Who Will Follow You Back? Reciprocal Relationship Prediction*

John Hopcroft
Department of Computer Science
Cornell University
Ithaca NY 14853
jeh@cs.cornell.edu

Tiancheng Lou
Institute for Interdisciplinary Information Sciences
Tsinghua University
Beijing 100084, China
ltc08@tsinghua.edu.cn

Jie Tang
Department of Computer Science
Tsinghua University
Beijing 100084, China
jietang@tsinghua.edu.cn

ABSTRACT

We study the extent to which the formation of a two-way relationship can be predicted in a dynamic social network. A two-way (called reciprocal) relationship, usually developed from a one-way (parasocial) relationship, represents a more trustful relationship between people. Understanding the formation of two-way relationships can provide us insights into the micro-level dynamics of the social network, such as what is the underlying community structure and how users influence each other.

Employing Twitter as a source for our experimental data, we propose a learning framework to formulate the problem of reciprocal relationship prediction into a graphical model. The framework incorporates social theories into a machine learning model. We demonstrate that it is possible to accurately infer 90% of reciprocal relationships in a dynamic network. Our study provides strong evidence of the existence of the structural balance among reciprocal relationships. In addition, we have some interesting findings, e.g., the likelihood of two "elite" users creating a reciprocal relationships is nearly 8 times higher than the likelihood of two ordinary users. More importantly, our findings have potential implications such as how social structures can be inferred from individuals' behaviors.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Data Mining; J.4 [**Social and Behavioral Sciences**]: Miscellaneous; H.4.m [**Information Systems**]: Miscellaneous

General Terms

Algorithms, Experimentation

Keywords

social network, reciprocal relationship, social influence, predictive model, link prediction, Twitter

*Authors are in alphabetic order. The work were done when the last two authors were visiting Cornell University.

1. INTRODUCTION

Online social networks (e.g., Twitter, Facebook, Myspace) significantly enlarge our social circles. One can follow any elites (celebrities), e.g., politicians, models, actors, and athletes, or close friends in her physical social network. An interesting question here is: when you follow a number of users, who will follow you back? A more specific question is: if you follow those celebrities (elite users) on Twitter, do you think they will follow you back? The answer is often No, but also Yes sometimes. There are a number of top users with tens of thousands of followers, who will follow everyone back. Some even use tools to do follow-back automatically, while others go through the following list and add their new followers manually.[1] Awareness of how these relationships are created can benefit many applications such as friend suggestion, community detection, and "word-of-mouth" product promotion.

In social science, relationships between individuals are classified into two categories: one-way (called parasocial) relationships and two-way (called reciprocal) relationships [9]. The most common form of the former are one-way relationships between celebrities and audience or fans, while the most common form of the latter are two-way relationships between close friends. Twitter and Facebook are respectively typical examples of the two types of social relationships. Social relationships form the basis of the social structure. Indeed, social relationships are always the basic object of analysis for social scientists, for instance, in Max Weber's theory of social action [29]. Understanding the formation of social relationships can give us insights into the micro-level dynamics of the social network, such as how an individual user influences her/his friends through different types of social relationships [26], and how the underlying social structure changes with the dynamics of relationship formation [23].

Employing Twitter as the basis of our analysis, we study how a two-way (reciprocal) relationship has been developed from a one-way (parasocial) relationship. Specifically, we try to answer: "when you follow a particular user (either an elite user or an ordinary user), how likely will she/he follow you back?". This problem also implicitly exists in other social networks such as Facebook and LinkedIn: when you send a friend request to somebody, how likely will she/he confirm your request?

Previous research on social relationships can be classified into three categories: link prediction [2, 16, 17, 23], relationship type inferring [4, 5, 27], and social behavior prediction [1, 25, 33]. Backstrom and Leskovec [2] proposed an approach called supervised random walks to predict and recommend links in social networks. Crandall et al. [4] investigated the problem of inferring social ties between people from co-occurrence in time and space.

[1]http://socialnewswatch.com/top-twitter-users/

Wang et al. [27] proposed an unsupervised algorithm to infer advisor-advisee relationships from a publication network. However, little research systematically studies how two-way relationships can be developed from one-way relationships. More fundamentally, what are the underlying factors that essentially influence the formation of two-way relationships? and how existing social theories (e.g., structural balance theory and homophily) can be connected to the formation process?

In this paper, we try to conduct a systematic investigation on the problem of two-way (reciprocal) relationship prediction. We precisely define the problem and propose a Triad Factor Graph (TriFG) model. The TriFG model incorporates social theories into a semi-supervised learning model, where we have some labeled training data (two-way relationships) but with low reciprocity [13]. Given a historic log of users' following actions from time 0 to t, we try to learn a predictive model to infer whether user A will add a follow-back link to user B at time $(t+1)$ if user B creates a new follow link to user A at time t. We evaluate the proposed model on a Twitter data consisting of 13,442,659 users and their profiles, tweets, following behaviors (new following or follow-back links) for nearly two months.

Results We show that incorporating social theories into the proposed factor graph model can significantly improve the performance (+22%-+27% by F1-Measure) of two-way (reciprocal) relationship prediction compared with several alternative methods. Our study also reveals several interesting phenomena:

1. Elite users tend to follow each other. The likelihood of an elite user following back another elite user is nearly 8 times higher than that of two ordinary users and 30 times that of an elite user and an ordinary user.
2. Two-way relationships on Twitter are balanced, but one-way relationships are not. More than 88% of social triads (groups of three people) with two-way relationships satisfy the social balance theory, while one-way relationships are unbalanced (merely 25% of them satisfy the balance theory).
3. Social networks are going global, but also stay local. No matter how far a user is from you, the likelihood that she/he follows you back is almost the same. While, on the other hand, the number of two-way relationships between users within the same time zone is 20 times higher than the number of users from different time zones.

Organization Section 2 formulates the problem. Section 3 introduces the data set and our analyses on the data set. Section 4 explains the proposed model and describes the algorithm for learning the model. Section 5 presents experimental results that validate the effectiveness of our methodology. Finally, Section 6 reviews the related work and Section 7 concludes this work.

2. PROBLEM DEFINITION

In this section, after presenting several definitions, we formally define the targeted problem in this work. We formulate the problem in the context of Twitter to keep things concrete, though adaptation of this framework to other social-network settings is straightforward.

The Twitter network can be modeled as a directed graph $G = \{V, E\}$, where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of users, and $E \subseteq V \times V$ is the set of directed links between users. Each directed link $e_{ij} = (v_i, v_j) \in E$ indicates that user v_i follows user v_j.

The Twitter network is dynamic in nature, with links added and removed from over time. However, our preliminary statistics on a large Twitter data set show that users tend to add new links much more frequently than to remove existing links (e.g., 97% of changes to links are adding new links). Therefore, adding new links forms the structure of the Twitter network. A new link results when a user performs a behavior of following another user (back) in Twitter. Particularly, we define two types of the link behavior:

Definition 1. **New-follow and Follow-back:** Suppose at time t, user v_i creates a link to v_j, who has no previous link to v_i, then we say v_i performs a new-follow behavior on v_j. When user v_i creates a link to v_j at time t, who already has a link to v_i before time t, we say v_i performs a follow-back behavior on v_j.

