Will Triadic Closure Strengthen Ties in Social Networks?*

HONG HUANG, Huazhong University of Science and Technology, China
YUXIAO DONG, Microsoft Research Redmond, USA
JIE TANG, Tsinghua University, China
HONGXIA YANG, Alibaba Group, China
NITESH V. CHAWLA, University of Notre Dame, USA
XIAOMING FU, University of Goettingen, Germany

The social triad—a group of three people—is one of the simplest and most fundamental social groups. Extensive network and social theories have been developed to understand its structure, such as triadic closure and social balance. Over the course of a triadic closure—the transition from two ties to three among three users, the strength dynamics of its social ties, however, are much less well understood. Using two dynamic networks from social media and mobile communication, we examine how the formation of the third tie in a triad affects the strength of the existing two ties. Surprisingly, we find that in about 80% social triads, the strength of the first two ties are weakened although averagely the tie strength in the two networks maintains an increasing or stable trend. We discover that 1) the decrease in tie strength among three males is more sharply than that among females, and 2) the tie strength between celebrities are more likely to be weakened as the closure of a triad than those between ordinary people. Further, we formalize a triadic tie strength dynamics prediction problem to infer whether social ties of a triad will become weakened after its closure. We propose a TRIST method—a kernel density estimation (KDE) based graphical model—to solve the problem by incorporating user demographics, temporal effects, and structural information. Extensive experiments demonstrate that TRIST offers a greater than 82% potential predictability for inferring triadic tie strength dynamics in both networks. The leveraging of the kernel density estimation and structural correlations enables TRIST to outperform baselines by up to 30% in terms of F1-score.

CCS Concepts: • **Information systems → Information systems applications**; *Information retrieval*; • **Applied computing → Law, social and behavioral sciences**;

Additional Key Words and Phrases: Social Triad, Tie strength, Dynamics, Predictive Model

*This work was primarily done when Hong Huang was a Postdoc researcher at University of Goettingen, Germany and Yuxiao Dong was a Ph.D. student at University of Notre Dame, IN. USA. Hong Huang is the corresponding author.

Authors' addresses: Hong Huang, Service Computing Technology and System Lab / Big Data Technology and System Lab, Huazhong University of Science and Technology, Wuhan, 430074, China, honghuang@hust.edu.cn; Yuxiao Dong, Microsoft Research Redmond, USA, yuxdong@microsoft.com; Jie Tang, Tsinghua University, Beijing, China, jietang@tsinghua.edu.cn; Hongxia Yang, Alibaba Group, Hanzhou, China, yang.yhx@alibaba-inc.com; Nitesh V. Chawla, University of Notre Dame, USA, nchawla@nd.edu; Xiaoming Fu, University of Goettingen, Goettingen, Germany, fu@cs.uni-goettingen.de.

1556-4681/2017/3-ART39
https://doi.org/0000001.0000001

ACM Reference Format:
Hong Huang, Yuxiao Dong, Jie Tang, Hongxia Yang, Nitesh V. Chawla, and Xiaoming Fu. 2017. Will Triadic Closure Strengthen Ties in Social Networks?. *ACM Trans. Knowl. Discov. Data.* 9, 4, Article 39 (March 2017), 27 pages. https://doi.org/0000001.0000001

1 INTRODUCTION

Social relationships are formed as individuals connect and interact with each other. Collectively, connected individuals further emerge as groups, communities, and societies, which manifest as networks. In this work, we focus on studying the simplest and most fundamental organizing unit among individuals in social networks — the social triad. As networks evolve, the strength of social ties does not keep constant over time. Some ties may become strong ties at first and then weaken over time, while other social ties begin as weak ties and become stronger. The dynamics of tie strength become even more complicated when we consider the interpersonal interactions, i.e., in a closed triad, will the formation of the third tie affect the strength of the existing two ties?

A significant amount of work has been devoted to investigating triadic relationships in social networks for decades. Simmel pioneered the study of "triad" and suggested that a social triad is fundamentally different from a dyad as interaction between members decreases, intimacy declines, strength and stability increase [Simmel 1950]. A common behavior was often observed in a triad that two of the members will tend to unite against the other one, which is known as two-against-one phenomenon [Caplow 1968]. Heider developed the balance theory [Heider 1958] in social triads that explains the proverbs — "A friend of my friend is my friend" and "The enemy of my enemy is my friend." Davis et al. proposed a status theory [Davis and Leinhardt 1972] that provides an organizing principle for directed networks of signed links. They addressed how the interplay between signed (i.e., positive and negative) relationships affects the structure of networks. However, in many online social networks until today (e.g., Twitter, Facebook, Weibo, etc.), it is hard to tell the sign of relationships in these networks where such theories, like balance theory and status theory would not be applied any more.

Besides social theories on triadic relationships, a large body of work [Fang and Tang 2015; Huang et al. 2014; Kossinets and Watts 2006; Lou et al. 2013; Romero and Kleinberg 2010; Zignani et al. 2014] has focused on modeling and predicting the process of triadic closure, i.e., the transition from open triads to closed triads. Furthermore, the process of triadic closure has been empirically demonstrated to be relevant for characterizing both social ties at the micro level [Dong et al. 2016; Sintos and Tsaparas 2014], and scaling laws at the macro level, in social and information networks [Klimek and Thurner 2013; Leskovec et al. 2008; Zhang et al. 2017].

Although amount of interesting and promising discoveries have been found in the field of social triads, little has been studied concerning what happens after triadic closure, especially the dynamics of tie strength within a triad. Essentially, previous attempts were limited by focusing on the closure transition, and ignoring the dynamics of triadic relationships after its closure. In other words, the interaction dynamics in a triad is still unclear, particularly, will triadic closure strengthen social ties? Further complications arise due to the complexity of scrutinizing various factors that drive the interaction dynamics of social triads, such as the demographics of a triad's three users, their tie strength, and the formation order of the links in a triad. Moreover, there is a lack of a basic understanding of the predictability of triadic tie strength dynamics in social networks.

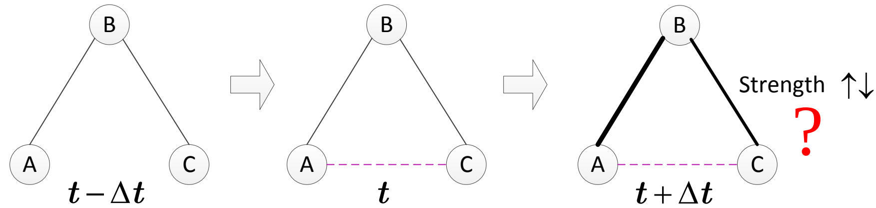


Fig. 1. Illustrative example of triadic tie strength dynamics in social networks. The formation of link e_{AC} makes an open triad $\mathcal{O}_{ABC}$ become a closed triad $\mathcal{T}_{ABC}$ at time t, and develop different levels of tie strength (numbers on each tie) within $\mathcal{T}_{ABC}$ at time $t+\Delta t$;

In light of these limitations, we aim to understand how people are embedded and interact within a closed social triad over time. Figure 1 shows an illustrative example of the evolution of triadic relationships in social networks. We observe that an open triad $\mathcal{O}_{ABC}$ becomes a closed one $\mathcal{T}_{ABC}$ with the formation of a new link e_{AC} at timestamp t. The goal is to trace the dynamics of triadic interactions within these three users for a short period Δt before and after t. Specifically, we aim to understand the extent to which triadic tie strength dynamics after the formation of the third link can be predicted from social networks.

Our studies could be beneficial to personal recommendation applications in OSN services. For example, if the tie strength dynamics between two users is known, one could recommend appropriate services for one user according to the usage of the other user. Similarity, when we invite a new user to OSN services, we could make more efficient friend recommendations as the estimation for tie strength dynamics would help us determine whether it is beneficial.

Contributions. Employing two types of large dynamic social networks — social media and mobile communication — as the basis of our study, we trace the dynamics of social interactions within social triads and systematically investigate the dynamics of tie strength in social triads over time. The strength of a social tie is measured by its interaction frequency.

Our study unveils the following interesting discoveries. We surprisingly find that although averagely the tie strength in the two dynamic social networks maintains an upward or stable trend (shown in Figure 2), in around 80% of closed social triads, the strength of the first two ties becomes weakened (shown in Figure 3) . We also discover that the stronger (as measured by interaction frequency and reciprocity) the third tie is, the less likely the first two ties are weakened; while the stronger the first two ties are, the more likely they are weakened. In addition, we find that the decrease in tie strength among three males is more sharply than that among females. Finally, we observe that in social media, tie strength between celebrities are more likely to be weakened as the closure of a triad than those between ordinary people.

We then formalize the question of whether tie strength of a triad after closure will become weakened as a triadic tie strength dynamics prediction problem. The prediction task is to infer whether the formation of the third link in a given triad will, within a predefined timeframe, make the interactions of the other two links infrequent. To solve this problem, we propose a triadic tie strength dynamics (TRIST) model — a kernel density estimation (KDE)-based factor graph. As a graphical model, TRIST incorporates not only attribute features but also structural features into a unified framework. Another advantage of the model comes from kernel density estimation, which smoothly models discrete attribute features. The TRIST-ST model is a reduced version of the TRIST model, which utilizes

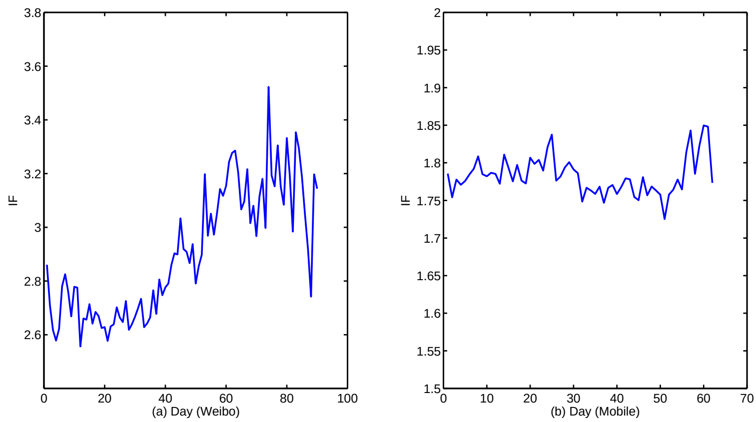


Fig. 2. Trends of the interaction frequency over time. y-axis: Interaction Frequency (#interactions per day). (a) Weibo network; (b) Mobile network.

