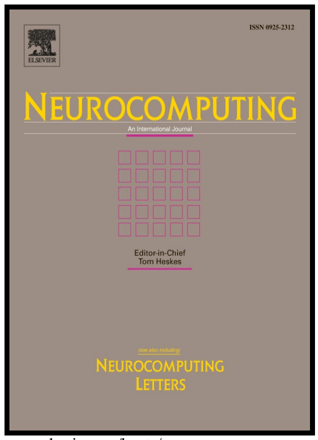
Author's Accepted Manuscript



www.elsevier.com/locate/neucom

TOSI: A Trust-Oriented Social Influence Evaluation Method in Contextual Social Networks

Guanfeng Liu, Feng Zhu, Kai Zheng, An Liu, Zhixu Li, Lei Zhao, Xiaofang Zhou

PII: S0925-2312(16)30607-5
DOI: http://dx.doi.org/10.1016/j.neucom.2015.11.129
Reference: NEUCOM17219

To appear in: *Neurocomputing*

Received date: 4 May 2015
Revised date: 5 November 2015
Accepted date: 9 November 2015

2015Published by Elsevier Ltd.

Procedia Computer Science

Procedia Computer Science 00 (2016) 1–16

TOSI: A Trust-Oriented Social Influence Evaluation Method in Contextual Social Networks

Guanfeng Liu[1], Feng Zhu[1], Kai Zheng[1], An Liu[1], Zhixu Li[1], Lei Zhao[1], and Xiaofang Zhou[1,2]

[1]*School of Computer Science and Technology*
Jiangsu Provincial Key Laboratory for Computer Information Processing Technology, Soochow University
Suzhou, China, 215006
[2]*School of Information Technology and Electrical Engineering*
University of Queensland
Brisbane, Australia, 4072
Email: {gfliu, kevin, anliu, zhixuli, zhaol}@suda.edu.cn; zxf@itee.uq.edu.au

Abstract

Online Social Networks (OSNs) have been used as the means for a variety of applications. For example, social networking platform has been used in employment system, e-Commerce and CRM system to improve the quality of recommendations with the assistance of social networks. In these applications, *social influence* acts as a significant role, affecting people's decision-making. However, the existing social influence evaluation methods do not fully consider the social contexts, i.e., the social relationships and the social trust between participants, and the preferences of participants, which have significant impact on social influence evaluation in OSNs. Thus, these existing methods cannot deliver accurate social influence evaluation results. In our paper, we propose a Trust-Oriented Social Influence evaluation method, called *TOSI*, with taking the social contexts into account. We conduct experiments onto two real social network datasets, i.e., *Epinions* and *DBLP*. The experimental results illustrate that our *TOSI* method greatly outperforms the state-of-the-art method *SoCap* in terms of effectiveness, efficiency and robustness.

Keywords: social network; social influence; trust.

1. Introduction

1.1. Background

Online Social Networks (OSNs) are becoming more and more popular and have been used as the means in a variety of applications, like employment, CRM and e-Commerce. In these applications, the *social influence* of a participant can affect others' decision-making. For example, at *Epinions* (epinions.com), an OSN based e-commerce platform, a buyer can write a product review to rate the products and corresponding seller. This review can be viewed by other buyers and thus can impact their decision making in purchasing the same products. As indicated in studies of *Social Psychology* [1, 2, 3] and *Computer Science* [4], a person is more likely to accept the recommendations given by participants with higher social influence (named as *Influencers*) in a specific domain. Therefore, it is significant to accurate evaluate the social influence of participants and identify those *Influencers* from social networks.

In the literature, many social influence evaluation methods have been proposed [5, 6, 7, 8, 9, 10, 11, 12, 13, 14], in which, *Independent Cascade* (*IC*) model [5] is a typical model to find the *Top-K* nodes who have the maximal social influence in a network. Subsequently, some important works [11, 9] are proposed to improve the scalability of

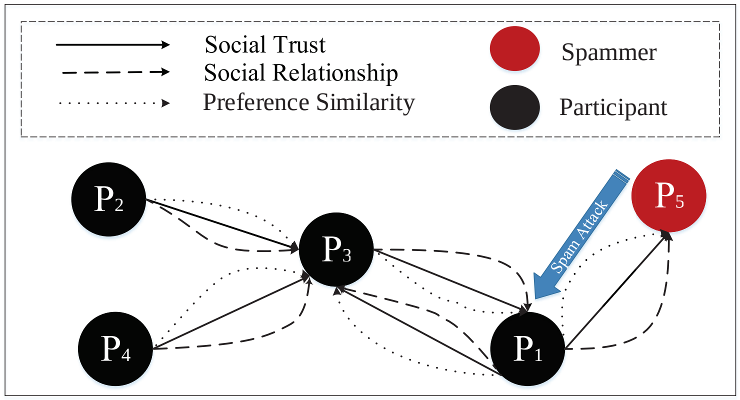


Figure 1: A social network of *Epinions*

IC model. In addition, in recent years, the *Local Influence Maximization* method [15] has been proposed to evaluate the social influence of a specific participant in OSNs. Furthermore, as some OSNs are becoming a large real-time generator of social data-streams, some streaming methods [16, 17] have proposed to evaluate the social influence of participants in OSNs.

1.2. The Problem and Motivation

As illustrated in *Social Psychology* [18, 19, 20], the social trust between participants (e.g., students trust their lecturers in a specific research area), the social relationship between participants (e.g., the relationship between a father and his song), and the preference similarity between participants (e.g., they all like to play basketball) have significant influence on participants' decision-making, and thus impact their social influence. However, these important social contexts are not fully considered by the existing social influence evaluation methods. Thus, these methods cannot deliver accurate social influence evaluation results. In addition, with the growth of network scale and complexity, social networks become susceptible to different types of unwanted and malicious spammer or hacker actions [21]. We propose a trust-oriented ranking strategy to defend against this kind of attack.

Example 1: Figure 1 depicts a social network from *Epinions*, which contains five participants (i.e., P_1 to P_5, they are all buyers). The trust relationship (represented as *arrows with solid lines*) between P_1 and P_3, P_2 and P_3, and P_4 and P_3 can be established based on the quality of the product review of P_3. Their social relationship and preferences can be mined from the their profiles and purchase history [22]. Suppose P_1 has closer social relationships, and has more similar preferences to P_3 than that of P_2, then P_3 can more likely affect the purchasing behavior of P_1 than P_2, which is not identified by the existing social influence evaluation methods. However, in traditional social influence evaluation methods [5, 6, 7, 8, 9, 10], the probability of the influence between two nodes is random in LT model or unified in IC model, which cannot reflect the realistic influence of participants. In addition, if P_5 is a spammer, in tradition social influence evaluation models, like the triggering model [5] and iterative model [23], he/she can utilize plenty of spam neighbors to establish fake strong social influence to affect P_1's decision-making.

The above mentioned problems motivate us to develop a social influence evaluation method to accurately evaluate participants' social influence in OSNs. In this paper, with considering the above mentioned important social contexts, we propose a Trust-Oriented Social Influence evaluation method, called *TOSI* by adopting iterative method. Since our method is convergent fast, thus we can deliver accurate social influence evaluation results with good efficiency.

1.3. Contributions

The main contributions of this paper can be summarised as follows:

- To the best of our knowledge, this is the first work that fully takes the social contexts into account in social influence evaluation.

