3-2011

Modeling link formation behaviors in dynamic social networks

Viet-An NGUYEN
Singapore Management University

Cane Wing-Ki LEUNG
Singapore Management University

Ee Peng LIM
Singapore Management University, eplim@smu.edu.sg

DOI: https://doi.org/10.1007/978-3-642-19656-0_48

Modeling Link Formation Behaviors in Dynamic Social Networks

Viet-An Nguyen[1], Cane Wing-Ki Leung[2], and Ee-Peng Lim[2]

[1] Department of Computer Science, University of Maryland, College Park, MD, USA
[2] School of Information Systems, Singapore Management University, Singapore

Abstract. Online social networks are dynamic in nature. While links between users are seemingly formed and removed randomly, there exists some interested link formation behaviors demonstrated by users performing link creation and removal activities. Uncovering these behaviors not only allows us to gain deep insights of the users, but also pave the way to decipher how social links are formed. In this paper, we propose a general framework to define user link formation behaviors using well studied local link structures (i.e., triads and dyads) in a dynamic social network where links are formed at different timestamps. Depending on the role a user plays in a link structure, we derive different types of *link formation behaviors*. We develop models for these behaviors and measure them for a set of users in an Epinions dataset.

Keywords: Link formation behaviors, link formation rules.

1 Introduction

Social links represent a rich set of structural knowledge about the linked users beyond their individual attributes (e.g., age, gender, etc.). At the network level, network properties such as density, diameter, user degree distribution, etc., have been well studied by both social and computer science researchers[6,3]. At the micro or node level, the social links reveal a user's preference of friends, and her preferred way to form social links with others. Links can be further utilized in a number of commercially interesting applications including item recommendation, information diffusion and community discovery.

In this paper, we examine link formation behaviors of each user and derive some models for measuring the behaviors. The link formation behaviors here are motivated by a set of dyadic and triadic pattern rules discovered from dynamic social network data. We define a *dynamic directed social network* to be a graph $G = (V, E, T, t)$. V is a set of vertices/nodes representing individuals in the network. E is a set of directed edges representing social links, such as friendship and trust links, between individuals. An element $(v_i, v_j) \in E$, where $v_i, v_j \in V$, is an edge from v_i to v_j. $t : E \rightarrow T$ is a mapping between edges and their timestamps. Without loss of generality, we represent timestamps as $T = \{t_h | h \geq 0\}$, such that $\forall t_{h_1}, t_{h_2} \in T, t_{h_1} < t_{h_2}$ *iff* $h_1 < h_2$. A graph may evolve with new

J. Salerno et al. (Eds.): SBP 2011, LNCS 6589, pp. 349–357, 2011.

nodes and edges joining at different time points. G can be viewed as a snapshot of a social network taken at a certain time point t_h, such that G contains nodes and edges that were formed at or before t_h. For simplicity, we assume that nodes and edges are not removed after creation.

We first define a representation for *link formation (LF) rules* (see Section 3). Unlike most earlier link patterns studied, we require each link formation rule to have pre-conditional link structure formed before the consequence link. For link formation rule, we then derive link formation behaviors based on the way a user forms links with other users in Section 4. In particular, we introduce several link formation behaviors including *rule usage* and *rule confidence* associated with different link formation rules. Such behaviors can be defined at both node (user) and instance levels. Here, an instance refers to a subgraph of users and their links following the rule structure. We assign a score function to each behavior so as to measure it quantitatively. We finally apply our proposed behaviors on the Epinions data (see Section 5).

2 Related Work

There are very little work on link formation behavior modeling and analysis for dynamic social networks. Most research in the past studied known dyadic and triadic structures in static social networks[10]. Recently, these local structures are extended with time order when analyzing dynamic social networks[2,8,7]. There are also new research on mining network specific local structures[4].

Leskovec, et al. studied the "edge destination selection process" based on some simple triangle closing models[2]. The networks they have considered are temporal but undirected networks, and the models introduced do not consider individual's link formation behaviors. Romero and Kleinberg, in a recent paper [8], showed that directed closure (also known as transitivity) is used by Twitter users in forming links with one another. A measure known as *closure ratio* was introduced to measure how likely the incoming links of a user node exhibits directed closure. Instead of examining links formed by transitivity only, this paper covers behaviors related to a variety of link formation rules. In our earlier paper [7], we introduced trust reciprocity related behaviors that are shown to help predicting reciprocal trust links. This paper enlarges the behavioral study to include other link formation rules.