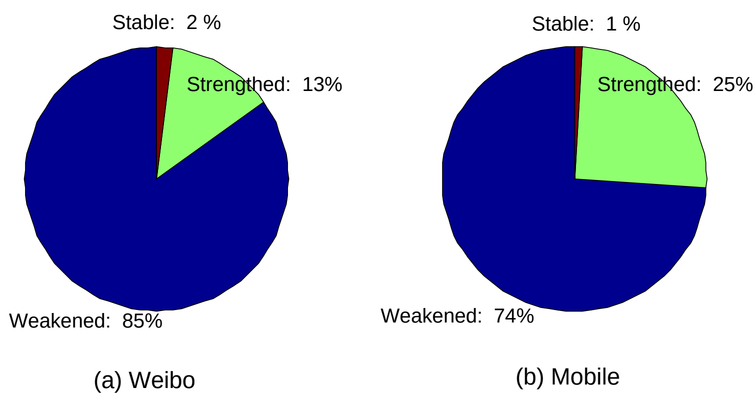


Fig. 3. Tie strength dynamics after triadic closure in social networks. In Weibo, ties in about 85% triads have been weakened while in Mobile ties in 74% triads have been weakened.

only attribute features by kernel density estimation, and ignores structural features. Our experimental results on both types of networks demonstrate that by using the same set of attribute features with logistic regression, SVM, decision trees, and naïve Bayes, the TRIST-ST model improves the prediction performance by up to 10% over the benchmarks due to the leverage of kernel density estimation. By leveraging structural features additionally, the proposed TRIST offers a greater-than-82% potential predictability for triadic tie strength dynamics, and outperforms alternative methods by up to 30%, in terms of F1-score.

Data. We use two types of networks — social media and mobile networks. The social media network comes from Weibo[1], which is the most popular microblogging service in China, with more than 560 million users. The Weibo dataset we use contains more than 1 million users and more than 308 million following relationships (links) [Huang et al. 2014].

[1] http://weibo.com

Table 1. Data statistics

Item	Weibo	Mobile
#Users	1,776,950	194,526
#Links	308,489,739	206,934
#Interactions	15,827,764	3,908,132
#Open Triads	241,364,986	35,314,058
#Closed Triads	21,622,013	2,259,480
#Closed Triads with reciprocal links	954,440	532,308

These users generated more than 12 million retweeting records from September 28^{th}, 2012 to October 29^{th}, 2012. For each user, we have the demographic information — gender and verified status — in their online profiles. The mobile network dataset is extracted from a subset of a collection of two-month mobile call detail records from an anonymous country. In this data, each user is anonymized by the data provider. We construct a sub-network by viewing each user as a node, and connecting a link between two users if they have at least one call from the two-month observation window. This resultant mobile sub-network (Mobile) contains 194,526 users and 206,934 links. Table 1 details the statistics of the two networks.

Organization. Section 2 formalizes the triadic tie strength dynamics prediction problem. Section 3 demonstrates the dynamics status of triadic relationships in social networks and investigates various factors that drive triadic tie strength dynamics. The proposed TRIST model is introduced in Section 4 and its prediction performance is presented in Section 5. Related work is summarized in Section 6, with our conclusion and recommendations for future work provided in Section 7.

2 PROBLEM DEFINITION

Let $G^t = (V^t, E^t, I^t)$ denote a directed and weighted network at time t, where $V^t = \{v_i\}$ is the set of users, $E^t \subset V^t \times V^t$ is the set of links between users, with each link denoted as $e_{ij} = (v_i, v_j) \in E^t$. Then we can define a dynamic network $G = \{G^t = (V^t, E^t), t \in \{1, \cdots, t, \cdots, T\}\}$ over a timeframe T. To make it more readable, we summarize all notations used in this paper in Table 2.

Definition 2.1. **Closed and Open Triads.** Given three users A, B, $C \in V^t$, if there exists a link between any two users – e_{AB}, e_{BC}, $e_{AC} \in E^t$, we say that A, B, and C form a closed triad $\mathcal{T}_{ABC}$ at timestamp t; if there are only two links – e.g., e_{AB}, $e_{BC} \in E^t$ – then A, B, and C form an open triad $\mathcal{O}_{ABC}$ in which users A and C are disconnected.

While extensive efforts have been devoted to explaining triadic relationships in social networks, such as social balance [Heider 1958] and social status theories [Davis and Leinhardt 1972], little has been done to understand triadic interaction dynamics. In this work, we aim to model *the interaction dynamics of links e_{AB} and e_{BC} – when and after the third link e_{AC} is formed by which an open triad $\mathcal{O}_{ABC}$ becomes a closed triad $\mathcal{T}_{ABC}$.*

Suppose at timestamp t, the formation of edge e_{AC} turns an open triad $\mathcal{O}_{ABC}$ into a closed triad $\mathcal{T}_{ABC}$. Our goal is to trace the evolution of interactions in this triad when and after the transition happens. More specifically, we investigate the changes of the interaction frequencies of links e_{AB} and e_{BC} within a timeframe $[t-\Delta t, t+\Delta t]$, where Δt is an observation window. The interaction frequency IF_e is simply defined as the average interaction times, i.e., $IF_e^{[t-\Delta t, t+\Delta t]}/2\Delta t$.

Definition 2.2. **Weakened or Strengthened Tie.** Given a closed triad $\mathcal{T}_{ABC}$ with the third link e_{AC} formed at timestamp t, and the interaction histories of e_{AB} and e_{BC}, in a future timestamp t' ($t' = t + \Delta t$, $\Delta t \geq 0$), we call tie IF_e is weakened within the timeframe $[t - \Delta t, t + \Delta t]$ if the interaction frequency $IF_e^{[t, t+\Delta t]}$ deceases significantly (with $p < 0.01$, t-test) compared with $IF_e^{[t-\Delta t, t]}$; otherwise, IF_e is strengthened.

In defining so, IF_e could be three different cases, including IF_{AB}, IF_{AC}, or IF_{AB+AC}. We formally define the problem of predicting triadic tie strength dynamics as follows.

PROBLEM 1. **Triadic Tie Strength Dynamics Prediction.** *Given the set of closed triads $\mathcal{T}^t$ who become closed at timestamp t and a future timestamp t' ($t' = t + \Delta t$, $\Delta t \geq 0$), the task is to learn a predictive function:*

$$f : (\{G, \mathcal{T}^t, \boldsymbol{X}\}) \rightarrow Y_{\mathcal{T}}^{t'}$$

where $Y_{\mathcal{T}}^{t'}$ denotes the dynamics status (strengthened vs. weakened) of IF_e within the timeframe $[t - \Delta t, t + \Delta t]$, with $y_i = 0$ indicating tie IF_e strengthened and $y_i = 1$ indicating that IF_e is weakened, and $\mathbf{X}$ is an attribute matrix associated with closed triads.

The problem is formalized in triads with directed links; there are in total 27 different types of directed triads. In this work we consider the representative directed triangles with limiting e_{AB} and e_{BC} as reciprocal links, and leave the remaining cases for future work. The reason comes from the fact that reciprocal relationships are considered friendships or social relationships in social media [Kwak et al. 2010; Lou et al. 2013] or mobile communications [Dong et al. 2014; Onnela et al. 2007].

3 TRIADIC TIE STRENGTH DYNAMICS

In this section, we first discern the degree to which triadic relationships in social networks are weakened. We then move on to examine how different factors influence the dynamics status of social triadic relationships in different networks. Specifically, given a closed triad $\mathcal{T}_{ABC}$ that becomes closed by the formation of e_{AC}, we investigate the following observations and factors:

- *Tie Dynamics in Triads:* What is the interaction dynamics of ties in social triads?
- *Social Tie:* How do the strength and reciprocity of the new tie e_{AC} and existing ties e_{AB} and e_{BC} affect the dynamics status of IF_e?
- *User Demographics:* How do users' demographic profiles — gender and social status — influence triadic tie strength dynamics?
- *Temporal Effects:* How long does the formation of link e_{AC} influence triadic tie strength dynamics?

3.1 Tie Dynamics in Triads

We examine the ratios of strengthened and weakened ties in social networks and the degree to which a network as a whole presents weakened. We compare the results with interaction dynamics in both open triads and a network as a whole. We consider all available data in our dataset except those with incomplete information. For example, for Weibo data, we consider 934,733 triads strength dynamics among all the 954,440 closed triads except those 19,707 triads ($\sim 2.06\%$) that lack of gender and verified status information..

Table 2. Notations

Notation	Meaning
G^t	A network at time t
V^t	A set of users
E^t	A set of links
e_{ij}	A link connecting node i and j
$\mathcal{T}_{ABC}$	A closed triad with three nodes being A, B and C
$\mathcal{O}_{ABC}$	An open triad with three nodes being A, B and C
IF_e	Interaction frequency of link e
Δt	An Observation window
$\mathbf{X}$	An attribute matrix associated with closed triads
$Y_{\mathcal{T}}^{t'}$	Dynamics status of triad $\mathcal{T}_{ABC}$ in time t'
$f(\mathbf{x_i^t}, y_i^t)$	The probability of a tie's dynamics state y_i^t at time t given $\mathbf{x_i^t}$ associated with $\mathcal{T}_i$
$g(y_i^t, y_i^{t'})$	The probability of a tie's dynamics state $y_i^{t'}$ at time t' given its state y_i^t at time t
$h(y_i^t, y_j^t)$	The probability of a tie's dynamics state y_i^t at time t given y_j^t of a triad $\mathcal{T}_j$
$f(\cdot)$	An attribute factor function
$g(\cdot)$	A temporal factor function
$h(\cdot)$	A social factor function
$\Phi(\cdot)$	An attribute feature function
K	The number of features
$\kappa(\cdot)$	A kernel function
$\Psi(\cdot)$	A temporal feature function
α, β, γ	The weight for feature functions
Z	Normalization factor
$\mathcal{R}$	Objective function
η	Learning rate

Suppose we have dynamic networks $G = \{G^t = (V^t, E^t), t \in \{1, \cdots, t, \cdots, T\}\}$ with T observed days in real datasets, where $T = 90$ in the Weibo network and $T = 60$ in the Mobile network. We enumerate all combinations $< t, \Delta t >$ of t and Δt with $t \geq \Delta t$ and $t + \Delta t \leq T$. Given $< t, \Delta t >$, we study the closed triads that become closed at t (Definition 2.1) and report the average percentage of strengthened and weakened ties conditioned on Δt (Definition 2.2). Figure 4 plots the interaction dynamics of social triads in Weibo and Mobile.

In Figure 4, we can clearly see that in around 80% of closed triads the ties maintain a weakened state, triggered by the formation of the third link. While the number of the two kinds of strengthened ties is small, the increasingly strengthened ties (red lines) are consistently more numerous than the non-changing strengthened ones (green lines). Specifically, we observe that the interaction dynamics of link e_{AB} in Figure 4 (b,e) is a little weaker than that of e_{BC} in Figure 4 (c,f). Overall, the triadic relationships reveal that tie strength tends to become weakened in both the social media and mobile social networks.

We now compare the observations with the interaction dynamics in open triads to examine whether the dynamics status of a closed triad $\mathcal{T}_{ABC}$ arises from the formation of link e_{AC}. Similar to the experiments in Figure 4, we study the interaction dynamics of open triads for each $< t, \Delta t >$. Figure 5 plots the interaction dynamics of open triads in the Weibo and Mobile.