The new-follow and follow-back behaviors respectively correspond to the one-way (parasocial) relationship and the two-way (reciprocal) relationship in sociology. In this work, we focus on investigating the formation of follow-back behaviors. For simplicity, let $y_{ij}^t = 1$ denote that user v_i follows back v_j at time t and $y_{ij}^t = 0$ denote user v_i does not follow back. We are concerned with the following prediction problem:

Problem 1. **Follow-back prediction.** Let $< 1, \ldots, t >$ be a sequence of time stamps with a particular time granularity (e.g., day, week etc.). Given Twitter networks from time 1 to t, $\{G^t = (V^t, E^t, Y^t)\}$, where Y^t is the set of follow-back behaviors at time t, the task is to find a predictive function:

$$f : (\{G^1, \cdots, G^t\}) \rightarrow Y^{(t+1)},$$

such that we can infer the follow-back behaviors at time $(t+1)$.

It bears pointing out that our problem is very different from existing link prediction [2, 17, 23] and social action prediction problems [25, 33]. First, as the twitter network is evolving over time, it is infeasible to collect a complete network at time t. Thus it is important to design a method that could take into consideration the unlabeled data as well. Second, it is unclear what are the fundamental factors that cause the formation of follow-back relationships. Finally, one needs to incorporate the different factors (e.g., social theories, statistics, and our intuitions) into a unified model to better predict the follow-back relationship.

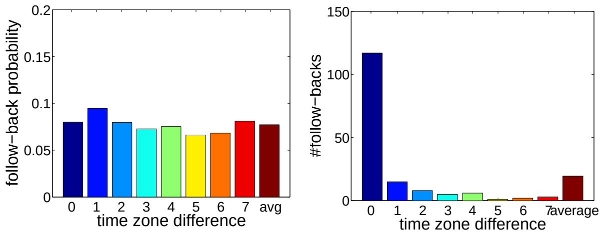
3. DATA AND OBSERVATIONS

3.1 Data Collection

We aim to find a large set of users and a continuously updated network among these users, so that we can use the data set as the gold-standard to evaluate different approaches for our prediction. To begin the collection process, we selected the most popular user on Twitter, i.e., "Lady Gaga", and randomly collected 10,000 of her followers. We took these users as seed users and used a crawler to collect all followers of these users by traversing following edges. We continue the traversing process, which produced in total 13,442,659 users and 56,893,234 following links, with an average of 728,509 new links per day. The crawler monitored the change of the network structure from 10/12/2010 to 12/23/2010. We also extracted all tweets posted by these users and in total there are 35,746,366 tweets.

In our analysis, we also consider the geographic location of each user. Specifically, we first extracted the location from the profile of each user[2], and then fed the location information to the Google Map API to fetch its corresponding longitude and latitude values. In this

[2]For example, Lady Gaga's location information is: "Location: New York, NY".



(a) Global (b) Local

Figure 1: Geographic distance correlation. X-axis: time zone difference (0 indicates that users are located in the same time zone); Y-axis: (a) probability that one user follows back another user, conditioned on the time zone difference of the two users. (b) number of two-way relationships among users from the same time zone or different time zones.

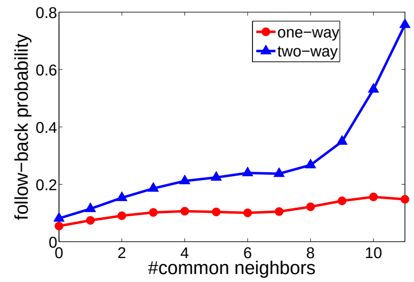


Figure 2: Link homophily. Y-axis: probability that two users follow back each other, conditioned on the number of common neighbors of two-way relationships (or one-way relationships).

way, we obtained the longitude and latitude of about 59% of users in our data set. More detailed analysis and an online demonstration is publicly available. http://arnetminer.org/reciprocal/

3.2 Observations

We first engage in some high-level investigation of how different factors influence the formation of follow-back (reciprocal) relationships, since a major motivation of our work is to find the underlying factors and their influence to this task. In particular, we study the interplay of the following factors with the formation of follow-backs: (1) *Geographic distance:* Do users have a higher probability to follow each other when they are located in the same region? (2) *Homophily:* Do similar users tend to follow each other? (3) *Implicit network:* How does the following network on Twitter correlate with other implicit networks, e.g., retweet and reply network? and (4) *Social balance:* Does the two-way relationship network on Twitter satisfy the social balance theory [6]? To which extent?

Geographic distance Figure 1 shows the correlation between geographic distance and the probability that two users create a two-way relationship (i.e., follow back each other). Interestingly, it seems that online social networks indeed go global: Figure 1(b) shows the likelihood of a user following another user back when they are from the same time zone or from different time zones. Clearly, the geographic distance is already not a factor to stop users from developing a trustful (reciprocal) relationship. Figure 1(a) shows another statistic which indicates a different perspective that the Twitter network (in some sense) still stays local: the average number of two-way (reciprocal) relationships between users from the same time zone is about 50 times higher than the number between users with a distance of three time zones.

Homophily The principle of homophily [15] suggests that users with similar characteristics (e.g., social status, age) tend to associate with each other. In particular, we study two kinds of homophilies on the Twitter network: link homophily and status homophily. For the link homophily, we test whether users who share common links (followers or followees) will have a tendency to associate with each other. Figure 2 clearly shows that the probability of two users following back each other when they share common neighbors is much higher than usual. When the number of common neighbors with two way relationships increases to 3, the likelihood of two users following back each other also triples. The effect is more pronounced when the number increases to 10. But it is worth noting that this only works for two-way (reciprocal) relationships and does not hold for the one-way (parasocial) relationship (as indicated in Figure 2).

For the status homophily, we test whether two users with similar social status are more likely to associate with each other. We categorize users into two groups (elite users and ordinary users) by three different algorithms: PageRank [22][3], #degree, and (α, β) algorithm [8][4]. Specifically, with PageRank, we estimate the importance of each user according to the network structure, and then select as elite users with the top 1% users[5] who have the highest PageRank scores and the rest as ordinary users; while with #degree, we select top 1% users with the highest number of indegree as elite users and the rest as ordinary users. For (α, β), we input the size of the core community as 200, and after running the algorithm, we use users selected in the core community as elite users and the rest as ordinary users. Then, we examine the difference of follow back behaviors among the two groups of users. Figure 3 clearly shows that, though the three algorithms present different statistics, "elite" users have a much stronger tendency to follow each other: the likelihood of two elite users following back each other is nearly 8 times higher than that of ordinary users (by the (α, β) algorithm). The (α, β) algorithm seems able to better distinguish elite users from ordinary users in our problem setting. This is because besides the global network structure, the (α, β) algorithm also considers the community structure among elite users.

