

A Survey of Signed Network Mining in Social Media

JILIANG TANG, Michigan State University

YI CHANG, Yahoo Research

CHARU AGGARWAL, IBM T.J. Watson Research Center

HUAN LIU, Arizona State University

Many real-world relations can be represented by signed networks with positive and negative links, as a result of which signed network analysis has attracted increasing attention from multiple disciplines. With the increasing prevalence of social media networks, signed network analysis has evolved from developing and measuring theories to mining tasks. In this article, we present a review of mining signed networks in the context of social media and discuss some promising research directions and new frontiers. We begin by giving basic concepts and unique properties and principles of signed networks. Then we classify and review tasks of signed network mining with representative algorithms. We also delineate some tasks that have not been extensively studied with formal definitions and also propose research directions to expand the field of signed network mining.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing**; **Social networks**;

Additional Key Words and Phrases: Negative links, signed networks, signed network mining, social media

ACM Reference Format:

Jiliang Tang, Yi Chang, Charu Aggarwal, and Huan Liu. 2016. A survey of signed network mining in social media. *ACM Comput. Surv.* 49, 3, Article 42 (August 2016), 37 pages.

DOI: <http://dx.doi.org/10.1145/2956185>

1. INTRODUCTION

In many real-world social systems, relations between two nodes can be represented as signed networks with positive and negative links. In the 1940s, Heider [1946] studied perception and attitude of individuals and introduced structural balance theory, which is an important social theory for signed networks. In the 1950s, Cartwright and Harary [1956] further developed the theory and introduced the notion of balanced signed graph to characterize forbidden patterns in social networks. With roots in social psychology, signed network analysis has attracted much attention from multiple disciplines such as physics and computer science and has evolved considerably from both data- and problem-centric perspectives.

This material is based upon work supported by, or in part by, the U.S. Army Research Office (ARO) under contract/grant number 025071 and W911NF-15-1-0328, the Office of Naval Research (ONR) under grant number N000141410095, and the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053.

Authors' addresses: J. Tang, Computer Science and Engineering, Michigan State University, East Lansing, MI 48824; email: Jiliang.Tang@cse.msu.edu; H. Liu, Computer Science and Engineering, Arizona State University, Tempe, AZ, 85281; email: Huan.Liu@asu.edu; C. Aggarwal, IBM T.J. Watson Research Center, 1101 Kitchawan Rd, Yorktown, NY 10598; email: charu@us.ibm.com; Y. Chang, Yahoo! Inc, Sunnyvale, CA 94089; email: yichang@yahoo-inc.com.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2016 ACM 0360-0300/2016/08-ART42 \$15.00

DOI: <http://dx.doi.org/10.1145/2956185>

The early work in the field was mainly based on signed networks derived from observations in the physical world such as the international relationships in Europe from 1872 to 1907 [Heider 1946], relationships among Allied and Axis powers during World War II [Axelrod and Bennett 1993], and the conflict over Bangladesh's separation from Pakistan in 1971 [Moore 1978, 1979]. These signed networks were typically characterized by a small number of nodes with dense relationships and were notable for their clean structure. With the development of social media, increasing attention has been focused on signed social networks observed in online worlds. Signed networks in social media represent relations among online users where positive links indicate friendships, trust, and like, whereas negative links indicate foes, distrust, dislike, and antagonism. Examples of signed networks in social media include trust/distrust in Epinions¹ [Massa and Avesani 2005; Leskovec et al. 2010a] and friends/foes in Slashdot² [Kunegis et al. 2009]. Signed networks in social media often have hundreds of thousands of users and millions of links, and they are usually very sparse and noisy. Data for signed network analysis has evolved from offline to social media networks.

Research problems have evolved together with the evolution of the nature of available datasets for signed network analysis. Signed networks observed in the physical world are often small but dense and clean. Therefore, early research about signed networks had mainly focused on developing theories to explain social phenomena in signed networks [Heider 1946; Cartwright and Harary 1956]. Later, studies were conducted on measurements [Harary 1959; Harary and Kimmel 1979; Harary and Kabell 1980; Frank and Harary 1980] and dynamics of social balance [Antal et al. 2005; Radicchi et al. 2007a, 2007b; Marvel et al. 2011]. The recent availability of large-scale, sparse, and noisy social media networks has encouraged increasing attention on leveraging data-mining, machine-learning, and optimization techniques [Kunegis et al. 2009; Leskovec et al. 2010a; Yang et al. 2012; Chiang et al. 2013; Tang et al. 2014a]. Research problems for signed network analysis have evolved from developing and measuring theories to mining tasks.

This survey mainly focuses on mining tasks for signed networks in social media. However, it should be pointed out that (a) we will review theories originating from signed networks in the physical world for mining signed networks, and (b) we will review measurements and dynamics of social balance as basis or objectives in mining signed networks. Note that since nodes represent users in social networks, we will use the terms “node” and “user” interchangeably in this article.

1.1. Mining Signed Networks in Social Media

The problem of mining unsigned networks in social media (or networks with only positive links) has been extensively studied for decades [Knoke and Yang 2008; Aggarwal 2011; Zafarani et al. 2014]. However, mining signed networks requires dedicated methods because we cannot simply use straightforward extensions of algorithms and theories in unsigned networks [Chiang et al. 2013]. First, the existence of negative links in signed networks challenges many concepts and algorithms for unsigned networks. For example, node ranking algorithms for unsigned networks such as PageRank [Page et al. 1999] and HITS [Kleinberg 1999] require all links to be positive. Similarly, spectral clustering algorithms for unsigned networks cannot, in general, be directly extended to signed networks [Kunegis et al. 2010], and the concept of a structural hole in unsigned networks is not applicable to signed networks [Zhang et al. 2016]. Second, some social theories such as balance theory [Heider 1946] and status theory [Leskovec et al. 2010b] are only applicable to signed networks, while social theories for

¹<http://www.epinions.com/>.

²<http://slashdot.org/>.

unsigned networks such as homophily may not be applicable to signed networks [Tang et al. 2014a]. In addition, the availability of negative links brings about unprecedented opportunities and potentials in mining signed networks. First, it is evident from recent research that negative links have significant added value over positive links in various analytical tasks. For example, a small number of negative links can significantly improve positive link prediction [Guha et al. 2004; Leskovec et al. 2010a], and they can also improve recommendation performance in social media [Victor et al. 2009; Ma et al. 2009]. Second, analogously to mining unsigned networks, we can have similar mining tasks for signed networks; however, negative links in signed networks make them applicable to a broader range of applications. For example, similar tasks for unsigned networks have new definitions for signed networks such as community detection and link prediction, while new tasks and applications emerged for only signed networks such as sign prediction and negative link prediction.

In this article, we present a comprehensive review of current research findings about mining signed networks and discuss some tasks that need further investigation. The major motivation of this article is twofold as follows:

- Negative links in signed networks present two unique types of properties: (1) distinct topological properties as opposed to positive links and (2) collective properties with positive links. These unique properties determine that the basic concepts, principles, and properties of signed networks substantially differ from those of unsigned networks. Therefore an overview of basic concepts, principles, and properties of signed social networks can facilitate a better understanding of the challenges, opportunities, and necessity of mining signed networks.
- The availability of large-scale signed networks from social media has encouraged a large body of literature in mining signed networks. On the one hand, a classification of typical tasks can promote a better understanding. On the other hand, the development of tasks of mining signed social networks is highly imbalanced—some tasks are extensively studied, whereas others have not been sufficiently investigated. For well-studied tasks, an overview will help users familiarize themselves with the state-of-the-art algorithms; for insufficiently studied tasks, it is necessary to give formal definitions with promising research directions that can enrich current research.

The organization and contributions of the article are summarized as follows:

- We give an overview of basic concepts, unique principles, and properties of signed networks in Section 2. We discuss approaches to represent signed networks, topological properties of the negative networks, and collective properties of positive and negative links with social theories.
- We classify the mining tasks of signed social networks into node-oriented, link-oriented, and application-oriented tasks. From Section 3 to Section 5, we review well-studied tasks in each category with representative algorithms.
- Mining signed networks is in the early stages of development. We discuss some tasks for each category that have not yet received sufficient attention in the literature. We discuss formal definitions and promising research directions.

The readers of this survey are expected to have some basic understanding of social network analysis such as adjacency matrices, reciprocity, and clustering coefficient; data-mining techniques such as clustering and classification; machine-learning techniques such as eigen-decomposition, mixture models, and matrix factorization; and optimization techniques such as gradient decent and expectation-maximization methods.

1.2. Related Surveys and Differences

A few surveys about signed networks analysis exist in the literature. One of the earliest surveys may be found in Taylor [1970]. This survey gives an overview of metrics to measure the degree of social balance for given signed networks. Very recently, Zheng et al. [2014] provided a comprehensive overview of social balance in signed networks. This survey gives an overview about recent metrics to measure the degree and the dynamics of social balance as well as the application of social balance in partitioning signed networks. With the increasing popularization of signed networks in social media, a large body of literature has emerged, which leverages machine-learning, data-mining, and optimization techniques. This survey provides a comprehensive overview of this emerging area, along with a discussion of applications and promising research directions.

Compared to signed networks, there are many more surveys about unsigned network analysis. These surveys cover various topics in unsigned network analysis, including community detection [Tang and Liu 2010], node classification [Bhagat et al. 2011], link prediction [Liben-Nowell and Kleinberg 2007], and network evolution [Aggarwal and Subbian 2014]. Surveys are also available about applications of unsigned networks such as data classification [Sen et al. 2008], recommendation [Tang et al. 2013], and information propagation [Chen et al. 2013a].

2. BASIS OF SIGNED NETWORKS

The basic concepts, principles, and properties of signed networks are related to but distinct from those of unsigned networks. In this section, we review the representations, distinct properties of negative links, and collective properties of positive and negative links with social theories.

2.1. Network Representation

A signed network \mathcal{G} consists of a set of N nodes $\mathcal{U} = \{u_1, u_2, \dots, u_N\}$, a set of positive links \mathcal{E}_p , and a set of negative links \mathcal{E}_n . There are two major ways to represent a signed network \mathcal{G} .

As suggested in Leskovec et al. [2010a], positive and negative links should be viewed as tightly related features in signed networks. One way is to represent both positive and negative links into one adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, where $\mathbf{A}_{ij} = 1$, $\mathbf{A}_{ij} = -1$ and $\mathbf{A}_{ij} = 0$ denote positive, negative, and missing links from u_i to u_j , respectively.

The independent analyses of the different networks in signed networks reveal distinct types of properties, and it is important to consider these distinct topological properties in modeling [Szell et al. 2010]. Therefore, we separate a signed network into a network with only positive links and a network with only negative links and then use two adjacency matrices to represent these two networks, respectively. In particular, it uses $\mathbf{A}^p \in \mathbb{R}^{N \times N}$ to represent positive links where $\mathbf{A}_{ij}^p = 1$ and $\mathbf{A}_{ij}^p = 0$ denote a positive link and a missing link from u_i to u_j . Similarly, $\mathbf{A}_{ij}^n \in \mathbb{R}^{N \times N}$ is used to represent negative links where $\mathbf{A}_{ij}^n = 1$ and $\mathbf{A}_{ij}^n = 0$ denote a negative link and a missing link from u_i to u_j .

It is easy to convert one representation into the other with the following rules: $\mathbf{A} = \mathbf{A}^p - \mathbf{A}^n$ and $\mathbf{A}^p = \frac{|\mathbf{A}| + \mathbf{A}}{2}$ and $\mathbf{A}^n = \frac{|\mathbf{A}| - \mathbf{A}}{2}$, where $|\mathbf{A}|$ is the component-wise absolute value of \mathbf{A} .