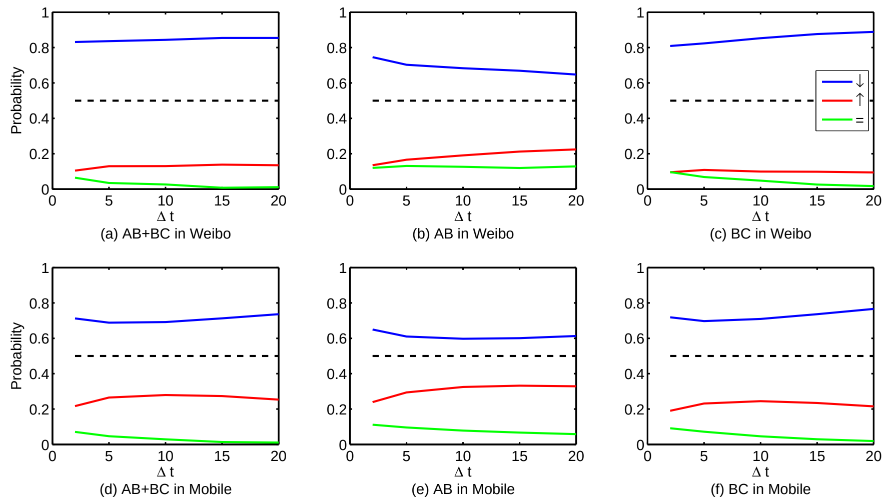


Fig. 4. Triadic tie strength dynamics of closed triads. x-axis: Δt in Definition 2.2; y-axis: Probability that a IF_e is strengthened (red and green lines) or weakened (blue line), conditioned on interaction dynamics of IF_{AB+BC} (a, d), IF_{AB} (b, e), and IF_{BC} (c, f).

We can see the differences of corresponding interaction dynamics between closed triads (Figure 4) and open triads (Figure 5). In Weibo, the probabilities that links are weakened in open triads vs. closed triads are 60% vs. 80% in Figures 4 (a) and 5 (a), 15% vs. 70% in Figures 4 (b) and 5 (b), 50% vs. 80% in Figures 4 (c) and 5 (c). We can also see that in Mobile, the weakened probabilities in Figures 4 (d, e, f) are higher than those in Figure 5 (d, e, f). Conversely, the remaining strengthened cases in open triads are more than those in closed triads. In this sense, we conclude that the formation of the third link e_{AC} activates weakened ties in triads.

3.2 Social Ties

We study how link e_{AC}'s attributes — such as tie strength and reciprocity — influence the dynamics status of ties in a closed triad $\mathcal{T}_{ABC}$. Although e_{AC} belongs to the triad $\mathcal{T}_{ABC}$, we refer to it as an external tie because our goal is to study e_{AC}'s influence on e_{AB} and e_{BC} with its establishment.

Tie Strength of e_{AC}. Tie strength represents the extent of closeness of social relationships. We measure the strength of social ties by the number of interactions between two users in Weibo (#retweets and #comments) and Mobile (#phone-calls) [Dong et al. 2014; Gilbert 2012; Onnela et al. 2007]. Such a definition suggests a way of answering the following question: How does the strength of the newly formed link e_{AC} affect the dynamics status of IF_e in $\mathcal{T}_{ABC}$? Figure 6 plots the influence of link e_{AC} with different tie strength (indicated by different colors) in Weibo and Mobile. First, we observe that both networks present similar patterns of triadic tie strength dynamics. Second, surprisingly, we find that as e_{AC}'s tie strength increases, the likelihood that tie strengthens in closed triads $\mathcal{T}_{ABC}$ increases (*blue* to *red* to *green* lines). In other words, frequent interactions of the newly formed link between A and C promote the stronger ties in the closed triad $\mathcal{T}_{ABC}$.

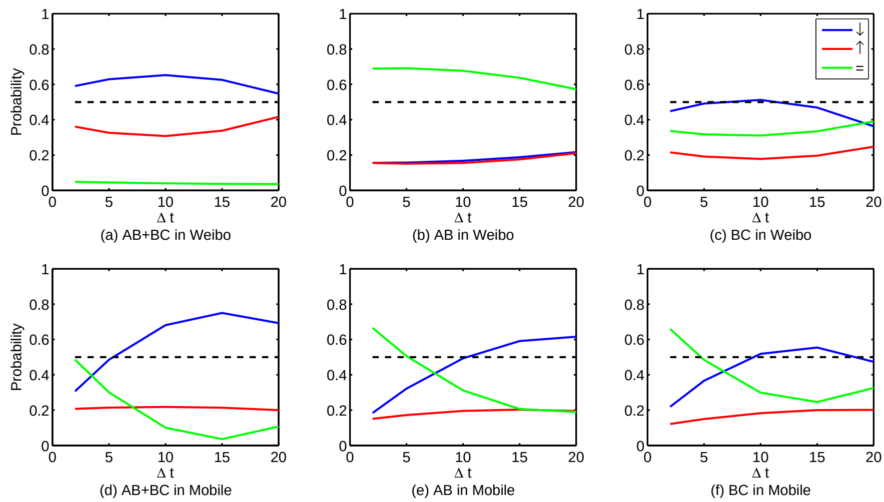


Fig. 5. Triadic tie strength dynamics of open triads. x-axis: Δt; y-axis: Probability that the interaction frequency of IF_e in an open triad is weakened (blue line), strengthened — increasing (red line), or strengthened — decreasing (green line), conditioned on the interaction dynamics of IF_{AB+BC} (a, d), IF_{AB} (b, e), and IF_{BC} (c, f).

Reciprocity of e_{AC}. A reciprocal (two-way) relationship, usually developed from a parasocial (one-way) relationship, represents a stronger or trustful relationship between users in social media [Kwak et al. 2010; Lou et al. 2013] and mobile communications [Dong et al. 2014; Onnela et al. 2007]. We examine the extent to which the formation of link e_{AC} as a parasocial (one-way) and a reciprocal (two-way) relationship can affect the dynamics status of the closed triad $\mathcal{T}_{ABC}$. Figure 7 reports the results of triadic tie strength dynamics conditioned on the reciprocity of e_{AC}. From this figure, we see that while there are only slight differences between parasocial and reciprocal relationships, the parasocial e_{AC} consistently shows more positive effects on activating weakened ties in $\mathcal{T}_{ABC}$.

Tie Strength of e_{AB} and e_{BC}. We investigate how the tie strength of e_{AB} and e_{BC} before the formation of link e_{AC} influence the tie dynamics in $\mathcal{T}_{ABC}$ after the establishment of e_{AC}. Relative to the external tie (e_{AC}), we refer to e_{AB} and e_{BC} as internal ties in $\mathcal{T}_{ABC}$. Our intuition is that with strong internal social ties, three users tend to maintain strengthened triadic relationships. Figure 8 details the results. Generally, we can see that in both Weibo and Mobile, internal tie strength and triadic tie strength dynamics have a negative correlation; that is, ties in a social triad have a high probability to transit to a weakened state if it has strong internal ties before its closeness. The observation that is against our intuition indicates that three people with strong connections in open triadic relationships have the tendency to disperse their social focus and investment to the newly connected social tie that actually makes this open triad closed.

3.3 User Demographics

We investigate the interplay of triadic tie strength dynamics and user demographic profiles in social networks. Due to the unavailability of mobile users' demographic information, in this

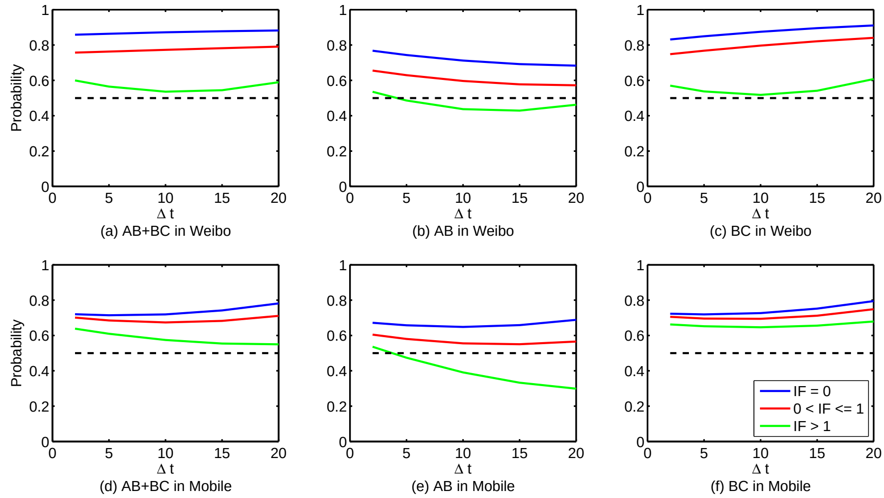


Fig. 6. External tie strength (e_{AC}). x-axis: Δt in Definition 2.2; y-axis: Probability that tie IF_e in $\mathcal{T}_{ABC}$ is weakened, conditioned on interaction dynamics of IF_{AB+BC} (a, d), IF_{AB} (b, e), and IF_{BC} (c, f). IF denotes the average number of interactions (#retweets and #comments in Weibo and #phone-calls in Mobile) per day.

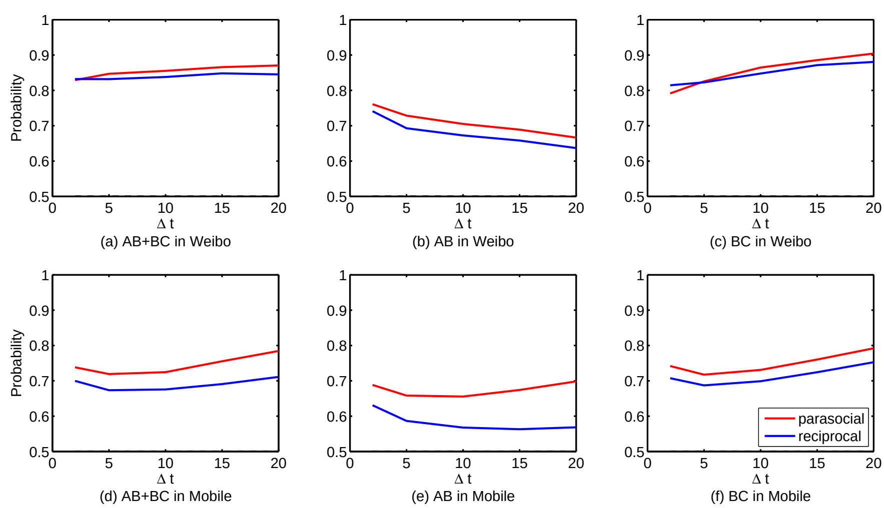


Fig. 7. Reciprocity of external tie e_{AC}. x-axis: Δt in Definition 2.2; y-axis: Probability that a tie IF_e in $\mathcal{T}_{ABC}$ is weakened, conditioned on interaction dynamics of IF_{AB+BC} (a, d), IF_{AB} (b, e), and IF_{BC} (c, f).

study we focus on Weibo. Specifically, we examine how users' gender and status correlate with the dynamics status of triadic relationships.