- In order to defend against the typical spam attack, i.e., *Spam Farm* [23], in the process of evaluating social influence, We propose a Trust-Oriented Social Influence evaluation (TOSI) method that adopt sspam mass [23] to measure the probability of a participant to be an attacker.
- We propose a novel social influence evaluation method, *TOSI*, which achieves $O(\lambda N^2)$ in computation cost, where N is the number of nodes in an OSN and λ is the iterative times in computation.
- We have conducted experiments on two real social network datasets, i.e., *Epinions* and *DBLP*. By comparing with the state-of-the-art individual social influence evaluation method, *SoCap* [12], our *TOSI* method greatly outperforms *SoCap* in effectiveness (on average, it improves precision by 225%) and in efficiency (on average, *TOSI* saves 89.2% execution time) for social influence evaluation.

This paper is organised as follows. *Section 2* discusses the related work. *Section 3* introduces the preliminaries. *Section 4* proposes our *TOSI* method. In *Section 5*, we investigate the effectiveness and efficiency of our proposed method by comparing with the state-of-the-art method, *SoCap*. *Section 6* concludes this paper.

2. Related Work

In the literature, existing social influence evaluation approaches can be categorised into four groups, i.e., (1) global influence maximization, (2) local influence maximization, (3) stream learning of influence, and (4) individual influence evaluation. We discuss these methods in details as below.

2.1. Global Influence Maximization

The global influence maximization is to find a group of nodes that can impact the maximal number of other nodes in an OSN. Kempe et al. [5] propose a greedy algorithm which guarantees $(1 - 1/e)$ approximation ratio. However, this algorithm has low efficiency in practice and thus it is not scalable with the network size. In order to improve the scalability, [9] propose an algorithm that has a simple turnable parameter, for users to control the balance between the running time and the influence spread of the algorithm. Jung et al. [7] propose an algorithm *IRIE* that integrates the advantages of influence ranking (*IR*) and influence estimation (*IE*) methods for the global influence maximization. [10] provide a scalable influence approximation algorithm, Independent Path Algorithm (*IPA*), for *IC* model. In the model, they study *IPA* efficiently approximates influence by considering an independent influence path as an influence evaluation unit. Moreover, in order to spend up the evaluation algorithm, [11] develop the *CELF* algorithm, which exploits sub-modularity to find near-optimal influencer selections. In addition, Leskovec et al. [24, 25] divide the relations between two participants into two types, i.e., positive influence and negative influence, and predict them via theories of balance and status from social psychology. Based on above study, Yanhua et al. [26] make the first attempt to investigate the influence diffusion and influence maximization in OSNs with both positive and negative relations.

2.2. Local Influence Maximization

The local influence maximization is to find a group of nodes that have the maximal impacts on a specified participant. Yeung et al. [13] have studied the relations between trust and product ratings in online consumer review sites. Moreover, they propose a method to estimate the strengths of trust relations so as to estimate the true influence among the trusted participants. In addition, Guo et al. [15] propose a method to find K nodes that have the maximal impacts on a specified participant. Furthermore, Iwata et al. [27] propose a probabilistic model to discover the latent influence between participants in *OSNs*. The model is used to find influential participants and discover relations between participants.

2.3. Stream Learning of Influence

In recent years, *OSN* is becoming a large real-time generator of social data-streams, like *Twitter* (twitter.com). The streaming methods of social influence become more and more popular. Kutzkov et al. [16] propose a streaming method, called *STRIP* for computing the influence strength along each link of an *OSN*. In addition, Karthik et al. [17] propose an approach to mine the flow patterns, following specific flow validity constraints. However, contrasting with microblogging platforms, the other OSNs cannot provide sufficient contexts to perform information flow pattern discovery. Thus, the streaming methods cannot be applied for the social influencer finding in the OSN based e-commerce platforms.

2.4. Individual Influence Evaluation Problem

In order to evaluate the social influence of a specific participant, Subbian et al. [12] propose an approach, called *SoCap*, to find influencers in an OSN by using the social capital values. They model the problem of finding influencers in an OSN as a value-allocation problem, where the allocated value represents the individual social capital. In addition, Franks et al. [14] propose a method to identify influential agents in open multi-agent systems by adopting matrix factorization method to measure the influence of nodes in a network.

Summary: The existing methods do not fully consider the social contexts, like social relationships and social trust between participants, and preferences of participants in OSNs. As indicated in *Social Psychology* [1, 2, 3] and *Computer Science* [4], such social contexts are significant for social influence evaluation. Therefore, these existing methods cannot deliver accurate social influence results.

3. Preliminary

Our method aims to evaluate the reasonable social influence of each participant in contextual social networks based on the social network structure and the social contexts.

3.1. Contextual Social Network

A Contextual Social Network (CSN) [28] is a labeled directed graph $G = (V, E, LV, LE)$, where

- V is a set of vertices;
- E is a set of edges, and $(v_i, v_j) \in E$ denotes a directed edge from vertex v_i to vertex v_j;
- LV is a function defined on V such that for each vertex v in V, $LV(v)$ is a set of labels for v. Intuitively, the vertex labels may for example represent social roles or social influence in a specific domain;
- LE is a function defined on E such that for each link (v_i, v_j) in E, $LE(v_i, v_j)$ is a set of labels for (v_i, v_j), like social relationships, social trust and preferences in a specific domain.

3.2. Social Contexts

Let P denote the set of participants, and R denote the set of social contexts vectors, $\overrightarrow{R} < t, s, p > \in R$ $(t, s, p \in [0, 1])$, where $\overrightarrow{R}_{i,j}(t)$, $\overrightarrow{R}_{i,j}(s)$ and $\overrightarrow{R}_{i,j}(p)$ represent social trust, social relationship and preference similarity between P_i and P_j respectively. In addition, we use IN_i to denote the incoming neighbors of P_i and ON_i to denote the outgoing neighbors of P_i.

- **Social Trust (ST)**: It is the belief from one to another, based on their interactions [28]. Let t denote the trust value between two participants. $\overrightarrow{R}_{i,j}(t) = 1$ indicates that P_i completely trusts P_j, and $\overrightarrow{R}_{i,j}(t) = 0$ indicates that P_i completely distrusts P_j. The value of ST could be given by one participant to another based on the experience of their interactions (e.g., the purchasing experience of a buyer in social e-commerce platforms).

- **Social Relationship (SR)**: It is the social intimacy degree between two participants [19]. Let s denote the intimacy of the *Social Relationship* between two participants. $\overrightarrow{R}_{i,j}(s) = 1$ indicates that P_i and P_j have intimate social relationship, and $\overrightarrow{R}_{i,j}(s) = 0$ indicates that P_i have not contacted with P_j. The intimacy of the social relationship between two participants can be computed based on their personal information (e.g., the father and his sone) or their online interactions (e.g., the frequency or the contents of their interactions).

- **Preference Similarity (PS)**: It reflects the differences of the preferences between two participants. Let p denote the value of *Preference Similarity* between two participants. $\overrightarrow{R}_{i,j}(p) = 1$ indicates that the preferences of P_i and P_j are exactly the same, and $\overrightarrow{R}_{i,j}(p) = 0$ indicates that there is nothing in common interest between P_i and P_j.

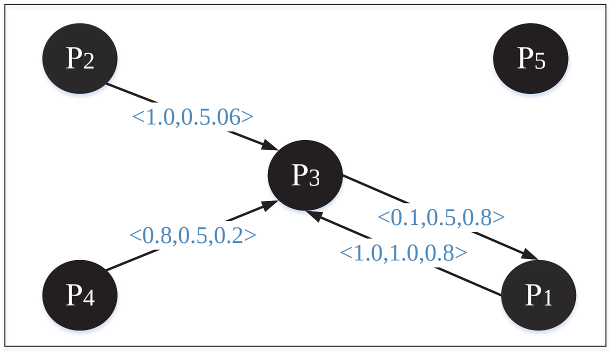


Figure 2: A contextual social network

Although it is difficult to build up comprehensive social trust, social relationship and preference similarity in all domains, it is feasible to build them up in some specific social communities by using data mining techniques [28]. Mining these social contexts' values is another challenging problem, which is out of the scope of this paper.