3 Link Formation Rules

Link Formation rule (LF-rule) is designed to describe an *observable social effect* facilitating or affecting the formation of a link from a certain node v_i to another node v_j, where v_i and v_j are called the *start* node and the *end* node of a LF-rule respectively. A LF-rule captures two important constraints. The first constraint is a certain *pre-condition* link structure related to the start node and/or the end node. The second constraint is the *temporal constraint* that the link from the start node to the end node must be formed after the pre-condition is formed.

Our LF-rules consist of *dyadic and triadic structures* which have long been recognized as interesting structures for understanding and predicting the dynamics of large, complex networks [5,1,9,8]. Figure 1 depicts five LF-rules (labeled r_1 to r_5) constructed from basic dyadic and triadic structures, which are the building blocks of local structures. In each LF-rule, the nodes labeled as s and e respectively correspond to its start node and end node. Recall that the (s, e) link (shown in blue in Figure 1) in a LF-rule is required to be formed later than other links in the same rule.

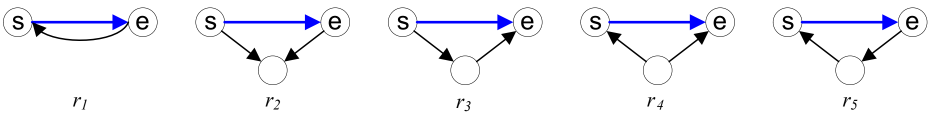


Fig. 1. LF-rules based on basic dyads and triads, including reciprocity (r_1), common out-neighbor (r_2), transitivity (r_3), common in-neighbor (r_4) and cycle (r_5)

Hence, the LF-rule set $\mathcal{R}$ in this study consists of the above five rules, i.e., $\mathcal{R} = \{r_1, \cdots, r_5\}$. As we are interested in link formation behaviors caused by local structural effect instead of some non-local (e.g., preferential attachment) and out-of-network (e.g., users already know each other before joining the network) effects, we focus on links formed with users within two-hop distance. Hence, users within two-hop distance apart must satisfy the pre-condition(s) of at least one of the given rules.

4 Link Formation Behaviors

Given the LF-rule set $\mathcal{R}$, we characterize individual nodes by their Link Formation behaviors (LF-behaviors), which describe the extent to which the nodes follow specific LF-rules in forming links. Consider that v_i takes on the role of start node, we can derive two LF-behaviors that corresponds to the usage and confidence of a rule r_l. We therefore define the following two kinds of LF-behaviors: *rule usage* and *rule confidence*.

To define the rule usage and confidence as measurable behaviors, we first introduce some important notations. Given an input dynamic social network $G = (V, E, T, t)$, an *instance* of the LF-rule r_l is a subgraph in G that: (a) is isomorphic to the graph of r_l; and (b) has the consequence edge formed after the pre-condition link structure. The set of instances of r_l with v_i and v_j taking the start node and end node roles respectively is known as the *instance set* of (v_i, v_j) w.r.t. r_l, and is denoted by $\mathbf{R}^l_{ij}$. The instance set of start node v_i w.r.t. r_l, $\mathbf{R}^l_{i*}$, is defined as $\cup_j \mathbf{R}^l_{ij}$. The node set of v_i w.r.t. r_l, $\mathbf{UR}^l_{i*}$, is defined as $\{v_j | \mathbf{R}^l_{ij} \neq \phi\}$. We define the instance of the pre-condition of r_l as a subgraph in G that is isomorphic to the pre-condition link structure of r_l. The set of instances of pre-condition of r_l with v_i and v_j taking the start node and end node roles respectively is known as the *pre-instance set* of (v_i, v_j) w.r.t. r_l, and is denoted by $\mathbf{P}^l_{ij}$. The pre-instance set of v_i w.r.t. r_l $\mathbf{P}^l_{i*}$ is therefore defined as $\cup_j \mathbf{P}^l_{ij}$. The pre-node set of v_i w.r.t. r_l, $\mathbf{UP}^l_{i*}$, is defined as $\{v_j | \mathbf{P}^l_{ij} \neq \phi\}$.