2.2. Properties of Negative Networks

There are some well-known properties of positive links such as power-law degree distributions, high clustering coefficient, high reciprocity, transitivity, and strong correlation with similarity. However, we cannot easily extend these properties of positive links to

negative links. In this subsection, we will review important properties of negative links in social media, which are analogous to those of positive links.

Power-Law Distributions. It is well known that the distributions of incoming or outgoing positive links for users usually follow power-law distributions—a few users with large degrees while most users have small degrees. In Tang et al. [2014a], incoming or outgoing negative links for each user are calculated, and there are two important findings as follows: (a) In a signed network, positive links are denser than negative links, and there are many users without any incoming and outgoing negative links; and (b) for users with negative links, the degree distributions also follow power-law distributions—a few users have a large number of negative links, while most users have few negative links.

Clustering Coefficient. Nodes in networks with positive links are often easy to cluster. This property is often reflected by their high clustering coefficients (CC). High values of CC are expected because of the inherently cohesive nature of positive links [Coleman 1988]. However, the values of clustering coefficients for negative links are significantly lower than those for positive links. This suggests that many useful properties such as triadic closure cannot be applied to negative links [Szell et al. 2010].

Reciprocity. Positive links show high reciprocity. Networks with positive links are strongly reciprocal, which indicates that pairs of nodes tend to form bi-directional connections, whereas networks with negative links show significantly lower reciprocity. Asymmetry in negative links is confirmed in the correlations between the in- and out-degrees of nodes. In- and out-degrees of positive links are almost balanced, while negative links show an obvious suppression of such reciprocity [Szell et al. 2010].

Transitivity. Positive links show strong transitivity, which can be explained as “friends’ friends are friends.” The authors of Tang et al. [2014a] examined the transitivity of negative links on two social media signed networks, Epinions and Slashdot, and found that negative links may be not transitive since they observed both “enemies’ enemies are friends” and “enemies’ enemies are enemies.”

Correlation with Similarity. Positive links have strong correlations with similarity, which can be explained by two important social theories, that is, homophily [McPherson et al. 2001] and social influence [Marsden and Friedkin 1993]. Homophily suggests that users are likely to connect to other similar users, while social influence indicates that users’ behaviors are likely to be influenced by their friends. Via analyzing two real-world signed social networks, Epinions and Slashdot, the authors in Tang et al. [2014a] found that users are likely to be more similar to users with negative links than those without any links, while users with positive links are likely to be more similar than those with negative links. These observations suggest that negative links in signed social networks may denote neither similarity nor dissimilarity.

In addition, a recent work conducted a comprehensive signed link analysis [Beigi et al. 2016] and found the following: (1) users with positive (negative) emotions are likely to establish positive (negative) links; (2) users are likely to like their friends’ friends and dislike their friends’ foes; and (3) users with higher optimism (pessimism) are more likely to create positive (negative) links.

2.3. Collective Properties of Positive and Negative Links

As shown in the previous subsection, distinct properties are observed for positive and negative links. When we consider positive and negative links together, they present collective properties, which can be explained by two important social theories in signed networks, that is, balance theory [Heider 1946] and status theory [Guha et al. 2004;

Leskovec et al. 2010b]. Next we present these collective properties by introducing these two social theories, which have been proven to be very helpful in mining signed social networks [Leskovec et al. 2010b; Yang et al. 2012; Zheng et al. 2014; Kunegis 2014]. For example, the signed clustering coefficient and the relative signed clustering coefficient [Kunegis et al. 2009] are defined based on the intuition “the enemy of my enemy is my friend” implied by balance theory. Note that balance theory is developed for undirected signed social networks, whereas status theory is developed for directed signed social networks

2.3.1. Balance Theory. Balance theory is originally introduced in Heider [1946] at the individual level and generalized by Cartwright and Harary [1956] in the graph-theoretical formation at the group level. When the signed network is not restricted to be complete, the network is balanced if all its cycles have an even number of negative links. Using this definition, it is proven in Harary et al. [1953] that “a signed graph is balanced if and only if nodes can be separated into two mutually exclusive subsets such that each positive link joins two nodes of the same subset and each negative link joins nodes from different subsets.” It is difficult to represent real-world signed networks by balanced structure. Therefore, Davis [1967] introduced the notion of a *clusterizable graph*—a signed graph is clusterizable if there exists a partition of the nodes such that nodes with positive links are in the same subset and nodes with negative links are between different subsets.

Later, researchers have proposed some important metrics to measure the degree of balance of signed networks. As mentioned above, the concept of balance has evolved and been generalized. Hence, these metrics can be categorized according to their adopted definitions of balance. Some metrics use the definition of balance by Cartwright and Harary [1956]; hence, they measure the number of balanced or unbalanced cycles. The ratio of balanced circles among all possible circles was calculate by using the adjacency matrix \mathbf{A} [Cartwright and Gleason 1966], which was modified to consider the length of cycles in Henley et al. [1969]. The time complexity of these metrics is $O(n^3)$, which is infeasible for large real-world signed networks. Terzi and Winkler proposed an efficient spectral algorithm to estimate the degree of balance for large signed networks [Terzi and Winkler 2011]. The definition of balance by Davis [1967] established the correlation between balance and clustering—clustering is partition of the nodes of a given signed network into k clusters, such that each pair of nodes in the same cluster has a positive link and a negative link exists between each pair of nodes from different clusters. Therefore, the metrics based on the definition by Davis [1967] measure the number of disagreements—the number of negative links inside clusters and the number of positive links between clusters [Bansal et al. 2004; Facchetti et al. 2011; Zheng et al. 2014]. Actually, these metrics later became criteria to partition signed networks into clusters (or communities) such that approximation algorithms were developed for minimizing disagreements by identifying the optimal number of clusters in Bansal et al. [2004]. More details about these clustering algorithms will be discussed in Section 3.2.1.

Balance theory generally implies that “the friend of my friend is my friend” and “the enemy of my enemy is my friend” [Heider 1946]. Let s_{ij} represent the sign of the link between the i th node and the j th node where $s_{ij} = 1$ and $s_{ij} = -1$ denote a positive link and a negative link are observed between u_i and u_j . Balance theory suggests that a triad $\langle u_i, u_j, u_k \rangle$ is balanced if $-(1) s_{ij} = 1$ and $s_{jk} = 1$, then $s_{ik} = 1$, or $(2) s_{ij} = -1$ and $s_{jk} = -1$, then $s_{ik} = 1$.

For a triad, four possible sign combinations exist as demonstrated in Figure 1. Among these four combinations, **A** and **C** are balanced. The way to measure the balance of signed networks in social media is to examine all these triads and then to compute the ratio of **A** and **C** over **A**, **B**, **C**, and **D**. Existing work reported that triads in signed

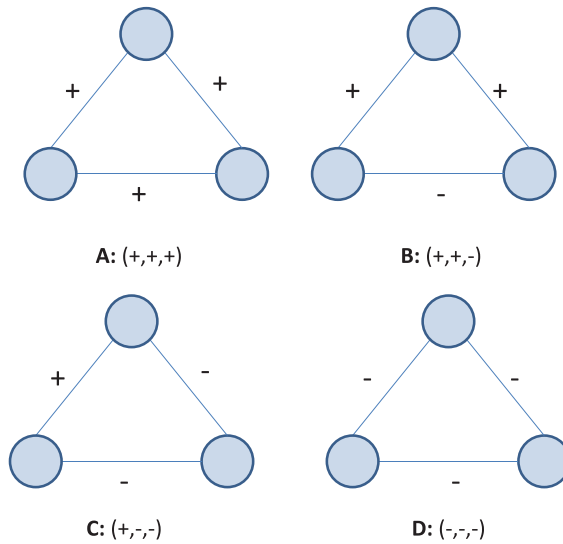


Fig. 1. Illustration of balance theory. As suggested by balance theory, triads **A** and **C** are balanced; while triads **B** and **D** are imbalanced.

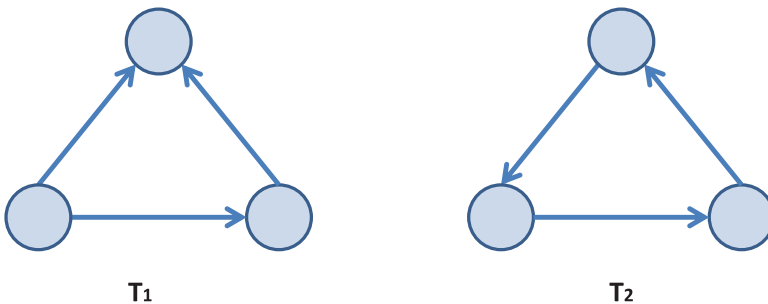


Fig. 2. Possible triads in directed social networks.

networks in social media are highly balanced. For example, Leskovec et al. [2010a] found that the ratios of balanced triads of signed networks in Epinions, Slashdot, and Wikipedia are 0.941, 0.912, and 0.909, respectively, and more than 90% of triads are balanced in other social media datasets [Yang et al. 2012]. Furthermore, the ratio of balanced triads increases while that of unbalanced triads decreases over time [Szell et al. 2010].

2.3.2. Status Theory. While balance theory is naturally defined for undirected networks, status theory [Guha et al. 2004; Leskovec et al. 2010b] is relevant for directed networks. Social status can be represented in a variety of ways, such as the rankings of nodes in social networks, and it represents the prestige of nodes. In its most basic form, status theory suggests that u_i has a higher status than u_j if there is a positive link from u_j to u_i or a negative link from u_i to u_j .

As shown in Figure 2, there are two types of triads in directed networks, which correspond to acyclic and cyclic triads. Note that flipping the directions of all the links has no impact on the type of the cyclic triad. Since there are four possible sign combinations, there are 4 types of cyclic signed triads for T_2 as shown in Figure 3. Each link in an acyclic triad can be positive or negative and the signs of links in an acyclic

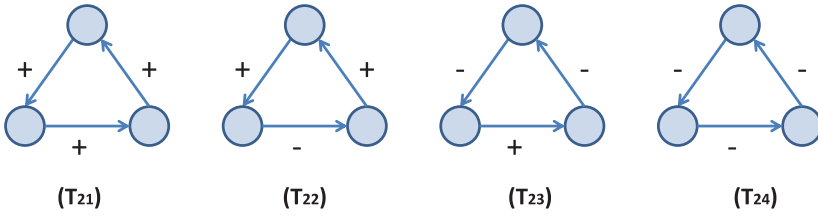


Fig. 3. An illustration of four types of cyclic signed triads.

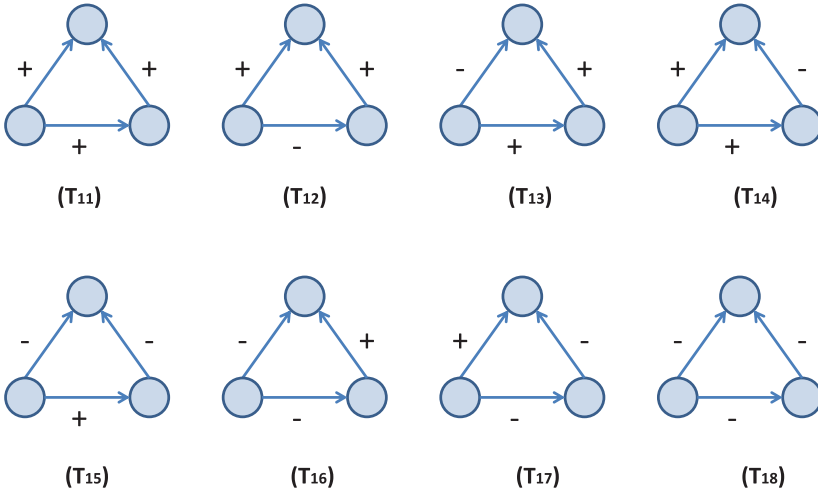


Fig. 4. Illustration of eight types of acyclic signed triads.