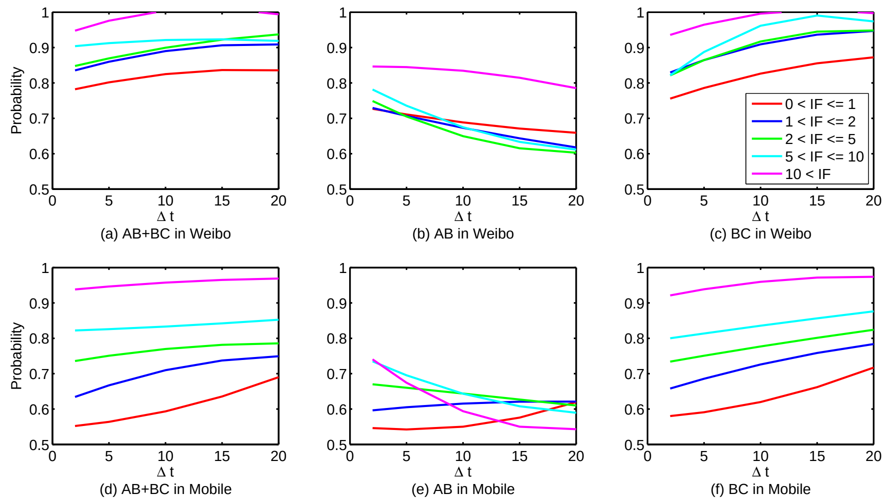


Fig. 8. Internal tie strength (e_{AB} and e_{BC}). x-axis: Δt in Definition 2.2; y-axis: Probability that a tie IF_e in $\mathcal{T}_{ABC}$ is weakened, conditioned on interaction dynamics of IF_{AB+BC} (a, d), IF_{AB} (b, e), and IF_{BC} (c, f). IF denotes the average number of interactions (#retweets and #comments in Weibo and #phone-calls in Mobile) per day.

Gender. Previous studies have revealed that females and males display different social behaviors and activities [Dong et al. 2014; Leskovec and Horvitz 2008]. Herein, we explore how people of different genders maintain their triadic social connections. Given a triad $\mathcal{T}_{ABC}$, we use a three-bit binary code XXX (X=F or M denotes a female or male user) to represent the gender information of its three users A, B, and C, respectively. We can enumerate eight different combinations of three users' gender. In Figure 9, we report the results of four special cases in our problem, i.e., *FFF, FMF, MFM,* and *MMM*. We can observe that different gender-based triads reveal different dynamics status. Generally, the decrease in tie strength among three males is more sharply than that among females. In opposite-gender triads, two females and one male (*FMF* in which the male serves as the bridge user B) have the tendency to maintain relatively strengthened relationships, compared with triads with two males and one female (*MFM*). Overall, we conclude that triadic relationships with more females (*FFF* and *FMF*) tend to be stronger compared with the relationships with more males (*MMM* and *MFM*).

Status. We now look at the effects of users' social status on triadic tie strength dynamics. Weibo.com provides a service for "celebrities" [2] to verify their real-world status, such as CEO, sports star, professor, and so on. We also use a three-bit binary code XXX ($X = 1$ or 0 denotes a verified celebrity or not) to represent the status information of three users A, B, and C in a triad. Figure 10 shows the dynamics status of triadic relationships conditioned on three users' social status. First, we can see that tie strength between celebrities are more likely to be weakened as the closure of a triad than those between ordinary people. Similar to the gender-based observations above, we examine the triadic tie strength dynamics conditioned on the status of the bridge user B (*101* vs. *010*). The results show that the

[2] We have also used degree and pagerank to category "celebrities" and got similar results.

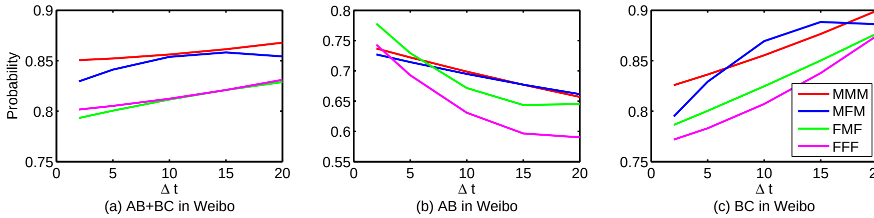


Fig. 9. User gender correlation in Weibo. F: female user; M: male user. x-axis: Δt in Definition 2.2; y-axis: Probability that a tie IF_e in $\mathcal{T}_{ABC}$ is weakened, conditioned on interaction dynamics of IF_{AB+BC} (a, d), IF_{AB} (b, e), and IF_{BC} (c, f).

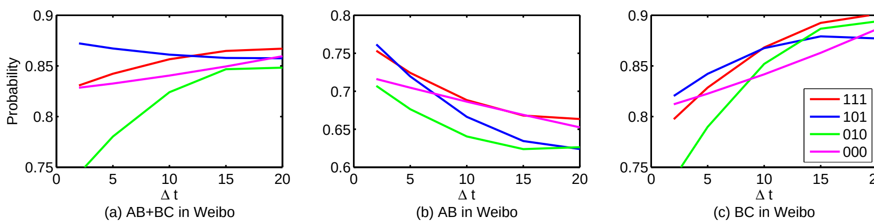


Fig. 10. User status correlation in Weibo. 1: verified celebrity; 0: ordinary user; x-axis: Δt in Definition 2.2; y-axis: Probability that a tie IF_e in $\mathcal{T}_{ABC}$ is weakened, conditioned on interaction dynamics of IF_{AB+BC} (a, d), IF_{AB} (b, e), and IF_{BC} (c, f).

triadic relationships maintained by one celebrity and two ordinary people (*010*) are stronger than those in the reverse case (*101*). Overall, we conclude that the triadic relationships among celebrities (*111* and *101*) tend to be weaker compared with the relationships among ordinary people (*000* and *010*).

3.4 Temporal Effects

We finally study how the length of the observation timeframe — Δt — influences the dynamics status of triadic relationships. Recall that in Definition 2.2, we use the interaction frequency within the timeframe Δt to determine the triadic tie strength dynamics.

From Figures 4, 3, 5 – 10, they are all plotted conditioned on the length of Δt (x-axis). Generally, we can see that when examining triadic tie strength dynamics conditioned on the interaction dynamics of e_{AB+BC} — subfigures (a, d) in Figures 4, 3, 5 – 10, the dynamics status of triadic relationships remains relatively horizontal as the increase of Δt. Specifically, we notice that in Figures 4 – 10 the overall increasing trends of triadic tie strength dynamics conditioned on e_{AB+BC} — subfigures (a, d) — come from the balance between the decreasing trends of triadic tie strength dynamics conditioned on e_{AB} — subfigures (b, e) — and the increasing trends conditioned on e_{BC} — subfigures (c, f) — as the length of Δt increases, respectively.

3.5 Summary

We conduct a regression analysis for the effects of different factors — user demographics and social ties — on triadic tie strength dynamics in Table 3. In this analysis, the ordinary least square model is used to model the relationships between the dependent variable (triadic tie strength dynamics) and the independent variables (Column 1). It can be seen that the reciprocity of link e_{AC} plays more important role in making ties a triad $\mathcal{T}_{ABC}$ stay weakened than its strength (2^{nd} line vs. 3^{rd} line). We also observe that in most cases, the demographics

Table 3. Regression Analysis for triadic tie strength dynamics

	Dynamics status on IF_{AB}	Dynamics status on IF_{BC}	Dynamics status on IF_{AB+BC}
Reciprocity of e_{AC}	0.796 (0.015)	$8.23e$-03** (0.013)	0.321 (0.013)
IF_{AC}	$6.32e$-09*** (0.011)	$5.02e$-09*** (0.010)	$1.87e$-11*** (0.009)
IF_{AB+BC}	$2.04e$-04*** (0.004)	$4.83e$-05*** (0.003)	$1.67e$-07*** (0.003)
Gender of A	$5.49e$-03** (0.019)	0.593 (0.017)	0.538 (0.016)
Gender of B	0.197 (0.020)	0.833 (0.018)	0.610 (0.017)
Gender of C	0.718 (0.019)	$2.55e$-03** (0.017)	$3.14e$-03** (0.016)
Status of A	0.048* (0.019)	0.875 (0.017)	0.087# (0.017)
Status of B	0.561 (0.020)	0.232 (0.018)	0.233 (0.017)
Status of C	0.695 (0.019)	$3.55e$-03** (0.017)	0.010* (0.016)
R^2	0.019	0.023	0.029

Note: Robust standard errors in parentheses.
R^2: the proportion of variance in the criterion that is explained by the estimated regression model.
Two-sided p-value are reported and its significant level at: 0.1 (#), 0.05 (*), 0.01 (**), 0.001(***).

(gender and status) of three users are highly correlated with the tie dynamics. The regression results are consistent with the observations above.

According to the correlation and regression analysis above, we provide the following intuitions related to triadic tie strength dynamics in social networks:

- The triadic relationships are strongly weakened in both online social media and mobile social networks.
- The stronger the third tie is, the less likely the first two ties are weakened; while the stronger the first two ties are, the more likely they are weakened.
- The decrease in tie strength among three males is more sharply than that among females.
- Tie strength between celebrities are more likely to be weakened as the closure of a triad than those between ordinary people.

4 TRIST MODEL FRAMEWORK

Our goal is to examine the extent to which the triadic tie dynamics status can be predicted in social networks. To do so, we propose a unified model to capture not only triads' attributes but also social and temporal correlations. In this section, the TRIST framework – a KDE-based Factor Graph (KFG) – is proposed for predicting triadic tie strength dynamics.

4.1 KDE-based Factor Graph (KFG)

Given a dynamic network $G = \{G^t = (V^t, E^t), t \in \{1, \cdots, t'\}\}$ and the attribute features $\mathbf{X}$ of candidate triads, we define an objective function by maximizing the conditional probability of triadic tie dynamics state Y, i.e., $P(Y|\mathbf{X}, G)$. In a factor graph [Kschischang et al. 2001], the global probability can be factored as a product of local factor functions that capture both the attribute features $\mathbf{X}$ and structural and temporal correlations in the dynamic network G. Herein, we design three types of factor functions to model the observations in Section 3.

- Attribute factor $f(\mathbf{x_i^t}, y_i^t)$: The probability of a tie's dynamics state y_i^t at time t given the attribute vector $\mathbf{x_i^t}$ associated with each triad $\mathcal{T}_i$.
- Temporal factor $g(y_i^t, y_i^{t'})$, $t < t'$: The probability of a tie's dynamics state $y_i^{t'}$ at time t' given its state y_i^t at time t.
- Social factor $h(y_i^t, y_j^t)$: The probability of a tie's dynamics state y_i^t at time t given the dynamics status state y_j^t of a triad $\mathcal{T}_j$.

Thus, we can define the joint distribution over the triadic tie strength dynamics Y given G as

$$P(Y|\mathbf{X}, G) \propto \prod P(\mathbf{X}|Y) \cdot P(Y|G) \tag{1}$$

where $P(Y|G)$ denotes the probability of labels, given the structure of the network and its temporal dynamics, and $P(\mathbf{X}|Y)$ denotes the probability of generating the attributes X associated with each triad $Y_{\mathcal{T}}^t$, given their label Y .