Example 2: Figure 2 depicts a contextual social network, which contains the social contexts as $\overrightarrow{R}_{2,3} =< 1.0, 0.5, 0.6 >$, $\overrightarrow{R}_{4,3} =< 0.8, 0.5, 0.2 >$, $\overrightarrow{R}_{1,3} =< 1.0, 1.0, 0.8 >$ and $\overrightarrow{R}_{3,1} =< 0.1, 0.5, 0.8 >$.

3.3. *Social Influence Evaluation Problem*

Given a contextual social network $G = (V, E, LV, LE)$ and a set of social contextual impact factors R, we evaluate the social influence of individual participant and deliver the social influence set SI.

4. Trust-Oriented Social Influence Evaluation Method

In this section, we propose a Trust-Oriented Social Influence evaluation method, called *TOSI*, by adopting the iterative method to evaluate social influence. *TOSI* takes the above important social contexts into consideration, and thus can deliver more accurate social influence evaluation results, and therefore can find more reliable *Influencers*.

4.1. *Algorithm Description*

In our *TOSI* method, the social influence of participants are constantly computed and replaced until the social influences achieve convergence by using iterative method. Next, we introduce the process of iteration and the details of *TOSI*.

The social influences at iteration time $t+1$ are based on the social influences delivered at the last iteration time t. In the process of evaluating new social influences, we consider the social contexts (*ST*, *SR* and *PS*) between a participant and his/her neighbors equally. Let SI_i^* denote the social influence of participant P_i, which can be computed by Eqs. (1) and (2) as below:

$$SI_i^{t+1} = \sum_{P_k \in IN_i} SI_k^t \cdot \rho_{k,i} \tag{1}$$

where $\sum_{P_i \in ON_k} \rho_{k,i} = 1$ and

$$\rho_{k,i} = \frac{\overrightarrow{R}_{k,i}(t)}{3 \cdot TTTR_k} + \frac{\overrightarrow{R}_{k,i}(s)}{3 \cdot TTSR_k} + \frac{\overrightarrow{R}_{k,i}(p)}{3 \cdot TTPS_k}. \tag{2}$$

$TTTR_k = \sum_{P_j \in ON_k} \overrightarrow{R}_{k,j}(t)$, $TTSR_k = \sum_{P_j \in ON_k} \overrightarrow{R}_{k,j}(s)$ and $TTPS_k = \sum_{P_j \in ON_k} \overrightarrow{R}_{k,j}(p)$. Here, $\rho_{k,i}$ reflects the whole influence probability from P_k to P_i.

Example 3: In Figure 3, for participant P_1, there are four social contexts from four participants, i.e., P_2 to P_5. The details of social contexts and social influence at iterative time t have been shown in Figure 3. At iterative time

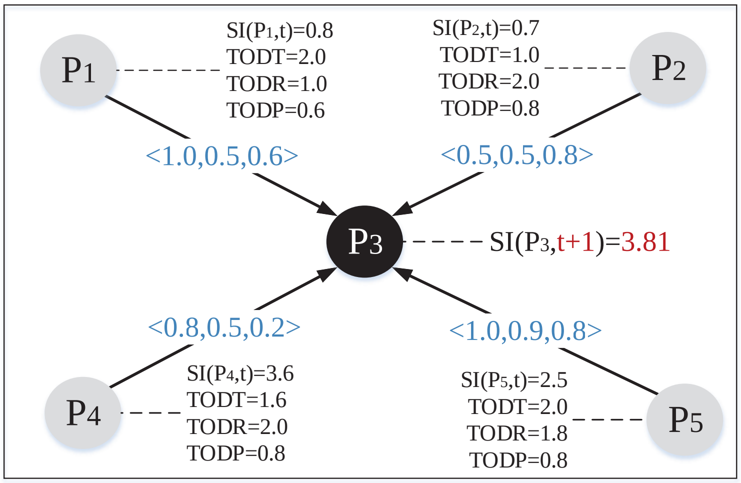


Figure 3: Computing social influence in iterative process

$t+1$, the social influence $SI_1^{t+1} = 0.8 \times (1.0/2.0 + 0.5/1.0 + 0.6/0.6)/3 + 0.7 \times (0.5/1.0 + 0.5/2.0 + 0.8/0.8)/3 + 3.6 \times (0.8/1.6 + 0.5/2.0 + 0.2/0.8)/3 + 2.5 \times (1.0/2.0 + 0.9/1.8 + 0.8/0.8)/3 = 3.81$.

4.2. Convergence of the Iteration

For an iterative method, a very key and basic task is to guarantee its convergence. Haveliwala et al. [29] have analysed the convergence of *PageRank*, stimulated by their ideas about iterative method, next we use an error function and prove the convergence of our *TOSI* method. Let SI_i^* denote the real social influence of participant P_i. And then we define the total error at iterative time t to be:

$$Error(t) = \sum_{i=1}^{N} |SI_i^t - SI_i^*| \tag{3}$$

where N is the number of participants.

Theorem 1: *TOSI is convergent, i.e., $Error(t) < Error(t-1)$.*

Proof 1: *Since SI_i^* is the real solution, according to eq.(1), it must satisfy following equation exactly:*

$$SI_i^* = \sum_{P_k \in IN_i} SI_k^* \cdot \rho_{k,i} \tag{4}$$

For a participant P_i, the error at iterative time t is:

$$SI_i^t - SI_i^* = \sum_{P_k \in IN_i} (SI_k^{t-1} - SI_k^*) \cdot \rho_{k,i} \tag{5}$$

Using the Triangle Inequality, we can obtain the expression as follows:

$$|SI_i^t - SI_i^*| \leq \sum_{P_k \in IN_i} |SI_k^{t-1} - SI_k^*| \cdot \rho_{k,i} \tag{6}$$

Next, we sum over all the errors of participants to obtain total error. Notice that $\sum_{P_i \in ON_k} \rho_{k,i} = 1$*:*

$$\begin{aligned} Error(t) &= \sum_{i=1}^{N} |SI_i^t - SI_i^*| \\ &\leq \sum_{i=1}^{N} \sum_{P_k \in IN_i} |SI_k^{t-1} - SI_k^*| \cdot \rho_{k,i} \\ &= \sum_{\vec{R}_{k,i} \in R} |SI_k^{t-1} - SI_k^*| \cdot \rho_{k,i} \\ &= \sum_{k=1}^{N} |SI_k^{t-1} - SI(P_k)^*| \cdot \sum_{P_i \in ON_k} \rho_{k,i} \\ &= Error(t-1) \end{aligned} \tag{7}$$

Recalling the eq.(6), we find that $Error(t) = Error(t-1)$ *if and only if* $\forall\, P_k \in P,\ SI_k^{t-1} - SI_k^* > 0$ *or* $\forall\, P_k \in P,\ SI_k^{t-1} - SI_k^* < 0$*. But, our iterative method reduce the total social influence to 1, it means that* $\sum_{k=1}^{N} SI_k^{t-1} = \sum_{k=1}^{N} SI_k^* = 1$*. It cannot satisfy the above condition, so* $Error(t) < Error(t-1)$*. Then* ***Theorem 1*** *is proved.* □