Table I. Balance Theory vs. Status Theory

	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}	T_{16}	T_{17}	T_{18}	T_{21}	T_{22}	T_{23}	T_{24}
Status Theory	✓	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✗
Balance Theory	✓	✗	✗	✗	✓	✓	✓	✗	✓	✗	✓	✗

triad are not exchangeable; hence, there are 8 types of acyclic signed triads as depicted in Figure 4. Overall, there are 12 types of triads in directed signed networks.

A popular approach to examine whether a given triad satisfies status theory or not is as follows. We reverse the directions of all negative links and flip their signs to positive. If the resulting triad is acyclic, then the triad satisfies status theory. It is easy to verify that (1) for a negative link $u_i \xrightarrow{-} u_j$, reversing its direction and flipping its sign simultaneously lead to a positive link $u_j \xrightarrow{+} u_i$, which preserves the status order of u_i and u_j according to status theory, and (2) for a positive and cyclic triad $u_i \xrightarrow{+} u_j \xrightarrow{+} u_k \xrightarrow{+} u_i$, their statuses should satisfy $u_i > u_j > u_k > u_i$ according to status theory, which leads to a logical contradiction $u_i > u_i$. Following the aforementioned approach, we find that 8 of the 12 types of triads in signed networks satisfy status theory as shown in the first row of Table I. Similarly to the approach for testing the balance of signed networks, we examine all 12 triads and then calculate the ratio of triads satisfying status theory. Examinations on signed networks in typical social media suggest that more than 90% of triads satisfy status theory [Leskovec et al. 2010b].

Table II. Statistics of Representative Signed Networks in Social Media

	Epinions	Slashdot	eEpinions	eSlashdot
# of Users	119,217	82,144	23,280	14,799
# of Links	841,200	549,202	332,214	232,471
Positive Link Percentage	85.0%	77.4%	87.7%	81.5%
Negative Link Percentage	15.0%	22.6%	12.3%	18.5%

As shown in Table I, status theory and balance theory do not always agree with one another. Note that we apply balance theory to directed signed networks by ignoring the directions of links. Some triads satisfy both theories such as the triad T_{11} . Some satisfy status theory but not balance theory such as the triad T_{12} . Some satisfy balance theory but not status theory such as the triad T_{21} . Others do not satisfy either such as the triad T_{24} .

2.4. Popular Data Sets for Benchmarking

In this subsection, we discuss some social media datasets widely used for benchmarking analytical algorithms in the signed network setting.

Epinions is a product review site. Users can create both positive (trust) and negative (distrust) links to other users. They can write reviews for various products with rating scores from 1 to 5. Other users can rate the helpfulness of reviews. There are several variants of datasets from Epinions publicly available [Massa and Avesani 2005; Leskovec et al. 2010a; Yang et al. 2012; Tang et al. 2015]. Statistics of two representative sets are illustrated in Table II. “Epinions” is from the Stanford large network dataset collection³ where only signed networks among users are available. In addition to signed networks, “eEpinion” [Tang et al. 2015] also provides item ratings, review content, helpfulness ratings, and categories of items. It also includes timestamps when links are established and ratings are created.

Slashdot is a technology news platform in which users can create friend (positive) and foe (negative) links to other users. They can also post news articles. Other users may annotate these articles with their comments. There also various variants of datasets from Slashdot [Kunegis et al. 2009; Leskovec et al. 2010a; Tang et al. 2015] and two of them are demonstrated in Table II. “Slashdot” is from the Stanford large network dataset collection with only signed networks among users, while the more detailed “eSlashdot” [Tang et al. 2015] provides signed networks, comments on articles, user tags, and groups in which users participate.

2.5. Tasks of Mining Signed Networks

There are similar tasks for mining unsigned and signed networks. However, the availability of negative links in signed networks determines that similar mining tasks for unsigned networks may have new definitions for signed networks, and there may be new tasks specific to signed networks. We category the tasks of mining signed networks as tasks that focus on nodes, links, and applications, that is, node-oriented, link-oriented, and application-oriented tasks as shown in Figure 5. Although a large body of work has emerged in recent years for mining signed social networks, the development of tasks in each category is highly imbalanced. Some of them are well studied, whereas others need further investigation. These tasks are highlighted in red in Figure 5. In the following sections, we give an overview of representative algorithms for well-studied tasks and also provide a detailed discussion of important and emerging tasks. Where needed, promising research directions are also highlighted. The notations used in this article are summarized in Table III. Any algorithms for directed signed

³<https://snap.stanford.edu/data/>.

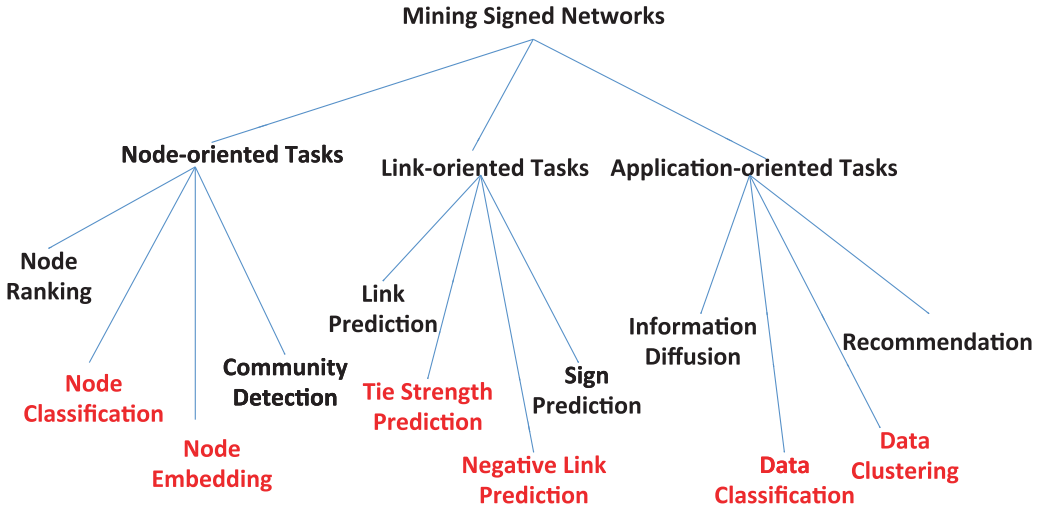


Fig. 5. An overview of tasks of mining signed networks in social media. Tasks highlighted in red have not been extensively studied.

Table III. Notation and Definitions

Notations	Descriptions
N	Number of Users
\mathbf{A}	Adjacency matrix of a signed network
\mathbf{A}^p	Adjacency matrix of a positive network
\mathbf{A}^n	Adjacency matrix of a negative network
\mathbf{D}^p	A diagonal matrix with $\mathbf{D}_{ii}^p = \sum_j \mathbf{A}_{ij}^p$
\mathbf{D}^n	A diagonal matrix with $\mathbf{D}_{ii}^n = \sum_j \mathbf{A}_{ij}^n$
I_i^+	The set of nodes that create positive links to u_i
$ I_i^+ $	In-degree of positive links of u_i
I_i^-	The set of nodes that create negative links to u_i
$ I_i^- $	In-degree of negative links of u_i
I_i	$I_i = I_i^+ \cup I_i^-$
$ I_i $	$ I_i = I_i^+ + I_i^- $
O_i^+	The set of users that u_i creates positive links to
$ O_i^+ $	Out-degree of positive links of u_i
O_i^-	The set of users that u_i creates negative links to
$ O_i^- $	Out-degree of negative links of u_i
O_i	$O_i = O_i^+ \cup O_i^-$
$ O_i $	$ O_i = O_i^+ + O_i^- $
d_i^+	$d_i^+ = I_i^+ + O_i^+ $
d_i^-	$d_i^- = I_i^- + O_i^- $
\mathcal{L}^p	Laplacian matrix for a positive network
\mathcal{L}^n	Laplacian matrix for a negative network
\mathcal{L}	Laplacian matrix for a signed network
c_i	Community of u_i
s_{ij}	Sign of the link from u_i to u_j
m	Number of links in a signed social network
m^+	Number of positive links in a signed social network
m^-	Number of negative links in a signed social network
\mathbf{X}_{ij}	the (i, j) entry of the matrix \mathbf{X}

networks are applicable to undirected signed networks by considering an undirected link as two bidirectional links. Hence, in the following sections, it can be assumed by default that an algorithm can be applied to both directed and undirected signed networks.

3. NODE-ORIENTED TASKS

As shown in Figure 5, important node-oriented tasks include node ranking, community detection, node classification, and node embedding, among which node ranking and community detection are extensively studied. On the other hand, node classification and node embedding need further investigation. In this section, we review node ranking and community detection with representative algorithms.

3.1. Node Ranking

The problem of node ranking for signed networks is that of exploiting the link structure of a network to order or prioritize the set of nodes within the network by considering both positive and negative links [Getoor and Diehl 2005]. Since negative links are usually not considered, most node ranking algorithms for unsigned networks cannot deal with negative values directly [Haveliwala 2002; Cohn and Chang 2000]. A straightforward solution is to apply node ranking algorithms of unsigned networks, such as EigenTrust [Kamvar et al. 2003], by ignoring negative links or zero the entries corresponding to the negative links in the matrix representation of the network [Richardson et al. 2003]. In other words, we only consider the positive network \mathbf{A}^p while ignoring the impact from \mathbf{A}^n in a signed network. This solution cannot distinguish between negative and missing links since both of them correspond to a zero entity in the representation matrix. Recent node ranking algorithms for signed networks fall into three themes as follows: (a) centrality measurements are used, (b) PageRank-like models are used [Page et al. 1999], and (c) HITS-like methods are used [Kleinberg 1999]. Next, we will introduce representative algorithms for each group.

3.1.1. Centrality-Based Algorithms. Centrality-based algorithms use certain centrality measurements to rank nodes in signed networks. If a node receives many positive incoming links, then it should have high prestige value, while nodes with many negative incoming links will have small values of prestige. A measure p_i of the status score of u_i based on the in-degree of positive and negative links is proposed in Zolfaghar and Aghaie [2010] as follows:

$$p_i = \frac{|I_i^+| - |I_i^-|}{|I_i^+| + |I_i^-|}, \quad (1)$$

where $|I_i^+|$, and $|I_i^-|$ are the in-degree of positive and negative links of u_i , respectively. A similar metric is used in Kunegis et al. [2009] as the subtraction of in-degree of negative links from in-degree of positive links, that is, $p_i = |I_i^+| - |I_i^-|$. An eigenvector centrality metric is proposed in Bonacich and Lloyd [2004] for balanced complete signed networks. We can divide nodes of a balanced complete signed network into two communities such that all positive links connect members of the same community and all negative links connect members of different communities. Thus, positive and negative scores in the eigenvector that correspond to the largest eigenvalue of the adjacency matrix \mathbf{A} reveal not only the clique structure but also status scores within each clique [Bonacich and Lloyd 2004].