Implicit structure On Twitter, besides the explicit network with following links, there are also some implicit network structure that can be induced from the textural information. For example, user A may mention user B in her tweet, i.e. "@B", which is called a reply link; user A may forward user B's tweet, which results in a retweet link. We study how the implicit links correlate with the formation of the follow-back relationship on Twitter. Figure 4 clearly shows that when users A and B retweet or reply each other's tweet, the likelihood of their following back each other is higher (3 times than chance). Another interesting phenomenon is that compared with replying someone's tweet, retweeting (forwarding) her tweet seems to be more helpful (15% vs. 9%) to win her follow-back.

Structural balance Now, we connect our work to a basic social psychological theory: structural balance theory [6]. Let us first explain the structural balance property. For every group of three users

[3]PageRank is an algorithm to estimate the importance of each node in a network.

[4](α, β) algorithm is designed to find core members (elite users) in a social network.

[5]Statistics have shown that less than 1% of the Twitter users produce 50% of its content [31].

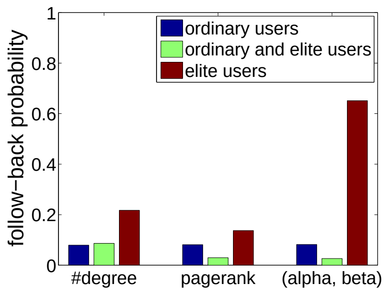


Figure 3: Status homophily by different algorithms. Y-axis: probability that two users follow back each other, conditioned on whether the two users are from the same group of elite/ordinary users or from different groups. #Degree, PageRank, and (α, β) are three algorithms to distinguish elite users from ordinary users.

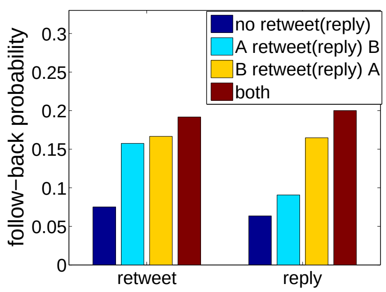


Figure 4: Implicit network correlation. Y-axis: probability that user B follows user A back, conditioned on one user (A or B) retweets or replies the other user's tweet.

(called triad), the balance property implies that either all three of these users are friends or only one pair of them are friends. Figure 5 shows such an example. To adapt the theory to our problem, we can map either the two-way relationship or the one-way relationship on the friendship. Then we examine how the Twitter network with (only two-way relationships or one-way relationships) satisfy the structural balance property. More precisely, we compare the probabilities of the resultant triads that satisfy the balance theory based on two-way relationships and one-way relationships on Twitter. Figure 6 clearly shows that it is much more likely (88%) for users to be connected with a balanced structure of two-way relationships. While with one-way relationships, the resultant structure is very unbalanced. This is because two users are very likely to follow a same movie star, but they do not know each other, which results in a unbalanced triad (Figure 5 (C)).

In summary, according to the statistics above, we have the following observations:

1. Geographic distance has a pronounced effect on the number of two-way relationships created between users, but little effect on the likelihood of users following back each other.
2. Users with common friends of two-way relationships have a tendency (link homophily) to follow each other.
3. Elite users have a much stronger tendency (status homophily) to follow each other than ordinary users.
4. The implicit networks of retweet or reply links have a strong correlation with the formation of two-way (reciprocal) relationships.

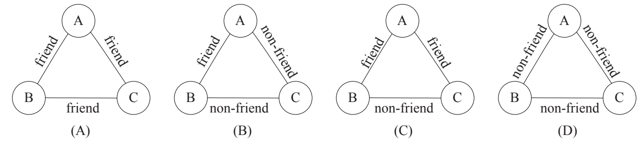


Figure 5: Illustration of structural balance theory. (A) and (B) are balanced, while (C) and (D) are not balanced.

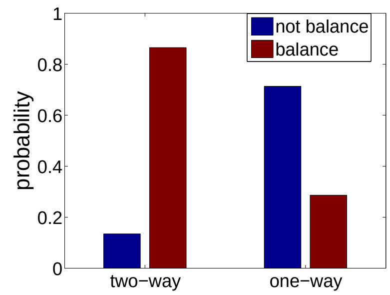


Figure 6: Structural balance correlation. Y-axis: probability that a triad creates two-way (reciprocal) relationships, conditioned on whether the resultant structure is balanced or not.

5. The network of two-way relationships on Twitter is balanced (88% of triads satisfying the structural balance property), while the network of one-way relationships is unbalanced (71% are unbalanced).

4. MODEL FRAMEWORK

In this section, we propose a novel Triad Factor Graph (TriFG) model to incorporate all the information within a single entity for better modeling and predicting the formation of two-way (follow-back) relationships.

For an edge $e_{ij} \in E$, if user v_j follows v_i at time t, our task is to predict whether user v_i will follow v_j back, i.e. $y_{ij} = 1$ or 0. For easy explanation, We introduce a slight change of notation. We write each edge as e_i with its two end users as v_i^s and v_i^u. For the follow-back prediction task, we assume that v_i^s follows v_i^u at time t, and our task is to predict whether v_i^u will follow v_i^s back at time $(t + 1)$. Based on the observations in §3, we define a number of attributes for each edge, denoted as $\mathbf{x}_i$. The $|E| \times d$ attribute matrix $\mathbf{X}$ describes edge-specific characteristics, where d is the number of attributes. For example, on Twitter, an attribute can be defined as whether two end users are from the same time zone. An element x_{ij} in the matrix $\mathbf{X}$ indicates the j^{th} attribute value of edge e_i.

4.1 The Proposed Model

We propose a Triad Factor Graph (TriFG) model. The name is derived from the idea that we incorporate social theories (structural balance and homophily) over triads into the factor graph model.