Assuming that the generative probability of attributes, given the label of each triad, is conditionally independent, then

$$P(\mathbf{X}|Y) \propto \prod_i f(\mathbf{x}_i^t | y_i^t) \tag{2}$$

where $f(\mathbf{x}_i^t | y_i^t)$ is the probability of generating attributes $\mathbf{x}_i$ given the label y_i.

Similarly, assuming that the probability of labels, given the structure of the network and its temporal dynamics, is also conditionally independent, then

$$P(Y|G) \propto \prod_{t=1}^{t'} \prod_{j}^{G} g(y_i^t, y_i^{t'}) h(y_i^t, y_j^t) \tag{3}$$

Thus, combining Eq.(2) and Eq. (3), we have

$$\begin{aligned} P(Y|\mathbf{X}, G) &\propto \prod P(\mathbf{X}|Y) \cdot P(Y|G) \\ &= \prod_{t=1}^{t'} \prod_{i,j}^{G} f(\mathbf{x_i^t}, y_i^t) g(y_i^t, y_i^{t'}) h(y_i^t, y_j^t) \end{aligned} \tag{4}$$

We illustrate the graphical representation of our proposed model in Figure 11. The bottom left part is the input network. From the input social network, we generate three closed candidate triads, including $\mathcal{T}_{v_1,v_2,v_3}$, $\mathcal{T}_{v_1,v_3,v_4}$ and $\mathcal{T}_{v_1,v_4,v_5}$. In the prediction model, these three candidate triads are modeled as yellow ellipses. The attribute features defined over the candidate triads at each timestamp are captured by the $f(\cdot)$ factor functions. The temporal correlations between triads at different timestamps are modeled by the $g(\cdot)$ factor functions. The social correlations between different triads are captured by the $h(\cdot)$ factor functions. Based on all the considerations, we construct the factor graph at the top level of Figure 11.

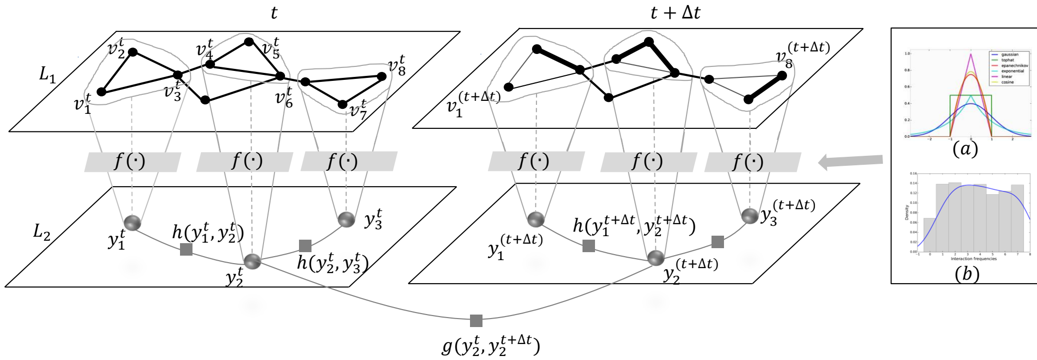


Fig. 11. Graphical representation of the TRIST model. Open triads become closed at time t. Three candidate triads in network layer L_1 are mapped into three hidden variable nodes $y_i^t, i = 1, 2, 3$ in factor graph layer L_2. We observe two timestamps t and $t + \Delta t$ ($\Delta t > 0$). The broader the line is, the stronger the tie is. $h(\cdot)$ represents the social correlation function between two triads, and $g(\cdot)$ represents the temporal correlation function between the two statuses of one triad at two timestamps. In order to get attribute factor function, we applied kernel density estimation $f(\cdot)$. An example of the histogram and kernel density estimation of the tie strength of e_{AB} and e_{BC} using a Gaussian kernel is shown in (b). y-axis: density; x-axis: interaction frequencies of ties e_{AB} and e_{BC}. In (a), we present different kernel functions.

KDE-based attribute factor. Straightforwardly, we initialize the defined factors in a Markov random field based on the Hammersley-Clifford theorem [Hammersley and Clifford 1971]. In particular, we initialize the attribute factor as

$$f(\mathbf{x_i^t}, y_i^t) = \frac{1}{Z_1} \exp\{\sum_{k=1}^{K} \alpha_k \Phi(x_{ik}^t, y_i^t)\} \tag{5}$$

where α_k is the weight of the k^{th} attribute, K is the number of features, and $\Phi(\cdot)$ is the attribute feature function.

There are various ways to instantiate the attribute feature function $\Phi(\cdot)$, i.e., binary function [Huang et al. 2014; Lou et al. 2013]. However, a binary feature function cannot accurately capture correlations and similarities among features and would lose important information. In order to better capture these similarities, we propose using kernel density estimation to instantiate attribute factor functions.

Kernel density estimation (KDE) [Wasserman 2004], without any priori information on the probability distribution of the dataset, using a non-parametric way to estimate random variables' density functions, is an attractive technique to obtain estimations [Gerber 2014; Li et al. 2013a; Zhang and Chow 2013]. To obtain a kernel density estimation, we first place a kernel — a smooth, strongly peaked function — at the position of each data point, and then add up the contributions from all the estimations to obtain a smooth curve. For instance, Figure 11 shows an example of the histogram (gray part) and kernel density estimation with a Gaussian kernel (blue curve) for the interaction-strength feature. Specifically, for a Gaussian kernel, we have the following definition:

$$\Phi(x_{ik}^t, y_i^t) = \sum_{n=1}^{N_k} \frac{1}{\lambda} \kappa(\frac{x_{ik} - x_n}{\lambda})$$

Table 4. Kernel functions

Kernel	Function
Gaussian kernel	$\kappa(x) = 1/\sqrt{(2\pi)}\exp(-1/2\ x^2)$
Tophat kernel	$\kappa(x) = 1/2$ if $\lvert x\rvert \leq 1$
Epanechnikov kernel	$\kappa(x) = 3/4(1-x^2)$ if $\lvert x\rvert \leq 1$
Exponential kernel	$\kappa(x) = \exp(-x)$
Linear kernel	$\kappa(x) = 1-x$ if $\lvert x\rvert \leq 1$
Cosine kernel	$\kappa(x) = \pi/4\cos(\pi/2\ x)$ if $\lvert x\rvert \leq 1$

By using kernel density estimation, the attribute feature function can be initialized as

$$f(\mathbf{x_i^t}, y_i^t) = \frac{1}{Z_1}\exp\{\sum_{k=1}^{K}\sum_{n=1}^{N_k}\alpha_k\frac{1}{\lambda}\kappa(\frac{x_{ik}-x_n}{\lambda})\} \tag{6}$$

where $\kappa(\cdot)$ is the kernel function with a peak at x_n , λ is the kernel bandwidth, and N is the number of data points of each feature. We thus get the real values of the attribute features.

Theoretically, we can use any smooth, strongly peaked function as a kernel, if the area under the curve of this kernel equals 1 — which is to make sure that the resulting KDE is normalized. In this work, we consider six commonly used kernel functions [Wasserman 2004] as shown in Table 4. The curves of these functions are shown in Figure 11. The Tophat kernel, the Epanechnikov kernel, the linear kernel, and the cosine kernel are zero outside a finite range, whereas the Gaussian kernel and exponential kernel are nonzero everywhere, but negligibly small outside a limited domain. By default, we use a Gaussian kernel in our TRIST framework and discuss the effects of different kernel choices in experiments.

It is worth mentioning that a KDE-based attribute factor can be used to study graph kernels [Vishwanathan et al. 2010] as well. Intuitively, graph kernels can be considered as functions measuring the similarity of pairs of graphs, making the whole family of kernel methods applicable to graphs [Bach 2008; Shervashidze et al. 2009; Vishwanathan et al. 2010]. In our work, we aim to measure similarity between different candidate closed triads, which also provides a way to measure graph similarity.

Temporal factor. For the temporal factor, we model the interrelations between different weakened states of one triad at different timestamps. Specifically, we define it as:

$$g(y_i^t, y_i^{t'}) = \frac{1}{Z_2}e^{\sigma_1(t'-t)}\exp\{\beta_i\Psi(y_i^{t'}, y_i^t)\} \tag{7}$$

where $e^{\sigma_1(t'-t)}$ is a triad-independent time-increase factor, σ_1 is a pre-defined parameter, $\Psi(\cdot)$ is a temporal feature function defined as $\Psi(\cdot) = (y_i^{t'} - y_i^t)^2$ and β is the weight of $\Psi(\cdot)$. In the model, we only consider the dependency of triadic tie strength dynamics between two subsequent timestamps as the same assumption in a hidden Markov model [Ghahramani and Jordan 1997], Markov Random field [Hammersley and Clifford 1971] and Kalman Filters [Haykin et al. 2001]. The triad-independent time-increase factor $e^{\sigma_1(t'-t)}$ is defined as an exponential function, so that its parameter σ_1 can be simply absorbed by combining it with β_i. Thus, the above temporal factor function can be rewritten as

$$g(y_i^t, y_i^{t'}) = \frac{1}{Z_2}\exp\{\beta_i(t'-t)(y_i^{t'} - y_i^t)^2\} \tag{8}$$

Social factor. Intuitively, a triad's dynamics status state may be influenced by other triads; e.g., its neighborhood. For example, closed triad $\mathcal{T}_{ABC}$ and closed triad $\mathcal{T}_{ABD}$ share a common link e_{AB}, then the dynamics status state of $\mathcal{T}_{ABC}$ is highly correlated with that of $\mathcal{T}_{ABD}$. In addition, individuals have a tendency to associate and bond with similar others according to homophily theory [McPherson et al. 2001]. Individuals in homophilic relationships share common characteristics, which makes similar triads have the similar dynamics status patterns [McPherson et al. 2001].

Similarly, in order to capture the correlations between two triads, we define the social factor function as:

$$h(y_i^t, y_j^t) = \frac{1}{Z_3} \exp\{\gamma_{ij}(y_i^t - y_j^t)^2\} \tag{9}$$

where γ_{ij} is the weight of the feature function, representing the influence degree of y_i on y_j.

Finally, by integrating Eqs. (2), (5), and (6) into Eq. (1), we can obtain joint probability as follows:

$$\begin{aligned} P(Y|\mathbf{X}, G) \propto & \frac{1}{Z} \exp\{\sum_{t=1}^{t'} \sum_{i}^{G} \sum_{k=1}^{K} \alpha_k \Phi_k(x_{ik}^t, y_i^t) \\ & + \sum_{t=1}^{t'} \sum_{i}^{G} \beta_i(t'-t)(y_i^{t'} - y_i^t)^2 + \sum_{t=1}^{t'} \sum_{i,j}^{G} \gamma_{ij}(y_i^t - y_j^t)^2\} \end{aligned} \tag{10}$$

where $Z = Z_1 Z_2 Z_3$ is a normalization factor to guarantee that the result is a valid probability.