3.1.2. PageRank-Based Algorithms. The original PageRank algorithm expresses the reputation score for the i th node as:

$$p_i = \sum_{u_j \in I_i^+} \frac{p_j}{|O_j^+|}, \quad (2)$$

where $|O_j^+|$ is the out-degree of positive links of u_j . The probability p_i can be computed in an iterative way:

$$p_i^{t+1} = \alpha \sum_{u_j \in I_i^+} \frac{p_j^t}{|O_j^+|} + (1 - \alpha) \frac{1}{N}, \quad (3)$$

where the term $(1 - \alpha) \frac{1}{N}$ is the restart component, N the total number of users, and α is a damping factor. In signed networks, mechanisms are also provided to handle negative links [Traag et al. 2010; Borgs et al. 2010; Chung et al. 2013]. Next, we detail three representative algorithms in this group [Shahriari and Jalili 2014; De Kerchove and Van Dooren 2008; Traag et al. 2010].

In Shahriari and Jalili [2014], two status scores are calculated by the original PageRank algorithm for the positive network and the negative network separately, and the difference of the two provides the final result. Therefore, this algorithm considers a signed network as two separate networks and completely ignores the interactions between positive and negative links. Furthermore, this approach does not have natural interpretations in terms of the reputation scores of nodes. In Wu et al. [2016], the Troll-Trust model is proposed that has a clear physical interpretation. An exponential node ranking algorithm based on discrete choice theory is proposed in Traag et al. [2010]. When the observed reputation is $k_i = \sum_{u_j \in I_i} \mathbf{A}_{ji} p_j$, the probability of u_i with the highest real reputation according to discrete choice theory is

$$p_i = \frac{\exp(k_i/\mu)}{\sum_j \exp(k_j/\mu)}. \quad (4)$$

An iterative approach is used to compute the status scores as follows:

$$\mathbf{p}^{t+1} = \frac{\exp(\frac{1}{\mu} \mathbf{A}^\top \mathbf{p}^t)}{\|\exp(\frac{1}{\mu} \mathbf{A}^\top \mathbf{p}^t)\|_1}. \quad (5)$$

Within a certain range of μ , the aforementioned formulation can achieve a global solution \mathbf{p}^* with arbitrary initializations.

The work in De Kerchove and Van Dooren [2008] and de Kerchove et al. [2009] uses the intuition that the random-walk process should be modified to avoid negative links. Therefore nodes receiving negative connections are visited less. This is formalized as follows:

$$\mathbf{p}_i^{t+1} = (1 - \hat{\mathbf{Q}}_{ii}^t) \left(\alpha \sum_{u_j \in I_i^+} \frac{\mathbf{p}_j^t}{|O_j^+|} + (1 - \alpha) \frac{1}{N} \right), \quad (6)$$

where $\hat{\mathbf{Q}}_{ii}^t$ gives the ratio of walkers that distrust the node they are in. In that manner, $(1 - \hat{\mathbf{Q}}_{ii}^t)$ represents the ratio of remaining walkers in u_i . The distrust matrix $\hat{\mathbf{Q}}$ is calculated as follows:

—A random walk according to the original PageRank formulation is used:

$$\hat{\mathbf{Q}}^{t+1} = \mathbf{T}^t \mathbf{Q}^t, \quad (7)$$

where \mathbf{T}^t is the transition matrix whose (i, j) th entry \mathbf{T}_{ij}^t indicates the ratio of walkers in u_i who were in u_j at time t as follows:

$$\mathbf{T}_{ij}^t = \frac{\alpha \mathbf{A}_{ij}^p \mathbf{p}_j^t / |O_j^+| + (1 - \alpha) \frac{1}{N}}{\alpha \sum_{u_k \in I_i^+} (\mathbf{p}_k^t / |O_k^+| + (1 - \alpha) \frac{1}{N})}. \quad (8)$$

—A walk in u_i automatically adopts negative opinions of u_i . In other words, he adds the nodes negatively pointed by u_i into his distrust list ($\mathbf{Q}_{ij}^{t+1} = -1$). A walker who distrusts a node leaves the graph if ever he visits the node ($\mathbf{Q}_{ij}^{t+1} = 0$). With the intuition, \mathbf{Q}_{ij}^{t+1} is updated as follows:

$$\mathbf{Q}_{ij}^{t+1} = \begin{cases} 1 & \text{if } \mathbf{A}_{ij} = -1, \\ 0 & \text{if } i = j, \\ \hat{\mathbf{Q}}_{ij}^{t+1} & \text{otherwise.} \end{cases} \quad (9)$$

3.1.3. HITS-Based Algorithms. The original HITS algorithm [Kleinberg 1999] calculates a hub score h_i and an authority score a_i for each node u_i as

$$h_i = \sum_{j \in I_i^+} a_j; \quad a_i = \sum_{j \in O_i^+} h_j. \quad (10)$$

HITS-based algorithms provide components to handle negative links based on the original HITS algorithm. In Shahriari and Jalili [2014], two strategies are proposed. The first applies the original HITS algorithm separately on the positive and negative networks as follows:

$$\begin{cases} h_i^+ = \sum_{j \in I_i^+} a_j^+; & a_i^+ = \sum_{j \in O_i^+} h_j^+ \\ h_i^- = \sum_{j \in I_i^-} a_j^-; & a_i^- = \sum_{j \in O_i^-} h_j^- \end{cases}. \quad (11)$$

Then, the final hub and authority scores are computed as follows:

$$a_i = a_i^+ - a_i^-; \quad h_i = h_i^+ - h_i^-. \quad (12)$$

The other way is to incorporate the signs directly as follows:

$$\begin{cases} h_i = \frac{\sum_{j \in I^+} a_j - \sum_{j \in I^-} a_j}{\sum_{j \in I^+} a_j + \sum_{j \in I^-} a_j} \\ a_i = \frac{\sum_{j \in O_i^+} h_j - \sum_{j \in O_i^-} h_j}{\sum_{j \in O_i^+} h_j + \sum_{j \in O_i^-} h_j} \end{cases}. \quad (13)$$

Instead of hub and authority scores in HITS, the concepts of *bias* and *deserve* are introduced in Mishra and Bhattacharya [2011]. Here, bias (or trustworthiness) of a link reflects the expected weight of an outgoing connection, and deserve (or prestige) of a link reflects the expected weight of an incoming connection from an unbiased node. Similarly to HITS, the deserve score DES_i for u_i is the aggregation of all unbiased votes from her incoming connections as:

$$DES_i^{t+1} = \frac{1}{|I_i|} \sum_j \mathbf{A}_{ji} (1 - X_{ji}^t), \quad (14)$$

where X_{ji} indicates the influence that bias of u_j has on its outgoing link to u_i ,

$$X_{ji} = \max\{0, BIAS_j * \mathbf{A}_{ji}\}, \quad (15)$$

while the bias score $BIAS_i$ for u_i is the aggregation of voting biases of her outgoing connections as:

$$BIAS_i^{t+1} = \frac{1}{2 * |O_i|} \sum_{u_j \in O_i} (A_{ji} - DES_j^t). \quad (16)$$

3.2. Community Detection in Signed Networks

The existence of negative links in signed networks makes the definition of community detection in signed networks differ substantially from that in unsigned networks. In unsigned networks, community detection identifies groups of densely connected nodes [Tang and Liu 2010; Papadopoulos et al. 2012; Ailon et al. 2013]. In signed networks, groups of users are identified, where users are densely connected by positive links within the group and negative links between groups. Based on the underlying methodology, clustering-based, modularity-based, mixture-model-based, and dynamic-model-based methods are used. Next we will give basic concepts for each group with representative algorithms.

3.2.1. Clustering-based Algorithms. Clustering-based algorithms transform a graph vertex clustering problem to one that can be addressed by traditional data clustering methods. If we consider a positive link or a negative link indicates whether two nodes are similar or different, then community detection in signed networks is boiled down to the correlation clustering problem [Bansal et al. 2004]. Bansal et al. proved non-deterministic polynomial-time hardness of the correlation clustering problem and gave constant-factor approximation algorithms for the special case in which the network is complete and undirected, and every edge has weight $+1$ or -1 [Bansal et al. 2004]. A two-phase clustering re-clustering algorithm is introduced in Sharma et al. [2009]: (1) the first phase is based on the Breadth First Search algorithm that forms clusters on the basis of the positive links only, and (2) the second phase is to reclassify the nodes with negative links on the basis of the participation level of the nodes having the negative links. In addition, there are two groups of clustering algorithms for community detection. One is based on k -balanced social theory and the other is based on spectral clustering. Note that algorithms based on spectral clustering are designed for undirected signed networks.

Algorithms based on k -balanced social theory aim to find k clusters with minimal positive links between clusters and minimal negative links inside clusters. In Doreian and Mrvar [1996], the objective function of clustering algorithms is defined as $E = \alpha N_n + (1 - \alpha)N_p$, where N_n is the number of negative links within clusters and N_p the number of positive links between clusters. The proposed algorithm in Doreian and Mrvar [1996] and Hassan et al. [2012a] first assigns the nodes to k clusters randomly and then optimizes the above objective function through reallocating the nodes. An alternative approach is to leverage simulated annealing to optimize the objective function E [Traag and Bruggeman 2009; Bogdanov et al. 2010].

One spectral clustering technique is introduced in Kunegis et al. [2010]. For a signed network \mathbf{A} , it first defines the signed Laplacian matrix [Hou 2005] as follows:

$$\mathcal{L} = \mathbf{D} - \mathbf{A}, \quad \mathbf{D}_{ii} = \sum_j |\mathbf{A}_{ij}|. \quad (17)$$

Similarly to the Laplacian matrix for unsigned networks, it can be proven that the signed Laplacian matrix \mathcal{L} is often positive semidefinite but it is positive definite if and only if the network is unbalanced. Spectral clustering algorithms on the signed Laplacian matrix can detect clusters of nodes within which there are only positive links. The Laplacian matrix in Equation (17) tends to separate pairs with negative

links rather than to force pairs with positive links closer. Hence, a balanced normalized signed Laplacian matrix is proposed in Zheng and Skillicorn [2015] as:

$$\mathcal{L} = (\mathbf{D}^p - \mathbf{A}^p + \mathbf{A}^n). \quad (18)$$

Another spectral clustering technique is balanced normalized cut [Chiang et al. 2012]. The objective of a positive ratio cut is to minimize the number of positive links between communities:

$$\min \sum_{c=1}^k \frac{x_c^\top \mathcal{L}^p x_c}{x_c^\top x_c}, \quad (19)$$

where $\{x_c\}_{c=1}^k$ are the community indicator vectors and \mathcal{L}^p is the Laplacian matrix of positive links. The objective of negative ratio association is to minimize the number of negative links in each cluster as:

$$\min \left(\sum_{c=1}^k \frac{x_c^\top \mathbf{A}^n x_c}{x_c^\top x_c} \right). \quad (20)$$

The balance normalized cut is to minimize the positive ratio cut and negative ratio association simultaneously as:

$$\min \left(\sum_{c=1}^k \frac{x_c^\top (\mathbf{D}^p - \mathbf{A}) x_c}{x_c^\top x_c} \right), \quad (21)$$

where the matrix of $\mathbf{D}^p - \mathbf{A}$ in balanced normalized cut is identical to the balanced normalized signed Laplacian matrix in Equation (18).

We can obtain $\{x_1, x_2, \dots, x_k\}$ by solving the optimization problems in Equations (19), (20), or (21). To generate k clusters, we can round $\{x_1, x_2, \dots, x_k\}$ to a valid indicator set [Chiang et al. 2012]; we consider $(x_1(i), x_2(i), \dots, x_k(i))$ as a k -dimensional vector representation of user i ($i \in \{1, 2, \dots, n\}$) and then perform k -means on these n vectors.