Figure 7 shows the graphical structure of the TriFG model. The left figure shows the following network of six users at time t. Blue arrows indicate new follow actions, black arrows indicate follow actions performed before time t, and blue $\nrightarrow$ indicates user v_i^u does not follow user v_i^s back at time t. The right figure is the factor graph model derived from the left input network. Each gray eclipse indicates an relationship (v_i^u, v_i^s) between users and each white circle indicates the hidden variable y_i, with $y_i = 1$ representing v_i^u performs a follow-back action, $y_i = 0$ not, and $y_i =?$ unknown, which actually is the variable we need to predict. Factor $h(.)$ represents

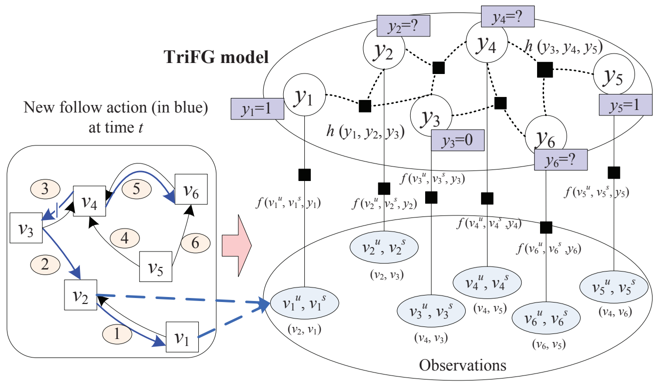


Figure 7: Graphical representation of the TriFG model. The left figure shows the follow network at time t. Blue arrows indicate new follow actions, black arrows indicate previously existing follow links, and blue $\nrightarrow$ indicates user v_i^u does not follow user v_i^s back. The right figure is the TriFG model derived from the following graph. Each gray eclipse indicates an relationship (v_i^u, v_i^s) between users and each white circle indicates the hidden variable y_i. $f(v_i^s, v_i^u, y_i)$ represents an attribute factor function and $h(.)$ represents a triad factor function.

a balance factor function defined on a triad; and $f(v_i^s, v_i^u, y_i)$ (or $f(\mathbf{x}_i, y_i)$) represents a factor to capture the information associated with edge e_i.

Given a network at time t, i.e., $G^t = (V^t, E^t, X^t)$ with some known variables $y = 1$ or 0 and some unknown variables $y =?$, our goal is to infer values of those unknown variables. For simplicity, we remove the superscript t for all variables if there is no ambiguity. We begin with the posterior probability of $P(Y|\mathbf{X}, G)$, according to the Bayes' theorem, we have

$$P(Y|\mathbf{X}, G) = \frac{P(\mathbf{X}, G|Y)P(Y)}{P(\mathbf{X}, G)} \propto P(\mathbf{X}|Y) \cdot P(Y|G) \quad (1)$$

where $P(Y|G)$ denotes the probability of labels given the structure of the network and $P(\mathbf{X}|Y)$ denotes the probability of generating the attributes $\mathbf{X}$ associated with each edge given their label Y. Assuming that the generative probability of attributes given the label of each edge is conditionally independent, we get

$$P(Y|\mathbf{X}, G) \propto P(Y|G) \prod_i P(\mathbf{x}_i|y_i) \quad (2)$$

where $P(\mathbf{x}_i|y_i)$ is the probability of generating attributes $\mathbf{x}_i$ given the label y_i. Now, the problem is how to instantiate the probabilities $P(Y|G)$ and $P(\mathbf{x}_i|y_i)$. In principle, they can be instantiated in different ways. In this work, we model them in a Markov random field, and thus by the Hammersley-Clifford theorem [7], the two probabilities can be instantiated as:

$$P(\mathbf{x}_i|y_i) = \frac{1}{Z_1} \exp\{\sum_{j=1}^{d} \alpha_j f_j(x_{ij}, y_i)\} \quad (3)$$

$$P(Y|G) = \frac{1}{Z_2} \exp\{\sum_c \sum_k \mu_k h_k(Y_c)\} \quad (4)$$

where Z_1 and Z_2 are normalization factors. Eq. 3 indicates that we define a feature function $f_j(x_{ij}, y_i)$ for each attribute x_{ij} associated with edge e_i and α_j is the weight of the j^{th} attribute; while Eq. 4 represents that we define a set of correlation feature functions

Input: network G^t, learning rate η
Output: estimated parameters θ

Initialize $\theta \leftarrow 0$;
repeat
- Perform LBP to calculate marginal distribution of unknown variables $P(y_i|x_i, G)$;
- Perform LBP to calculate the marginal distribution of triad c, i.e., $P(y_c|\mathbf{X}_c, G)$;
- Calculate the gradient of μ_k according to Eq. 7 (for α_j with a similar formula):

$$\frac{\mathcal{O}(\theta)}{\mu_k} = \mathbb{E}[h_k(Y_c)] - \mathbb{E}_{P_{\mu_k}(Y_c|\mathbf{X}, G)}[h_k(Y_c)]$$

- Update parameter θ with the learning rate η:

$$\theta_{\text{new}} = \theta_{\text{old}} + \eta \cdot \frac{\mathcal{O}(\theta)}{\theta}$$

until *Convergence*;

Algorithm 1: Learning algorithm for the TriFG model.

$\{h_k(Y_c)\}_k$ over each triad Y_c in the network. Here μ_k is the weight of the k^{th} correlation feature function.

Based on Eqs. 2-4, we define the following log-likelihood objective function $\mathcal{O}(\theta) = \log P_\theta(Y|\mathbf{X}, G)$:

$$\mathcal{O}(\theta) = \sum_{i=1}^{|E|} \sum_{j=1}^{d} \alpha_j f_j(x_{ij}, y_i) + \sum_c \sum_k \mu_k h_k(Y_c) - \log Z \quad (5)$$

where Y_c is a triad derived from the input network, $Z = Z_1 Z_2$ is a normalization factor and $\theta = (\{\alpha\}, \{\mu\})$ indicates a parameter configuration. One example of factor decomposition is shown in Figure 7. There are six edges, three with known variables (two $y = 1$ and one $y = 0$) and three with unknown values ($y =?$). We have four triads (e.g., $Y_c = (y_1, y_2, y_3)$) based on the structure of the input network. For each edge, we define a set of factor functions $f(v_i^s, v_i^u, y_i)$ (also written as $f(\mathbf{x}_i, y_i)$).

We now briefly introduce possible ways to define the factor functions $f_j(x_{ij}, y_i)$ and $h_k(Y_c)$. $f_j(x_{ij}, y_i)$ is an attribute factor function. It can be defined as either a binary function or a real-valued function. For example, for the implicit network feature, we simply define it as a binary feature, that is if user v_i^s forwarded (retweeted) v_i^u's tweet before time t and user v_i^u follows user v_i^s back, then a feature $f_j(x_{ij} = 1, y_i = 1)$ is defined and its value is 1; otherwise 0. (Such a feature definition is often used in graphical models such as Conditional Random Fields [14]. For the triad factor function $h(Y_c)$, we define four features, two balanced and two unbalanced factor functions, as depicted in Figure 5. The triad function is defined as a binary function, that is, if a triad satisfies the structural balance property, then the value of a corresponding triad factor function is 1, otherwise 0. More details of the factor function definition are given in Appendix.