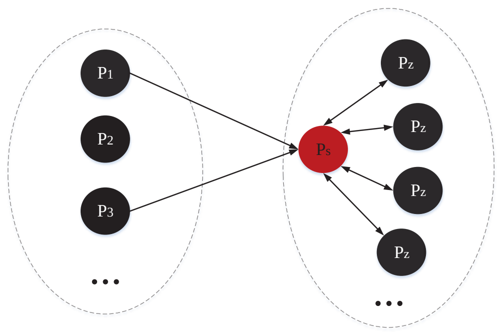


Figure 4: Spam attack

4.3. Spam Attack

In our method, we consider to defend against a typical spam attack, i.e., *Spam Farm* [23], in the process of evaluating social influence. The attack way of *Spam Farm* has been shown in Figure 4. The spammer P_s has partial honest in-coming participants, i.e., P_1 and P_3, and many own participants, i.e., P_z. Such own participants only have relations with the spammer P_s. For example, at Facebook, zombie fans always play the role of own participant. Such network structure like Figure 4 makes the spammer have the unreal social influence. Based on the TrustRank [30], we adopt a trust-oriented ranking strategy to defend against the spam attack, and use TR_i to denote the TrustRank of P_i.

Firstly, TrustRanks TR^{t+1} at iterative time $t+1$ are based on the trust ranks TR^t delivered at the last iterative time t. Just like the process of evaluating social influence, we also consider the social contexts (*ST*, *SR* and *PS*) between a participant and his/her neighbors equally. The detailed equation as follows:

$$TR_i^{t+1} = \sum_{P_k \in IN_i} TR_k^t \cdot \rho_{k,i} \tag{8}$$

where $\rho_{k,i}$ has been defined in eq.(2).

Next, in order to defend against the *Spam Farm* attack, we select a set of trustworthy nodes that are widely trusted by the other participants in the OSN. For example, at Weibo (www.weibo.com), we can select the participants who have been certified as the trustworthy participants. Here, we use TN to denote the set of trust nodes and J to denote the

Algorithm 1 *TOSI* Algorithm

Require: The set of participants P, the set of relation vectors R, iterative times λ, decay factor α for TrustRank, trustworthy nodes TN, the number of trustworthy nodes J;

Ensure: The social influence set of all participants SI, the spam mass set of all participants SM;

$SI, TR, SM \leftarrow \{rand(1)\}$;
$NewSI, NewTR \leftarrow \emptyset$;
$TotalSI, TotalTR, i \leftarrow 0$;
while $i < \lambda$ **do** /* Iterative evaluate the social influences which are based on last social influences */
    $i \leftarrow i + 1$;
    $TotalSI \leftarrow 0$;
    for *each* P_j *in* P **do**
        if P_j is not isolated node **then**
            $v, tr \leftarrow 0$;
            for each node P_k in the incoming neighbors of P_j **do**
                $\rho_{k,j} \leftarrow (\overrightarrow{R}_{k,j}(t)\ TTTR_k + \overrightarrow{R}_{k,j}(r)\ /\ TTSR_k + \overrightarrow{R}_{k,j}(p)\ /\ TTPS_k)\ /\ 3$; /* Calculate the influence probability from P_k to P_j */
                $v \leftarrow v + SI[k] * \rho_{k,j}$; /* Add up the social influence propagating from P_k to P_j */
                if P_j is trustworthy participant **then** /* Add up the TrustRank propagating from P_k to P_j */
                    $tr \leftarrow tr + \alpha * TR[k] * \rho_{k,j} + (1-\alpha)/J$;
                else
                    $tr \leftarrow tr + \alpha * TR[k] * \rho_{k,j}$;
                end if
            end for
            $NewSI[j] \leftarrow v, NewTR[j] \leftarrow tr$;
            $TotalSI \leftarrow TotalSI\ +\ v, TotalTR \leftarrow TotalTR\ +\ tr$;
        end if
    end for
    $NewSI \leftarrow NewSI\ /\ TotalSI$; /* Reduce the sum of social influences to 1 */
    $NewTR \leftarrow NewTR\ /\ TotalTR$; /* Reduce the sum of TrustRanks to 1 */
    $SM \leftarrow (NewSI\ -\ NewTR)\ /\ NewSI$;
    Replace SI with $NewSI$, TR with $NewTR$;
end while
Return SI and SM.

number of the set. Otherwise, we adopt a decay factor α for biased TrustRank [23]. Then the eq.(8) can be improved as:

$$TR_i^{t+1} = \begin{cases} \alpha \cdot (\sum_{P_k \in IN_i} TR_k^t \cdot \rho_{k,i}) + (1-\alpha)/J & (if\ P_i\ is\ a\ trustworthy\ participant), \\ \alpha \cdot (\sum_{P_k \in IN_i} TR_k^t \cdot \rho_{k,i}) & (Otherwise). \end{cases} \tag{9}$$

Finally, according to the *TrustRank* scores, TOSI computes the spam mass value [23] that reflects the probability of a participant to be a *Spam Farm* attacker. The larger the spam mass is, the higher the probability of a participant to be an attacker. Let SM_i to denote the spam mass of participant P_i. The equation of calculating the spam mass as follows:

$$SM_i = (SI_i - TR_i)/SI_i \tag{10}$$

So, our TASI method can recommend those realible social influencers who have strong social influence and identify the participants who have high probability to be a *Spam Farm* attacker.

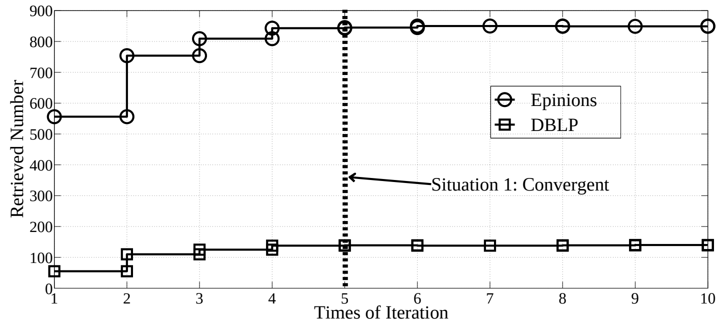


Figure 5: The convergence of our *TOSI* method. X axis is iteration times and Y axis is retrieved number. We use stairs lines to show the trends of retrieved number with the increasing of iteration times.

4.4. Algorithm

The pseudo-code of the algorithm is given in Algorithm 1. As *TOSI* is convergent fast, and the iteration times of is a constant. Therefore, the time complexity of the *TOSI* method is $O(\lambda N^2)$, where N is the number of participants and λ is iterative times.

5. Experiments

In our experiments, we compare our proposed *TOSI* method with the state-of-the-art method, *SoCap* [12]. In order to investigate effectiveness, we compare the accuracy of the two methods in *Exp-1* and the number of nodes that are influenced by *Top-K* influencers delivered by the two methods in two classical diffusion models, i.e., *Linear Threshold (LT)* model [31] and *Independent Cascade (IC)* model [5] in *Exp-2*. In order to investigate the robustness of our method under the *Spam Farm* attack, we explain the process of how to distinguish the *Spam Farm* attacker in *Exp-3*. In order to investigate efficiency, we compare the execution time of the two methods in social influence evaluation in *Exp-4*.