3.2.2. Modularity-Based Algorithms. These algorithms are to detect communities by optimizing modularity or its variants for signed networks [Li et al. 2014a]. The original modularity [Newman and Girvan 2004] is developed for unsigned networks and it measures how far the real positive connections deviates from the expected random connections, which is formally defined as follows:

$$Q^+ = \frac{1}{2m^+} \sum_{ij} \left(\mathbf{A}_{ij}^p - \frac{d_i^+ d_j^+}{2m^+} \right) \delta(i, j), \quad (22)$$

where $\delta(c_i, c_j)$ is the Kronecker δ function, which is 1 if u_i and u_j are in the same community and 0 otherwise. In Gómez et al. [2009], modularity of networks with only negative links Q^- is defined in a similar as Q^+ :

$$Q^- = \frac{1}{2m^-} \sum_{ij} \left(\mathbf{A}_{ij}^n - \frac{d_i^- d_j^-}{2m^-} \right) \delta(i, j). \quad (23)$$

Modularity for signed network Q should balance the tendency of users with positive links to form communities and that of users with negative links to destroy them and the mathematical expression of Q is

$$Q = \frac{2m^+}{2m^+ + 2m^-} Q^+ - \frac{2m^-}{2m^+ + 2m^-} Q^-. \quad (24)$$

Equation (24) can be rewritten as:

$$Q = \frac{1}{2m} \sum_{ij} \left(\mathbf{A}_{ij} + \frac{d_i^- d_j^-}{2m^-} - \frac{d_i^+ d_j^+}{2m^+} \right) \delta(i, j). \quad (25)$$

The definition of Q in Equation (24) has three properties as follows [Li et al. 2014a]: (1) Q boils down to Q^+ if no negative link exists, (2) $Q = 0$ if all nodes are assigned to the same community, and (3) Q is anti-symmetric in weighted signed networks. Based on Q in Equation (24), several variants of modularity are developed such as modularity density [Li et al. 2014a] and frustration [Anchuri and Magdon-Ismail 2012]. Community structure can be obtained by either minimizing frustration [Anchuri and Magdon-Ismail 2012] or maximizing modularity, both of which have been proven to be a general eigenvector problem [Anchuri and Magdon-Ismail 2012]. In Amelio and Pizzuti [2013], a community detection framework is proposed by using non-dominated sorting genetic [Srinivas and Deb 1994; Pizzuti 2009] to minimize frustration and maximize signed modularity simultaneously.

3.2.3. Mixture-Model-Based Algorithms. Mixture-model-based algorithms generate the division of the network into communities based on generative graphical models [Chen et al. 2013]. In general, there are two advantages of mixture-model-based algorithms. First, they provide soft-partition solutions in signed networks. Second, they provide soft-memberships that indicate the strength of a node belonging to a community. These two advantages determine that they can identify overlapping communities. Stochastic block-based models and probabilistic mixture-based models are two types of mixture models widely adopted for community detection in signed networks. Stochastic block-based models generate a network from a node perspective where each node is assigned to a block or community and links are independently generated for pairs of nodes. In Jiang [2015], a generalized stochastic model, that is, signed stochastic block model (SSBM), is proposed to identify communities for signed networks where nodes within a community are more similar in terms of positive and negative connection patterns than those from other communities. SSBM represents the memberships of each node as hidden variables and uses two matrices to explicitly characterize positive and negative links among groups, respectively. While probabilistic mixture-based models generate a network from the link perspective [Shen 2013]. In Chen et al. [2013], a signed probabilistic mixture (SPM) model is proposed to detect overlapping communities in undirected signed networks. A link from u_i to u_j is generated by SPM as follows:

—If the link from u_i to u_j is positive, that is, $\mathbf{A}_{ij} > 0$, then:

- (1) Choose a community c for the link with probability W_{cc}
- (2) Select u_i from c with probability θ_{ci}
- (3) Select u_j from c with probability θ_{cj}

—If the link from u_i to u_j is negative, that is, $\mathbf{A}_{ij} < 0$, then:

- (1) Choose two different communities c and s for the link with probability $W_{cs(c \neq s)}$
- (2) Select u_i from c with probability θ_{ci}
- (3) Select u_j from s with probability θ_{sj}

Overall, the probability of the link from u_i to u_j can be rewritten as:

$$P(\mathbf{A}_{ij}|W, \theta) = \left(\sum_{cc} W_{cc} \theta_{ci} \theta_{cj} \right)^{\mathbf{A}_{ij}^p} \left(\sum_{cs(c \neq s)} W_{cs} \theta_{ci} \theta_{sj} \right)^{\mathbf{A}_{ij}^n}. \quad (26)$$

3.2.4. Dynamic-Model-Based Algorithms. Dynamic-model-based algorithms consider a dynamic process taking place on the network, which reveals its communities. One type of

algorithm in this group is based on discrete-time and continuous-time dynamic models of social balance, and a review of these algorithms can be found in Zheng et al. [2014]. A framework based on agent-based random walk model is proposed in Yang et al. [2007] to extract communities for signed networks. Generally, links are much denser within a community than between communities. The intuition behind this framework is that an agent, starting from any node, should have higher chances to stay in the same community than to go to a different community after a number of walks. The framework has two advantages: (1) it is very efficient with linear time complexity in terms of the number of nodes, and (2) it considers both the density of links and signs, which provides a unified framework for community detection for unsigned and signed networks. Some additional steps are added by Kong and Yang [2011] to further advance the framework such as introducing a method to detect random walk steps automatically.

3.3. Promising Directions for Node-oriented Tasks

In this subsection, we discuss two node-oriented tasks including node classification and node embedding, which need further investigations to help us gain a better understanding of nodes in signed networks.

3.3.1. Node Classification in Signed Networks. User information such as demographic values, interest beliefs, or other characteristics plays an important role in helping social media sites provide better services for their users such as recommendations and content filtering. However, most social media users do not share too much of their information [Zheleva and Getoor 2009]. For example, more than 90% of users in Facebook do not reveal their political views [Abbasi et al. 2014]. One way of bridging this knowledge gap is to infer missing user information by leveraging the pervasively available network structures in social media. An example of such inference is that of node classification in social networks. The node classification problem has been extensively studied in the literature [Getoor and Diehl 2005]. The vast majority of these algorithms have focused on unsigned social networks (or social networks with only positive links) [Sen et al. 2008]. Evidence from recent achievements in signed networks suggests that negative links may be also potentially helpful in the task of node classification.

Let $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$ be the set of m class labels. Assume that $\mathcal{U}^L = \{u_1, u_2, \dots, u_n\}$ is the set of n labeled users, where $n < N$ and $\mathcal{U}^U = \mathcal{U} \setminus \mathcal{U}^L$ is the set of $N - n$ unlabeled users. We use $\mathbf{Y} \in \mathbb{R}^{n \times m}$ to denote the label indicator matrix for \mathcal{U}^L , where $\mathbf{Y}_{ik} = 1$ if u_i is labeled as c_k , and $\mathbf{Y}_{ik} = 0$ otherwise. With the above notations and definitions, the problem of user classification in a signed social network can be formally stated as follows: *Given a signed social network \mathcal{G} with \mathbf{A}^p and \mathbf{A}^n , and labels \mathbf{Y} for some users \mathcal{U}^L , user classification in a signed social network aims to infer labels for \mathcal{U}^U by leveraging \mathbf{A}^p , \mathbf{A}^n , and \mathbf{Y} .*

There are two possible research directions for node classification in signed networks. Since node classification has been extensively studied for unsigned networks, one way is to transform algorithms from unsigned to signed networks. Negative links present distinct properties from positive links [Szell et al. 2010]. As suggested in Leskovec et al. [2010a], positive and negative links should also be viewed as tightly related features in signed social networks. Meanwhile links could have different semantics in different social media sites. Therefore, an alternative approach is to develop novel models based on the understandings about signed networks. Very recently, a framework is proposed to capture both single- and multi-view information from signed networks for node classification that significantly improves the classification performance [Tang et al. 2016a].

3.3.2. Node Embedding. Node embedding (or network embedding), which aims to learn low-dimensional vector representations for nodes, has been proven to be useful in many

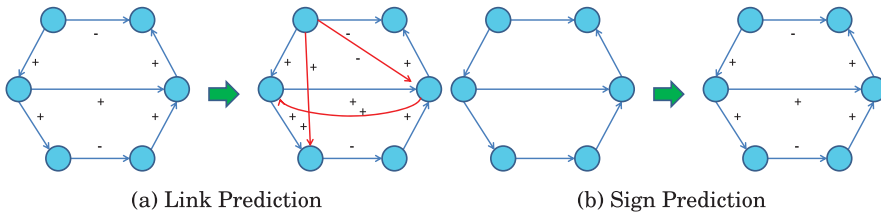


Fig. 6. Illustration of the differences of link prediction and sign prediction.

tasks of social network analysis such as link prediction [Liben-Nowell and Kleinberg 2007], community detection [Papadopoulos et al. 2012], and node classification [Bhagat et al. 2011]. The vast majority of existing algorithms have been designed for social networks with only positive links while the work on signed network embedding is rather limited.

Given a signed network $\mathcal{G}(\mathcal{N}, \mathbf{A}^n, \mathbf{A}^p)$, the task of signed-network embedding is to learn a low-dimensional vector representation $\mathbf{u}_i \in \mathbb{R}^d$ for each user u_i , where d is the embedding dimension. Similarly to unsigned network embedding, a signed network embedding algorithm needs (1) an objective function for signed network embedding and (2) a representation learning algorithm to optimize the objective function. Social theories for unsigned social networks have been widely used to design objective functions for unsigned social network embedding. For example, social correlation theories such as homophily and social influence suggest that two positively connected users are likely to share similar interests, which are the foundations of many objective functions of unsigned network embedding [Belkin and Niyogi 2001]. Many social theories such as balance and status theories are developed for signed social networks and they provide fundamental understandings about signed social networks, which could pave us a way to develop objective functions for signed network embedding. Meanwhile, recently deep learning techniques provide powerful tools for representation learning that have enhanced various domains such as speech recognition, natural language processing, and computer vision [Lecun et al. 2015]. Therefore, a useful future direction is to harness the power of deep learning techniques to learn low-dimensional vector representations of nodes while preserving the fundamental understanding about signed social networks from social theories.

4. LINK-ORIENTED TASKS

The objects of link-oriented tasks are links among nodes, which aim to reveal fine-grained and comprehensive understandings of links. The availability of negative links in signed networks not only enriches the existing link-oriented tasks for unsigned networks such as link prediction and tie strength prediction but also encourages novel link-oriented tasks specific to signed networks such as sign prediction and negative link prediction. In this section, we review two extensively investigated link-oriented tasks in signed networks including link prediction and sign prediction. We would like to clarify the differences of these two tasks since they are used interchangeably in some literature. The differences of link prediction and sign prediction are demonstrated in Figure 6:

- In link prediction, signs of old links are given, while no signs are given to links in sign prediction; and
- Link prediction predicts new positive and negative links, while sign prediction predicts signs of existing links.

4.1. Link Prediction in Signed Networks

Link prediction infers new positive and negative links by giving old positive and negative links [Leskovec et al. 2010a; Chiang et al. 2011]. Existing link prediction algorithms can be roughly divided into two groups, which correspond to supervised and unsupervised methods. Supervised methods consider the link prediction problem as a classification problem by using the existence of links as labels, while unsupervised methods make use of the topological properties of the snapshot of the network. Next, we will review each group with representative algorithms.

4.1.1. Supervised Methods. Supervised methods treat link prediction as a classification problem and usually consist of two important steps. One is to prepare labeled data and the other is to construct features for each pair of users. The first step is trivial since the signs of links can be naturally treated as labels. Therefore different algorithms in this family provide different approaches to construct features.