4.2 Model Learning and Prediction

We now address the problem of estimating the free parameters and inferring users' follow-back behaviors. Learning the TriFG model is to estimate a parameter configuration $\theta = (\{\alpha\}, \{\mu\})$ to maximize the log-likelihood objective function $\mathcal{O}(\theta) = \log P_\theta(Y|\mathbf{X}, G)$, i.e.,

$$\theta^\star = \arg\max \mathcal{O}(\theta) \quad (6)$$

To solve the objective function, we adopt a gradient decent method (or a Newton-Raphson method). We use μ as the example to explain how we learn the parameters. Specifically, we first write the gradient of each μ_k with regard to the objective function (Eq. 5):

$$\frac{\mathcal{O}(\theta)}{\mu_k} = \mathbb{E}[h_k(Y_c)] - \mathbb{E}_{P_{\mu_k}(Y_c|\mathbf{X},G)}[h_k(Y_c)] \tag{7}$$

where $\mathbb{E}[h_k(Y_c)]$ is the expectation of factor function $h_k(Y_c)$ given the data distribution (essentially it can be considered as the average value of the factor function $h_k(Y_c)$ over all triads in the training data); and $\mathbb{E}_{P_{\mu_k}(Y_c|\mathbf{X},G)}[h_k(Y_c)]$ is the expectation of factor function $h_k(Y_c)$ under the distribution $P_{\mu_k}(Y_c|\mathbf{X}, G)$ given by the estimated model. A similar gradient can be derived for parameter α_j.

One challenge here is that the graphical structure in the TriFG model can be arbitrary and may contain cycles, which makes it intractable to directly calculate the marginal distribution $P_{\mu_k}(Y_c|\mathbf{X}, G)$. A number of approximate algorithms can be considered, such as Loopy Belief Propagation (LBP) [20] and Meanfield [32]. We chose Loopy Belief Propagation due to its ease of implementation and effectiveness. Specifically, we approximate the marginal distribution $P_{\mu_k}(Y_c|\mathbf{X}, G)$ using LBP. With the marginal probabilities, the gradient can be obtained by summing over all triads. It is worth noting that we need to perform the LBP process twice in each iteration, one time for estimating the marginal distribution of unknown variables $y_i =?$ and the other time for marginal distribution over all triads. Finally with the gradient, we update each parameter with a learning rate η. The learning algorithm is summarized in Algorithm 1.

Predicting Follow-back With the estimated parameters θ, we can predict the label of unknown variables $\{y_i =?\}$ by finding a label configuration which maximizes the objective function, i.e., $Y^* = \text{argmax}\mathcal{O}(Y|\mathbf{X}, G, \theta)$. It is still intractable to obtain the exact solution. Again, we utilize the loopy belief propagation to approximate the solution, i.e., to calculate the marginal distribution of each relationship with unknown variable $P(y_i|\mathbf{x}_i, G)$ and finally assign each relationship with label of the maximal probability.

5. EXPERIMENTS

In this section, we first describe our experimental setup. We then present the performance results for different approaches in different settings. Next, we present several analyses and discussions. Finally, we use a case study further to demonstrate the advantage of the proposed model.

5.1 Experimental Setup

Prediction Setting We use the data set described in §3 in our experiments. To quantitatively evaluate the effectiveness of the proposed model and compare with other alternative methods, we carefully select a sub network from the data set, which has a completely historic log of link formation among all users, i.e., each user is associated with a complete list of followers and users they are following at each time stamp. The sub network is comprised of 112,044 users, 468,238 following links among them, and 2,409,768 tweets. Averagely there are 3,337 new follow-back links per day. We divide the sub network into 13 time stamps by viewing every four days as a time stamp.

Our general task is to predict whether a user will follow another user back at the next time stamp when she received a new following link from the other user. By a more careful study however, we found that it is very challenging if we restrict the prediction just for the next time stamp. Figure 8 shows the distribution of time span in which a user performs the follow-back action, which indicates that 60% of follow-backs are performed in the next time stamp though, 37% of the follow-backs would be still performed in the following three time stamps. A further data analysis, shows that active users often either perform an immediate follow-back (at the next time stamp) or reject to follow-back; while some other (inactive) users may not frequently login into Twitter, thus the time span of follow-backs varies a lot. According to this observation, in our first experiment, we use a network of the first 8 time stamps for training and predicate follow-back actions in the following 4 (9th-12th) time stamps (Test Case 1). Then we incrementally add the network of the 9th time stamp into the training data and again use the following 4 (10th-13th) time stamps for prediction (Test Case 2). We respectively report the prediction performance of different approaches for the two test cases.

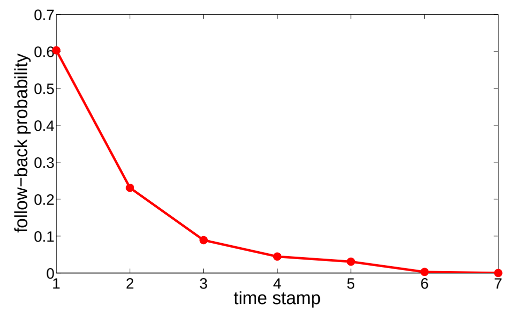


Figure 8: Follow-back probability for different time stamps.

Comparison Methods We compare the proposed TriFG model with the following methods:

SVM: it uses the same attributes associated with each edge as features to train a classification model and then employs the classification model to predict edges' label in the test data. For SVM, we employ SVM-light.

LRC: it uses the same attributes associated with each edge as features to train to train a logistic regression classification model [16] and then predict edges' label in the test data.

CRF-balance: it trains a Conditional Random Field [14] model with attributes associated with each edge. The difference of this method from our model is that it does not consider structural balance factors.

CRF: it trains a Conditional Random Field model all factors (including attributes and structural balance factors) and predicts edges' label in the test data.

TriFG: the proposed model, which trains a factor graph model with unlabeled data and all factors we defined in §4.

Weak TriFG (wTriFG): the difference of wTriFG from TriFG is that we do not consider status homophily and structural balance here. We use this method to evaluate how social theories can help this task.

In the five methods, SVM and CRF-balance only considers attribute factors; wTriFG further considers unlabeled data. CRF considers all factors we defined, but does not consider unlabeled data. Our proposed TriFG model considers all factors as well as the unlabeled data.

Evaluation Measures We evaluate the performance of different approaches in terms of Precision (Prec.), Recall (Rec.), F1-Measure (F1), and Accuracy (Accu.).

Table 1: Follow-back prediction performance of different methods in the two test cases. **Test Case 1: predicting follow-back actions in the 9th-12th time stamps; and Test Case 2 for the 10th-13th time stamps.**

Data	Algorithm	Prec.	Rec.	F1	Accu.
Test Case 1	SVM	0.6908	0.6129	0.6495	0.9590
	LRC	0.6957	0.2581	0.3765	0.9510
	CRF-balance	0.9968	0.5161	0.6801	0.9670
	CRF	**1.0000**	0.6290	0.7723	0.9770
	wTriFG	0.9691	0.5483	0.7004	0.9430
	TriFG	**1.0000**	**0.8548**	**0.9217**	**0.9910**
Test Case 2	SVM	0.7323	0.6212	0.6722	0.9534
	LRC	0.8333	0.3030	0.4444	0.9417
	CRF-balance	0.9444	0.5151	0.6667	0.9114
	CRF	**1.0000**	0.6333	0.7755	0.9717
	wTriFG	0.9697	0.5697	0.7177	0.9389
	TriFG	**1.0000**	**0.8788**	**0.9355**	**0.9907**

Table 2: Follow-back prediction performance of TriFG with three different algorithms (#degree, PageRank and (α, β)) for finding elite users from ordinary users.

