CASINO: Towards Conformity-aware Social Influence Analysis in Online Social Networks

Hui Li **Sourav S Bhowmick** **Aixin Sun**

School of Computer Engineering
Nanyang Technological University, Singapore
assourav|axsun@ntu.edu.sg

Abstract

Social influence analysis in online social networks is the study of people's influence by analyzing the social interactions between individuals. In recent years, there have been increasing research efforts to understand the influence propagation phenomenon due to its importance to viral marketing and information dissemination among others. Despite the progress achieved by state-of-the-art social influence analysis techniques, a key limitation of these techniques is that they *only* utilize *positive* interactions (*e.g.,* agreement, trust) between individuals, ignoring two equally important factors, namely, *negative* relationships (*e.g.,* distrust, disagreement) between individuals and *conformity* of people, which refers to a person's inclination to be influenced.

In this paper, we propose a novel algorithm for social influence analysis called CASINO (**C**onformity-**A**ware **S**ocial **IN**fluence c**O**mputation), which quantitatively studies the interplay between *influence* and *conformity* of each individual by exploiting the positive and negative relationships between individuals. Given a social network, CASINO first extracts a set of *topic-based subgraphs* where each subgraph depicts the social interactions between individuals associated with a specific topic. Then it optionally labels the edges (relationships) between individuals with positive or negative signs. Finally, it iteratively computes the *influence* and *conformity indices* of each individual in each *signed* topic-based subgraph. Our exhaustive empirical study with several real-world social networks demonstrates superior effectiveness and accuracy of CASINO for social influence analysis compared to state-of-the-art methods. Furthermore, our investigation revealed several interesting characteristics of "influentials" and "conformers" in these social networks.

1 Introduction

Though the field of social network analysis (SNA) has developed over the past 50 or more years, it is with the recent emergence of large-scale online social networking applications (e.g., Facebook, LinkedIn, MySpace) that techniques from this area have received a great deal of attention. We are now faced with the opportunity to analyze social network data at unprecedented levels of scale and temporal resolution for marketing, health, politics, communication, education and other applications. However, translating the research techniques of traditional SNA to these large-scale online data-intensive applications is a daunting task. In this paper, we present our work towards addressing one of the challenges, namely the *social influence analysis* problem.

The goal of social influence analysis is to study individuals' influence by analyzing the social interactions between people. Recently, it has attracted tremendous research interest due to its role in governing the interactions in social networks as well as understanding the spread patterns of social influence. By identifying the "influentials" in a social network, users may be able to maximize the influence of a piece of information [8, 9, 15]. Informally, influentials are those individuals whose opinions or advices are often accepted and supported by others. For instance, Domingos and Richardson [9, 23] are the first to study influence maximization as an algorithmic problem. They proposed a probabilistic model of interaction and heuristics were given for choosing individuals with a large overall effect on the network. Kempe et al. [15] studied this problem from discrete optimization perspective and proposed three cascade models for influence propagation. They proposed a greedy algorithm which aimed to find a limited number of influentials from whom the information diffusion can be maximized.

Recently, Leskovec et al. [16, 17] viewed online social networks as *signed* networks, where social interactions involve both positive and negative relationships (edges). For instance, consider the signed network in Figure 1 depicting interactions between a set of individuals. An edge pointing from u to v denotes that person u trust/agree (*resp.,* distrust/disagree) person v. An edge representing agreement or trust relationship is labeled as positive (*e.g.,* edge $\overrightarrow{uv}$) whereas the one representing disagreement or distrust is labeled as negative (*e.g.,* edge $\overrightarrow{wv}$). Note that social influence flows in the opposite direction of the edges (*i.e.,* v influenced u and w).

The representation of social interactions as positive and negative relationships demanded a revisit of the social influence analysis problem as general influence analysis techniques overwhelmingly assumed only positive edges among individuals. However, negative relationships between individuals are valuable in several real-world scenarios (*i.e.,* politics) as they often carry as much information as the positive ones. For instance, in 2006, 31% of US residents used the Internet for gathering or sharing political information (above 60 millions people). 28% of them mentioned that most sites they use share their point of views while 29% of the Internet users mentioned that most challenge their point of views [20]. Thus, it is imperative to reconsider social influence analysis problem by taking into account

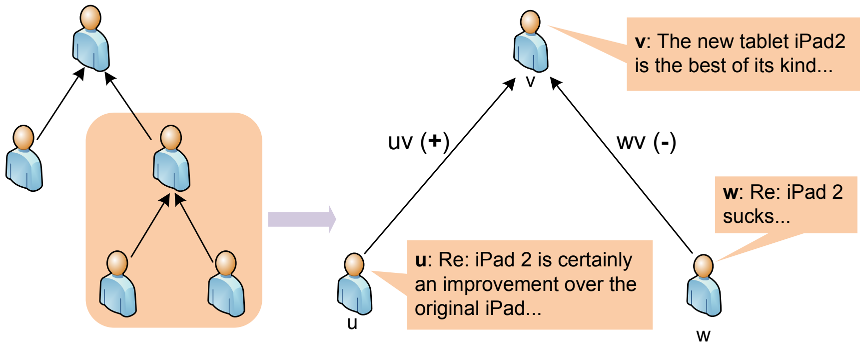


Figure 1: Social influence propagation scenario.

the signed interactions between individuals.

1.1 Motivation

A closer analysis of social influence phenomenon in signed networks reveals that there are three important factors that play a key role. Firstly, an individual's ability to influence others (*e.g.,* v in Figure 1). Secondly, the nature of social interactions (positive or negative) between individuals (*e.g.,* $\overrightarrow{uv}$ and $\overrightarrow{wv}$). Lastly, the degree of *conformity* of an individual, which is a person's inclination to be influenced [3]. The last factor is important as empirical studies have shown that large conversation are not only driven by the limited number of influentials but also by the large population of the early adopters who *accept* the influentials' opinion [24]. Observe that an individual's ability to influence or conform is *context-sensitive*. For example, reconsider Figure 1, where u conforms to v's opinion on `iPad 2`. However, it does not necessarily mean that u will always conform to v on *any* topic. For instance, u may not agree with v on conversation related to `salsa dancing` as u may believe that she is a better dancer than v. Additionally, v may not even be considered as influential on this topic.

Despite the benefits of the state-of-the-art social influence analysis techniques, a key limitation is their inability to systematically exploit the aforementioned second and third factors for superior analysis. Majority of existing research have overwhelmingly focused on considering only the positive relationships. Only very recently Cai et al. [7] took a step towards classifying influential individuals into *Positive Persona*, *Negative Persona*, and *Controversy Persona*, by exploiting both positive and negative relationships in the network. However, they ignored the effect of conformity of individuals on influentials. To elaborate further, consider the two signed networks in Figure 2 where an edge $\overrightarrow{uv}$ with positive (*resp.,* negative) sign indicates u trust (*resp.,* distrust) v. The shadowed part depicts the conformity of individuals u_1 and u_2. Specifically, u_1 is easily convinced by others whereas u_2 is not. Thus, it is easier for v_1 to influence u_1 than v_2 to influence u_2. However, state-of-the-art approaches have ignored this issue and failed to differentiate between these two cases. Consequently, these conformity-unaware techniques compute the

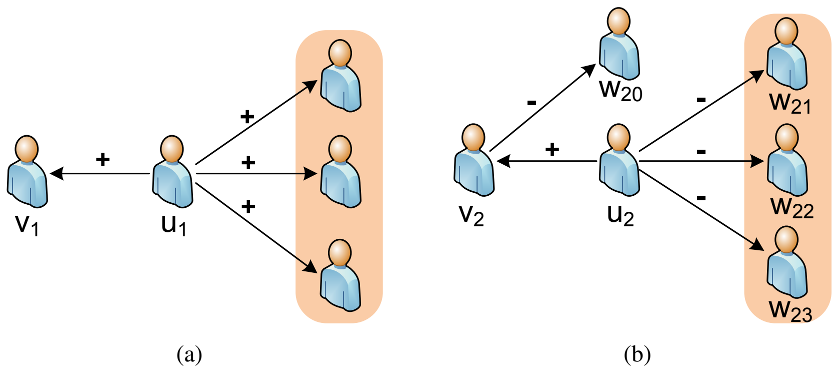


Figure 2: Conformity and negative edge effect.

same *influence score* for v_1 and v_2. However, an individual's influence should be increased if a larger number of users conform to her (positive interactions) but decreased if she is distrusted by individuals who are not conforming to her (negative interactions). *Is it possible to design a conformity-aware social influence analysis strategy that can address the aforementioned limitation?* In this paper, we provide an affirmative answer to this question.