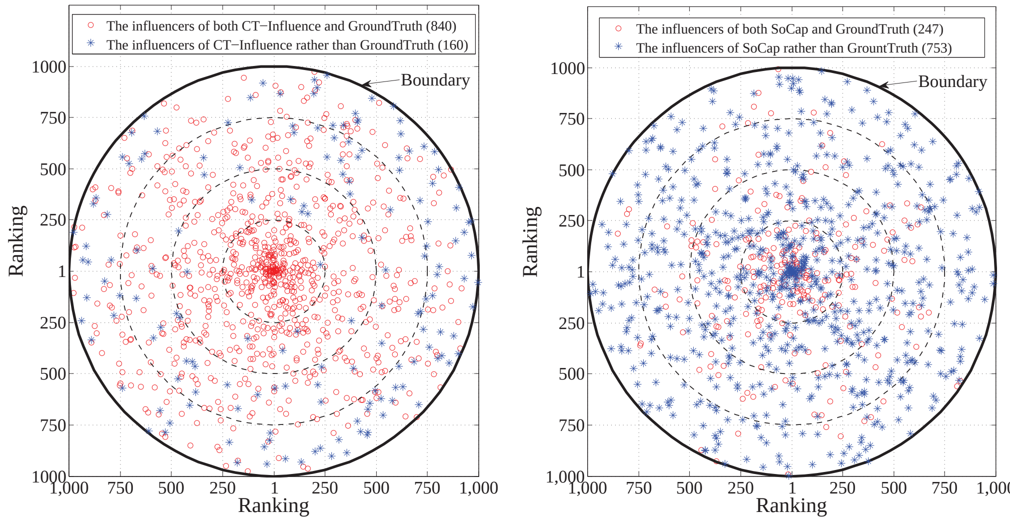
5.1. Experimental Setting

Table 1: Experimental Datasets

Dataset	Epinions	DBLP
Nodes	75,879	317,080
Links	508,837	1,049,866
Average Indegree	6.706	3.311
High Indegree Nodes (Indegree ≥ 50)	2032	170
The Ratio of High Indegree Nodes	2.679 %	0.054 %

Table 2: The performances of *TOSI* and *SoCap* with *Ground Truth Top-1000*

Method	**DataSet**	**Retrieved Number**	**Precision**	**Average Execution Time**
TOSI	Epinions	840	0.84	444 ms
SoCap	Epinions	247	0.247	6436 ms
TOSI	DBLP	138	0.138	1244 ms
SoCap	DBLP	44	0.044	8364 ms



(a) *TOSI* (*Epinions*) (b) *SoCap* (*Epinions*)

Figure 6: Top 1000 influencers of each method on *Epinions*. *X* axis and *Y* axis are all ranking of social influence, and the large circle is the boundary which contains the *Top-1000* influencers (small red circles and blue stars) delivered by each method. Small red circles are both *Ground Truth Top-1000* influencers and *Top-1000* influencers delivered by each method respectively. Blue stars are *Top-1000* influencers delivered by each method rather than *Ground Truth Top-1000* influencers. The more close to the center of the large circle, the higher influence ranking the influencers have.

5.1.1. *Datasets*

We adopt two real social network datasets, *Epinions* [32] and *DBLP* [33]. The *Epinions* dataset has 75,879 nodes and 508,837 links, where each node represents a buyer, and each link corresponds to the relationships between buyers. The *DBLP* dataset has 317,080 nodes and 1,049,866 links, where each node represents an author, and each link corresponds to the co-author relationships between authors. The details of the two datasets are listed in Table 1.

5.1.2. *Ground Truth*

As indicated in *Social Psychology* [34], if a participant can influence the maximal number of participants who have a high social influence, then such a participant has high social influence as well. Therefore, we rank the influencers based on the number of influenced participants as the *Ground Truth* in *Exp-1* and *Exp-3*.

5.1.3. *Diffusion Models*

In *Exp-2*, we adopt two classical diffusion models, i.e., *Linear Threshold (LT)* model [31] and *Independent Cascade (IC)* model [5]. These models have been widely used to investigate the effectiveness of social influence evaluation methods in [35, 36, 37] by comparing the number of nodes that are influenced by the seeds in these diffusion models.

- **Linear Threshold (LT) Model:** *LT* model is the first model to imitate the diffusion process of information. The approach is based on the node-specific thresholds [31]. In the model, at time step t, all nodes that were influenced in step $t-1$ remain being influenced. A participant P_i is influenced based on a monotonic function of its influenced neighbors $f(In(i,t)) \in [0,1]$ (see Eq.(8)) and a threshold $\theta_i \in [0,1]$, i.e., P_i is influenced at time t if $f(In(i,t)) \geq \theta_i$.

$$f(In(i,t)) = \sum_{P_j \in In(i,t)} b_{i,j} \tag{11}$$

where $In(i,t)$ is the influenced neighbors of P_i at time step t. Here, we set

$$b_{i,j} = \frac{\overrightarrow{R}_{i,j}(t) + \overrightarrow{R}_{i,j}(s) + \overrightarrow{R}_{i,j}(p)}{\sum_{P_k \in On_i}(\overrightarrow{R}_{i,k}(t) + \overrightarrow{R}_{i,k}(s) + \overrightarrow{R}_{i,k}(p))}, \tag{12}$$

Table 3: The comparison of *TOSI* and *SoCap* with *Ground Truth Top-10* on *Epinions*

Nodes' ID	Ground Truth Ranking	TOSI Ranking	SoCap Ranking
18	1	**1**	**108**
737	2	**2**	**269**
401	3	**3**	**308**
40	4	**4**	**631**
118	5	**6**	**669**
34	6	**7**	1184 (missing)
550	7	**8**	6226 (missing)
136	8	**12**	6442 (missing)
143	9	**23**	6448 (missing)
1719	10	**32**	23842 (missing)

On_i is the outgoing neighbors of P_i and $\sum_{P_j \in On_i} b_{i,j} \leq 1$. In our experiments, in order to investigate the effectiveness of our method based on different thresholds, for each P_i, we set $\theta_i \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

- **Independent Cascade (IC) Model:** *IC* model is a dynamic cascade model for the diffusion process. The model is based on the interacting particle system from probability theory [5]. At each time step t, each participant is either influenced or susceptible. A participant P_j that was influenced at time step $t-1$ has a single chance to influence each of its incoming neighbors P_i. The influence succeeds with probability $p_{i,j}$ (see Eq.(10)). Therefore, for participant P_i, if at least one of its influenced outgoing neighbors succeeds, P_i gets influenced. The probability of participant P_i getting influence at time step t is:

$$f(i,t) = 1 - \prod_{P_j \in In(i,t-1)} (1 - p_{i,j}) \tag{13}$$

where $In(i, t-1)$ is the influenced incoming neighbors of P_i at time step $t-1$. Here, we set $p_{i,j} = (\overrightarrow{R}_{i,j}(t) + \overrightarrow{R}_{i,j}(s) + \overrightarrow{R}_{i,j}(p))/3$.

In our experiments, we select the *Top-K* influencers delivered by our *TOSI* and *SoCap* to act as seeds in the different diffusion models respectively, here, $K \in \{1, 5, 10, 20, 50, 100\}$. Based on the properties of the diffusion models, the number of nodes that are influenced by the seeds delivered by the diffusion models can illustrate the influence of the *Top-K* influencers [35, 36]. The more the number is, the higher the effectiveness of corresponding method is.