Data	Algorithm	Prec.	Rec.	F1	Accu.
Test Case 1	(α, β)	**1.0000**	**0.8548**	**0.9217**	**0.9910**
	#degree	**1.0000**	0.7903	0.8829	0.9870
	pagerank	**1.0000**	0.7581	0.8624	0.9850
Test Case 2	(α, β)	**1.0000**	**0.8788**	**0.9355**	**0.9907**
	#degree	**1.0000**	0.8363	0.9109	0.9874
	pagerank	**1.0000**	0.8181	0.9000	0.9860

All algorithms are implemented in C++, and all experiments are performed on a PC running Windows 7 with Intel(R) Core(TM) 2 CPU 6600 (2.4GHz and 2.39GHz) and 4GB memory. All algorithms have a good efficiency performance: the CPU time needed for training and prediction by all methods on the Twitter network ranges from 2 to 5 minutes.

5.2 Prediction Performance

We now describe the performance results for the different methods we considered. Table 1 show the results in the two test cases (prediction performance for the 9th-12th time stamps and that for the 10th-13th time stamps).

It can be clearly seen that our proposed TriFG model significantly outperforms the four comparison methods. In terms of F1-Measure, TriFG achieves a +27% improvement compared with the (SVM). Comparing with the other three graph-based methods, TriFG also results in an improvement of 22-25%. The advantage of TriFG mainly comes from the improvement on recall. One important reason here is that TriFG can detect some difficult cases by leveraging the structural balance correlation and homophily correlation. For example, without considering the two kinds of social correlations, the performance of wTriFG decreases to 70-72% in terms of F1-Measure in the two test cases. Another advantage of TriFG is that it makes use of the unlabeled data. Essentially, it further considers some latent correlation in the data set, which cannot be leveraged with only the labeled training data.

5.3 Analysis and Discussions

Now, we perform several analyses to examine the following aspects of the TriFG model: (1) contribution of different factors in the TriFG model; (2) convergence property of the learning algorithm; (3) Effect of different settings for the time span; and (4) Effect of different algorithms for elite user finding.

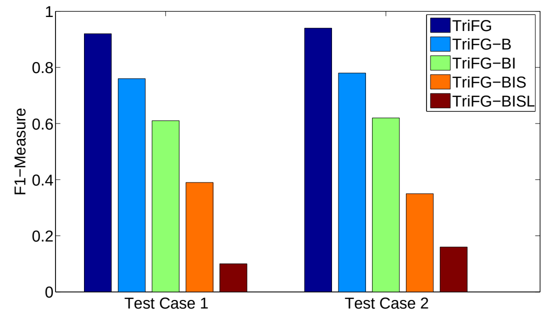


Figure 9: Factor contribution analysis. **TriFG-B stands for ignoring structural balance correlation. TriFG-BI stands for ignoring both structural balance correlation and implicit network correlation. TriFG-BIS stands for further ignoring status homophily and TriFG-BISL stands for further ignoring link homophily.**

Factor Contribution Analysis In TriFG, we consider five different factor functions: Geographic distance (G), Link homophily (L), Status homophily (S), Implicit network correlation (I), and structural Balance correlation (B). Here we examine the contribution of the different factors defined in our model. We first rank the individual factors by their predictive power[6], then remove them one by one in reversing order of their prediction power. In particular, we first remove structural balance correlation denoted as TriFG-B, followed by further removing the implicit network correlation denoted as TriFG-BI, status homophily denoted as TriFG-BIS, and finally removing link homophily denoted as TriFG-BISL. We train and evaluate the prediction performance of the different versions of TriFG. Figure shows the average F1-Measure score of the different versions of the TriFG model. We can observe clear drop on the performance when ignoring each of the factors. This indicates that our method works well by combining the different factor functions and each factor in our method contributes improvement in the performance.

Convergence Property We conduct an experiment to see the effect of the number of the loopy belief propagation iterations. Figure 10 illustrates the convergence analysis results of the learning algorithm. We see on both test cases, the BLP-based learning algorithm can converges in less than 10 iterations. After only seven learning iterations, the prediction performance of TriFG on both test cases becomes stable. This suggests that learning algorithm is very efficient and has a good convergence property.

Effect of Time Span Figure 8 already shows the distribution of follow-backs in different time stamps. Now, we quantitatively examine how different settings for the time span will affect the prediction performance. Figure 11 lists the average prediction performance of TriFG in the two test cases with different settings of the time span. It shows that when setting the time span as two or less time stamps, the prediction performance of TriFG drops sharply;

[6] We did this by respectively removing each particular factor from our model and evaluated the decrease of the prediction performance by the TriFG model. A larger decrease means a higher predictive power.

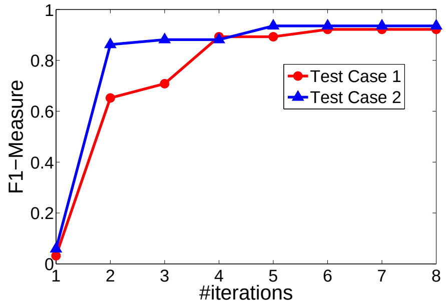


Figure 10: Convergence analysis of the learning algorithm.

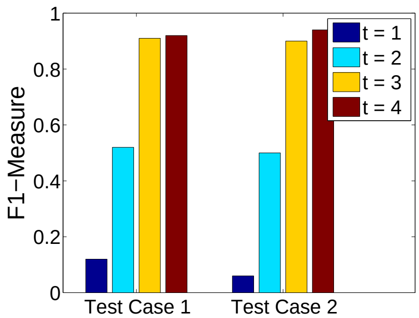


Figure 11: Follow-back prediction for different time stamps.

while when setting it as three time stamps, the performance is acceptable. The results are consistent with the statistics in Figure 8: more than 90% of follow-back actions are performed in the first three time stamps, and only about 80% of the follow-back actions are in the first two time stamps.

Effect of different algorithms for elite user finding The status homophily factor depends on results of elite user finding. We use three different algorithms, i.e., PageRank, #degree, and (α, β) algorithm, to find elite users. Now we examine how the different algorithms would affect the prediction performance. Table 2 shows the prediction performance of TriFG with different elite user finding algorithms in the two test cases. Interestingly, though TriFG with the (α, β) algorithm achieves the best performance, the difference of performance among the three algorithms, especially in the second test case is not that pronounced (with a difference of 1%-4% in terms of F1-measure score). This confirms the effectiveness and generalization of incorporating the status homophily factor into our TriFG model.