1.2 Overview and Contributions

We propose a novel algorithm called CASINO (**C**onformity-**A**ware **S**ocial **IN**fluence c**O**mputation) that somewhat departs from existing influence analysis techniques in the following way: where existing strategies essentially ignore conformity of individuals, CASINO focuses on integrating the interplay of influence and conformity of individuals for social influence analysis by exploiting the positive and negative signs of edges. Additionally, it is *context-aware*, allowing the same individual to exhibit different influence and conformity over different topics of social interactions.

Given a social network, if it is a *context-aware* one, then CASINO first extracts a set of *topic-based subgraphs*. Each subgraph depicts the social interactions between individuals associated with a specific topic. Since the edges of a social network may not be always explicitly labeled with positive or negative signs, CASINO exploits an existing sentiment analysis technique to label the edges in each topic-based subgraph. Finally, given a set of signed topic-based subgraphs, the algorithm iteratively computes the *influence* and *conformity indices* of each individual in each subgraph.

To validate the proposed algorithm, a series of experiments have been conducted on five public datasets from three popular online social media sites, *i.e., Slashdot*[1], *Epinions*[2], and *Twitter*. *Epinions* and *Slashdot* are *context-free* networks where individuals trust (distrust) each other regardless of any specific topic. The

[1] http://slashdot.org/
[2] http://www.epinions.com

sign of each edge in these networks is explicitly provided. *Twitter* is a *context-aware* network where each conversation is based on a specific topic. Note that the signs of edges are not explicit in *Twitter* and hence they need to be extracted using a sentiment analysis technique.

There is no direct mechanism to justify the list of influentials generated by a generic social influence analysis technique. Hence, to quantitatively evaluate the performance of CASINO, we borrow the experimental framework articulated by Leskovec et al. [16, 17]. In this framework, machine-learning based approach is used for discovering the presence of unknown edges and edge sign prediction, wherein the conformity and influential indices of individuals are used as additional features to describe the signed edges. The intuition for using this framework is as follows. When an individual v influences u, it results in an edge $\overrightarrow{uv}$ in a social network graph. Moreover, if v has more influence on u (u is a conformer), then there is a greater probability of the existence of $\overrightarrow{uv}$. Thus, presence of a positive edge in a social network can be interpreted as the result of influence of v and conformity of u. In summary, the main contributions in this paper are as follows.

- In Section 3, we discuss the roles of influentials and conformers in the context of social influence analysis and describe how to quantify influence and conformity of each individual in a social network. *To the best of our knowledge, we are the first to study the interplay of influentials and conformers with the goal of social influence analysis.*
- In Section 4, we propose an iterative algorithm called CASINO that utilizes signs of social interactions (edges) to compute the influence and conformity indices of each vertex in an arbitrary social network. Moreover, we prove that the proposed algorithm is guaranteed to converge.
- By applying CASINO to real-world online social media sites, in Section 5, we strongly demonstrate the effectiveness and superiority of CASINO compared to state-of-the-art approaches and at the same time reveal several interesting characteristics of influentials and conformers in these sites.

In the next section, we review related work.

2 Related Work

We address related work from a number of relevant research areas, including: influence within online social media; conformity in social psychology; and analysis and prediction of signed networks.

Social influence within social media. Several recent works have focused on developing effective techniques to find a limited number of individuals who have significant influence on others within an online social network. These efforts use graph-theoretic approaches where influence analysis is transferred to a link analysis problem and the directed links are viewed as influences between pairs of individuals. For instance, Agarwal et al. [1] proposed an influence-based ranking function for blogs which exploits the *influence* of posts in each blog. The influence of a post is computed based on its length, comments, and a *propagation factor* which is the aggregated influence from the posts that linked to and from the current one. Each blogger is associated with an influence score. Interestingly, this work demonstrated that the most influential bloggers are not necessarily those who are most *active* (large number of posts). Ma et al. [19] used heat diffusion models to find a set of k influential candidates as target for marketing strategy in social networks. Particularly, the influence propagation is modeled as a heat diffusion process within social networks where the influence each node receives at a particular timepoint follows a heat diffusion formula. Based on this idea, the top-k candidates whose heat diffused to the largest scope are selected using a greedy algorithm. Recently, Bao et al. [5] proposed an influence-based advertising model, which diffuses *hint words* of influential users to others and then matches advertisements for each user with aggregated hints. They tested their model on a large online Q & A community and showed that it can improve the advertisement *click through rate*. In a different media, Pal et al. [21] used the count of original tweets, conversational tweets, and re-tweets of a tweeter as features to rank the *authority* of each tweeter in the context of different topics. They employed a Gaussian Mixture Model to compute the authority score of each tweeter. To validate their results, they conducted a survey to rate the authority of the tweeters and use it as the ground truth for authority ranking. Bakshy et al. [4] also studied social influence in *Twitter* by assigning a score to the tweet first appeared in a conversation and then diffuse the score to other tweets along the conversation. Each tweet in the conversation receives a portion of the score. A regression tree-based approach is used to predict the unknown influence scores using a set of features including number of followers, number of tweets, etc. The predicted influence score was validated using an online survey, which asked people to rate the interestingness of the tweets.

The aforementioned research have exclusively emphasized on the influence capability of the individuals and ignored the conformity of those who are influenced. However as discussed in the preceding section, the interplay between influence and conformity are indispensable for investigating the social influence phenomenon. In contrast, CASINO seamlessly incorporates both factors for social influence analysis. Specifically, we demonstrate that by considering the interplay between influence and conformity, the social influence process can be modeled more accurately.

Conformity in social psychology. The notion of *conformity* originated in social psychology in the context of social networks. It is defined as yielding to perceived group pressure by copying the behavior and beliefs of others [3]. Several lines of work in social psychology have focused on conducting experiments on groups of people in order to find the cause of conformity from the aspect of human nature [6, 11, 14]. These efforts identified two major causes of conformity, namely, *influential* and *normative*. The former claims that people conform to others' opinions for advice [6]; the latter claims that people conform to others' opinions in order to be consistent with the social group regardless whether the opinion is right or not [11]. In this paper, we are inspired by the conformity study in social psychology and utilize it to enhance social influence analysis in online social networks. We take into account the conformity in social influence and propose a model to evaluate the conformity of each individual in a social group. However, classification of the causes of conformity within online social networks is beyond the scope of this paper.

Characteristics of signed social networks. In a different direction, there are also large bodies of work involving positive (friendly) and negative (antagonistic) links in social media. In [17], the authors analyzed two structural properties of positive edges and negative edges, referred to as *balance* and *status*, in three real-world networks (*Slashdot*, *Epinions* and *Wiki*[3]). *Balance* is a classical theory from social psychology, which postulates that when considering the relationships between three people, either only one or all three of the relations should be positive. *Status* is a theory of directed signed networks which postulates that when person A makes a positive link to person B, then A is asserting that B has higher status – with a negative link from A analogously implying that A believes B has lower status. The authors use these two theories to explain the observed edge signs in undirected and directed social networks.

Leskovec et al. [16] used logistic regression to predict the signs of edges in signed networks by exploiting 7-dimensional degree features and 16-dimensional triad features of the networks. They showed that their logistic regression-based method yields higher accuracy in predicting the signs of edges compared to the state-of-the-art. Besides, the authors demonstrated that the accuracy of predicting only the positive edges can be improved by taking into account the negative interactions. In other words, it is often important to consider the interplay between positive and negative interactions. Recently Cai et al. [7] proposed another feature (*i.e.,* influence) aside from the 7-dimensional degree features in [16]. A PageRank-like algorithm was developed to compute the influence of individual users and then use it as another feature in an SVM classifier to predict the signs of edges. They showed that by taking into account the social influence of individual users the accuracy of edge sign prediction can be significantly improved. Based on this, they categorized influence personae into *Positive Persona*, *Negative Persona*, and *Controversy Persona*. *Positive* and *Negative Personae* represent users with high positive and negative influence, respectively. The last kind of *Controversy Persona* represents a group of individuals who are liable to be challenged or supported by many.

In the aforementioned research, different structural features of signed networks (*i.e.,* positive/negative edges, in/out degree, influence, triad frequency) are used in machine learning methods in order to predict the signs of edges. However, these research do not address the following two issues. Firstly, although [7] has taken into account the influence of node u for prediction of the sign of edge $\overrightarrow{uv}$,

[3] http://www.wikipedia.org/

Table 1: Symbols.