5.1.4. Experimental Environments

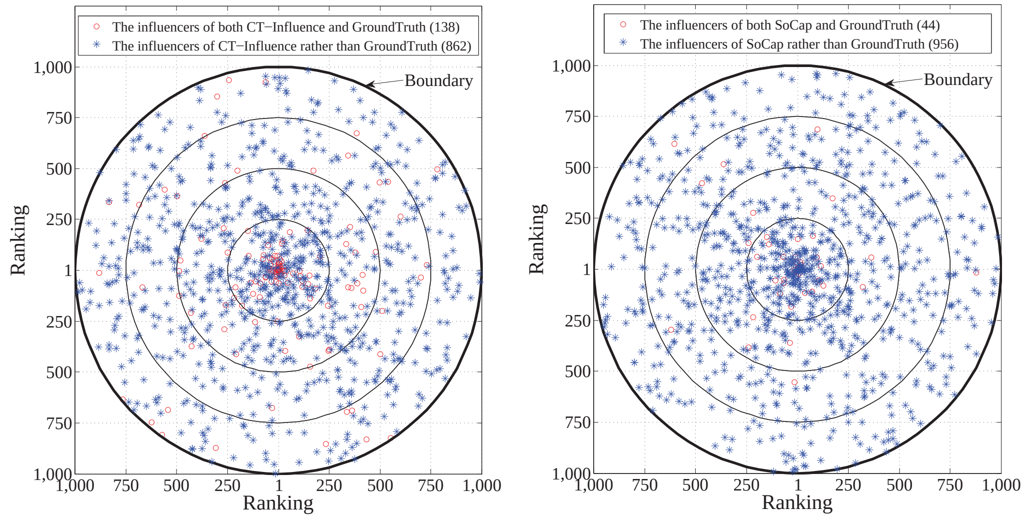
All experiments were run on a PC powered by two Intel Core i5-3470 CPU 3.20 GHz processors with 8 GB of memory, using Windows 7 Professional. The code was implemented by using Visual C++ 2012 and the experimental data was managed by MySQL Server 5.6. All the experimental results are averaged based on five independent runs.

5.2. Experimental Results and Analyses

5.2.1. Exp-1. Effectiveness (by Ground Truth)

We measured the precision by varying the *Top-1000* influencers retrieved by each method against the *Ground Truth Top-1000* influencers.

- Firstly, we observe the trend of *Retrieved Number* with the increasing of times of iteration to investigate the convergence of our *TOSI* method. Here, *Retrieved Number* is the number of retrieved influencers, which are both the *Ground Truth Top-1000* influencers and the *Top-1000* influencers delivered by our *TOSI* method. The experimental results delivered based on *Epinions* dataset and *DBLP* datasetare are shown in Figure 5, where we can see that the *Retrieved Numbers* of our *TOSI* method keep stable after 5 times of iterations for both datasets. Then, in the following experiments, we set the *Iterative times* λ as 5.



(a) *TOSI* (*DBLP*) (b) *SoCap* (*DBLP*)

Figure 7: Top 1000 influencers of each method on *DBLP*. *X* axis and *Y* axis are all ranking of social influence, and the large circle is the boundary which contains the *Top-1000* influencers (small red circles and blue stars) delivered by each method. Small red circles are both *Ground Truth Top-1000* influencers and *Top-1000* influencers delivered by each method respectively. Blue stars are *Top-1000* influencers delivered by each method rather than *Ground Truth Top-1000* influencers. The more close to the center of the large circle, the higher influence ranking the influencers have.

- Secondly, after five iterations, the experimental results are listed in Table 2. For *Epinions*, our *TOSI* method finds 840 out of the *Ground Truth Top-1000* influencers, while *SoCap* can only find 247 influencers. Based on the precision function in Eq. (8) [12], the precision of our *TOSI* method is 84%. In contrast, it is only 24.7% for *SoCap* method. Therefore, comparing with *SoCap*, on average, our method greatly improves the precision of social influence evaluation by 240% in *Epinions* dataset. For *DBLP*, our *TOSI* method finds 138 (precision is 13.8%) out of the *Ground Truth Top-1000* influencers, but *SoCap* method only finds out 44 (precision is 4.4%). Therefore, on average, our method improves the precision of social influence evaluation by 210% in *DBLP* dataset.

$$Precision = \frac{|Relevant \bigcap Retrieved|}{|Retrieved|} \tag{14}$$

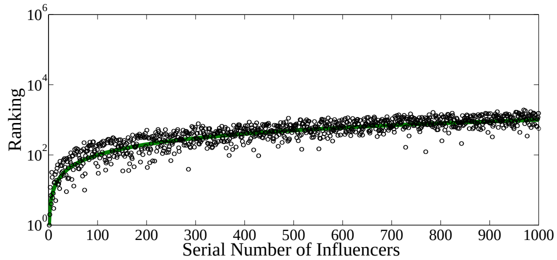
- Next, we list the results of the *Top-10* influencers retrieved by each method against the *Ground Truth Top-10* influencers in Table 3. From Table 3, our *TOSI* method can find all 10 influencers, and the *TOSI Ranking* is very close to the *Ground Truth Ranking*. But the influencers delivered by *SoCap* is far away from the *Ground Truth Ranking*, and 5 out of 10 influencers are missing in the *Top-10* list.

- The experimental results of *Epinions* and *DBLP* are plotted in Figure 6 and Figure 7, where we can see that the number of the *Ground Truth Top-1000* influencers retrieved by our *TOSI* method is more than *SoCap*'s with higher rankings (the small red circles of our *TOSI* method are closer to the center of the large circle). Therefore, our *TOSI* can deliver more accurate social influence evaluation results than *SoCap*.

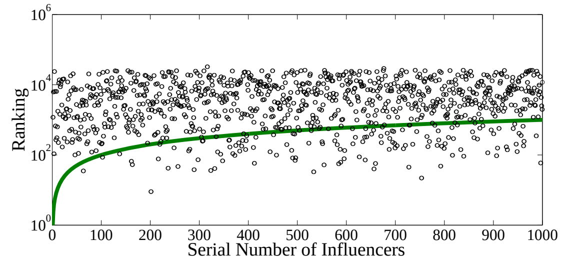
- In addition, we use an *Absolute Error* (*AE*) function to measure the error of each method. The error is the absolute value between the influence ranking delivered by each method and the *Ground Truth Ranking*. The detailed calculation is as follows:

$$AE(method, dataset) = \sum_{P_i \in GT} |RA(P_i) - GTRA(P_i)| \tag{15}$$

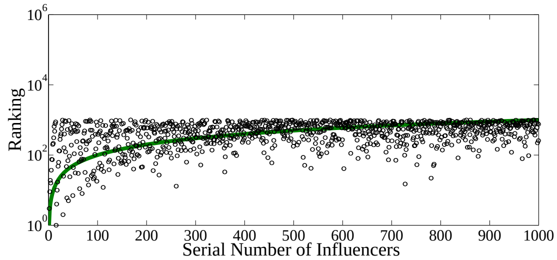
where GT is the set of *Ground Truth Top-1000* influencers, $RA(P_i)$ is the influence ranking of P_i evaluated by each method, and $GTRA(P_i)$ is the *Ground Truth Ranking* of P_i. From Figure 8, $AE(SoCap, Epinions)$ is much



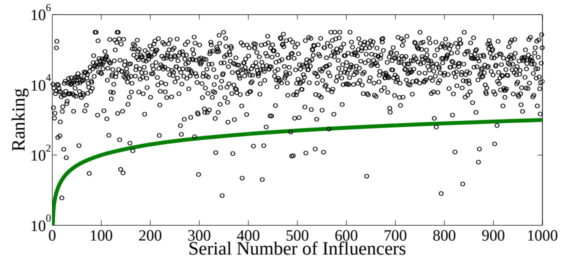
(a) $AE(TOSI, Epinions) = 171310$



(b) $AE(SoCap, Epinions) = 5967223$

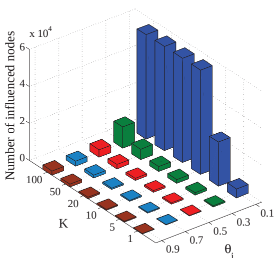


(c) $AE(TOSI, DBLP) = 263866$

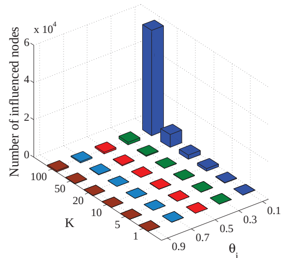


(d) $AE(SoCap, DBLP) = 51327091$

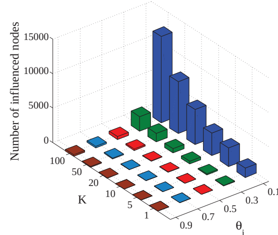
Figure 8: The errors of each method on each dataset. X axis is the serial number of influencers, Y axis is the ranking of social influence. Green curve is *Ground Truth Ranking*, and small black circles are the rankings of influencers delivered by each method.