In addition to in-degree and out-degree of positive (or negative) links, triangle-based features according to balance theory are extracted in Leskovec et al. [2010a]. Since signed social networks are usually very sparse and most users have few of in-degree or out-degree, many users could have no triangle-based features and triangle-based features may not be robust [Chiang et al. 2011]. A link prediction algorithm can be developed based on any quantitative social imbalance measure of a signed network. Hence, k -cycle-based features are proposed in Chiang et al. [2011], where triangle-based features are special cases of k -cycle-based features when $k = 3$. In addition to k -cycle-based features, incoming local bias (or the percentage of negative reviews it receives in all the incoming reviews) and outgoing local bias (or the percentage of negative reviews it gives to all of its outgoing reviews) are also reported to be helpful for the performance improvement in link prediction [Zhang et al. 2013]. In chemical and biological sciences, the quantitative structure-activity relationship hypothesis suggests that “similar molecules” show “similar activities,” for example, the toxicity of a molecule can be predicted by the alignment of its atoms in the three-dimensional space. This hypothesis may be applicable to social networks—the structure and network patterns of the ego-networks are strongly associated with the signs of their generated links. Therefore, frequent sub-networks from the ego-networks are used as features in Papaoikonomou et al. [2014]. Besides features extracted from topological information, attributes of users such as gender, career interest, hometown, movies, and thinking are also used as features in Patidar et al. [2012] where it first trains a classifier based on these features and then suggests new links and finally refines them which either maintain or enhance the balance index according to balance theory. Other types of features are also used for the problem of link prediction in signed networks including user interaction features [DuBois et al. 2011] and review-based features [Borzemek and Sydow 2010]. Interaction features are reported to be more useful than node attribute features in DuBois et al. [2011].

4.1.2. Unsupervised Methods. Unsupervised methods are usually based on certain topological properties of signed networks. Algorithms in this family can be categorized into similarity-based, propagation-based, and low-rank approximation-based methods.

Similarity-Based Methods: Similarity-based methods predict the signs of links based on node similarity. Note that similarity-based methods are typically designed for undirected signed networks. A typical similarity-based method consists of two steps. First, it defines a similarity metric to calculate node similarities. Then, it provides a way to predict positive and negative links based on these node similarities.

One popular way of calculating node similarity is based on user clustering. We discuss two representative approaches below:

- The network is partitioned into a number of clusters using the method in Doreian and Mrvar [1996]. Then, the conditional similarity for two clusters A and B with a third cluster C is defined according to Javari and Jalili [2014]:

$$Sim_{A,B|C} = \frac{\sum_{s \in S_{A,B|C}} m_{A,s} m_{B,s}}{\sqrt{s \in S_{A,B|C}} m_{A,s}^2 \sqrt{s \in S_{A,B|C}} m_{B,s}^2}, \quad (27)$$

where $S_{A,B|C}$ is the set of nodes in the cluster C, which are linked by nodes in A and B, and $m_{A,s}$ is the average signs of links from nodes in cluster A to node s . Node similarity is calculated as the similarity between clusters where these two nodes are assigned.

- Spectral clustering based on the Laplacian matrix for signed networks is performed [Symeonidis and Mantas 2013]. Then, two similarities are defined. The first is the similarity of nodes that are assigned to the same cluster:

$$simSC(i, j) = 1 - \|D(i, c_i) - D(j, c_j)\|. \quad (28)$$

The second is the similarity of nodes that are assigned to different clusters:

$$simDC(i, j) = \frac{1}{1 + D(i, c_i) + D(j, c_j)}, \quad (29)$$

where $D(\cdot, \cdot)$ is a distance metric.

Another way of calculating node similarity is based on status theory. According to status theory, the positive in-degree $|I^+|$ and the negative out-degree $|O^-|$ of a node increase its status. In contrast, the positive out-degree $|O^+|$, and negative in-degree $|I^-|$ decrease its status. According to this intuition, similarity is defined as follows [Symeonidis and Tiakas 2013]:

$$sim(i, j) = \frac{1}{\sigma(i) + \sigma(j) - 1} \quad (30)$$

$$\sigma(i) = |I_i^+| + |O_i^-| - |O_i^+| - |I_i^-|.$$

With node similarity, the second step is to determine the signs of links. Since we have pairwise node similarities, user-oriented collaborative filtering are used to aggregate signs from similar nodes to predict positive and negative links [Javari and Jalili 2014]. Another approach is based on status theory and the sign from i to j is predicted as the sign of the sum of $sign(sim(i, k) + sim(k, j))$ over all triplets (i, j, k) [Symeonidis and Tiakas 2013].

Propagation-Based Methods: The vast majority of propagation-based methods are proposed for trust-distrust networks, which are a special (and important) class of signed networks. The adjacency matrix \mathbf{A} is very sparse and many entries in \mathbf{A} are zero. The basic idea of propagation-based methods is to compute a dense matrix $\hat{\mathbf{A}}$ with the same size of \mathbf{A} by performing certain propagation operators on \mathbf{A} . Then the sign of a link from u_i to u_j is predicted as $sign(\hat{\mathbf{A}}_{ij})$ and the likelihood is $|\hat{\mathbf{A}}_{ij}|$. In Guha et al. [2004], trust propagation is treated as a repeating sequence of matrix operations, which consists of four types of atomic trust propagations. These four types are direct propagation, trust coupling, co-citation, and transpose trust, as shown in Figure 7. Two strategies are studied for incorporating distrust. The first is that of *one-step distrust propagation*, in which we propagate multiple-step trust and then propagate one-step distrust. The second is that of *multiple-step distrust propagation*, in which trust and distrust propagate together. One-step distrust propagation often outperforms multiple-step distrust propagation [Guha et al. 2004]. However, one-step distrust propagation

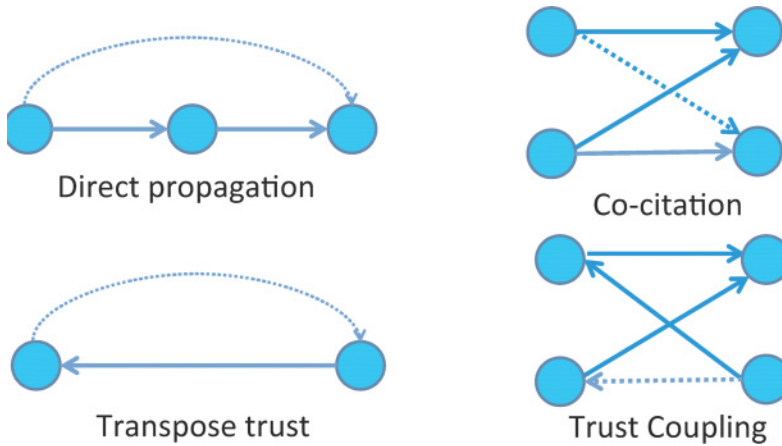


Fig. 7. Four types of atomic trust propagations.

might not converge, when the network is dominated by distrust links. On the other hand, multiple-step distrust propagation may yield some unexpected behaviors [Ziegler and Lausen 2005]. To mitigate these two problems, Ziegler and Lausen [2005] propose to integrate distrust into the process of the Applesseed trust metric computation instead of superimposing distrust afterwards. Methods in Guha et al. [2004] and Ziegler and Lausen [2005] are based on the matrix representation. There are methods in this family investigating other representations such as subjective logic [Knapskog 1998], intuitionistic fuzzy relations [De Cock and Da Silva 2006], and bilattice [Victor et al. 2006], which can naturally perform both trust and distrust propagation by defining corresponding operators.

Low-rank approximation methods: The notion of balance was generalized by Davis [1967] to weak balance, which allows triads with all negative links. Low-rank approximation methods are based on weak structural balance, as suggested in Hsieh et al. [2012] that weakly balanced networks have a low-rank structure, and weak structural balance in signed networks naturally suggests low-rank models for signed networks. Low-rank approximation methods compute the dense matrix $\hat{\mathbf{A}}$ via the low-rank approximation of \mathbf{A} instead of propagation operators for propagation-based methods. With $\hat{\mathbf{A}}$, the sign and the likelihood of a link from u_i to u_j are predicted as $\text{sign}(\hat{\mathbf{A}}_{ij})$ and $|\hat{\mathbf{A}}_{ij}|$, respectively. In Hsieh et al. [2012], the link prediction problem in signed networks is mathematically modeled as a low-rank matrix factorization problem as follows:

$$\min_{\mathbf{W}, \mathbf{H}} \sum_{i,j} (\mathbf{A}_{ij} - (\mathbf{W}^T \mathbf{H})_{ij})^2 + \lambda (\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2), \quad (31)$$

where $\mathbf{W}^T \mathbf{H}$ is the low-rank matrix to approximate \mathbf{A} . The square function is chosen as the loss function in $(\mathbf{A}_{ij} - (\mathbf{W}^T \mathbf{H})_{ij})^2$. Pairwise empirical error, similar to the hinge loss convex surrogate for 0/1 loss in classification, is used in Agrawal et al. [2013]. They use this particular variation since it elegantly captures the correlations amongst the users and thereby makes the technique more robust to fluctuations in individual behaviors. In Cen et al. [2013], a low-rank tensor model is proposed for link prediction in dynamic signed networks.

4.2. Sign Prediction

Most social media services provide unsigned social networks such as the friendship network in Facebook and the following network in Twitter, while only a few services provide signed social networks. The task of sign prediction is to infer the signs of existing links in the given unsigned network. It is difficult, if not impossible, to predict signs of existing links by only utilizing the given unsigned network [Yang et al. 2012]. Therefore, most of the existing sign predictors use additional sources of information. The most widely used sources are user interaction information and cross-media information.

4.2.1. Sign Prediction with Interaction Data. In reality, we are likely to adopt the opinions from our friends while fighting the opinions of our foes. As a consequence, decisions of users with positive links are more likely to agree, whereas for users with negative connections, the chance of disagreement would be considerably higher. In social media, users can perform positive or negative interactions with other users. Positive interactions show agreement and support, while negative interactions show disagreement and antagonism. There are strong correlations between positive (or negative) links and positive (or negative) interactions [Yang et al. 2012]. Tang et al. suggest a straightforward algorithm for sign prediction based on the correlation between interactions and links. The first step is to initialize signs of links based on interactions. Positive signs are used for positive interactions, whereas negative signs are used for negative interactions. Next, the signs of links are refined according to status theory or balance theory [Tang et al. 2015]. More sophisticated algorithms incorporate link and interaction information into coherent frameworks. In Yang et al. [2012], a framework with a set of latent factor models is proposed to infer signs for unsigned links, which capture user interaction behavior, social relations, as well as their interplay. It also models the principles of balance and status theories for signed networks. A one-dimensional latent factor β_i is introduced for u_i and then we model the sign between u_i and u_j as $s_{ij} = \beta_i \beta_j$, which can capture balance theory. The vector parameter η is introduced for users to capture their partial ordering, and then the status ℓ_i of u_i is modeled as $\ell_i = \eta \gamma_i$, where γ_i is the latent factor vector of u_i . Status theory characterizes the sign from u_i to u_j as their relative status difference $\ell_{ij} = \ell_i - \ell_j$. Yu and Xie find significant correlations and mutual influence between user interactions and signs of links. They propose a mutual latent random graph framework to flexibly model evidence from user interactions and signs. This approach is used to perform user interaction prediction and sign prediction simultaneously [Yu and Xie 2014a, 2014b].