4.2 Learning and Prediction

Model Learning. Given the joint probability in Eq. (10), we have the following log-likelihood objective function

$$\mathcal{R}(\theta) = \log P_\theta(Y|\mathbf{X}, G) \tag{11}$$

The task of model learning is to estimate a parameter configuration $\theta = (\{\alpha_k\}, \{\beta_i\}, \{\gamma_{ij}\})$ that maximizes the log-likelihood objective function — i.e.,

$$\theta = \arg\max \mathcal{R}(\theta)$$

To solve the maximization problem, we employ a gradient descent method. The idea is that each parameter θ is assigned an initial value, and then the gradient of each parameter with regard to the objective function is calculated. For example, the gradient of the parameter α_k with regard to Eq. (11) is written as:

$$\frac{\mathcal{R}(\theta)}{\alpha_k} = \mathbb{E}[\Phi_k(x_{ik}^t, y_i^t)] - \mathbb{E}_{P_{\alpha_k}(Y|\mathbf{X})}[\Phi_k(x_{ik}^t, y_i^t)] \tag{12}$$

where $\mathbb{E}[\Phi_j(x_{ik}^t, y_i^t)]$ is the expectation of factor function $\Phi_k(x_{ik}^t, y_i^t)$ given the data distribution; and $\mathbb{E}_{P_{\alpha_k}(Y|\mathbf{X})}[\Phi_k(x_{ik}^t, y_i^t)]$ is the expectation of factor function $\Phi_k(x_{ik}^t, y_i^t)$ under the distribution $P_{\alpha_k}(Y|\mathbf{X})$ estimated by the model. Before calculating variables' marginal distribution, we need first get factor functions for different factors; i.e., we run KDE to get estimates for attribute factors, and get temporal factors from two subsequent timestamps in the whole period we observed.

Similarly, the gradients of parameters β_i and γ_{ij} can be derived. Finally each parameter is updated with a learning rate η:

Algorithm 1: Learning algorithm for the TRIST model.

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

Initialize $\theta \leftarrow 0$;
repeat
  repeat
    Get temporal factors between timestamp t and its last timestamp $t-1$;
    Update t;
  until $t \geqslant T$;
  Run KDE to get estimates of attribute factors:

$$f(\mathbf{x_i^t}, y_i^t) = \frac{1}{Z_1} \exp\{\sum_{k=1}^{K} \sum_{n=1}^{N_k} \alpha_k \frac{1}{\lambda} \kappa(\frac{x_{ik} - x_n}{\lambda})\}$$

  Perform LBP to calculate marginal distribution of unknown variables $P(Y, Y^L|G)$;
  Perform LBP to calculate the marginal distribution of known variables $P(Y|G)$;
  Calculate the gradient of α_j according to Eq. (12) (for β_i and γ_{ij} with a similar formula);
  Update parameter θ with the learning rate η with Eq. (13);
until *Convergence*;

$$\theta_{m+1} = \theta_m + \eta \cdot \frac{\mathcal{R}(\theta)}{\theta} \tag{13}$$

where m is the iteration time. The learning algorithm is summarized in Algorithm 1.

Feature Definition. We now describe how we define the factor functions in our model. According to the analysis in the previous section, we define factor functions of three categories: an attribute factor, a temporal factor and a social factor.

Attribute factor. Attribute factors include social tie and user demographics. For social tie, we define a feature for e_{AC}'s reciprocity indicating whether e_{AC} is reciprocal or not, a feature for the tie strength of e_{AC} denoting the interaction frequency between users A and C, and also a feature for tie strength of e_{AB} and e_{BC}. The tie strength features are modeled via kernel density estimation. For Weibo dataset, we also define six features based on user demographics — gender and status of three users — with their kernel density estimates respectively (Cf. Eq. (3)).

Temporal factor. Temporal factors are used to model the interrelations between the states of one triad at different timestamps (Cf. Eq. (5)).

Social factor. We define one correlation function for social factors to see whether any two triads have the same triadic tie strength dynamics patterns (Cf. Eq. (6)).

Prediction. With the estimated parameters θ, we can predict the labels of unknown variables $y_i =?$ by finding a label configuration that maximizes the objective function — i.e., $Y^\star = \arg\max \mathcal{R}(Y|\mathbf{X}, G, \theta)$. Specifically, we use the learned model to calculate the marginal distribution of each candidate triad with unknown variable and finally assign each candidate triad with a label of the maximal probability.

5 TRIADIC TIE STRENGTH DYNAMICS PREDICTION

In this section, we conduct various experiments to demonstrate the effectiveness of our proposed TRIST model and the predictability of triadic tie strength dynamics in social networks.

5.1 Experiment Setup

Settings. The task is to predict whether ties in the candidate triads that become closed at t will become weakened within a timeframe Δt. We set up two prediction cases (Weibo1 and Weibo2) using the Weibo dataset and one case (Mobile) using the Mobile dataset. In Weibo1, we use a network over the first 10 days of the window for experiments ($t = 5$ and $\Delta t = 5$) and predict the dynamics status in day t' ($t' = 10$, $\Delta t = 5$) of newly closed triads (formed on the 5^{th} day). In Weibo2, we incrementally observe the network and use the first 20 days for prediction experiments ($t = 10$ and $\Delta t = 10$), predicting the dynamics status on day t' ($t' = 20$ and $\Delta t = 10$) of closed triads (formed on the 10^{th} day). Using the Mobile dataset, we construct an experimental case Mobile by using the same setting as that of the Weibo1 case. Similar to the observations in Section 3, for each prediction case, we use three different measures to determine the dynamics status of social triads, including the interaction frequencies of both links IF_{AB+BC} together, and also these two links separately IF_{AB} and IF_{BC}.

Evaluation. We use 50% of the observed closed triads as a training set and the remaining triads as a test set. We repeat the prediction experiments ten times, and report the average performance in terms of Accuracy, Precision, Recall, and F1-score.

Comparisons. We compare the proposed TRIST model with four classical classification baselines, including logistic regression, support vector machine (SVM), decision tree (C4.5), and Naïve Bayes. The TRIST model is implemented in C++, and all experiments are performed on a PC running Windows 7 with an AMD Opteron (TM) Processor 6276 (2.3GHz) and 4GB memory. As to the baseline methods, we employ the open-source software Weka with default parameters.

SVM uses all the same features defined in the attribute factors of our TRIST model for each triad to train a corresponding classification model, and then use the classification model to predict triadic tie strength dynamics in the test data.

Logistic Regression uses all the same features defined in the attribute factors of our TRIST model for each triad to train a corresponding classification model, and then use the classification model to predict triadic tie strength dynamics in the test data.

Decision Tree uses all the same features defined in the attribute factors of our TRIST model for each triad to train a corresponding classification model, and then use the classification model to predict triadic tie strength dynamics in the test data.

Naïve Bayes uses all the same features defined in the attribute factors of our TRIST model for each triad to train a corresponding classification model, and then use the classification model to predict triadic tie strength dynamics in the test data.

TRIST represents the proposed model that trains a factor graph model with all three types of factors — attribute factors, social factors, and temporal factors.

TRIST-ST uses all the same features defined in the attribute factors of our TRIST model for each triad. The difference lies in that TRIST-ST leverages kernel-based attribute factors to capture the feature correlations.

Table 5. Triadic tie strength dynamics prediction performance.

Data	Method	IF_{AB+BC}				IF_{AB}				IF_{BC}			
		Accu.	Prec.	Rec.	F1	Accu.	Prec.	Rec.	F1	Accu.	Prec.	Rec.	F1
Weibo1	Logistic	.538	.541	.538	.538	.557	.595	.557	.403	.591	.560	.591	.516
	SVM	.540	.547	.540	.537	.557	.632	.557	.402	.592	.351	.592	.440
	Decision Tree	.527	.530	.527	.527	.553	.545	.553	.438	.589	.565	.589	.555
	NaiveBayes	.535	.538	.535	.535	.556	.563	.556	.405	.566	.551	.566	.554
	TRIST	**.844**	**.940**	**.756**	**.838**	**.859**	**.944**	**.756**	**.840**	**.864**	**.972**	**.782**	**.867**
	TRIST-ST	.541	.559	.579	.569	.554	.727	.006	.013	.580	.585	.949	.724
Weibo2	Logistic	.535	.534	.535	.534	.575	.634	.575	.424	.583	.540	.583	.531
	SVM	.537	.540	.537	.537	.573	.328	.573	.417	.614	.604	.614	.491
	Decision Tree	.532	.532	.532	.532	.556	.502	.556	.462	.590	.558	.590	.553
	NaiveBayes	.545	.544	.545	.544	.570	.520	.570	.432	.574	.566	.574	.569
	TRIST	**.809**	**.889**	**.723**	**.797**	**.797**	**.821**	**.702**	**.757**	**.826**	**.947**	**.749**	**.836**
	TRIST-ST	.540	.562	.599	.580	.553	.818	.007	.014	.597	.603	.939	.735
Mobile	Logistic	.601	.558	.601	.510	.574	.572	.574	.573	.608	.557	.608	.516
	SVM	.605	.366	.605	.456	.568	.560	.568	.478	.615	.378	.615	.468
	Decision Tree	.605	.366	.605	.456	.570	.563	.570	.479	.615	.378	.615	.468
	NaiveBayes	.604	.365	.604	.455	.568	.562	.568	.562	.615	.570	.615	.468
	TRIST	**.750**	**.738**	**.911**	**.815**	**.705**	**.750**	**.714**	**.732**	**.764**	**.756**	**.907**	**.824**
	TRIST-ST	.604	.614	.937	.742	.573	.618	.630	.626	.610	.620	.937	.746

TRIST-A, *TRIST-T*, and *TRIST-S* are three reduced versions of our TRIST model that ignore the corresponding factors.

5.2 Prediction Results

Performance. Table 5 reports the prediction performance of triadic tie strength dynamics as measured by interaction frequencies IF_{AB+BC}, IF_{AB}, and IF_{BC}. Generally, the results in both Weibo and Mobile show that our TRIST model clearly outperforms the baseline methods in each case.

In terms of F1-score, the proposed TRIST model achieves a $(25\%, 44\%)$ improvement compared with baselines when triadic tie strength dynamics is measured by IF_{AB+BC}, a $(16\%, 44\%)$ improvement when measured by IF_{AB}, and a $(26\%, 42\%)$ improvement when measured by IF_{BC}. Similarly, the three tables show that the proposed TRIST model also outperforms other methods significantly in terms of other evaluation metrics — Precision, Recall, and Accuracy.

On average, by using the same prediction settings, the results in the Weibo1 prediction case in Table 5 reveal about $(10\%, 15\%)$ greater predictability of triadic tie strength dynamics in terms of Accuracy than those of Mobile cases. Meanwhile, we also note that the prediction performance with a short Δt (5 days) in the Weibo1 case is better than that with a relatively long Δt (10 days) in the Weibo2 case, in terms of both Accuracy and F1-score.

The reasons that TRIST outperforms Logistic Regression, SVM, Decision Tree, and naïve Bayes come from 1) the modeling of structural correlations of triadic tie strength dynamics captured by temporal and social factors in our TRIST model, and 2) the use of kernel densities for modeling attribute features.