We define the *rule usage at node level*, *rule usage at instance level*, *rule confidence at node level*, *rule confidence at instance level* behaviors of a user v_i w.r.t. rule r_l in Equations 1, 2, 3, and 4 respectively.

$$\mathbf{NUsage}^l_{i*} = |\mathbf{UR}^l_{i*}| / \sum_{r_k \in \mathcal{R}} |\mathbf{UR}^k_{i*}| \quad (1) \qquad \mathbf{Usage}^l_{i*} = |\mathbf{R}^l_{i*}| / \sum_{r_k \in \mathcal{R}} |\mathbf{R}^k_{i*}| \quad (2)$$

$$\mathbf{NConf}^l_{i*} = |\mathbf{UR}^l_{i*}| / |\mathbf{UP}^l_{i*}| \quad (3) \qquad \mathbf{Conf}^l_{i*} = \sum_j |\mathbf{R}^l_{ij}| / \sum_j |\mathbf{P}^l_{ij}| \quad (4)$$

The rule usage at node level in Eq 1 reveals the proportion of users v_i has links to are based on rule r_l. Note that it is possible that v_i may link to a user v_j using different link formation rules, i.e., $\mathbf{UR}^l_{ij} \cap \mathbf{UR}^k_{ij} \neq \phi$ for $r_l \neq r_k$. The rule usage at instance level in Eq 2 measures the proportion of v_i's rule instances are based on rule r_l. Multiple rule instances based on a rule can be associated with a link from v_i to another user v_j, as there can be multiple instances of the rule's pre-condition occurring before the (v_i, v_j) link.

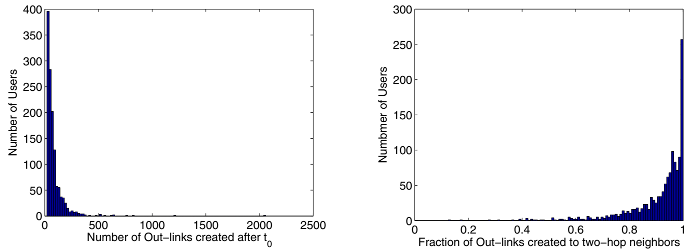
The rule confidence at node level in Eq 3 measures the proportion of users v_i has connections to based on rule r_l are subsequently linked from v_i directly. The rule confidence at instance level in Eq 4 measures the proportion of v_i's pre-condition instances based on rule r_l are subsequently linked from v_i directly. $\mathbf{NConf}^l_{i*}$ may not be equal to $\mathbf{Conf}^l_{i*}$ except for reciprocity rule which has one pre-condition user instance corresponding to one pre-condition graph instance.

5 Link Formation Behavior Analysis of Web of Trust

5.1 Dataset

We conduct an experimental study on the *Epinions* dataset available at `http://www.trustlet.org/wiki/Extended_Epinions_dataset`. Epinions contains a directed and time-stamped Web of Trust (WOT) with *trust* edges. About 69% of edges come with an initial timestamp of 2001/01/10 (t_0), which represents *all timestamps on t_0 or prior to t_0*. The formation date and order of all edges formed after t_0 are known. As temporal information is important in characterizing LF-behaviors, we discarded an edge (v_i, v_j) with timestamp t_0 unless both users v_i and v_j were involved in at least one edge formed after t_0. As we are interested in users with sufficient link formation history, we require each user v_i to creating at least 20 out-links and in-links. There are 1295 users meeting this selection criteria.

Figure 2 shows the distribution of outdegree of our users. Note that these out-links may be created to users more than two hops away but such links are the minority. On average, more than 90% of the out-links are directed to users within two-hop away which is shown in Figure 2(b). This also illustrates the dominance of local structural effect in forming links.

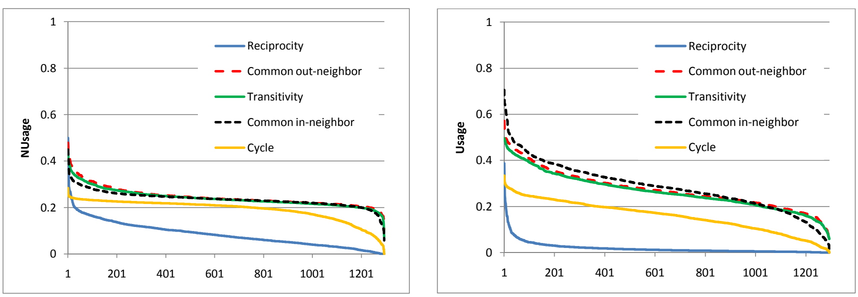


(a) Outdegree Distribution (b) Fraction of Outlinks created to two-hop neighbors

Fig. 2. Statistics of Outlinks

5.2 Link Formation Behavior Distribution

We first examine the distribution of LF-behavior scores among users. Figure 3 shows the rule usage behaviors scores of all users ordered from highest to lowest score values. The values on the x axis are the rank positions of users. For different rules, the same rank may refer to different users.