4.2.2. Sign Prediction with Cross-Media Data. In the task of link prediction in signed networks, Leskovec et al. find that the learned link predictors have very good generalization power across social media sites. This observation suggests that general guiding principles might exist for sign inference across different networks, even when links have different semantic interpretations in different networks [Leskovec et al. 2010a]. Another useful source for sign prediction is cross-media information. The goal is to predict signs of a target network with a source signed network. The basic approach is to learn knowledge or patterns from the source signed network and use it to predict link signs in the target network. The vast majority of algorithms in this family use transfer learning to achieve this goal. One representative way is to construct generalizable features that can transfer patterns from the source network to the target network for sign prediction. Since some social theories such as status and balance theories are applicable for all signed networks, it is possible to extract generalizable features suggested by social theories, such as balance and status theory. In Tang et al. [2012], a factor-graph model is learned with features from the source network to infer signs of the target

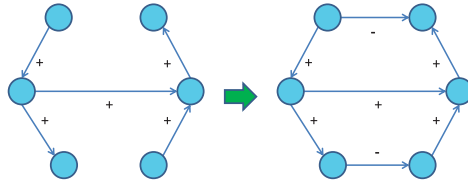


Fig. 8. Illustration of the problem of negative link prediction.

network. Although links in different signed networks may have different semantics, a certain degree of similarity always exists across domains, for example, similar degree distributions and diameters. With this intuition, an alternative way is to project the source and target networks into the same latent space. Latent topological features are constructed to capture the common patterns between the source and target networks. This is obtained through the following optimization problem [Ye et al. 2013]:

$$\min_{\mathbf{U}_s, \Sigma, \mathbf{V}_s, \mathbf{U}_t, \mathbf{V}_t} \|\mathbf{A}_s - \mathbf{U}_s \Sigma \mathbf{V}_s^\top\|_F^2 + \|\mathbf{A}_t - \mathbf{U}_t \Sigma \mathbf{V}_t^\top\|_F^2 + \alpha \|\Sigma\|_F^2, \quad (32)$$

where \mathbf{A}_s and \mathbf{A}_t denote the adjacency matrices for the source and target network, respectively. \mathbf{U}_s , \mathbf{V}_s , \mathbf{U}_t , and \mathbf{V}_t are four latent topological feature matrices [Ye et al. 2013]. Σ is the common latent space for both networks, which ensures that the extracted topological features of both graphs are expressed in the same space. With the latent topological features, a transfer learning with instance weighting algorithm is proposed to predict signs of the target unsigned network \mathbf{A}_t by learning knowledge from the source signed network \mathbf{A}_s .

4.3. Promising Directions for Link-Oriented Tasks

For many social media sites, negative links are usually unavailable, which might limit the applications of mining signed networks. Therefore, it is helpful to predict negative links. Furthermore, for most signed social networks in social media, only binary relations are available and strengths of the relations are not available. In other words, we would like to perform *tie strength prediction*. In this subsection, we discuss these two link-oriented tasks.

4.3.1. Negative Link Prediction. It is evident from recent work that negative links have significant added value over positive links in various analytical tasks such as positive link prediction [Guha et al. 2004; Leskovec et al. 2010a] and recommender systems [Victor et al. 2009; Ma et al. 2009]. However, it is generally not very desirable for online social networks to explicitly collect negative links [Hardin 2004; Kunegis et al. 2013]. As a consequence, the vast majority of social media sites such as Twitter and Facebook do not enable users to explicitly specify negative links. Therefore, it is natural to question whether one can predict negative links automatically from the available data in social networks. While this problem is very challenging [Chiang et al. 2013], the results of such an approach have the potential to improve the quality of the results of a vast array of applications. The negative link prediction problem is illustrated in Figure 8. The negative link prediction problem differs from both the link prediction and sign prediction problems as follows:

- Link prediction in signed networks predicts positive and negative links from existing positive and negative links. On the other hand, negative link prediction does not assume the existence of negative links.

—Sign prediction predicts signs of *already existing* links. While the negative link prediction problem needs to identify the pairs of nodes between which negative links are predicted to exist.

A recent work in Tang et al. [2015] found that negative links can be predicted with user interaction data by using the correlation between negative interactions and negative links. Furthermore, the proposed negative link predictor in Tang et al. [2015] has very good generalization across social media sites, which suggests that cross-media data might be also helpful in the problem. It is possible to build signed networks via sentiment analysis of texts [Hassan et al. 2012b; Wang et al. 2014], which suggests that user-generated content has significant potential in predicting negative links in social media.

4.3.2. Tie-Strength Prediction. The cost of forming links in social media is very low, as a result of which many weak ties are formed [Xiang et al. 2010]. The authors of Huberman et al. [2008] show that users can have many followers and followers in Twitter with whom they are only weakly associated in the physical world. Users with strong ties tend to be more similar than those with weak ties. Since homophily is a useful property from the perspective of mining tasks, such as recommendation and friend management, it suggests that tie-strength prediction can also be very useful. For unsigned networks in social media, such as friendship in Facebook and Twitter, we often choose a binary adjacency matrix representation where 1 denotes a positive link from u_i to u_j and 0 otherwise. The tie-strength prediction task in unsigned networks is to infer a strength in $[0, 1]$ for a given positive link. The original binary matrix representation with values in $\{0, 1\}$ is converted into a continuous valued matrix representation with values in $[0, 1]$ by tie-strength prediction in unsigned networks.

If we choose one adjacency matrix \mathbf{A} to represent a signed network with $\{-1, 0, 1\}$ to denote negative, missing, and positive links, then a tie strength predictor infers strength values in $[-1, 0]$ for negative links and $[0, 1]$ for positive links. If we choose two adjacency matrices \mathbf{A}^p and \mathbf{A}^n in $\{0, 1\}$ to represent positive and negative links separately, then a tie strength predictor infers strength values in $[0, 1]$ for positive and negative links.

Previous studies in positive tie-strength prediction problem suggest that pairwise user similarity is reflected in strong ties. Therefore, the strengths of positive ties are modeled as the hidden impacts of node similarities. Furthermore, the strengths of positive ties are modeled as the hidden causes of user interactions since they affect the frequency and nature of user interactions [Xiang et al. 2010]. A preliminary work in Tang et al. [2014b] finds that it is more likely for two users to have negative links if they have more negative interactions. Analogously, this suggests the following directions for tie-strength prediction: (a) What is the relation between negative tie strength and node-node similarities and how negative tie strength impacts user interactions, and (b) how negative and positive tie strength affect one another.

5. APPLICATION-ORIENTED TASKS

Just as unsigned networks are used frequently in the context of various applications such as data classification [Zhu et al. 2007], data clustering [Long et al. 2006], information propagation [Kempe et al. 2003], and recommendation [Tang et al. 2013], signed networks can be leveraged as well. Application-oriented tasks augment traditional algorithms with signed networks. For example, in addition to rating information, recommender systems with signed networks can also make use of signed networks. In this section, we review the recommendation and information diffusion applications and discuss promising research directions.

5.1. Recommendation with Signed Networks

Assume that \mathbf{R} is the user-item ratings matrix where \mathbf{R}_{ij} is the rating from the i th user to the j th item. In a typical recommender system, most of the entries are missing. Traditional recommender systems aim to predict these missing values by using observed values in \mathbf{R} . In the physical world, we always seek recommendations from our friends, which suggests that social information may be useful to improve recommendation performance. Many recommender systems are proposed to incorporate ones' friends for recommendation and gain performance improvement. A comprehensive review about social recommendation can be found in Tang et al. [2013, 2014]. Scholars have noted that negative links may be more noticeable and credible than the positive links with a similar magnitude [Cho 2006]. Negative links may be as important as positive links for recommendation. In recent years, systems based on collaborative filtering (CF) are proposed to incorporate both positive and negative links for recommendation. Typically, a CF-based recommender system with signed networks contains two components corresponding to the basic CF model and the model extracted from the signed network. The basic CF models are categorized into memory-based and model-based systems.

5.1.1. Memory-Based Methods. Memory-based recommender systems with signed networks choose memory-based collaborative filtering and especially user-oriented models [Victor et al. 2009, 2013; Chen et al. 2013b; Nalluri 2014]. A typical user-oriented model first calculates pairwise user similarity based on some similarity metrics such as Pearson's correlation coefficient or cosine similarity. Then, a missing rating of user i for item j is predicted by aggregating ratings from the similar peers of user i as follows:

$$\hat{\mathbf{R}}_{ij} = \hat{r}_i + \frac{\sum_{v \in N_i} W_{iv} (\mathbf{R}_{vj} - \hat{r}_v)}{\sum_{v \in N_i} W_{iv}}, \quad (33)$$

where N_i is the set of similar users of u_i , \hat{r}_i is the average rating from u_i , and W_{iv} is the connection strength between u_i and u_v . There are several strategies for incorporating negative links into the above user-oriented model as:

—One is to use negative links to avoid recommendations from these “unwanted” users as Victor et al. [2009]:

$$\hat{\mathbf{R}}_{ij} = \hat{r}_i + \frac{\sum_{v \in N_i \setminus D_i} W_{iv} (\mathbf{R}_{vj} - \hat{r}_v)}{\sum_{v \in N^+} W_{iv}}. \quad (34)$$

D_i is the set of users to whom u_i has negative links.

—Another way is to consider negative links as negative weights, that is, considering negative links as dissimilarity measurements, as Victor et al. [2013]:

$$\hat{\mathbf{R}}_{ij} = \hat{r}_i + \frac{\sum_{v \in N_i} W_{iv} (\mathbf{R}_{vj} - \hat{r}_v)}{\sum_{v \in N_i} W_{iv}} - \frac{\sum_{v \in D_i} d_{iv} (\mathbf{R}_{vj} - \hat{r}_v)}{\sum_{v \in D_i} d_{iv}}, \quad (35)$$

where d_{iv} is the dissimilarity between u_i and u_v .

—In reality, positive and negative links in signed networks are very sparse; therefore, Nalluri [2014] proposes a recommender system that first propagates positive and negative values in signed networks and then reduces the influence from negative values as:

$$\hat{\mathbf{R}}_{ij} = \hat{r}_i + \frac{\sum_{v \in N_i} (W_{iv} - d_{iv}) (\mathbf{R}_{vj} - \hat{r}_v)}{\sum_{v \in N_i} (W_{iv} - d_{iv})}. \quad (36)$$

5.1.2. Model-Based Methods. Model-based recommender systems with negative links use model-based collaborative filtering. Matrix factorization methods are particularly

popular [Ma et al. 2009; Forsati et al. 2014]. Assume that \mathbf{U}_i is the k -dimensional preference latent factor of u_i and \mathbf{V}_j is the k -dimensional characteristic latent factor of item j . A typical matrix factorization-based collaborative filtering method models the rating from u_i to the j th item \mathbf{R}_{ij} as the interaction between their latent factors, that is, $\mathbf{R}_{ij} = \mathbf{U}_i^\top \mathbf{V}_j$, where \mathbf{U}_i and \mathbf{V}_j can be obtained by solving the following optimization problem:

$$\min_{\mathbf{U}, \mathbf{V}} \sum_{i=1}^n \sum_{j=1}^m \mathbf{W}_{ij} (\mathbf{R}_{ij} - \mathbf{U}_i \mathbf{V}_j^\top)^2 + \alpha (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2), \quad (37)$$

where $\mathbf{U} = [\mathbf{U}_1^\top, \mathbf{U}_2^\top, \dots, \mathbf{U}_N^\top]^\top \in \mathbb{R}^{n \times K}$ and $\mathbf{V} = [\mathbf{V}_1^\top, \mathbf{V}_2^\top, \dots, \mathbf{V}_M^\top]^\top \in \mathbb{R}^{m \times K}$, where N and M are the numbers of users and items in a recommender system. The term $\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2$ is introduced to avoid over-fitting, controlled by the parameter α . $\mathbf{W} \in \mathbb{R}^{n \times m}$ is a weight matrix, where \mathbf{W}_{ij} is the weight for the rating for u_i to v_j . A common way to set \mathbf{W} is $\mathbf{W}_{ij} = 1$ if we observe a rating from u_i to the j th item, and $\mathbf{W}_{ij} = 0$ otherwise. The optimization problem in Equation (37) is convex for \mathbf{U} and \mathbf{V} , respectively. Therefore it is typically solved by gradient decent methods or alternating least squares. If u_i positively link to u_j , then u_i and u_j are likely to share similar preferences. Therefore, to capture positive links, Ma et al. [2011] added a term to minimize the distance of the preference vectors of two users with a positive link based on Equation (37) as follows:

$$\min_{\mathbf{U}, \mathbf{V}} \sum_{i=1}^n \sum_{j=1}^m \mathbf{W}_{ij} (\mathbf{R}_{ij} - \mathbf{U}_i \mathbf{V}_j^\top)^2 + \alpha (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) + \beta \sum_i \sum_{j \in N_i^+} S_{ij}^p \|\mathbf{U}_i - \mathbf{U}_j\|_2^2, \quad (38)$$

where S_{ij}^p is the strength of the positive link from u_i to u_j and β controls the contribution from positive links.