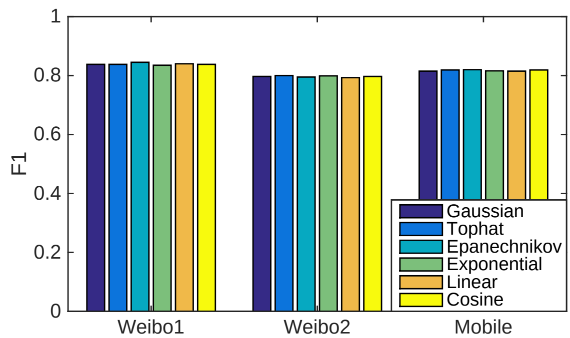
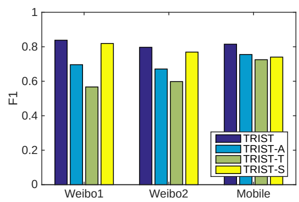


Fig. 12. Effects of different kernel functions in TRIST. X-axis: Different test cases. Y-axis: F1-score. The TRIST model with different kernel functions yields similar prediction performance.

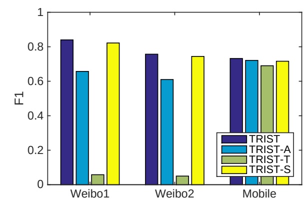
Effects of Kernel Density. We demonstrate the power of TRIST's kernel density by comparing the reduced version of our model TRIST-ST with baseline methods that use the same set of features. The difference lies in that in TRIST-ST, the kernel density is incorporated to smooth the discrete feature values. In Table 5, we can see that in most cases, TRIST-ST yields better prediction performance than the alternative methods. For example, by our methodology, the predictive power of TRIST-ST significantly outperforms the other four methods, with a $(5\%, 23\%)$ increase of F1-score in the Mobile case.

We also examine the effects of different kernel density functions on the prediction power of our method. Specifically, we use six different kernels: Gaussian, Tophat, Epanechnikov, exponential, linear, and cosine. The details of these six kernels are introduced in Section 4.1. The prediction results on triadic tie strength dynamics as measured by IF_{AB+BC} in Weibo1, Weibo2, and Mobile cases are shown in Figure 12. Clearly, we can see that the TRIST model with different kernel functions yields similar prediction performance. We conclude that our TRIST model is robust with respect to the choice of different kernel functions.

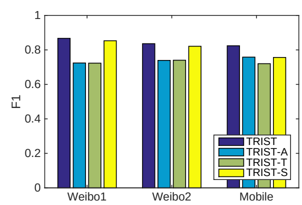
Factor Contribution Analysis. To predict the dynamics status of triadic relationships, we devise three kinds of factors in our proposed model TRIST. To explore the contributions of different factors to the prediction task, we remove each type of factor and keep the remaining two in a series of experiments. For this purpose, we have three versions of our model, i.e., TRIST-A (removing attribute factors), TRIST-T (removing temporal factors), and TRIST-S (removing social factors). The results of the three reduced methods and TRIST are shown in Figure 13. Clearly, we can see that the removal of each type of factor results in a clear drop in the prediction performance. By removing social factors, the F1-score of TRIST-S model deceases slightly compared to the original model TRIST in all prediction cases. However, the drops of F1-score when removing attribute or temporal factors (TRIST-A or TRIST-T) are much more significant than TRIST-S's performance drops, which indicates the importance of attribute and temporal factors in predicting triadic tie strength dynamics. Specifically, TRIST-A achieves better prediction performance than TRIST-T, which means that temporal factors are more telling than attribute factors for predicting the dynamics status of social triads in TRIST model. These experimental results further demonstrate the effectiveness of our TRIST model by modeling the attribute, temporal, and social correlations examined in Section 3.



(a) IF_{AB+BC}



(b) IF_{AB}



(c) IF_{BC}

Fig. 13. Factor contributions for triadic tie strength dynamics Prediction. x-axis: different prediction cases; y-axis: Prediction performance in terms of F1-score. TRIST is the proposed model; TRIST-A is the reduced version of TRIST without modeling attributed factors; TRIST-T is the reduced version of TRIST without modeling temporal factors; TRIST-S is the reduced version of TRIST without modeling social factors.

Training/Test Ratio. We analyze the effects of different training samples on the prediction performance of triadic tie strength dynamics. Figure 14 shows the prediction performance with different ratios of training samples in Weibo1, Weibo2, and Mobile. First, we observe that as training ratios increase, the TRIST model gradually achieves better performance in terms of both F1-score and Accuracy and reaches a stable performance when using only 5% to 10% training samples. Second, we can see that the results in different prediction cases show similar trends to those obtained by increasing the number of training samples, in three sub-figures. The results indicate a positive effect of the size of training data on predicting triadic tie strength dynamics in social networks. At the same time, we also conclude that the predictability of triadic tie strength dynamics can be largely revealed by a small set of labeled triads.

Convergence. We conduct experiments to see the convergence of our TRIST model — the number of iterations of our learning algorithm. The convergence properties of TRIST with different kernel density functions by using the default training data (50%) are plotted in Figures 15 (a), (b), and (c). From Figures 15 (a) and (b) we observe similar convergence patterns, that is, the TRIST model gradually reaches convergence states within 50 to 60 iterations. From Figure 15 (c) we observe that the TRIST model immediately converges when the number of iterations approaches around 70. From the top three sub-figures of Figure 15 we also notice that different kernels have limited effects on TRIST's convergence. By using different ratios of training data, we report the convergence of our proposed model with the default kernel function (Gaussian) in Figures 15 (d), (e), and (f). In the bottom three sub-figures of Figure 15, we again find that the model converges gradually in Weibo cases, and reaches convergence suddenly in Mobile case within 100 iterations. We also find

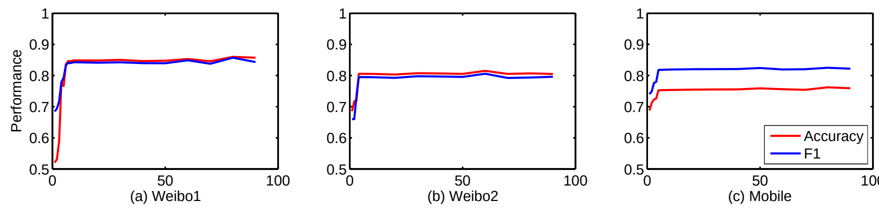


Fig. 14. Performance of triadic tie strength dynamics prediction with different percentage of training data. x-axis: different ratios of training set; y-axis: prediction performance when measuring IF_{AB+BC} using TRIST model in terms of both F1-score (blue line) and Accuracy (red line).

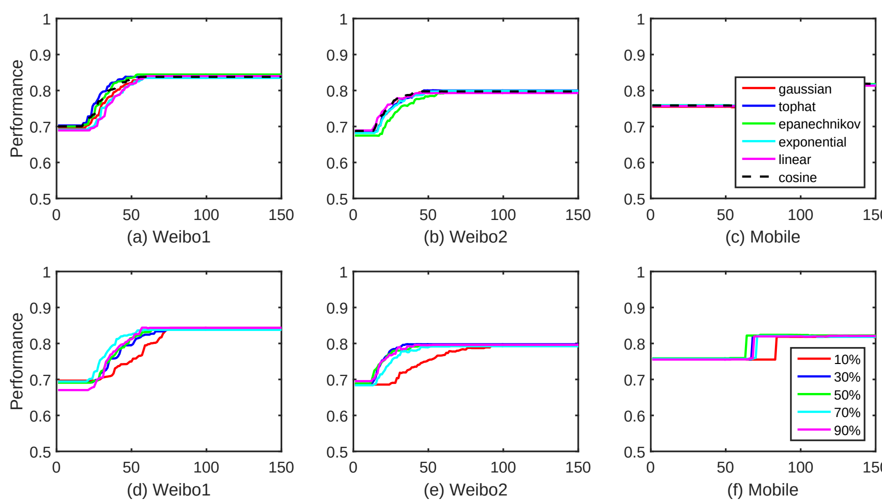


Fig. 15. Convergence of TRIST. x-axis: the iteration time m in Eq. 13 of our learning algorithm; y-axis: F1-score. (a,b,c). Convergence rates with different kernel functions in different prediction cases (50% training data); (d,e,f). Convergence rates with different training data in different prediction cases (Gaussian kernel).

that the TRIST model with 10% training data (red line) has the lowest convergence rate compared with the model with more training data. Overall, Figure 15 demonstrates that the proposed TRIST model can reach the convergence state quickly.

6 RELATED WORK

The social triad, one of the simplest groupings of individuals, serves as the basis of social network analysis, and has been studied extensively in both sociology and network science. The study of social triads can help understand emergent network distributions and social phenomena at the macroscopic level.

The study of "triad" was pioneered by Simmel in 1950 [Simmel 1950]. Since then, sociologists have worked out several profound theories on triadic relationships in social networks. Heider [Heider 1958] developed the theory of social balance, in which the balance state is reached when there are three positive relationships, or two negatives with one positive, in a social triad. Essentially, the balance theory explains the real-world social phenomenon, that is “A friend of my friend is my friend” and “The enemy of my enemy is my friend.” Davis et al. [Davis and Leinhardt 1972] took the theory of social status in directed networks further. Status theory posits that by reversing the direction and flipping the sign to positive of each negative edge in a triad, a social triad complies with status theory if the resulting triad is acyclic. Recently, Leskovec et al. [Leskovec et al. 2010] suggested an alternate theory of status that provides a different organizing principle for signed networks. The social theories developed on triad structure successfully characterize the nature of social behaviors among three people in social networks. However, previous studies are limited by explaining the phenomena of triadic relationships in a static way or relying on the sign of relationships. This work is the first to propose to examine the dynamics status of triadic relationships among three people over time without the sign of relationships.

Besides the well-known social theories, a large body of work has been devoted to modeling and predicting the process of triadic closure [Huang et al. 2015, 2014; Kossinets and Watts 2006; Zignani et al. 2014]. Romero and Kleinberg [Romero and Kleinberg 2010] developed a framework to understand the closure process of social triads in directed networks. Lou et al. [Lou et al. 2013] presented a machine learning model to predict whether an open triad will become closed in Twitter. Fang and Tang [Fang and Tang 2015] recovered the formation process of a closed social triad in social networks. Moreover, much work has demonstrated that triadic closure can be identified as one of the fundamental dynamical principles in network formation and evolution [Klimek and Thurner 2013; Leskovec et al. 2008; Li et al. 2013b]. For example, preferential triadic closure has been proposed to infer link formation across different social networks [Dong et al. 2012]. In addition, triadic closure can benefit many applications in social networks, such as characterizing tie strength [Sintos and Tsaparas 2014], influence diffusion [Zhang et al. 2015], and spam detection [Becchetti et al. 2008]. The major difference between our work and previous work lies in that we focus on the dynamics of triadic relationships over time after a closed triad is formed, while the modeling of triadic closure focuses on the transition from an open triad to a closed one.

