Status and Friendship: Mechanisms of Social Network Evolution

Christina Brandt*
Stanford University
cbrandt@cs.stanford.edu

Jure Leskovec
Stanford University
jure@cs.stanford.edu

ABSTRACT

We examine the evolution of five social networking sites where complex networks of social relationships developed: Twitter, Flickr, DeviantArt, Delicious, and Yahoo! Answers. We study the differences and similarities in edge creation mechanisms in these social networks. We find large differences in edge reciprocation rates and overall structure of the underlying networks. We demonstrate that two mechanisms can explain these disparities: directed triadic closure, which leads to networks that show characteristics of status-oriented behavior, and reciprocation, which leads to friendship-oriented behavior. We develop a model that demonstrates how variances in these mechanisms lead to characteristic differences in the expression of network subgraph motifs. Lastly, we show how a user's future popularity, her indegree, can be predicted based on her initial edge creation behavior.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database applications—*Data mining*
General Terms: Experimentation, Measurement.
Keywords: Network evolution; triadic closure; reciprocation.

1. INTRODUCTION

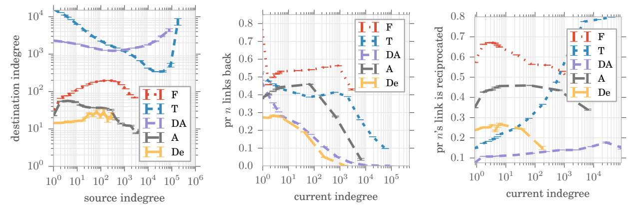
As people use social networking services, an intricate network of complex social relationships between the users of these service develops. A natural question, then, is: *In what aspects do various online social networks differ and why?* Here we tackle this question by examining mechanisms of link creation in five different social networks with the goal of understanding the roles that different social mechanisms play in the formation of the underlying networks.

In contrast to most work on graph evolution, we study networks as *directed, evolving* graphs, which allows us to study the *context* of link formation in which relationships between users, expressed as edge formation events, occur. Studying networks as directed allows us to account for asymmetries in relationships between nodes as well as consider finer details of triadic closure.

We organize our investigations as follows: First, we study how node's degree and local network structure (characterized via small induced subgraphs, i.e., network motifs) affects its edge creation and edge reciprocation behavior. Second, we build a network formation model based on *friendship-oriented* behavior and *status-oriented* behavior, and we show that variability in network motif creation can be explained by these two simple mechanisms. Last, we show how node's local network structure can predict node's future prominence. We find that friendship- and status-oriented linking behaviors provide insight into the evolving structures of social networks.

*Supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1147470.

Copyright is held by the author/owner(s).
WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.257732.



(a) d_{in}(link recipient) vs. $\overline{d}_{in}$(link initiator) (b) $d_{in}(n)$ vs. pr(n reciprocates) (c) $d_{in}(n)$ vs. pr(reciprocation to n)

Figure 1: Edge creation and reciprocation with node's indegree.

2. EDGE FORMATION IN CONTEXT

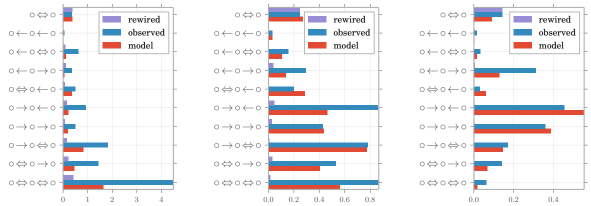
We examine the directed temporal dynamics of five widely used online social networks: photo-sharing Flickr (photo-sharing), DeviantArt (art-sharing), Twitter (microblogging), Delicious (link-sharing), and Yahoo!Answers (Q&A). All networks allow users to create *directed* edges through which content can be consumed by "followers." For each network we have access to the timestamps of edge creation events. Further description of networks and their properties can be found in the extended version of the paper [1].

Edge creation and reciprocation. We found differences in reciprocation and edge clustering that cannot be explained solely by differences in networks' average degree [1]. Thus we examine linking behavior as a function of user's indegree, which can be considered a proxy for her *status* in the community [2] We find that the behavior of DeviantArt (DA) in Figure 1 can be explained by a status factor: low-status users create links to high-status users (1a), but as the status of a node grows, it tends to reciprocate its edges less and less (1b) while the links it creates tend to be reciprocated more (1c). Flickr (F) is the opposite: high-status nodes tend to reciprocate links, and edges initiated by low-status nodes are usually reciprocated. Twitter (T) falls between these two extremes.

Motif formation. A node's local network structure provides a skeleton for edge formation as neighboring nodes act as 'facilitators' for edge creation. Table 1 illustrates a subset of possible network contexts (*network motifs*) just before the dashed edge is created. We observe (see [1] for details) that, consistent with the status theory, directed transitive links (○→○→○) tend to be created to high-indegree nodes. In DA and Twitter, the sibling motif (○→○←○) also conforms to a status hypothesis, while for Flickr, the motif ap-

mutual-followed		feed-forward transitive	
sibling		mutual-friends	

Table 1: Subset of edge formation motifs. Dashed arrow indicates the newly created edge. Double lines denote reciprocated links.