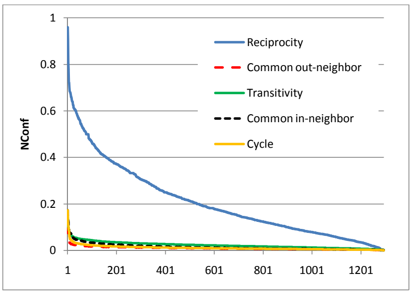


(a) At Node Level (b) At Instance Level

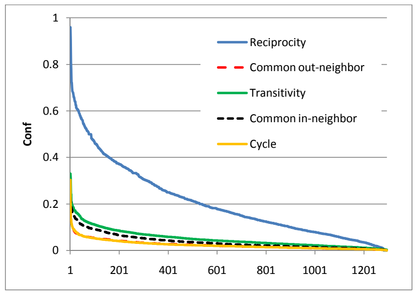
Fig. 3. Distribution of Rule Usage Behaviors

Figure 3 shows that common out-neighbor, transitivity, and common in-neighbor are the rules mostly used among the users. This is followed by cycle and reciprocity rules. Based on node level rule usage behavior values in Figure 3(a), most users have about 24% of their links formed involving each of these three rules. On average, users have only 18% and 8% of their links formed using cycle and reciprocity respectively. Similar observations can be made for the instance level rule usage behaviors (see Figure 3(b)) except that the proportions of using reciprocity and cycle are even lower. These behavior scores suggest that users may either have less distinctive opportunity or confidence to use the reciprocity and cycle rules. The exact reasons are revealed in Figure 4.

Figure 4 shows that users have relatively higher confidence for using reciprocity rule compared with other rules. On average, users have 19% confidence for this rule but not more than 5% confidence for other rules at the node and instance levels. With this, we conclude that users have low usage for reciprocity because of less distinctive opportunity instead of low confidence. While users may have more links formed with common out-neighbor, transitivity, and common in-neighbor rule, their confidence of using these rules is actually very low.



(a) At Node Level



(b) At Instance Level

Fig. 4. Distribution of Rule Confidence Behaviors

5.3 Node Level vs. Instance Level Behaviors

Figures 5 and 6 depict the scatterplots of users' behavior scores at node and instance levels. Each figure has x and y axes representing the node and instance level behaviors respectively, and each point represents a user's node and instance level behavior scores. From the two figures, some interesting observations are: (1) The behavior scores at node and instance levels are well correlated. Hence, when a rule is used often for linked users, it is likely that the rule is also used often for the link formation structures involving the linked users; (2) Except for reciprocity, the behavior scores at node level tend to be smaller that the corresponding scores at instance level. This shows that for triadic rules, users usually form multiple pre-condition instances before actually forming the direct link.

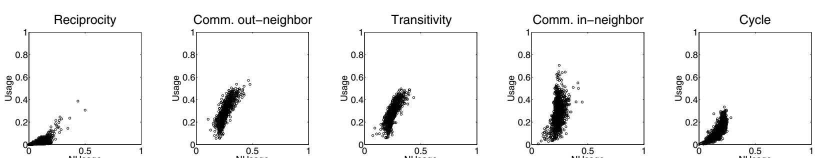


Fig. 5. Correlation of Node and Instance Level Rule Usage

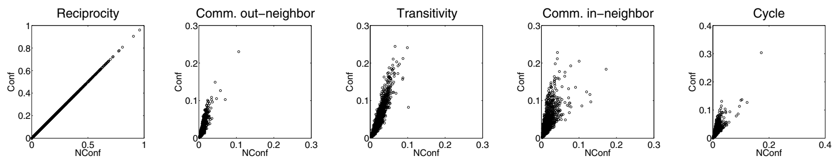
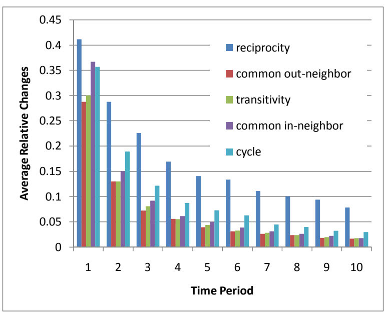


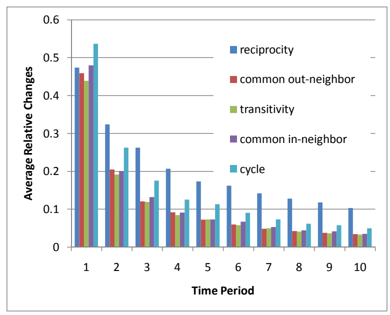
Fig. 6. Correlation of Node and Instance Level Rule Confidence

