was reciprocated, i.e. if the edge $s \rightarrow f$ was also present in the network in Time 1. For each dyad, we also counted the number of transitive relationships the two nodes belonged to with any third node *X*. For example, we use Right Transitivity to denote the number of nodes *X* such that there is a path $s \rightarrow X \rightarrow f$ at Time 1. The dyadic properties also include the number of common friends, followers, and neighbors between *s* and *f*. Finally, using the follower's perspective we calculated a simple "prestige ratio" between the two nodes in the dyad: the ratio of the number of followers of *s* to that of *f*.

METHOD

We use multilevel logistic regression to study the influence of the Time 1 properties of the seed, follower, and dyad on the breaking of the follow edge between follower and seed by Time 2. Our data includes many dyads that share the same seed node. In addition, various seed properties are unaccounted for, and may have a significant impact on likelihood of tie breaks (e.g., how "interesting" the seeds updates are). Therefore, we use multilevel logistic regression model, in which dyads *(s,f)* (level-1 units) are nested within seed nodes *s* (level-2 units). We calculated per-seed random intercept in order to account for seed nodes' inherent propensity to generate tie breaks. The analysis follows Golder & Yardi that used similar methods to investigate tie formation in experimental settings [5]. Importantly, we assume that beyond seed-level tendency, tie break decisions by different followers of the same seed are independent of each other; this assumption is likely to hold on Twitter where tie breaks are hard to detect by third parties. Finally, following [5], some variables had a power-law distribution and were log-transformed. We create three models to analyze the influence of the characteristics of seed, follower, and dyads on breaking of each follower-seed edge (our dependent variable). With Model 1 we look at the contribution of seed characteristics for explaining the tie breaks. We add follower properties as independent variables to construct Model 2, and the dyadic characteristics to create Model 3. The follower count of *f* is excluded from Model 3 as it had little effect in Model 2, and is directly determined by the newly added Prestige Ratio and *s*'s followers count.

Results

The results are summarized in Table 1, reporting non-standardized coefficients, standard errors, and odds ratios for each model. Model 3 performed the best according to the Akaike's Information Criterion, and is our focus in the discussion below. Odds ratios are reported for easier interpretation of the magnitude of contribution of each variable. An odds ratio provides the increase in tie break odds for every one unit increase in the respective variable. For instance, odds ratio of 0.8 suggests that as the variable increases one unit, the odds of tie break decreases by 20% (e.g., from 1:2 to 1:2.5).

Predictors	Model 1		Model 2		Model 3	
Intercept	.80 *(.12)*	2.23 ***	1.63 *(.12)*	5.10 ***	.56 *(.13)*	.57 ***
Seed Properties						
^ Follower Count	-.30 *(.02)*	.74 ***	-.19 *(.02)*	.82 ***	.08 *(.02)*	1.08 ***
^ Ratio	-.09 *(.1)*	.91	-.08 *(.09)*	.92	-.07 *(.11)*	.93
^ Network Density	-.78 *(.07)*	.45 ***	-.55 *(.07)*	.57 ***	-.27 *(.08)*	.76 ***
Reciprocity Rate	-1.23 *(.19)*	.29 ***	-.70 *(.19)*	.49 ***	.50 *(.21)*	1.64 *
Follow-Back Rate	-.37 *(.16)*	.69 *	-.28 *(.15)*	.75 ◊	-.69 *(.17)*	.50 ***
Follower Properties						
^ Follower Count			-.06 *(.00)*	.94 ***	NA	NA
^ Ratio			-2.18 *(.02)*	.11 ***	-.81 *(.02)*	.44 ***
Edge (Dyad) Properties						
✦ Reciprocity					-.79 *(.01)*	.45 ***
^ Common Neighbors					-.33 *(.00)*	.71 ***
^ Common Followers					.05 *(.00)*	1.05 ***
^ Common Friends					.04 *(.00)*	1.04 ***
^ Right Transitivity					-.08 *(.00)*	.92 ***
^ Left Transitivity					-.03 *(.00)*	.97 ***
^ Mutual Transitivity					-.05 *(.00)*	.95 ***
^ Prestige Ratio					.06 *(.00)*	1.06 ***
Intercept Variance	.19 (SD=.44)		.17 (SD=.40)		.21 (SD=.46)	

◊ p<.10 * p<.05 ** p<.01 *** p<.001 (Standard errors are shown in parentheses.)
^ - variable was log-normalized ✦ - binary variable

Table 1. Effects (standardized coefficients and odds ratio) of seed, follower and dyadic properties on edge break (unfollow).

A number of variables had a negative effect on tie breaks in our results. Notably, the binary indicator for reciprocity had a large significant negative effect on tie breaks, the odds ratio indicating that *s* following back *f* decreases the odds of the breaking of the edge from *f* to *s* by 55%. (In the raw data, the break rate for reciprocated ties was 16.4%, while as many as 44.5% of unreciprocated ties broke by Time 2). The seed's follow-back rate also contributed significantly and negatively to all models, as well as the follower's follower-to-followee ratio. Further major negative effects were found for the seed's network density and the dyad's number of common neighbors.

The variables that had a positive affect on tie breaks included the seed's number of followers (the more followers *s* has, the more likely an edge to s is to be broken). However, that effect was minor and reversed from Model 1 and 2, suggesting that a negative effect of the number of followers might be mitigated by some of the dyadic features. Similarly, in Model 3 the seed's reciprocity rate had significant positive effect on tie break, reversed from Model 1 and 2, as the dyad-level reciprocity explains away reciprocity rate's negative effect. We discuss next what these findings suggest for our theoretical background.