Symbol	*Semantics*
G	social network graph
V	vertex set
E	edge set
E^+	positive edge set
E^-	negative edge set
A	topic ID
E_A	edge correlated with topic A
G_A	subgraph correlated with topic A
$\mathcal{G}$	topic-based subgraph set
$\Omega(\cdot)$	conformity index
$\Phi(\cdot)$	influence index
$\Omega_A(\cdot)$	conformity index with respect to topic A
$\Phi_A(\cdot)$	influence index with respect to topic A
$\overrightarrow{uv}$	the edge pointing from u to v
$\overrightarrow{uAv}$	the edge pointing from u to v on context topic A

it does not consider the conformity of v. That is, whether v accepts the opinion of u. In contrast, we not only study the influence of u on the sign of edge $\overrightarrow{uv}$ but also investigate the conformity of v and its effect on u's influence. Secondly, an individual v may agree with u on a certain topic as u maybe more knowledgeable than her in that topic. However, v may not agree with u in some other topic where u may not be as proficient as v. In this paper, we propose a *context-aware* strategy which allows the same individual to exhibit different influence and conformity in different topics.

3 Influence and Conformity

In this section, we formally introduce the notion of *influence* and *conformity* in the context of signed social networks. We begin by briefly introducing signed social networks, which lie at the foundation of our proposed strategy. In the sequel, we shall use the notations shown in Table 1 to represent different concepts.

3.1 Signed Social Networks

Social interactions in online social networks can be either positive (indicating relations such as friendship) or negative (indicating relations such as distrust and opposition). For instance, in online rating sites such as *Epinions*, people can give both positive and negative ratings not only to items but also to other users. In online discussion sites such as *Slashdot*, users can tag other users as “friends” (positive) and “foes” (negative). In blogosphere and *Twitter*, the reply relationship among users can be a positive or a negative one. In our following discussion, we treat such social interaction as signed directed graph.

In a *signed* social network $G(V,E)$, each edge has a positive or negative sign depending on whether it expresses a positive or negative attitude from the generator of the edge to the recipient [16]. Specifically in this paper, a positive sign indicates that the recipient supports the opinion of the generator whereas the negative sign represents otherwise. For example, Figure 2(b) depicts a signed social network. The positive edge $E^+ = \{\overrightarrow{u_2 v_2}\}$ represents trust relationship while the negative ones ($E^- = \{\overrightarrow{w_{20} v_2}, \overrightarrow{u_2 w_{21}}, \overrightarrow{u_2 w_{22}}, \overrightarrow{u_2 w_{23}}\}$) represent distrust relationships. Note that the signs on the edges are not always available explicitly. In networks such as *Epinions* and *Slashdot*, the sign of each edge is explicitly provided. However, in other networks such as blogosphere and *Twitter* the sign of each edge is not explicitly available. In this case, we need to preprocess the network using text mining methods to discover signs associated with the links (detailed in Section 4.2). Consequently, a social network $G(V,E)$ containing both positive and negative edges can be represented using a pair of graphs $G^+(V,E^+)$ and $G^-(V,E^-)$ such that the following hold.

$$\forall \overrightarrow{uv} \in E, \begin{cases} (\overrightarrow{uv} \in E^+) \cap (\overrightarrow{uv} \in E^-) = 0, \\ (\overrightarrow{uv} \in E^+) \cup (\overrightarrow{uv} \in E^-) = 1 \end{cases}$$

In other words, $G^+(V,E^+)$ denotes the induced graph of positive edges E^+ (trust/agreement relationship) and $G^-(V,E^-)$ denotes that of negative edges E^- (distrust/disagreement relationship).

3.2 Definitions

In our approach, each individual (vertex) in a signed network is associated with a pair of *influence index* and *conformity index* to describe the power of influence and conformity of the individual, respectively. Reconsider the signed network in Figure 2(b). Intuitively, the influence of v_2 should increase as aggregated conformity of those who trust v_2 (*i.e.,* u_2) increases. On the other hand, the influence of v_2 should decrease if the aggregated conformity of those who distrust v_2 (*i.e.,* w_{20}) increases. Thus, the *influence index* of an individual should capture this interplay of influence and conformity and penalize her whenever necessary.

Definition 1 ***[Influence Index]*** *Let $G^+(V,E^+)$ and $G^-(V,E^-)$ be the induced graphs of the signed social network $G(V,E)$. The influence index of vertex $v \in V$, denoted as $\Phi(v)$, is defined as follows.*

$$\Phi(v) \quad = \quad \sum_{\overrightarrow{uv} \in E^+} \Omega(u) - \sum_{\overrightarrow{uv} \in E^-} \Omega(u)$$

where $\Omega(u)$ represents the conformity index of vertex $u \in V$. ∎

Similarly, the *conformity index* of u_2 in Figure 2(b) depends on the influences of vertices which are trusted or distrusted by u_2. Intuitively, as the aggregated influence of those vertices which u_2 trust (*e.g.,* v_2) increases, u_2 is more inclined to conform to others. On the other hand, when the aggregated influence of vertices which u_2 distrust (*e.g.,*, w_{21}, w_{22}, w_{23}) increases, u_2 is less inclined to conform to others. This intuition is captured by *conformity index* which is defined as follows.

Definition 2 ***[Conformity Index]*** *Let $G^+(V, E^+)$ and $G^-(V, E^-)$ be the induced graphs of the signed social network $G(V, E)$. The conformity index of vertex $u \in V$, denoted as $\Omega(u)$, is defined as follows.*

$$\Omega(u) \quad = \quad \sum_{\overrightarrow{uv} \in E^+} \Phi(v) - \sum_{\overrightarrow{uv} \in E^-} \Phi(v)$$

where $\Phi(v)$ is the influence index of vertex $v \in V$. ∎

Thus, according to the above definition the influence index of v_2 in Figure 2(b) can be computed as $\Phi(v_2) = \Omega(u_2) - \Omega(w_{20})$. The conformity index of u_2 is computed as $\Omega(u_2) = \Phi(v_2) - \Phi(w_{21}) - \Phi(w_{22}) - \Phi(w_{23})$. Observe that the aforementioned definitions of influence and conformity are mutually dependent on each other. Consequently, a recursive computation framework is necessary to compute these two indices. In the next section, we shall present an algorithm called CASINO that computes the influence index and conformity index of every vertex iteratively.

Remark. Observe that the aforementioned definitions aim to measure each individual's ability to positively influence her neighbors. The absolute number of incoming edges play a lesser role compared to relative number of positive and negative edges. This is because an individual A with 500 positive and 500 negative incoming edges may have strong social ties but not necessarily good influence, as half of A's neighbors do not trust her. In contrast, another user B with 9 positive and 1 negative incoming edges may not have strong social ties compared to A but definitely exhibits good influence within his friends. Having said this, in our definitions we do consider the influence of node degree as a node with 100 positive and 10 negative incoming edges may exhibit higher index value compared to another with 10 positive and 1 negative incoming edges.

4 Conformity-aware Influence Computation

In this section, we formally describe the algorithm CASINO for computing conformity and influence indices of individuals in a social network containing positive and negative edges. We begin by briefly describing the notion of *context-aware* and *context-free* signed social networks to represent real-world online networks.

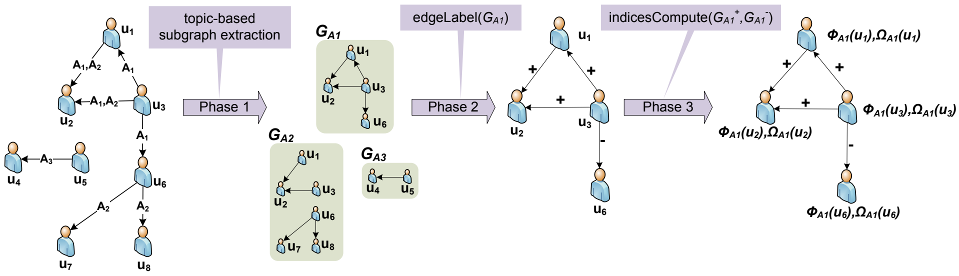


Figure 3: Overview of CASINO.

Algorithm 1: The CASINO algorithm.