Many researchers have adopted tie strength as an analytic framework for studying individuals and organizations [Granovetter 1995; Schaefer et al. 1981] and paid a lot of attention to measuring the tie strength of social relations. Using survey data on friendship ties, Marsden et al. constructed and validated measures of tie strength [Marsden and Campbell 1984]. Krackhardt validated that a ”Simmelian tie” can strengthen the relationships between the individuals in social triads or groups [Krackhardt 1999]. Gilbert and Karahalios proposed a predictive model to map social media data to tie strength and distinguished them into strong and weak ties [Gilbert and Karahalios 2009]. Jones et al. used online interaction data (specifically, Facebook interactions) to successfully identify real-world strong ties [Jones et al. 2013]. Xiang et al. developed an unsupervised model to estimate relationship strength from interactions [Xiang et al. 2010].

On the other hand, less efforts are devoted to tie strength dynamics. Saramäki et al. found that the distribution of people that distributed their social investment over different social ties among their ego networks tended to persist over time [Saramäki et al. 2014]. Patil et al. presented a model to predict whether a group will remain stable or shrink over time [Patil et al. 2013]. Burke and Kraut investigated the factors that associated with tie strength dynamics. They found that tie strength increased with both one-on-one communication, such as posts, comments, and messages, and through reading friends’ broadcasted content, such as status updates and photos [Burke and Kraut 2014]. However, most of them focus on understanding the dynamics status of social ties and communities, or measuring the tie strength in social networks, the structural factors associated with tie strength dynamics are not well addressed. As far as we know, our work is the first to investigate the dynamics status of triadic relationships from a microscopic view in social networks.

7 CONCLUSION

In this work, we discover the interaction dynamics in a triad after closure. By tracing the interaction dynamics in two social networks — Weibo and Mobile — we find that the formation of the third link in a triad will demote the interaction strength of the other two links. We demonstrate that in around 80% of closed social triads, the strength of the first two ties become weakened in both online social media and mobile social networks. We also uncover an interesting phenomenon, that is, both males and celebrities tend to maintain more weakened triadic relationships than females and ordinary users.

By experimenting on these two datasets, we formalize the triadic tie strength dynamics prediction problem, and present a TRIST model to solve it by incorporating user demographics and temporal and structural correlations. Extensive experimental results show that our proposed model outperforms comparison methods by up to 30% in terms of F1-score. Furthermore, we demonstrate that our methodology offers a greater-than-82% potential predictability for inferring the dynamics status of social triads in both networks.

Despite the rich set of results on triadic tie strength dynamics, there is still much room for future work. First, the underlying mechanism of the evolution of triadic relationships is still largely untouched. Second, we need to connect the microscopic triadic principles with network scaling phenomena at the macro level. Finally, it is also necessary to examine the dynamics status of social triads in other types of networks, such as collaboration networks, Facebook, location-based social networks, and so on.

ACKNOWLEDGMENTS

Hong Huang is supported by the International Science and Technology Cooperation Program of China (No. 2015DFE12860), NSFC (No. 61433019, U1435217). Jie Tang is supported by National Natural Science Foundation of China (No. 61631013, No. 61561130160). Xiaoming Fu is supported by Lindemann Foundation.

REFERENCES

Francis R Bach. 2008. Graph kernels between point clouds. In *Proceedings of the 25th international conference on Machine learning*. ACM, 25–32.

Luca Becchetti, Paolo Boldi, Carlos Castillo, and Aristides Gionis. 2008. Efficient semi-streaming algorithms for local triangle counting in massive graphs. In *KDD'08*. 16–24.

Moira Burke and Robert E Kraut. 2014. Growing closer on facebook: changes in tie strength through social network site use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 4187–4196.

Theodore Caplow. 1968. Two against one: Coalitions in triads. (1968).

James A. Davis and Samuel Leinhardt. 1972. The structure of positive interpersonal relations in small groups. In *Sociological Theories in Progress*, J. Berger (Ed.). Vol. 2. Houghton Mifflin, 218–251.

Yuxiao Dong, Reid A Johnson, Jian Xu, and Nitesh V Chawla. 2016. Structural Diversity and Homophily: A Study Across More than One Hundred Large-Scale Networks. *arXiv preprint arXiv:1602.07048* (2016).

Yuxiao Dong, Jie Tang, Sen Wu, Jilei Tian, Nitesh V. Chawla, Jinghai Rao, and Huanhuan Cao. 2012. Link Prediction and Recommendation across Heterogeneous Social Networks. In *IEEE ICDM'12*. 181–190.

Yuxiao Dong, Yang Yang, Jie Tang, Yang Yang, and Nitesh V. Chawla. 2014. Inferring User Demographics and Social Strategies in Mobile Social Networks. In *KDD'14*. ACM, 15–24.

Zhanpeng Fang and Jie Tang. 2015. Uncovering the Formation of Triadic Closure in Social Networks.. In *IJCAI*. Citeseer, 2062–2068.

Matthew S Gerber. 2014. Predicting crime using Twitter and kernel density estimation. *Decision Support Systems* 61 (2014), 115–125.

Zoubin Ghahramani and Michael I Jordan. 1997. Factorial hidden Markov models. *Machine learning* 29, 2-3 (1997), 245–273.

Eric Gilbert. 2012. Predicting Tie Strength in a New Medium. In *CSCW '12*. ACM, 1047–1056.

Eric Gilbert and Karrie Karahalios. 2009. Predicting tie strength with social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 211–220.

Mark S Granovetter. 1995. *Getting a job: A study of contacts and careers*. University of Chicago Press.

J. M. Hammersley and P. Clifford. 1971. Markov field on finite graphs and lattices. *Unpublished manuscript* (1971).

Simon S Haykin, Simon S Haykin, and Simon S Haykin. 2001. *Kalman filtering and neural networks*. Wiley Online Library.

F. Heider. 1958. *The psychology of interpersonal relations*. Wiley.

Hong Huang, Jie Tang, Lu Liu, JarDer Luo, and Xiaoming Fu. 2015. Triadic closure pattern analysis and prediction in social networks. *IEEE Transactions on Knowledge and Data Engineering* 27, 12 (2015), 3374–3389.

Hong Huang, Jie Tang, Sen Wu, Lu Liu, et al. 2014. Mining triadic closure patterns in social networks. In *WWW'14*. 499–504.

Jason J Jones, Jaime E Settle, Robert M Bond, Christopher J Fariss, Cameron Marlow, and James H Fowler. 2013. Inferring tie strength from online directed behavior. *PloS one* 8, 1 (2013), e52168.

Peter Klimek and Stefan Thurner. 2013. Triadic closure dynamics drives scaling laws in social multiplex networks. *New Journal of Physics* 15, 6 (2013), 063008.

Gueorgi Kossinets and Duncan J Watts. 2006. Empirical analysis of an evolving social network. *Science* 311, 5757 (2006), 88–90.

David Krackhardt. 1999. The ties that torture: Simmelian tie analysis in organizations. *Research in the Sociology of Organizations* 16, 1 (1999), 183–210.

Frank R. Kschischang, Brendan J. Frey, and Hans A. Loeliger. 2001. Factor graphs and the sum-product algorithm. *IEEE TOIT* 47 (2001), 498–519.

Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a social network or a news media?. In *WWW'10*. 591–600.

Jure Leskovec, Lars Backstrom, Ravi Kumar, and Andrew Tomkins. 2008. Microscopic evolution of social networks. In *KDD'08*. 462–470.

Jure Leskovec and Eric Horvitz. 2008. Planetary-scale views on a large instant-messaging network. In *WWW '08*. ACM, 915–924.

Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Signed Networks in Social Media. In *CHI '10*. ACM, 1361–1370.

Linna Li, Michael F Goodchild, and Bo Xu. 2013a. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *cartography and geographic information science* 40, 2 (2013), 61–77.

Menghui Li, Hailin Zou, Shuguang Guan, Xiaofeng Gong, Kun Li, Zengru Di, and Choy-Heng Lai. 2013b. A coevolving model based on preferential triadic closure for social media networks. *Scientific reports* 3 (2013).

Tiancheng Lou, Jie Tang, John Hopcroft, Zhanpeng Fang, and Xiaowen Ding. 2013. Learning to predict reciprocity and triadic closure in social networks. *TKDD* 7, 2 (2013), 5.

Peter V Marsden and Karen E Campbell. 1984. Measuring tie strength. *Social forces* 63, 2 (1984), 482–501.

Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* (2001), 415–444.

J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. 2007. Structure and tie strengths in mobile communication networks. *PNAS* (2007).

Akshay Patil, Juan Liu, and Jie Gao. 2013. Predicting group stability in online social networks. In *WWW'13*. 1021–1030.

Daniel M. Romero and Jon Kleinberg. 2010. The directed closure process in hybrid social-information networks, with an analysis of link formation on Twitter. In *ICWSM '10*.

Jari Saramäki, E. A. Leicht, Eduardo Lpez, Sam G. B. Roberts, Felix Reed-Tsochas, and Robin I. M. Dunbar. 2014. Persistence of social signatures in human communication. *PNAS* 111, 3 (2014), 942–947.

Catherine Schaefer, James C Coyne, and Richard S Lazarus. 1981. The health-related functions of social support. *Journal of behavioral medicine* 4, 4 (1981), 381–406.

Nino Shervashidze, Tobias Petri, Kurt Mehlhorn, Karsten M Borgwardt, and Svn Vishwanathan. 2009. Efficient graphlet kernels for large graph comparison. In *International conference on artificial intelligence and statistics*. 488–495.

Georg Simmel. 1950. *The sociology of georg simmel*. Vol. 92892. Simon and Schuster.

Stavros Sintos and Panayiotis Tsaparas. 2014. Using strong triadic closure to characterize ties in social networks. In *KDD'14*. 1466–1475.

S Vichy N Vishwanathan, Nicol N Schraudolph, Risi Kondor, and Karsten M Borgwardt. 2010. Graph kernels. *The Journal of Machine Learning Research* 11 (2010), 1201–1242.

Larry Wasserman. 2004. *All of statistics: a concise course in statistical inference*. Springer.

Rongjing Xiang, Jennifer Neville, and Monica Rogati. 2010. Modeling relationship strength in online social networks. In *WWW'10*. 981–990.

Jing Zhang, Zhanpeng Fang, Wei Chen, and Jie Tang. 2015. Diffusion of Following Links in Microblogging Networks. *TKDE* (2015).

Jing Zhang, Jie Tang, Yuanyi Zhong, Yuchen Mo, Juanzi Li, Guojie Song, Wendy Hall, and Jimeng Sun. 2017. StructInf: Mining Structural Influence from Social Streams.. In *AAAI*. 73–80.

Jia-Dong Zhang and Chi-Yin Chow. 2013. iGSLR: personalized geo-social location recommendation: a kernel density estimation approach. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 334–343.

Matteo Zignani, Sabrina Gaito, Gian Paolo Rossi, Xiaohan Zhao, Haitao Zheng, and Ben Y Zhao. 2014. Link and Triadic Closure Delay: Temporal Metrics for Social Network Dynamics. In *ICWSM'14*. 564–573.