DISCUSSION

Our results provide initial evidence that tie strength, status and embeddedness influence the breaking of ties.

A number of cues in our model illustrate the positive effect of tie strength on tie persistence. The large and negative effect of the dyadic reciprocity variable on tie break indicates that the reciprocity in dyadic relationship, which may indicate a stronger tie, is one of the most important factors for the persistence of the follower's edge. (We note, though, that while reciprocity *may* indicate strength of a tie [6], it is not a fully convincing measure in online networks [4, 7].) In addition, the results show that the odds of tie break decreases as *s*'s network gets denser. High density indicates intimate, tightly-knit networks, and ties between individuals in denser network are more likely to be strong.

Several indicators illustrate the connection between power and status, and breaking of ties. The prestige ratio between the follower and the seed contributes positively to the likelihood of a broken edge, pointing that the more "important" a follower is than the seed, the more likely the follower is to stop following that seed. The Follow-back rate is a potential measure of seed's status (in that users reciprocate connections to them), and its significant negative contribution in the final model suggests that the higher status of *s* mitigates tie break, even after controlling for *s*'s network size. Indeed, the negative effect of the follow-back rate on tie breaks may hint that this measure is a solid marker of status on Twitter. The two follower-to-followee ratio measures (for *s* and *f*) can also be considered proxies for status. While *s*'s follower-to-followee ratio did not significantly contribute to the model, the same ratio for *f* had a significant large negative effect, indicating that the higher the "status" of *f*, the less likely *f* is to break a tie, once he/she is already following *s*. Finally, a lack of a reciprocated edge to $f \leftarrow s$ may imply that *s* is in a more powerful location in the network than *f* (*s* may be an information source, or an authoritative figure). As reciprocity may be socially expected for relationships in Twitter, this imbalance might have even stronger consequences. The reciprocity effect, therefore, may also suggest that tie breaks are influenced by status differences between individuals in the dyad.

The influence of embeddedness on tie persistence is primarily marked by the effect of the common neighbors: the more common neighbors a dyad shares, the less likely the follow edge is to break. However, the effect of inlink and outlink equivalence is not quite evident, with minor and conflicting effects for the number of common followers and friends of the seed and follower. This outcome is perhaps due to the relationships between the equivalence measure variables. The minor, but significant negative effect of the transitivity variables may be because of other factors such as network density and reciprocity being more influential over tie breaks, and thus eliminating the effect of the various embeddedness and transitivity dyadic measures.

SUMMARY AND CONCLUSIONS

We have shown that network structure alone can significantly help explain the breaking and persistence of ties in Twitter's online social network. We connected these findings to sociological concepts that may be at the root of the tie break phenomenon, such as strength of ties and status. In future work, we would like to further explore the phenomenon of tie breaks using additional variables such as content characteristics, for a better operationalization of the sociological concepts. (We may even do a qualitative study!) At the same time, our findings already provide insights for systems or applications that build on online social networks, which account for a large proportion of activity on the Web today.

ACKNOWLEDGEMENTS

We would like to thank Scott Golder for his help and feedback. We also thank Yotam Margalit for support that cannot be quantitatively measured. This work is partially supported by a Google Research Award (thanks, Google).

REFERENCES

1. Burt R.S. Decay functions. *Soc. Networks* 22:1–28. 2000.
2. Burt, R.S. Bridge decay. *Soc. Networks* 24,333–63. 2002.
3. Emerson, R.M. Power-dependence relations. *American Sociological Review*, 27(1):31–41. 1962.
4. Gilbert, E. and Karahalios, K. Predicting tie strength with social media. *Proc., CHI 2009*.
5. Golder, S. A. and Yardi, S. Structural predictors of tie formation in twitter: Transitivity and mutuality. *Proc., IEEE Social Computing 2010*.
6. Granovetter, M. S. The strength of weak ties. *American Journal of Sociology,* 78(6):1360–1380. 1973.
7. Huberman, B., Romero, D. M., and Wu, F. Social networks that matter: Twitter under the microscope. *First Monday*, 14(1). 2009.
8. Kwak, H., Lee, C., Park, H., and Moon, S. What is Twitter, a social network or a news media? *Proc., WWW 2010*.
9. Leskovec, J., Huttenlocher, D., and Kleinberg, J. Signed networks in social media. *Proc., CHI 2010*.
10. Liben-Nowell, D. and Kleinberg, J. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.*, 58(7):1019–1031. 2007.
11. Martin, J. L. and Yeung, K.-T. Persistence of close personal ties over a 12-year period. *Soc. Networks* 28:331–362. 2006.
12. Naaman, M., Boase, J., and Lai, C.-H. Is it really about me? Message content in social awareness streams. *Proc., CSCW 2010*.
13. Romero, D.M., and Kleinberg, J. The directed closure process in information networks with an analysis of link formation on twitter. *Proc., ICWSM 2010.*
14. Wasserman, S. and Faust, K. Social network analysis: Methods and applications. Cambridge Univ. Press. 1994
15. Zheleva, E., Getoor, L., Golbeck, J., and Kuter, U. Using friendship ties and family circles for link prediction. *Proc., SNA-KDD 2008*.