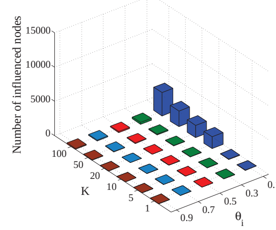
(a) *TOSI* (*LT*,*Epinions*)



(b) *SoCap* (*LT*,*Epinions*)



(c) *TOSI* (*LT*,*DBLP*)



(d) *SoCap* (*LT*,*DBLP*)

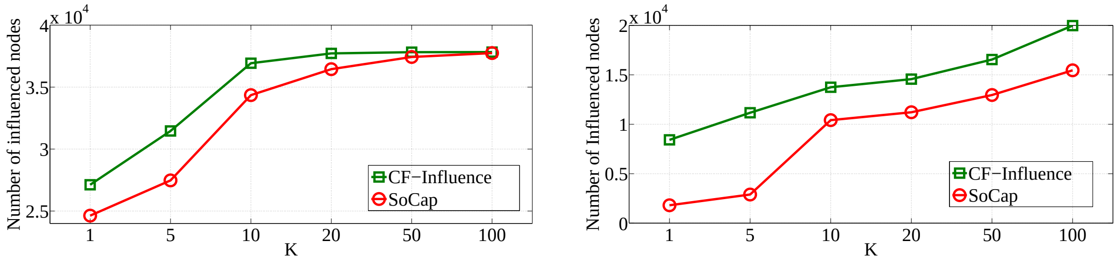
Figure 9: The number of influenced nodes on *Linear Threshold* model

greater than $AE(TOSI, Epinions)$ and $AE(SoCap, DBLP)$ is much greater than $AE(TOSI, DBLP)$, *so the error level of SoCap is high*.

- Finally, we study the accuracy and absolute error of our *TOSI* method by comparing the *Top-1000* influencers identified by each method against the *Ground Truth Top-1000* influencers. Since *SoCap* ignore the social relationship and preference similarity between participants, it cannot deliver accurate social influence evaluation results. *Therefore, our TOSI method outperforms SoCap in Effectiveness based on Ground Truth results*.

5.2.2. Exp-2. Effectiveness (by Diffusion Models)

Figure 9 depicts the experimental results of *LT* model, where we can see that in all cases, the number of influenced nodes identified by our *TOSI* with different K and θ_i are more than that of *SoCap*. The average number of influenced nodes identified by our *TOSI* is 5,609.18, while that of *SoCap* is 1,371.02 which is 75.56% less than that of *TOSI*. In addition, the number of influenced nodes identified by the two methods increases with the increase of K. This is because that with the increase of K, the number of sources for the spread of information increases, which leads to the *Top-K* influencers identified by both of *TOSI* and *SoCap* can influence more nodes in *LT* model. Furthermore, the number of influenced nodes identified by the two methods decreases with the decrease of θ_i. This is because that the



(a) *IC,Epinions* (b) *IC,DBLP*

Figure 10: The number of influenced nodes on *Independent Cascade* model

Table 4: The spam mass for attackers

Tag	**Dataset**	**TOSI Ranking, SM(Attacker Probability)**	**SoCap Ranking**
Honest Influencer	Epinions	1, -1.13(low)	1199
SF-Attacker-50	Epinions	53, 0.94(high)	2127
SF-Attacker-100	Epinions	48, 0.93(high)	809
SF-Attacker-200	Epinions	51, 0.96(high)	1170
SF-Attacker-500	Epinions	42, 1.06(high)	423
SF-Attacker-1000	Epinions	58, 1.97(very high)	230
Honest Influencer	DBLP	1, -2.24(very low)	65256
SF-Attacker-50	DBLP	20, 0.86(high)	19754
SF-Attacker-100	DBLP	17, 0.95(high)	5165
SF-Attacker-200	DBLP	18, 1.21(high)	30
SF-Attacker-500	DBLP	9, 1.93(very high)	856
SF-Attacker-1000	DBLP	19, 2.23(very high)	82

limit for the spread of information decreases with the decrease of θ_i, which leads to the *Top-K* influencers identified by both of *TOSI* and *SoCap* can influence more nodes in *LT* model. Therefore, based on the properties of diffusion models [5], the experimental results illustrate that the *Top-K* influencers identified by our *TOSI* have *more influences* than that of *SoCap* in *LT* model.

Figure 10 depicts the number of influenced nodes identified by our *TOSI* and *SoCap*, where we can see that with the increase of K in *IC* model respectively, where we can see that the number of influenced nodes of our *TOSI* are more than that of *SoCap* in all 6 cases on the two datasets. The average number of influenced nodes identified by our *TOSI* is 24,441.92, and that of *SoCap* is 21,069.5 which is 13.8% less than the former. This is because that based on the properties of the *IC* model introduced in the Section *Diffusion Models*, with taking the three social contexts into consideration, the *Top-K* influencers identified by our *TOSI* have higher probability to influence their neighbor nodes. In addition, with the increase of K, the number of nodes influenced by the *Top-K* nodes identified by both of the two methods increases.

From the experimental results in the two classical diffusion models, i.e., *LT* model and *IC* model, we can see that the *Top-K* influencers identified by our *TOSI* have more influences than that of the state-of-the-art method, *SoCap*. Based on the properties of diffusion models, on average, our *TOSI* improves the effectiveness of *SoCap* by 90%. Thus *our TOSI method outperforms SoCap in effectiveness based on the two classical diffusion models.*

5.2.3. Exp-3. Robustness

Firstly, we set 5 *Spam Farm* attackers who have 500 randomly honest in-coming participants and M own participants respectively in the two datasets. We give them tags “SF-Attacker-M”, here, $M \in \{50, 100, 200, 500, 1000\}$. In addition, we select the $Top-1$ influencer delivered by our *TOSI* as “Honest Influencer” to compare with these attackers. Next, we select the *Ground Truth Top-1000* influencers as trustworthy nodes, and set α as 0.85 [23] to evaluate the TrustRank scores. At the end, we investigate the *Spam Mass* of participants based on Eq.(10) From Table 4, compared with the "Honest Influencer", we can see that the attackers have much higher *Spam Mass* in the two datasets. As indicated in Section 4.3, the larger the *Spam Mass* is, the higher probability of a participant to be a *Spam Farm* attacker. Therefore, our TOSI method can identify the *Spam Farm* attackers well via *Spam Mass*, which improve the robustness of the social influence. However, *SoCap* cannot find these attackers.