Input: Social network $G(V,E)$
Output: the influence index $\mathbb{I}_A = (\Phi_A(u_1), \Phi_A(u_2), \ldots, \Phi_A(u_\ell))$ and conformity index $\mathbb{C}_A = (\Omega_A(u_1), \Omega_A(u_2), \ldots, \Omega_A(u_\ell))$ for $V = \{u_1, u_2, \ldots, u_\ell\}$ and for each topic A

begin
  if *G is context-aware* **then**
    $\mathcal{G} \leftarrow$ **extractSubgraph**(G);
  else
    $\mathcal{G} = \{G\}$;
  foreach $G_A \in \mathcal{G}$ **do**
    if *G_A is not a signed network* **then**
      $(G_A^+(V_A, E_A^+), G_A^-(V_A, E_A^-)) \leftarrow$ **edgeLabel**(G_A);
    $(\mathbb{I}_A, \mathbb{C}_A) \leftarrow$ **indicesCompute**($G_A^+(V_A, E_A^+), G_A^-(V_A, E_A^-)$);

4.1 Context-aware and Context-free Networks

Online social networks can be classified into *context-aware* and *context-free* networks. The former represent networks where the edges are associated with topics (context) as social interactions may often involve conversations on specific topics. For example, each conversation in *Twitter* is based on a specific topic. Figure 1 depicts interactions between three users on the topic `iPad 2`. On the other hand, interactions in context-free networks do not involve specific topics. For example, in *Epinions* and *Slashdot* individuals trust (distrust) each other regardless of any specific topic.

The leftmost social network in Figure 3 is an example of context-aware social network where an edge labeled as A_1,A_2 indicates that the pair of individuals communicate with each other on topics A_1 and A_2.

4.2 The Algorithm CASINO

The CASINO (**C**onformity-**A**ware **S**ocial **IN**fluence c**O**mputation) algorithm is outlined in Algorithm 1 and consists of three phases, namely the *topic-based subgraph* *extraction* phase (Line 3), the *edge labeling* phase (Line 8), and the *indices computation* phase (Line 9).

Figure 3 depicts an overview of the CASINO algorithm. Given a social network $G(V,E)$, if it is a context-aware network then the *topic-based subgraph extraction* phase extracts a set of subgraphs of G (denoted by $\mathcal{G}$) where each subgraph $G_A(V_A, E_A) \in \mathcal{G}$ contains all the vertices and edges in G associated with a specific topic A. Each subgraph G_A represents positive or negative attitudes of individuals toward opinions of others in G with respect to the topic A. For instance, in Figure 3, this phase generates three topic-based subgraphs, namely, G_{A_1}, G_{A_2}, and G_{A_3}, for topics A_1, A_2, and A_3, respectively. Recall that edges of a social network may not be explicitly labeled with positive or negative signs. This is especially true for context-aware networks (*e.g., Twitter*). On the other hand, links in many context-free networks (*e.g., Slashdot* and *Epinions*) are explicitly labeled with signs. Hence, it is important to label the edges in each topic-based subgraph G_A. The objective of the *edge labeling* phase is to assign sign to each edge by analyzing the sentiment expressed by the generator and recipient of the edge. Figure 3 depicts the labeling of G_{A_1}. Finally, given a set of signed topic-based subgraphs $\mathcal{G}$, the goal of the *indices computation* phase is to iteratively compute the influence and conformity indices of each individual in each $G_A \in \mathcal{G}$. Observe that a vertex v in G may have multiple pairs of indices if v is involved in more than one topic-based subgraph. Since the first phase is straightforward, we now elaborate on the remaining two phases in turn.

The edge labeling phase. The edge labeling method varies with dataset. In this paper, we adopt the method described in Algorithm 2. We denote each edge $\overrightarrow{uv}$ associated with topic A as $\overrightarrow{uAv}$. This enables us to differentiate between an edge which shares the same generator and recipient for more than one topic. For each edge $\overrightarrow{uAv}$ in a topic-based subgraph G_A, we identify 5-level sentiment (*i.e.,* like, somewhat like, neutral, somewhat dislike, dislike)[4] expressed at both ends using *LingPipe* [2] which is a popular sentiment mining package adopted in several recent research [12,18,22,25] (Lines 4-5). Note that *LingPipe* has been tested to provide very promising results (*i.e.,* with accuracy over 85% in most cases [12,18]) on sentiment extraction. If the sentiments at both ends are similar (*sentiment similarity threshold* is less than ε), we denote the edge as positive (Lines 6-7). Otherwise, we denote it as negative (Lines 8-9).

Indices computation phase. Given a topic A and topic-based subgraph G_A, the preceding phase generates G_A^+ and G_A^-. Without loss of generality, assume that there are $|\mathcal{G}|$ different topics. Then, we are able to compute an individual's influence and conformity indices for each topic (*i.e.,* $\Phi_A(u)$ and $\Omega_A(u)$). We now elaborate on the algorithm for computing these indices.

Algorithm 3 outlines the strategy for computing a pair of influence and confor-

[4] In this paper to simplify the discussion, we adopt a 5-level system of sentiment which is widely used in many rating networks such as Yahoo! Answer.

Algorithm 2: The *edgeLabel* procedure.

Input: Topic-based subgraph $G_A(V_A, E_A)$ induced by topic A,
Output: $G_A^+(V_A, E_A^+)$ and $G_A^-(V_A, E_A^-)$ such that: $E_A^+ \cup E_A^- = E_A$ and $E_A^+ \cap E_A^- = \emptyset$

begin
  $E_A^+ = E_A^- = \emptyset$;
  foreach $\overrightarrow{uAv} \in E_A$ **do**
    $u.sentiment \leftarrow LingPipe.\mathbf{sentExtr}(u)$;
    $v.sentiment \leftarrow LingPipe.\mathbf{sentExtr}(v)$;
    if $|u.sentiment - v.sentiment| < \varepsilon$ **then**
      $E_A^+ = E_A^+ \cup \{\overrightarrow{uAv}\}$
    else
      $E_A^- = E_A^- \cup \{\overrightarrow{uAv}\}$

mity indices ($\Phi(u)$, $\Omega(u)$) for each vertex u. It first initializes the influence index and conformity index of all vertices to be 1 (Lines 1-4). Subsequently, in each iteration it computes them for each vertex by using the values of the indices in previous iteration (Lines 6-8) and normalizing these values using the square root of the summation of all vertices' index values (Lines 9-13). The algorithm terminates when both indices converge. We shall now prove that the proposed algorithm is guaranteed to converge after a fixed number of iterations n. In other words, the difference between an arbitrary node's indices between n and $n+1$ rounds of iteration is insignificant and hence we do not need to consider additional iterations.

Theorem 1 *The indicesCompute procedure described in Algorithm 3 converges.*

Proof 1 *(Sketch)*
According to Definition 1, for each vertex u its influence index $\Phi(u)$ can be computed as the following.

$$\Phi(u) \quad = \quad \sum_{\overrightarrow{u'u} \in E^+} \Omega(u') - \sum_{\overrightarrow{u'u} \in E^-} \Omega(u')$$

If we denote $\mathbb{I} = (\Phi(u_1), \Phi(u_2), \ldots, \Phi(u_\ell))^\top$ and $\mathbb{C} = (\Omega(u_1), \Omega(u_2), \ldots, \Omega(u_\ell))^\top$ for $V = \{u_1, u_2, \ldots, u_\ell\}$, then the computation of both indices in each iteration can be represented as:

$$\begin{cases} \mathbb{I} = \mathbb{A}_+^\top \mathbb{C} - \mathbb{A}_-^\top \mathbb{C} \\ \mathbb{C} = \mathbb{A}_+ \mathbb{I} - \mathbb{A}_- \mathbb{I} \end{cases}$$

where $\mathbb{A}_+$ and $\mathbb{A}_-$ represent the adjacency matrices for G^+ and G^-, respectively. If we substitute $\mathbb{C}$ in the first equation using the second equation, then the first line

Algorithm 3: The *indicesCompute* procedure.