(a) Flickr (b) Twitter (c) DA

Figure 2: Comparison between model motif distribution, observed motif distribution, and random rewiring motif distribution.

pears to be due to high node activity. While uncommon in DA, the mutual-friend motif (○⇔○⇔○) was common in Flickr and tended to be initiated by higher-status nodes. Since the mutual-followed motif (○←○→○) occurred primarily between high-indegree nodes and may be primarily due to node indegree. Our results suggest that motifs such as sibling, the follower-of-friend, and mutual-followed may be natural consequences of status-based links.

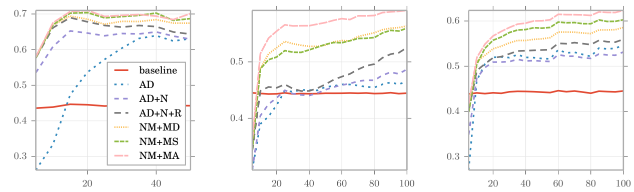
3. MODELING EDGE FORMATION

Our analyses [1] suggest that link formation may be caused by two competing mechanisms: *friendship*, seen as edge reciprocation, and *status*, visible as feed-forward transitive links. To test this, we define a simple conceptual model $(S, \alpha, \beta, \gamma)$ of network formation.

Model definition. Let $S = \{(u, t)\}$ be the schedule of link initiations, where (u, t) denotes that u creates an edge at time t. Parameter α controls the degree of reciprocation, β the feed-forward triadic closure, and γ the random linking behavior. Node u wakes up at a specified time t according to schedule S and creates an edge via the following process (WPR indicates 'with probability'):

- WPR β, node u creates a feed-forward transitive motif by choosing one of its outlinks v, choosing one of v's outlinks w, then creating a link $u \rightarrow w$.
- WPR$(1-\beta)$, u chooses a node r from the network.
  - WPR$1-\gamma$, u stops and constructs a link $u \rightarrow r$
  - Else u chooses one of r's outlinks s and links $u \rightarrow s$.
- Independent of link type, the edge is reciprocated WPR α.

Experimental setup and results. We fit our model to the three larger networks: Flickr, Twitter, and DA [1]. We estimate parameters via coarse-grained grid search, minimizing squared error over motifs. Figure 2 gives a comparison of motif occurrences. Motifs such as sibling and follower-of-friend emerge naturally in our model, but are not simply a product of degree or reciprocation.We investigate the parameter values over various metrics. Both Twitter and DA have high values of γ, which leads to high-indegree "celebrity" users. Combined with low β, this leads to behavior in DA similar in spirit to [3] and [4] and can be shown using mean field analysis to result in preferential attachment [1]. Flickr and Twitter have high β, but in Flickr, reciprocation leads to high clustering, while in Twitter the transitivity leads to status-oriented network. (Exact parameter values can be found in [1].) To better handle reciprocal edges, the model can also be extended by adding a parameter $\rho > 1$ so that the probability of choosing bilateral edges in the search process is increased $\propto \rho$.



(a) Flickr (b) Twitter (c) DA

Figure 3: k vs F1-score of indegree prediction.

4. PREDICTING FUTURE PROMINENCE

Since motifs help to characterize network evolution, they may be useful in predicting the node's future behavior. Given the initial k edges received by the node, we seek to predict whether a node will be 'popular"/"unpopular" (high/low future indegree) in the future.

Features. Rather than maximizing performance, our goal is to examine how structural features and mechanisms of edge formation indicate future prominence in the community. When each node u receives its k^{th} incoming edge, we compute following features: **(AD)** *Age and degree*: captures node's temporal activity, e.g. age, birth, outdegree, and rate to in-/out-degrees $k_0 < k$. **(N)** *Neighbor status and activity*: follower/friend in-/out-degree. **(R)** *Reciprocation*: reciprocated edges initiated by u, edges u did/did not reciprocate, etc. **(NM)** *Non-motif*: union of AD, N, and R. **(MS)** *Motif-Source*: capture the way u links to other nodes via motifs (Table 1) of u's outgoing edges **(MD)** *Motif-Destination*: motifs created with u as an edge destination. **(MA)** *Motif-All*: union of MS and MD.

Experimental setup and results. Given a node's situation after its k incoming edge, we aim to predict whether it will have high/low indegree at the end of a timespan L (90 120, and 180 days for Flickr, Twitter, and DA, respectively). Indegrees at 80th and 50th percentile, to indicate "high" and "low." We use logistic regression for the learning task and compare the results over 10-fold cross-validation. Figure 3 shows how the relative $F1$ values vary with k, as well as a no-knowledge prediction baseline. The F1 score is around 0.6-0.7, which means that network structure (not activity) provides signal for future user behavior.

In all cases, the full feature set (NM+MA) performs best over all k. In DA, the ordering and gaps between the results are maintained as k increases. In Flickr, non-motif features become more useful at higher degrees, but the relative importance of motif-based features also increases. In Twitter, the gap between motif- and nonmotif-features eventually peaks, while reciprocation-based features are increasingly important at high k. The motif features improve precision in Flickr and Twitter, and in all cases, they improve recall. This means that in DA, while degree features successfully identify a few popular nodes, especially at high values of k, motif features allow the algorithm to distinguish many more.

5. REFERENCES

[1] For details, refer to the extended version on authors' websites.
[2] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec. Effects of user similarity in social media. In *WSDM '12*, 2012.
[3] M. O. Jackson and B. W. Rogers. Meeting strangers and friends of friends: How random are social networks? *The American economic review*, pages 890–915, 2007.
[4] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins. *Microscopic Evolution of Social Networks*. 2008.