5.2.4. Exp-4. Efficiency

Table 2 lists the corresponding execution times of social influence evaluation (except the time of "loading all data into memory") of two methods. On *Epinions* dataset, the average execution time is 444 ms for our *TOSI*. By contrast, it is 6,436 ms for *SoCap*. On average, our method can save 93.1% of the execution time. On *DBLP* dataset, it is 1,244 ms for our *TOSI* and 8,364 ms for *SoCap*. On average, our method can save 85.1% of the execution time. This is because that based on *Theorem 1*, the convergence of our *TOSI* is fast. Therefore, *our TOSI method greatly outperforms SoCap in efficiency*.

6. Conclusion and Future Work

In this paper, we have proposed a Trust-Oriented Social Influence evaluation (*TOSI*) method which is based on three social contexts, i.e., social trust, social relationship and preference similarity, between two participants to evaluate the social influence of each participant. The experiments conducted on two real social network datasets (*Epinions* and *DBLP*) have demonstrated our *TOSI* method greatly outperforms the state-of-the-art method, *SoCap*, and can deliver more accurate social influence evaluation results with less execution time.

In future work, we plan to apply our *TOSI* method into real OSNs to deliver accurate social influence evaluation results and recommend reliable influencers.

Acknowledgement

This work was partially supported by Natural Science Foundation of China (Grant Nos. 61303019, 61402312, 61402313, 61232006, 61003044, 61440053), Doctoral Fund of Ministry of Education of China (20133201120012), Postdoctoral Science Foundation of China (2015M571805), and Collaborative Innovation Center of Novel Software Technology and Industrialization, Jiangsu, China.

[1] E. Berscheid, H. T. Reis, et al., Attraction and close relationships, The handbook of social psychology 2 (1998) 193–281.
[2] I. Yaniv, Receiving other peoples advice: Influence and benefit, Organizational Behavior and Human Decision Processes 93 (1) (2004) 1–13.
[3] S. T. Fiske, Social beings: Core motives in social psychology, John Wiley & Sons, 2009.
[4] P. Bedi, H. Kaur, S. Marwaha, Trust based recommender system for semantic web., in: IJCAI, 2007, pp. 2677–2682.
[5] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: KDD, 2003, pp. 137–146.
[6] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in: KDD, 2009, pp. 199–208.
[7] K. Jung, W. Heo, W. Chen, Irie: Scalable and robust influence maximization in social networks, in: ICDM, 2012, pp. 918–923.
[8] W. Chen, W. Lu, N. Zhang, Time-critical influence maximization in social networks with time-delayed diffusion process., in: AAAI, 2012, pp. 592–598.
[9] W. Chen, C. Wang, Y. Wang, Scalable influence maximization for prevalent viral marketing in large-scale social networks, in: KDD, 2010, pp. 1029–1038.
[10] J. Kim, S.-K. Kim, H. Yu, Scalable and parallelizable processing of influence maximization for large-scale social networks?, in: ICDE, 2013, pp. 266–277.
[11] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, N. Glance, Cost-effective outbreak detection in networks, in: KDD, 2007, pp. 420–429.
[12] K. Subbian, D. Sharma, Z. Wen, J. Srivastava, Finding influencers in networks using social capital, in: ASONAM, 2013, pp. 592–599.
[13] C.-m. Au Yeung, T. Iwata, Strength of social influence in trust networks in product review sites, in: WSDM, 2011, pp. 495–504.
[14] H. Franks, N. Griffiths, S. S. Anand, Learning influence in complex social networks, in: AAMAS, 2013, pp. 447–454.
[15] J. Guo, P. Zhang, C. Zhou, Y. Cao, L. Guo, Personalized influence maximization on social networks, in: CIKM, 2013, pp. 199–208.
[16] K. Kutzkov, A. Bifet, F. Bonchi, A. Gionis, Strip: stream learning of influence probabilities, in: KDD, 2013, pp. 275–283.
[17] K. Subbian, C. C. Aggarwal, J. Srivastava, Content-centric flow mining for influence analysis in social streams, in: CIKM, 2013, pp. 841–846.
[18] R. Mansell, B. S. Collins, Introduction: Trust and crime in information societies, Edward Elgar, 2005.
[19] S. S. Brehm, Intimate relationships., Mcgraw-Hill Book Company, 1992.
[20] N. Luhmann, H. Davis, J. Raffan, K. Rooney, Trust; and, Power: two works by Niklas Luhmann, Wiley Chichester, 1979.
[21] X. Jin, C. Lin, J. Luo, J. Han, A data mining-based spam detection system for social media networks, Proceedings of the VLDB Endowment 4 (12).

[22] F. Chua, E.-P. Lim, Trust network inference for online rating data using generative models, in: KDD'10, pp. 889–898.
[23] Z. Gyongyi, P. Berkhin, H. Garcia-Molina, J. Pedersen, Link spam detection based on mass estimation, in: Proceedings of the 32nd international conference on Very large data bases, VLDB Endowment, 2006, pp. 439–450.
[24] J. Leskovec, D. Huttenlocher, J. Kleinberg, Predicting positive and negative links in online social networks, in: Proceedings of the 19th international conference on World wide web, 2010, pp. 641–650.
[25] J. Leskovec, D. P. Huttenlocher, J. M. Kleinberg, Signed networks in social media, in: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Atlanta, Georgia, USA, April 10-15, 2010, 2010, pp. 1361–1370.
[26] Y. Li, W. Chen, Y. Wang, Z. Zhang, Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships, in: Sixth ACM International Conference on Web Search and Data Mining, WSDM 2013, Rome, Italy, February 4-8, 2013, 2013, pp. 657–666.
[27] T. Iwata, A. Shah, Z. Ghahramani, Discovering latent influence in online social activities via shared cascade poisson processes, in: KDD, 2013, pp. 266–274.
[28] G. Liu, Y. Wang, M. A. Orgun, Trust transitivity in complex social networks., in: AAAI, Vol. 11, 2011, pp. 1222–1229.
[29] T. Haveliwala, Efficient computation of pagerank.
[30] Z. Gyöngyi, H. Garcia-Molina, J. Pedersen, Combating web spam with trustrank, in: Proceedings of the Thirtieth international conference on Very large data bases-Volume 30, VLDB Endowment, 2004, pp. 576–587.
[31] M. Granovetter, Threshold models of collective behavior, American journal of Sociology 83 (6) (1978) 1420–1443.
[32] M. Richardson, R. Agrawal, P. Domingos, Trust management for the semantic web, in: ISWC, 2003.
[33] J. L. Jaewon Yang, Defining and evaluating network communities based on ground-truth, in: ICDM, 2012.
[34] R. J. Deluga, The relation between trust in the supervisor and subordinate organizational citizenship behavior., Military Psychology 7 (1) (1995) 1.
[35] S. Feng, X. Chen, G. Cong, Y. Zeng, Y. M. Chee, Y. Xiang, Influence maximization with novelty decay in social networks, in: AAAI, 2014, pp. 37–43.
URL `http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8485`
[36] M. Purohit, B. A. Prakash, C. Kang, Y. Zhang, V. S. Subrahmanian, Fast influence-based coarsening for large networks, in: KDD, 2014, pp. 1296–1305. doi:10.1145/2623330.2623701.
URL `http://doi.acm.org/10.1145/2623330.2623701`
[37] A. Srivastava, C. Chelmis, V. K. Prasanna, Influence in social networks: A unified model?, in: ASONAM, 2014, pp. 451–454. doi:10.1109/ASONAM.2014.6921624.
URL `http://dx.doi.org/10.1109/ASONAM.2014.6921624`