Input: $G(V,E) = G^{+}(V,E^{+}) \cup G^{-}(V,E^{-})$

Output: the influence index $\mathbb{I} = (\Phi(u_1), \Phi(u_2), \ldots, \Phi(u_\ell))$ and conformity index $\mathbb{C} = (\Omega(u_1), \Omega(u_2), \ldots, \Omega(u_\ell))$ for $V = \{u_1, u_2, \ldots, u_\ell\}$

begin
- $k = 1$ /*initialize iteration counter*/ ;
- **foreach** $u \in V$ **do**
  - $\Phi^{k}(u) = \Omega^{k}(u) = 1$
- **while** $\mathbb{I}$ *or* $\mathbb{C}$ *not converged* **do**
  - **foreach** $u \in V$ **do**
    - $\Phi_0^{k+1}(u) = \sum_{\overrightarrow{vu} \in E^{+}} \Omega^{k}(v) - \sum_{\overrightarrow{vu} \in E^{-}} \Omega(v);$
    - $\Omega_0^{k+1}(u) = \sum_{\overrightarrow{uv} \in E^{+}} \Phi^{k}(v) - \sum_{\overrightarrow{uv} \in E^{-}} \Phi(v);$
  - **foreach** $u \in V$ **do**
    - $\Phi^{k+1}(u) = \frac{\Phi_0^{k+1}(u)}{\sqrt{\sum_{v \in V} \Phi_0^{k+1}(v)^2}};$
    - $\Omega^{k+1}(u) = \frac{\Omega_0^{k+1}(u)}{\sqrt{\sum_{v \in V} \Omega_0^{k+1}(v)^2}};$
  - $\mathbb{I}^{k+1} = (\Phi^{k+1}(u_1), \Phi^{k+1}(u_2), \ldots, \Phi^{k+1}(u_l));$
  - $\mathbb{C}^{k+1} = (\Omega^{k+1}(u_1), \Omega^{k+1}(u_2), \ldots, \Omega^{k+1}(u_l));$
  - $k = k + 1;$

turns into the following:

$$\begin{aligned} \mathbb{I}_{k+1} &= \frac{1}{Z}(\mathbb{A}_{+}^{\top} - \mathbb{A}_{-}^{\top})(\mathbb{A}_{+} - \mathbb{A}_{-})\mathbb{I}_k \\ &= \frac{1}{Z}(\mathbb{A}_{+} - \mathbb{A}_{-})^{\top}(\mathbb{A}_{+} - \mathbb{A}_{-})\mathbb{I}_k \end{aligned}$$

where Z is a normalizing factor such that $\|\mathbb{I}_{k+1}\| = 1$. If we compute $\mathbb{I}_{k+1}$ using $\mathbb{I}_k$ for $k = 1, 2, \ldots, n$ recursively, then $\mathbb{I}_{n+1}$ should be the unit vector along the direction of

$$((\mathbb{A}_{+} - \mathbb{A}_{-})^{\top}(\mathbb{A}_{+} - \mathbb{A}_{-}))^{n}(\mathbb{A}_{+} - \mathbb{A}_{-})^{\top}(1, 1, \ldots, 1)^{\top}.$$

Similarly, $\mathbb{C}_{n+1}$ should be the unit vector along the direction of

$$((\mathbb{A}_{+} - \mathbb{A}_{-})(\mathbb{A}_{+} - \mathbb{A}_{-})^{\top})^{n+1}(1, 1, \ldots, 1)^{\top}.$$

According to the result in [10], if M is a symmetric matrix, and v is a vector not orthogonal to the principal eigenvector $\omega_1(M)$, then the unit vector in the direction of $M^k v$ converges to $\omega_1(M)$ as k increases.

Comparing with our case, $(1, 1, \ldots, 1)^{\top}$ is not orthogonal to $\omega_1((\mathbb{A}_{+} - \mathbb{A}_{-})(\mathbb{A}_{+} - \mathbb{A}_{-})^{\top})$, thus $\mathbb{C}_k$ converges. Similarly, $\mathbb{I}_k$ also converges.

In summary, both $\Phi(u)$ and $\Omega(u)$ converge.

Table 2: Statistics of context-free datasets.

Dataset	*#nodes*	*#edges*	*#positive edges*	*#negative edges*
Slashdot1	77,357	516,575	396,378	120,197
Slashdot2	81,871	545,671	422,349	123,322
Slashdot3	82,144	549,202	425,072	124,130
Epinions	131,828	841,372	717,667	123,705

Table 3: Statistics of the context-aware dataset.

Dataset	*#tweets*	*#trends*	*#tweeters*	*#edges*
Twitter	1,054,261	21,917	576,894	1,230,748

Observe that the aforementioned technique can easily be extended to compute the aggregated indices of an individual by taking into account the entire social network G over all topics $A = 1, \ldots, |\mathcal{G}|$. In this case, E^+ and E^- in Definitions 1 and 2 are replaced by $\bigcup_{A=1}^{|\mathcal{G}|} E_A^+$ and $\bigcup_{A=1}^{|\mathcal{G}|} E_A^-$, respectively.

5 Experimental Study

In this section, we report experimental results to quantitatively measure the performance of CASINO for discovering influential individuals by considering the conformity of others. To this end, we borrow the experimental framework articulated by Leskovec et al. [16, 17], in which machine-learning based approach is used for discovering the presence of unknown edges and edge sign prediction by exploiting various edge features. By considering conformity and influence indices as new features for classification, it is expected that these additional features will provide more concrete evidence for edge prediction compared to state-of-the-art strategies.

5.1 Datasets

We consider the following context-free and context-aware networks where each link is explicitly or implicitly labeled as positive or negative.

Context-free network data. In order to compare the performance of CASINO with state-of-the-art efforts on influence evaluation in signed networks [7, 16, 17], we adopt the same datasets that have been used in these work: *Slashdot* and *Epinions* [5]. Recall that these network data are context-free and contain explicit signs of edges to indicate the attitudes of individuals towards one another. Specifically, we obtained three *Slashdot* datasets at different timepoints. The statistics of each dataset is reported in Table 2.

[5] http://snap.stanford.edu/data/

Context-aware network data. We use the *Twitter* dataset to investigate the performance on a context-aware network. The dataset was crawled using the *Twitter* API [6] during Dec 2010 to Feb 2011. We extracted top 20 trends keywords at hourly duration and retrieved up to 1500 tweets for each trend. Then we identified the relationships between all the tweeters in the dataset. Table 3 reports the statistics associated with this dataset. Note that these statistics are computed after removing non-English tweets (using *Twitter* API). In order to compute accurate influential and conformity indices, we need to have large context-aware interaction graphs. We removed spam trend keywords which contain only meaningless IDs. Thus, we selected top 492 trends that contain more than 1,000 tweets to compute the indices. For each trend (topic), we identified all the tweets associated to it. Then the edges connecting different tweets using `'@'` tag are extracted and their signs are assigned as positive or negative using Algorithm 2. Additionally, there exists another tag `'RT'` in many tweets indicating that a tweet author supports another author's opinion by re-tweeting it. That is, if an author u directly re-tweets another twitter v, then it indicates that u wants to distribute this tweet to her followers. Hence, we assign positive signs to such re-tweet edges.

5.2 Experimental Setup

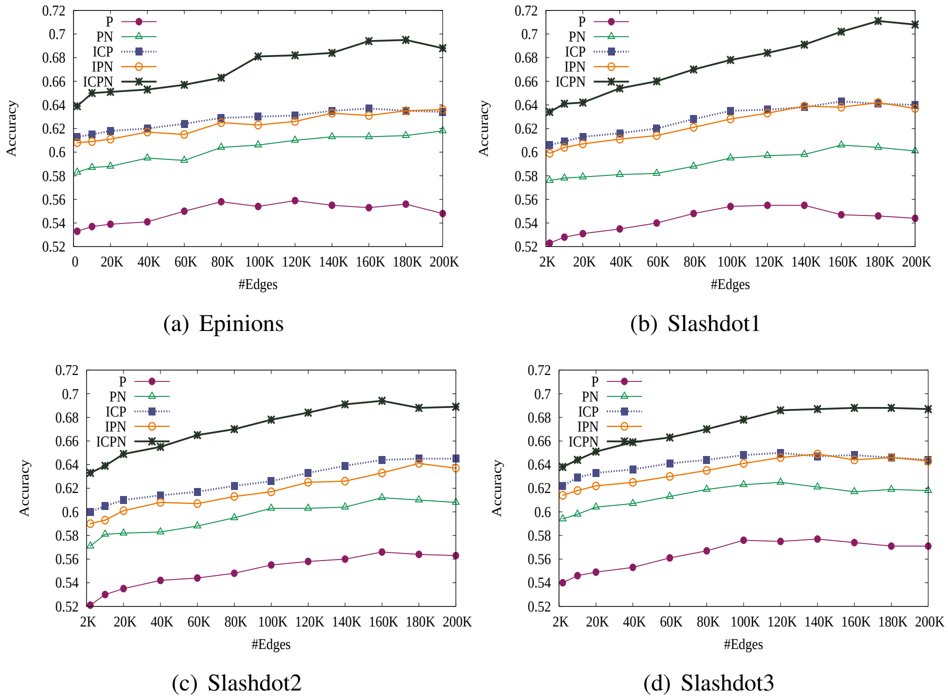
We have implemented CASINO using Java and run all the experiments on a 1.86GHz Intel 6300 machine with 4GB RAM with Windows XP. The algorithm converges after average 30 iterations. For the *Twitter* dataset, the sentiments for all the tweets are classified into 5 different levels (*i.e.,* 1-5) using the *LingPipe* library [2]. We set the *sentiment similarity threshold* ε (Algorithm 2) to 1. Note that it is not effective to set $\varepsilon > 1$. For instance, if ε is set to 2 or 3 then it indicates that the sentiment is 'somewhat dislike' or 'neutral', respectively. Consequently, higher value of ε will ignore negative edges. Obviously, the selection of ε depends on how many levels of sentiment we have classified.