5.4 Qualitative Case Study

Now we present a case study to demonstrate the effectiveness of the proposed model. Figure 12 shows an example generated from our experiments. It represents a portion of the Twitter network from the 10th-13th time stamps. Black arrows indicate following links created 4 time stamps (we use 4 time stamps as the time span for prediction) before. Blue arrows indicate new following link in the past 4 time stamps. Dash arrows indicate follow-back links in our data set (a), predicted by SVM (b), and predicted by our model TriFG (c), with green color denoting a correct one and red color denoting a mistake one. Red colored $\nrightarrow$ indicates there should be a follow-back link, which the approach did not detect.

We look at specific examples to study why the proposed model can outperform the comparison methods. "A", "B", and "C" are three elite users identified using the (α, β) algorithm [8]. SVM correctly predicts that there is a follow-back link from "C" to "B", but misses predicting the follow-back link from "C" to "A". Our model TriFG correctly predicted both the follow-back links. This is because TriFG leverages the structural balance factor. The resulting structure among the three users by SVM is unbalanced. TriFG leverages the structural balance factor and tends to result in a balanced structure.

It is also worth looking at the situation of user 9 and 10. TriFG made a mistake here: it does not predict the follow-back link, while the link was correctly predicted by SVM. User 9 and user 10 have a similar social status (similar indegree) and also they are from the same time zone, thus SVM successfully predicted the follow-back link. However, as the resulting structure is unbalanced, TriFG made a compromise and finally resulted in a mistaken prediction.

6. RELATED WORKS

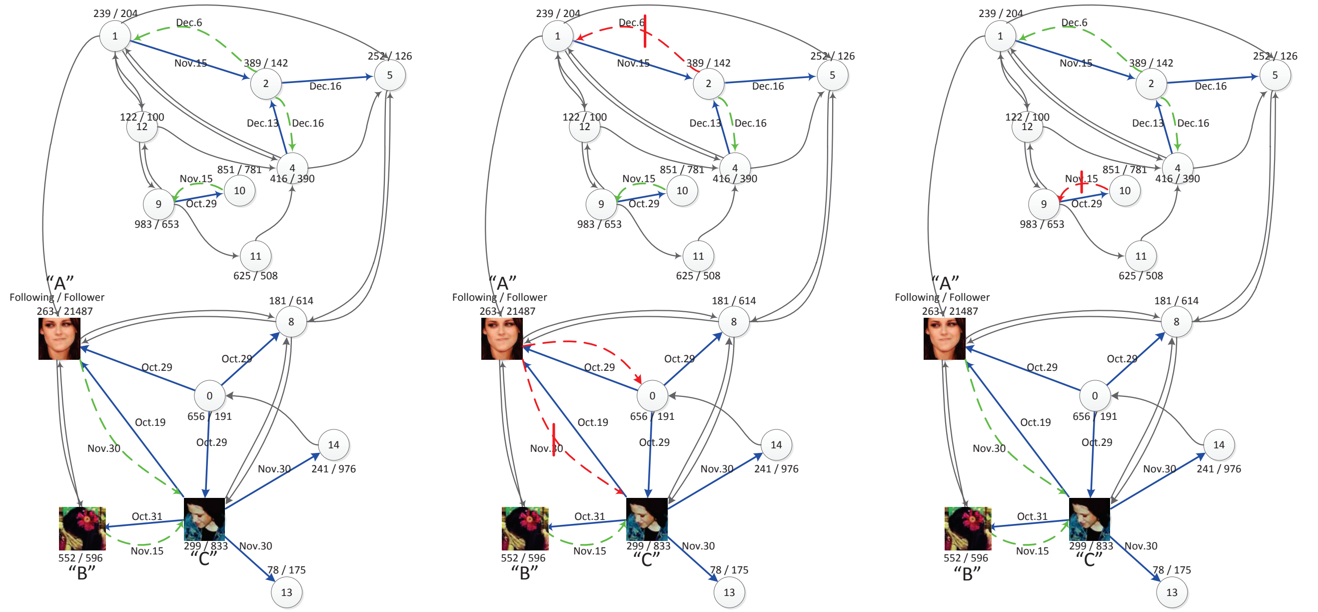
In this section, we review related work on link prediction and Twitter study in social networks.

Our work is related with link prediction, which is one of the core tasks in social networks. Existing work on link prediction can be broadly grouped into two categories based on the learning methods employed: unsupervised link prediction and supervised link prediction. Unsupervised link predictions usually assign scores to potential links based on the intuition - the more similar the pair of users are, the more likely they are linked. Various similarity measures of users are considered, such as the *preferential attachment* [21], and the Katz measure [12]. A survey of unsupervised link prediction can be found in [17]. Recently, [18] designs a flow based method for link prediction.

There are also a number of works which employ supervised approaches to predict links in social networks, such as [28, 18, 2, 16]. Backstrom et al. [2] propose a supervised random walk algorithm to estimate the strength of social links. Leskovec et al. [16] employ a logistic regression model to predict positive and negative links in online social networks. The main differences between existing work on link prediction and our work are about two aspects. First, existing work handles undirected social networks, while we address the directed nature of the Twitter network and predict a directed link between a pair of users given an existing link in the another direction. Secondly, most existing models for link prediction are static. In contrast, our model is dynamic and learned from the evolution of the Twitter network. Moreover, we combine social theories (such as homophily and structural balance theory) into a semi-supervised learning model.

Another type of related work is social behavior analysis. Tang et al. [26] study the difference of the social influence on different topics and propose Topical Affinity Propagation (TAP) to model the topic-level social influence in social networks and develop a parallel model learning algorithm based on the map-reduce programming model. Tan et al. [25] investigate how social actions evolve in a dynamic social network and propose a time-varying factor graph model for modeling and predicting users' social behaviors. The proposed methods in these work can be utilized in the problem defined in this work, but the problem is fundamentally different.

There is little doubt that Twitter has intrigued worldwide netizens, and the research communities alike. Existing Twitter study is mainly centered around the following three aspects: 1) *the Twitter network*. Java et al. [11] study the topological and geographical properties of the Twitter network. Their findings verify the *homophily* phenomenon that users with similar intentions connect with each other. Kwak et al. [13] conduct a similar study on the entire Twittersphere and they observe some notable properties of Twitter, such as a non-power-law follower distribution, a short ef-



(a) Ground Truth (b) SVM (c) Our approach (TriFG)

Figure 12: Case study. Portion of the Twitter network during the 10th-13th time stamps. The two numbers associated with each user are respectively the number of followees and that of followers. Black arrows indicate following links created 4 time stamps (we use 4 time stamps as the time span for prediction) before. Blue arrows indicate new following link in the past 4 time stamps. Dash arrows indicate follow-back links in our data set (a), predicted by SVM (b), and predicted by our model TriFG (c), with green color denoting a correct one and red color denoting a mistake one. Red colored ↛ indicates there should be a follow-back link, which the approach did not predict.

fective diameter, and low reciprocity, marking a deviation from known characteristics of human social networks. 2) *the Twitter users*. Work of this category mainly focus on identifying influential users in Twitter [30, 3, 13] or examining and predicting tweeting behaviors of users [10, 25]. 3) *the Tweets*. Sakaki et al. [24] propose to utilize the real-time nature of Twitter to detect a target event; while Mathioudakis and Koudas [19] present a system, TwitterMonitor, to detect emerging topics from the Twitter content.