5.4 Stability of User Behaviors

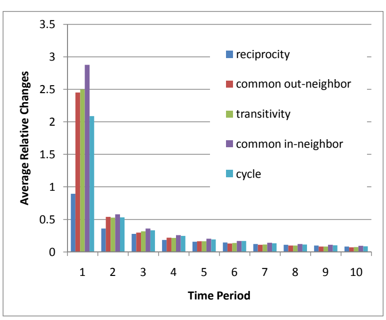
Next we would like to examine the stability of user behaviors over time. The rule usage and rule confidence of each user defined in Equations 1 to 4 can be computed at different time points where $\mathbf{UR}^l_{i*}$, $\mathbf{UP}^l_{i*}$, $\mathbf{R}^l_{i*}$ and $\mathbf{P}^l_{i*}$ contain the corresponding set of users and instances up to the given time. For simplicity, we compute for each user his/her behavior scores every time the user creates an out-link (instead of for every possible time point). Thus, given a user for each type of



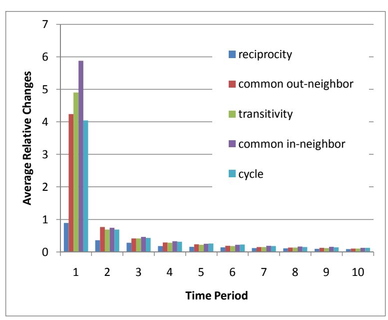
(a) Node level rule usage



(b) Instance level rule usage



(c) Node level rule conf.



(d) Instance level rule conf.

Fig. 7. Relative changes of different behaviors averaged over all users

behavior there is a time series of scores. Given a series of user behavior scores, we split it into 10 equal partitions and obtain a sub-series of scores at each splitting time. Each series of behavior scores of a user now is characterized by 11 representative values $x_0, x_1, \cdots, x_{10}$.

For each partition $i \in [1, 10]$ we compute the relative change as $\frac{x_i - x_{i-1}}{x_{i-1}}$. Figure 7 shows the average relative changes over all users for different behavior scores. As we can see, the average relative changes of all behavior scores decrease as time increases which shows that the users tend to follow the behaviors defined by our rules more consistently over time. In the case of rule confidence, the stability is even more obvious after the 1st partition. Thus, our proposed behavior scores can be used to effectively characterize how users form links in dynamic social networks which is useful in various tasks including user clustering and link prediction.

6 Conclusion

In this paper, we introduce a set of link formation behaviors derived from a set of link formation rules for dynamic social networks. Using an Epinions dataset, we show that active users have their links formed using rules with different rule usage and confidence behaviors. Reciprocity rule may not enjoy high usage but most users have relatively higher confidence using it. The node and instance level behaviors are found to be correlated. We further find the behaviors become more stable as users establish more links. With this knowledge, we believe that link behaviors can be a part of user profile which can be useful in social network applications such as link prediction and recommendation. This will also be the direction for our future research.

Acknowledgement

This work is supported by Singapore's National Research Foundation's research grant, NRF2008IDM-IDM004-036.

References

1. Faust, K.: Very local structure in social networks. Sociological Methodology 31(1), 209–256 (2007)
2. Leskovec, J., Backstrom, L., Kumar, R., Tomkins, A.: Microscopic evolution of social networks. In: SIGKDD (2008)
3. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graphs over time: densification laws, shrinking diameters and possible explanations. In: SIGKDD (2005)
4. Leung, C.W.-K., Lim, E.-P., Lo, D., Weng, J.: Mining interesting link formation rules in social networks. In: CIKM (2010)
5. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: Simple building blocks of complex networks. Science 298(5594), 824–827 (2002)
6. Newman, M.E.J.: The structure and function of complex networks. SIAM Review 45, 167–256 (2003)

7. Nguyen, V.-A., Lim, E.-P., Tan, H.-H., Jiang, J., Sun, A.: Do you trust to get trust? A study of trust reciprocity behaviors and reciprocal trust prediction. In: SIAM SDM (2010)
8. Romero, D.M., Kleinberg, J.: The directed closure process in hybrid social-information networks, with an analysis of link formation on Twitter. In: ICWSM (2010)
9. Snijders, T.A.B., Van de Bunt, G.G., Steglich, C.E.G.: Introduction to stochastic actor-based models for network dynamics. Social Networks, Special Issue on Dynamics of Social Networks 32(1), 44–60 (2010)
10. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)