Similar to [7,16,17], in our experiments, the task of edge sign prediction is considered as a binary classification problem. We adopted SVM^{light} classifier [13] and classification *accuracy* is taken as the main measure for evaluation. Specifically, accuracy is defined as follows.

$$accuracy = \frac{\#TP + \#TN}{\#TP + \#FP + \#TN + \#FN}$$

In the above equation TP and TN stand for "true positives" and "true negatives", respectively; FP and FN stand for "false positives" and "false negatives", respectively. The reason for adopting accuracy as evaluation measure over precision and recall is that the former is suitable for quantifying the prediction performances of both positive and negative samples. This is crucial in our framework as in most experiments both positive and negative edges are being predicted.

[6] `http://dev.twitter.com/doc`



(a) Epinions (b) Slashdot1

(c) Slashdot2 (d) Slashdot3

Figure 4: Positive edge presence prediction.

As mentioned in [16], the overwhelming majority of the edges in *Slashdot* and *Epinions* are positive. Consequently, random guessing can achieve approximately 80% accuracy [7]. In order to avoid such biased classification, we adopt the same strategy used in [7, 16, 17]. Specifically, we create a balanced dataset with same number of positive and negative edges for training and testing.

In all experiments we report the average accuracy and perform 5-folds cross validation. For example, for the dataset with 200K edges, we first randomly select 100K negative edges and separate them into 5 parts each of which include 20K negative edges. Then we iteratively select 20K random positive edges and add them into each of the 5 negative edge sets. Based on the learned model (discussed below), we predict a label 1 or -1 for each target edge indicating its possibility to be positive or negative, respectively.

In order to compare the prediction accuracy of the proposed approach with state-of-the-art efforts, a baseline classifier is constructed by referring to the structural features discussed in [16]. Specifically, given an edge from vertex u to v, a 7-dimensional feature vector $\{d^+_{in}(v), d^-_{in}(v), d^+_{out}(u), d^-_{out}(u), d_{in}(v), d_{out}(u), C(u,v)\}$ is constructed, where features $d^+_{in}(v)$ and $d^-_{in}(v)$ denote the number of positive and negative incoming edges to v, features $d^+_{out}(u)$ and $d^-_{out}(u)$ represent the number of positive and negative outgoing edges from u, features $d_{in}(v)$ and $d_{out}(u)$ denote the number of in-degree and out-degree of v and u, and $C(u,v)$ denotes the total number of common neighbors of u and v without considering the edge direction. Then,

Table 4: Features involved in different approaches. (P: positive, N: negative, I: influence, C: conformity, A: topic)

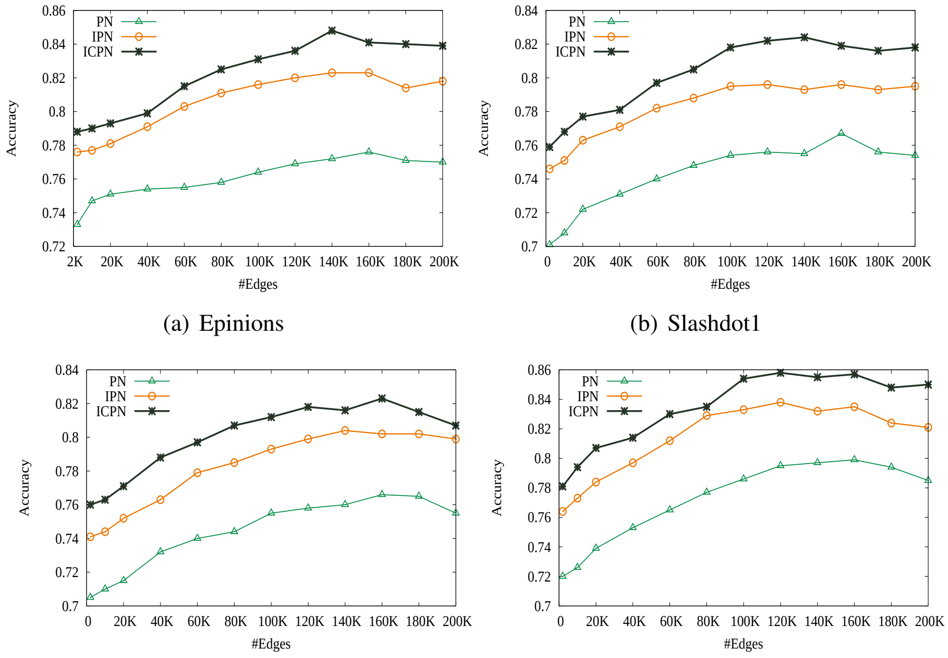
Features \ Approaches	P	PN	IPN	ICP	ICPN	ICAPN
$d^+_{in}(v)$	✓	✓	✓	✓	✓	✓
$d^-_{in}(v)$	-	✓	✓	-	✓	✓
$d^+_{out}(u)$	✓	✓	✓	✓	✓	✓
$d^-_{out}(u)$	-	✓	✓	-	✓	✓
$d_{in}(v)$	-	✓	✓	-	✓	✓
$d_{out}(u)$	-	✓	✓	-	✓	✓
$C(u,v)$	✓	✓	✓	✓	✓	✓
$\Phi(u)$	-	-	✓	-	-	-
$\Phi(v)$	-	-	✓	✓	✓	✓
$\Omega(u)$	-	-	-	✓	✓	✓
context A	-	-	-	-	-	✓
$\Phi_A(v)$	-	-	-	-	-	✓
$\Omega_A(u)$	-	-	-	-	-	✓

we create variants of this baseline classifier by adding influence index and conformity index of vertices as new features in the feature vector. Table 4 describes the features involved in the feature vector for each edge $\overrightarrow{uv}$ for different variations of the baseline classifier.

5.3 Experimental Results

The goals of the evaluation were to establish whether the proposed approach can reliably predict presence of edges; predict signs of edges; and seek to understand the characteristics and importance of influence and conformity indices for these prediction tasks.

Positive edge presence prediction. We first conduct a series of experiments to predict the presence of positive edges. Note that in order to test the effect of negative edges in predicting the presence of positive edges, we adopted a regression function of SVM^{light} where presence of positive edges are labeled as 1 and negative edges are labeled as -1. We investigate how each of the aforementioned classifiers perform in predicting the presence of positive edges. Figure 4 shows the average accuracy for the benchmark datasets, where the results are compared between different classifiers for different size of training and testing data. We can make the following observations. First, for each dataset PN performs better than P indicating that information related to negative edges enhance the quality of edge presence prediction. Second, the prediction accuracy is further improved when we incorporate the influence index or conformity index as a feature (ICP and IPN). Note that ICP performs slightly better than IPN but the improvement is not significant. When the training set is large both approaches exhibit similar performance. Third, ICPN consistently reports the best prediction performance. That is, edge presence prediction



(a) Epinions (b) Slashdot1

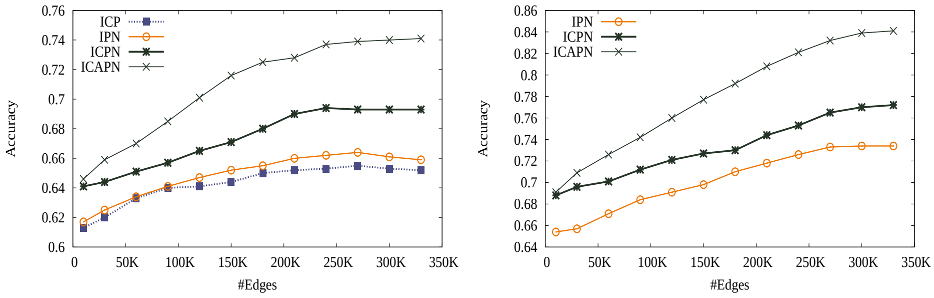
(c) Slashdot2 (d) Slashdot3

Figure 5: Signed edge prediction.

is enhanced when we consider the interplay between influentials and conformers.