7. CONCLUSION

In this paper, we study the novel problem of two-way relationship prediction in social networks. We formally define the problem and propose a Triad Factor Graph (TriFG) model, which incorporates social theories into a semi-supervised learning model. We evaluate the proposed model on a large Twitter network. We show the proposed factor graph model can significantly improve the performance (+22%-+27% by F1-Measure) for two-way relationship prediction comparing with several alternative methods. Our study also reveals several interesting phenomena.

The general problem of reciprocal relationship prediction represents a new and interesting research direction in social network analysis. There are many potential future directions of this work. First, some other social theories can be further explored and validated for reciprocal relationship prediction. Looking farther ahead, it is also interesting to develop a real friend suggestion system based on the proposed method. We can validate the proposed method based on user feedbacks. We can also further study theoretical methodologies for improving the predictive performance by incorporating user interactions. Finally, building a theory of why and how users create relationships with each other in different kinds of networks is an intriguing direction for further research.

Acknowledgements John Hopcroft was partially supported by the U.S. Air Force Office of Scientific Research under Grant FA9550-09-1-0675. Jie Tang is supported by the Natural Science Foundation of China (No. 61073073) and Chinese National Key Foundation Research (No. 60933013, No.61035004). Tiancheng Lou is supported in part by the National Basic Research Program of China Grant 2007CB807900, 2007CB807901, the National Natural Science Foundation of China Grant 61033001, 61061130540, 61073174.

8. REFERENCES

[1] L. Backstrom, R. Kumar, C. Marlow, J. Novak, and A. Tomkins. Preferential behavior in online groups. In *WSDM'08*, pages 117–128, 2008.

[2] L. Backstrom and J. Leskovec. Supervised random walks: predicting and recommending links in social networks. In *WSDM'11*, pages 635–644, 2011.

[3] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *ICWSM'10*, 2010.

[4] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. *PNAS*, 107:22436–22441, Dec. 2010.

[5] N. Eagle, A. S. Pentland, and D. Lazer. Inferring social network structure using mobile phone data. *PNAS*, 106(36), 2009.

[6] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.

[7] J. M. Hammersley and P. Clifford. Markov field on finite

graphs and lattices. *Unpublished manuscript*, 1971.

[8] J. He, J. E. Hopcroft, H. Liang, S. Suwajanakorn, and L. Wang. Detecting the structure of social networks using (α, β)-communities. In *WAW'11*, 2011.

[9] D. Horton and R. R. Wohl. Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry*, pages 215–229, 1956.

[10] B. Huberman, D. M. Romero, and F. Wu. Social networks that matter: Twitter under microscope. In *First Monday*, volume 14, pages 118–138, 2009.

[11] A. Java, X. Song, T. Finin, and B. L. Tseng. Why we twitter: An analysis of a microblogging community. In *WebKDD/SNA-KDD*, pages 118–138, 2007.

[12] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[13] H. Kwak, C. Lee, H. Park, and S. B. Moon. What is twitter, a social network or a news media? In *WWW'10*, pages 591–600, 2010.

[14] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML'01*, pages 282–289, 2001.

[15] P. F. Lazarsfeld and R. K. Merton. Friendship as a social process: A substantive and methodological analysis. *M. Berger, T. Abel, and C. H. Page, editors, Freedom and control in modern society, New York: Van Nostrand*, pages 8–66, 1954.

[16] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. In *WWW'10*, pages 641–650, 2010.

[17] D. Liben-Nowell and J. M. Kleinberg. The link-prediction problem for social networks. *JASIST*, 58(7):1019–1031, 2007.

[18] R. Lichtenwalter, J. T. Lussier, and N. V. Chawla. New perspectives and methods in link prediction. In *KDD'10*, pages 243–252, 2010.

[19] M. Mathioudakis and N. Koudas. Twittermonitor: trend detection over the twitter stream. In *SIGMOD'10*, pages 1155–1158, 2010.

[20] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *UAI'99*, pages 467–475, 1999.

[21] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Phys. Rev. E*, 64(2):025102, 2001.

[22] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report SIDL-WP-1999-0120, Stanford University, 1999.

[23] D. M. Romero and J. M. Kleinberg. The directed closure process in hybrid social-information networks, with an analysis of link formation on twitter. In *ICWSM'10*, 2010.

[24] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *WWW'10*, pages 851–860, 2010.

[25] C. Tan, J. Tang, J. Sun, Q. Lin, and F. Wang. Social action tracking via noise tolerant time-varying factor graphs. In *KDD'10*, pages 1049–1058, 2010.

[26] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *KDD'09*, pages 807–816, 2009.

[27] C. Wang, J. Han, Y. Jia, J. Tang, D. Zhang, Y. Yu, and J. Guo. Mining advisor-advisee relationships from research publication networks. In *KDD'10*, pages 203–212, 2010.

[28] C. Wang, V. Satuluri, and S. Parthasarathy. Local probabilistic models for link prediction. In *ICDM'07*, pages 322–331, 2007.

[29] M. Weber. *The Nature of Social Action in Runciman, W.G. 'Weber: Selections in Translation'*. Cambridge University Press, 1991.

[30] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In *WSDM'10*, pages 261–270, 2010.

[31] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts. Who says what to whom on twitter. In *WWW'11*, pages 705–714, 2011.

[32] E. P. Xing, M. I. Jordan, and S. Russell. A generalized mean field algorithm for variational inference in exponential families. In *UAI'03*, pages 583–591, 2003.

[33] Z. Yang, J. Guo, K. Cai, J. Tang, J. Li, L. Zhang, and Z. Su. Understanding retweeting behaviors in social networks. In *CIKM'10*, pages 1633–1636, 2010.

Appendix: Factor function definition

This section depicts how we define the factor functions in our experiments. In total, we define 25 features of five categories: Geographic distance, Link homophily, Status homophily, Structural balance, and Implicit network correlation.

Geographic distance We use Google Map API to get the exact locations (longitude and latitude) of some users. Based on the two values, we define the following three features : the absolute distance and the time zone difference between two users, and whether or not the two users are from the same country.

Link homophily First, we treat each link as undirected link, and define the following four features : the number of common neighbors, percentage of common neighbors of the two users(respectively) and the average percentage.

Then we consider directed links and define another three features : the number of common two-way links, number of common followers and number of common followees.

Status homophily We also test whether two users have similar social status, and define the following three features : whether or not the two users are both elite users, an ordinary and an elite, and both ordinary users.

Implicit network correlation We consider the interaction between user A and user B, and define the following four features respectively represent the number of retweets(replies) from A to B and from B to A.

Structural balance Based on the structural balance theory, as in Figure 5, we define eight features capturing all situations of structural balance theory for each triad.