Signed edge prediction. Next, we undertake a series of experiments to predict the signs of edges. Similar to our earlier experiments, we ensure that the training set and test set both contain equal number of positive and negative edges. We vary the size of training set to test the prediction accuracy. Note that in a binary classification, positive edges and negative edges belong to two different classes. Our goal is to predict the signs of edges which maybe either positive or negative. Figure 5 reports the prediction accuracies of the classifiers. Observe that among the three approaches involving both positive and negative edges (PN, IPN, and ICPN), ICPN performs the best, followed by IPN and PN, respectively. Thus, by taking into account the influence and conformity of vertices, the accuracy of sign prediction task can improve significantly.

Edge presence and signed edge prediction in context-aware networks. We now report the performance of our model on a context-aware network (*Twitter*). In this experiment we adopt the classifier ICAPN which takes into account the topic information associated with each edge. That is, the following features for each edge $\overrightarrow{uAv}$ are used to train the model: $d^+_{in}(v)$, $d^-_{in}(v)$, $d^+_{out}(u)$, $d^-_{out}(u)$, $d_{in}(v)$, $d_{out}(u)$, $C(u,v)$, $\Phi(v)$, $\Omega(u)$, $\Phi_A(v)$ and $\Omega_A(u)$. Figure 6 plots the prediction accuracies of the relevant classifiers. Observe that in both figures ICAPN outperforms the rest. Note that the performances of ICAPN and ICPN are similar when the training set is very



(a) Positive edge presence (b) Edge sign prediction

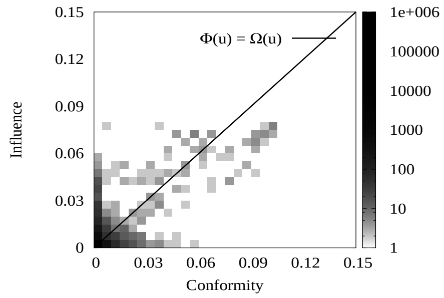
Figure 6: Context-aware prediction accuracy.

small. This is because there may not be enough training edges in each topic-based subgraphs G_A when the training set is very small. Consequently, not enough information is available to accurately compute $\Phi_A(v)$ and $\Omega_A(u)$. All these evidences demonstrate that by leveraging on the influence and conformity indices in topic-based subgraphs, the proposed model leads to superior prediction performance for both positive edge presence and edge sign prediction tasks.

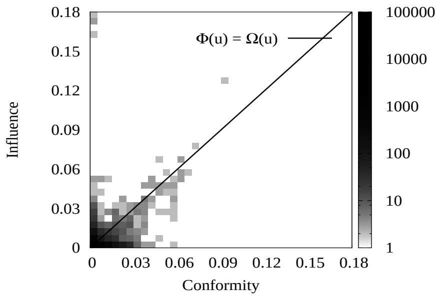
Influentials and conformers. Lastly, we analyze the list of influentials and conformers detected by the CASINO algorithm. Figures 7 and 8 depict the distribution heatmap of influence index versus conformity index for each benchmark dataset. For each individual u in a network we compute her influence index $\Phi(u)$ and conformity index $\Omega(u)$ and represent it as a point in the influence-conformity 2-D plane. Then we separate the plane into grids of size 0.005×0.005 and count the number of points in each grid. The color shade of a grid denotes the number of points residing in it. Note that both influence index and conformity index are normalized into the range of $[0, 1)$. For each figure, we explicitly draw a boundary line along which the vertices exhibit identical influence index and conformity index. Observe that the line separates the influence-conformity plane into two areas. In the sequel, we refer to the top area as '*Area I*' and the down one as '*Area II*'. The points belonging to '*Area I*' exhibit higher influence index compared to conformity index, indicating that individuals in this area are more prone to influence others than being influenced. We refer to them as *influence-biased*. On the other hand, the points in '*Area II*' represent individuals who are conforming in nature. That is, they are more prone to be influenced than influencing others. We refer to these individuals as *conformity-biased*.

We first analyze the context-free networks (Figures 7(a)-(d)). Consider Figure 7(a) related to *Epinions* dataset. Observe that 31% of all individuals belong to '*Area I*'. Consequently, fewer number of individuals in this network are influence-biased. That is, majority of individuals in *Epinions* are often conforming to the others. Similar phenomenon also exists in the *Slashdot* datasets where the percentage of individuals in '*Area I*' is between 36% to 37%.

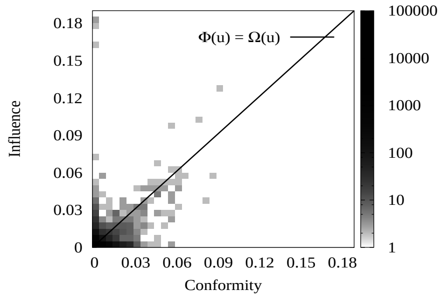
Observe that those vertices in *Epinions* which exhibit very high conformity in-



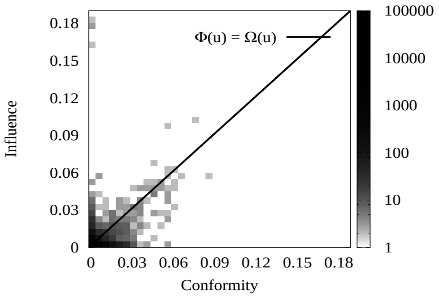
(a) Epinions



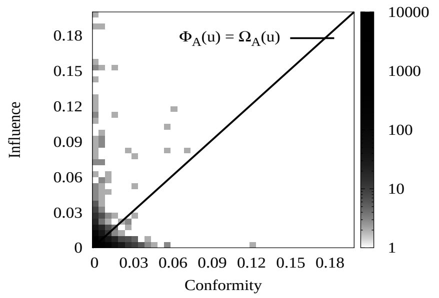
(b) Slashdot1



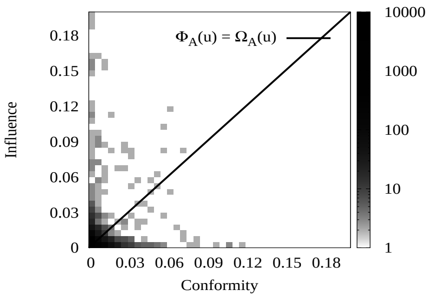
(c) Slashdot2



(d) Slashdot3



(e) Twitter (Top-1)



(f) Twitter (Top-2)

Figure 7: Influence vs. conformity distribution heatmap.

dex values also have high influence index values. On the other hand, vertices with highest influence index values have a wider range of conformity indices (*i.e.,* from 0 to 0.11). Such phenomenon indicates that in *Epinions* most influence-biased individuals may also conform to others whereas the most conformity-biased individuals are always influencing others. Interestingly, the phenomenon is different in the three *Slashdot* datasets (Figures 7(b)- 7(d)). Specifically, individuals who are associated with highest influence index values have very small conformity index (*i.e.,* less than 0.02). But individuals with highest conformity index may not exhibit small influence index values. In fact, the conformity index values of these conformity-biased individuals are distributed along the boundary line ($\Phi(u) = \Omega(u)$). Thus, we can make the following observations regarding *Slashdot*. Firstly, there are a few influence-biased individuals who exhibit very high influence but are not easily influenced by others. Secondly, there do not exist conformity-biased individuals who are not influencing others at all.

Next, we analyze the context-aware network (Figures 7(e)-(f) and 8). Figures 7(e)-(f) show the distributions of influence and conformity indices for the top-2 topics (`Mumford & Sons` and `BornThisWayFriday`) with the most number of tweets (4390 and 4046, resp.). Observe that 40% and 45% of all the individuals fall in '*Area I*' for Figure 7(e) and (f), respectively. Notably, both these figures exhibit *certain* influence-biased characteristics similar to *Slashdot*. That is, there are a few influence-biased individuals who exhibit very high influence but are not easily influenced by others. This similarity may be due to the fact that both *Slashdot* and *Twitter* are driven by user conversations where majority individuals are commenting or following a few individuals who started the conversations. Figure 8 plots the distribution of indices computed over *all* topics. In this case, 41% of all individuals belong to '*Area I*'. The most influential author has influence index of 0.138 whereas the most conforming individual has a conformity index of 0.082.

Table 5 shows IDs of top-10 authors who exhibit the highest influence index and conformity index for the top-2 topics as well as for all topics. Consider the top two twitters for all topics. The author `'142987924'` who has the highest influence index receives 66 conforming edges out of 73 in-links over 22 topics. Similarly, the author `'49276778'` receives 61 conforming edges out of 82 in-links over 24 topics. On the other hand, the author `'51389816'` who exhibits the highest conformity index initiates 35 conforming edges out of 37 out-links over 37 topics indicating that she has high chance to conform to others' opinions in almost all the topics she is involved in. Furthermore, we can make the following observations. Firstly, none of the top-10 authors occupies a position in *both* indices for each category (*all*, *top-1*, and *top-2*). Secondly, the top-10 individuals having highest influence and conformity indices are different for different topics. This confirms our hypothesis that social influence phenomenon is context-sensitive as same individual may exhibit different influence and conformity over different topics of social interactions.

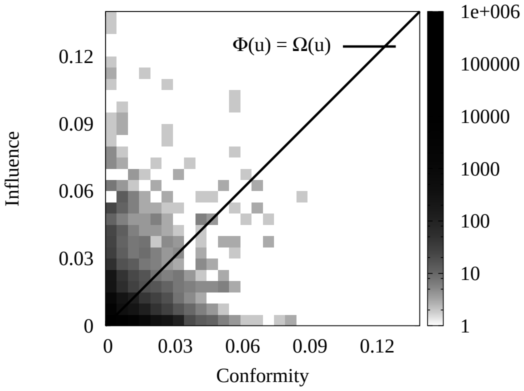


Figure 8: Heatmap of *Twitter* for all topics.

Table 5: Top 10 authors with the highest influence index and conformity index.

Rank	Influential twitter (#positive in-links/#in-links)			Conformer twitter (#positive out-links/#out-links)		
	All	*Top-1*	*Top-2*	*All*	*Top-1*	*Top-2*
1	142987924 (66/73)	3453454 (13/13)	950596 (31/34)	51389816 (35/37)	121836131 (14/16)	49276778 (101/144)
2	49276778 (61/82)	56068621 (11/11)	190108655 (11/11)	172039151 (31/34)	105332925 (13/14)	202346609 (45/61)
3	119394881 (60/77)	3984874 (10/10)	3498571 (8/9)	177173204 (30/35)	177255919 (11/12)	197538544 (26/30)
4	231134989 (55/71)	133282617 (11/11)	147327886 (5/5)	143062806 (27/34)	193206052 (11/12)	184930795 (22/26)
5	2109823 (56/72)	199855121 (7/7)	49126931 (5/5)	128118710 (25/33)	36525648 (9/10)	148335502 (21/23)
6	92503401 (55/78)	8234375 (5/5)	121158546 (5/5)	130414633 (30/41)	90723076 (7/8)	171387567 (17/20)
7	206661373 (51/66)	1465130 (3/3)	129009252 (5/5)	4782790 (23/30)	123606641 (6/8)	126407259 (18/22)
8	220490093 (46/60)	2894822 (3/3)	79897503 (4/4)	125551983 (22/34)	51513825 (6/6)	114455733 (14/20)
9	168175236 (40/51)	21755211 (2/2)	83629945 (4/4)	91930055 (21/28)	203774695 (5/6)	217826740 (15/20)
10	171287044 (41/62)	4051581 (2/2)	166830172 (4/4)	145339829 (22/31)	203780314 (4/4)	159724683 (12/17)

6 Conclusions and Future Work

The social influence analysis problem for online social networks, which focuses on studying people's influence by analyzing social interactions between individuals, is considered important with applications to viral marketing and information dissemination among others. Recently, several techniques have been proposed to address this problem by exploiting the positive interactions (*e.g.,* trust, agreement, friendship) between individuals. However, these techniques ignore two equally important factors that play a key role in social influence propagation, namely, negative relationships (*e.g.,* distrust, disagreement, antagonism) between individuals and conformity of people who are being influenced. In this paper, we propose a novel algorithm for social influence analysis called CASINO, which quantifies the influence and conformity of each individual in a network by utilizing the positive and negative relationships between individuals.

Our exhaustive experimental study using several online social media sites demonstrates the effectiveness and superior accuracy of CASINO compared to state-of-the-art methods. Specifically, our investigation revealed that the knowledge of conformity of individuals enhance the accuracy of social influence analysis. We also observed several interesting characteristics of influentials and conformers in *Slashdot*, *Epinions*, and *Twitter*. Particularly, in *Slashdot* and *Twitter*, there are a few individuals who exhibit very high influence but are not easily influenced by others. However, in *Epinions*, individuals who exhibit high influence are often conforming to others. Besides, in *Slashdots* and *Twitter* there do not exist individuals who are always influenced but they are not influentials. In contrasts, such individuals can be found in *Epinions*.

There are a number of further directions suggested by this work. A first one is of course to explore methods for maximizing the *spread* of influence [15] by incorporating conformity characteristics of individuals. We are also interested in investigating the role of the two causes of conformity, namely, influence and normative, to the influence analysis problem. In summary, the results of this paper are an important first step in this regard.

References

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In *WSDM*, pages 207–218, 2008.

[2] Alias-i. Lingpipe 4.0.1. http://alias-i.com/lingpipe, 2008.

[3] E. Aronson, T. D. Wilson, and R. M. Akert. *Social Psychology*. Prentice Hall, fifth edition, Feb. 2004.

[4] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: quantifying influence on twitter. In *WSDM*, pages 65–74, 2011.

[5] H. Bao and E. Y. Chang. Adheat: an influence-based diffusion model for propagating hints to match ads. In *WWW*, pages 71–80, 2010.

[6] R. S. Baron, J. A. Vandello, and B. Brunsman. The forgotten variable in conformity research: Impact of task importance on social influence. *Journal of Personality and Social Psychology*, 71:915–927, 1996.

[7] K. Cai, S. Bao, Z. Yang, J. Tang, R. Ma, L. Zhang, and Z. Su. Oolam: an opinion oriented link analysis model for influence persona discovery. In *WSDM*, pages 645–654, 2011.

[8] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *KDD*, pages 199–208, 2009.

[9] P. Domingos and M. Richardson. Mining the network value of customers. In *KDD*, pages 57–66. ACM Press, 2001.

[10] G. H. Golub and C. F. V. Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.

[11] A. L. Hodges, B. H.; Geyer. A nonconformist account of the asch experiments: Values, pragmatics, and moral dilemmas. *Personality and Social Psychology Review*, 10(1):2–19, 2006.

[12] Y. Jo and A. H. Oh. Aspect and sentiment unification model for online review analysis. In *WSDM*, pages 815–824. ACM, 2011.

[13] T. Joachims. Making large-scale SVM learning practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press, Cambridge, MA, USA, 1999.

[14] H. C. Kelman. Compliance, identification, and internalization: Three processes of attitude change. *The Journal of Conflict Resolution*, 2(1), 1958.

[15] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD*, pages 137–146, 2003.

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

[17] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg. Signed networks in social media. In *CHI*, pages 1361–1370, 2010.

[18] C. Lin and Y. He. Joint sentiment/topic model for sentiment analysis. In *CIKM*, pages 375–384. ACM, 2009.

[19] H. Ma, H. Yang, M. R. Lyu, and I. King. Mining social networks using heat diffusion processes for marketing candidates selection. In *CIKM*, pages 233–242, 2008.

[20] S. A. Munson and P. Resnick. Presenting diverse political opinions: how and how much. In *CHI*, pages 1457–1466. ACM, 2010.

[21] A. Pal and S. Counts. Identifying topical authorities in microblogs. In *WSDM*, pages 45–54, 2011.

[22] R. K. Pon, A. F. Cardenas, D. Buttler, and T. Critchlow. Tracking multiple topics for finding interesting articles. In *KDD*, pages 560–569. ACM, 2007.

[23] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *KDD*, pages 61–70. ACM Press, 2002.

[24] D. J. Watts and P. S. Dodds. Influentials, networks, and public opinion formation. *Journal of Consumer Research*, 34:441–458, 2007.

[25] X. Yu, Y. Liu, X. Huang, and A. An. A quality-aware model for sales prediction using reviews. In *WWW*, pages 1217–1218. ACM, 2010.