If u_i has a negative link to u_j , then it is likely that u_i thinks that u_j has totally different tastes. With this intuition, for a negative link from u_i to u_j , Ma et al. [2009] introduce a term to maximize the distance of their latent factors based on the matrix factorization model as follows:

$$\min_{\mathbf{U}, \mathbf{V}} \sum_{i=1}^n \sum_{j=1}^m \mathbf{W}_{ij} (\mathbf{R}_{ij} - \mathbf{U}_i \mathbf{V}_j^\top)^2 + \alpha (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) - \beta \sum_i \sum_{j \in D_i} S_{ij}^n \|\mathbf{U}_i - \mathbf{U}_j\|_2^2, \quad (39)$$

where S_{ij}^n is the strength of the negative link for u_i to u_j . The underlying assumption of Equation (39) is to consider negative links as dissimilarity measurements. Gradient descent is performed in Ma et al. [2009] to obtain a local minimum of the objective function given by Equation (39). However, recent research suggests that negative links may not denote dissimilarity and users with negative links tend to be more similar than randomly selected pairs [Tang et al. 2014a]. It also observes that users with positive links are likely to be more similar than pairs of users with negative links, which is very consistent with the extension of the notion of structural balance in Cygan et al. [2012]—a structure in signed network should ensure that users are able to have their “friends” closer than their “enemies,” that is, users should sit closer to their “friends” (or users with positive links) than their “enemies” (or users with negative links). With this intuition, for (i, j, k) where u_i has a positive link to u_j while has a negative link to u_k , the latent factor of u_i should be more similar to the latent factor of u_j than that of u_k to capture negative links. In particular, for each triple as (i, j, k) , a regularization term is added as follows:

$$\ell(d(\mathbf{U}_i, \mathbf{U}_j), d(\mathbf{U}_i, \mathbf{U}_k)), \quad (40)$$

where d is a similarity metric and ℓ is a penalty function that assesses the violation of latent factors of users with positive and negative links [Forsati et al. 2014]. Possible choices of $\ell(z)$ are the hinge loss function $\ell(z) = \max(0, 1 - z)$ and the logistic loss function $\ell(z) = \log(1 + e^{-z})$. In Forsati et al. [2014], a stochastic gradient descent (SGD) method is employed to optimize Equation (40). For a signed network with N users, there could be N^3 triples that indicates we need to introduce N^3 possible regularization terms as Equation (40) to capture the signed network for recommendations [Forsati et al. 2014]. Therefore, the computational cost of the system is very high. In Tang et al. [2016b], a system with only N extra regularization terms is proposed that is much more efficient. A sophisticated recommender system is proposed in Yang et al. 2012]. This system has several advantages: (1) It can perform recommendation and sign prediction simultaneously, and (2) it is the first framework to model balance theory and status theory explicitly for recommendation with signed networks.

5.2. Information Diffusion

Information diffusion can enable various online applications such as effective viral marketing and has attracted increasing attention [Kempe et al. 2003; Chen et al. 2009]. There are many information diffusion models for unsigned social networks including the classic voter model [Clifford and Sudbury 1973], the susceptible-infected-recovered (SIR) epidemic model [May and Lloyd 2001], the independent cascade (IC) model [Goldenberg et al. 2001a, 2001b], and the threshold model [Granovetter 1978; Schelling 2006]. One can apply these models of unsigned networks to signed networks by ignoring negative links. However, ignoring negative links might result in over-estimation of the impact of positive links [Li et al. 2013]. Therefore, studying information diffusion and maximization in signed networks can not only help us understand the impact of user interactions on information diversity but also push the boundaries of researches about dynamical process in complex networks. In addition, empirical results on real-world signed networks demonstrate that incorporating link signs into information diffusion models usually gains influence [Li et al. 2013, 2014b; Shafaei and Jalili 2014]. For example, the voter model with negative links generates at maximum of 38% and 21% more influence in the Epinions dataset compared to the model with only positive links [Li et al. 2013]. In the rest of this section, we will review representative diffusion models for signed networks

5.2.1. Voter Model for Signed Networks. A typical scenario of the application of the voter model is when users' opinions switch forth and back according to their interactions with other users in networks. The authors of Li et al. [2013, 2014] investigate how two opposite opinions diffuse in signed networks based on the voter model proposed in Clifford and Sudbury [1973]. It is more likely for users to adopt and trust opinions from their friends, while users are likely to adopt the opposite opinions of their foes. This intuition corresponds to the principles of "enemies' enemies are my friends" and "my enemies' friends are my enemies." Hence, each node u_i selects one user u_j from his/her outgoing social networks randomly and performs two possible actions: (1) If u_i has a positive link to the selected user u_j , then u_i adopts u_j 's opinion, and (2) if u_i has a negative link to u_j , then u_i chooses the opinion opposite to u_j 's.

5.2.2. SIR Epidemic Model for Signed Networks. Using epidemiology to study information spread has become increasingly popular in recent years [May and Lloyd 2001] because the information spread mechanisms are qualitatively similar to those of the biological disease spread [Volz and Meyers 2007]. The standard SIR model assigns one of three states (susceptible, infected, or recovered) to each user. Based on SIR, the authors of Li et al. [2013] and Fan et al. [2012] define five states for signed networks: (1) S_0 : susceptible with neutral opinions; (2) I_- : infected with negative opinions;

Table IV. SIR for Signed Networks

Actions	Probabilities	Relationship
$I_+ + S_0 \rightarrow I_+ + I_+$	λ_s	+1
$I_+ + S_0 \rightarrow I_+ + I_-$	λ_o	-1
$I_- + S_0 \rightarrow I_- + I_-$	λ_s	+1
$I_- + S_0 \rightarrow I_- + I_+$	λ_o	-1

(3) I_+ : infected with positive opinions; (4) R_- : recovered with negative opinions; and (5) R_+ : recovered with positive opinions. Users with S_0 can be infected by users with I_- or I_+ ; and users with R_+ or R_- do not spread their opinions any more. With the same intuition in Li et al. [2013], users are likely to adopt and trust opinions from their friends, while adopting the opposite opinions of their foes. At each step, users with state I_+ (or I_-) pick up one user from their social networks to interact with, and they can perform four possible actions depending on probabilities and the sign of links as shown in Table IV.

5.2.3. Independent Cascade Model for Signed Networks. Nodes in the network are assigned one of two states, active or inactive, by the independent cascade model [Goldenberg et al. 2001a]. At the t th step, every active node u_i has one single opportunity to activate inactive users u_j in his/her network with an independently successful probability p_{ij} . u_j becomes active in the $t + 1$ -th step if u_i succeeds. After this opportunity, u_i cannot take actions on u_j any more in subsequent steps. The authors of Li et al. [2014b] propose a Polarity-related Independent Cascade (ICP) model for signed networks. Each node in the ICP model is assigned to one of three states: (1) negative: adopting but opposing the spreading opinion; (2) positive: adopting and supporting the opinion; and (3) inactive: not adopting the opinion. There are two major differences between the ICP model and the standard IC model. First, each user can be only activated once in each step for ICP. Second, if u_i activates u_j , then the state S_j of u_j depends on u_i 's state S_i and the sign of their link as $S_j = S_i \times s_{ij}$.

5.2.4. Threshold Model for Signed Networks. The node u_i becomes active in the threshold model if and only if his/her active neighbors are more than a threshold θ_i as: $\sum_{u_j \text{ active neighbor of } u_i} b_{ij} > \theta_i$, where b_{ij} is a weight between u_i and u_j . The authors of Shafaei and Jalili [2014] introduce an information diffusion model based on the threshold model for signed networks where each node maintains a payoff matrix. If the payoff matrices for all nodes are the same, then the proposed model boils down to the standard threshold model. We assume that there are two behaviors "B" and "A"; all nodes start with "B" and then some randomly selected nodes change to "A." In each iteration, every node observes his/her social network, calculates the payoff matrix, and then adopts the behavior maximizing the benefits to him/her. Note that the payoff matrix is calculated only for these nodes with behavior "B." If many friends have the same behavior, then doing the behavior changes can increase the payoff gain, which also increases if few foes have the behavior.

5.3. Promising Directions for Application-Oriented Tasks

Unsigned networks are exploited to help various real-world applications such as data classification [Sindhwani et al. 2005], data clustering [Long et al. 2006], active learning [Bilgic et al. 2010], information propagation [Kempe et al. 2003], recommendation [Tang et al. 2013], sentiment analysis [Speriosu et al. 2011], and feature selection [Tang and Liu 2012]. Therefore, there are many opportunities in the signed network setting. In this subsection, we focus our discussions on two application-oriented tasks, which are data classification and clustering. We focus on these tasks because these

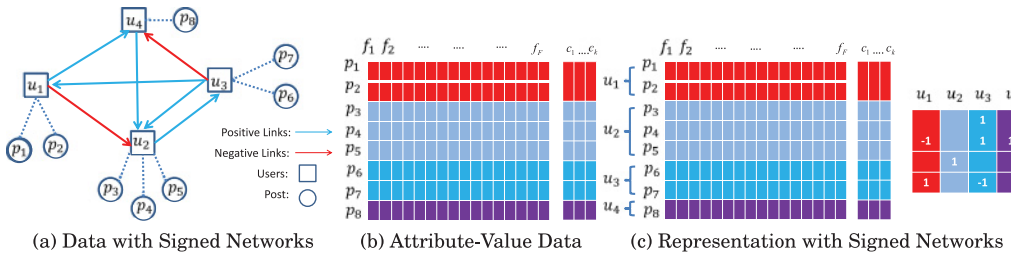


Fig. 9. A simple example of data in the problem of data classification with signed networks.

problems are very general and have applicability to many problems such as sentiment analysis [Tan et al. 2011; Hu et al. 2013]. Furthermore, we can follow similar ways for data classification and clustering problems to define other application-oriented tasks such as active learning and feature selection.

5.3.1. Data Classification with Signed Networks. Figure 9 demonstrates a simple example for data classification with signed networks. The signed network in Figure 9(a) has four users (u_1, \dots, u_4) and each user has some posts (e.g., u_1 has two posts p_1 and p_2). We use posts in a loose way to cover various types of user-generated content such as posts, tweets, or images. In data classification with signed networks, there is additional link information such as user-post and user-user links as shown in Figure 9(c). Let $\mathcal{F} = \{f_1, f_2, \dots, f_F\}$ be a set of F features and $\mathcal{P} = \{p_1, p_2, \dots, p_M\}$ be the set of M posts. $\mathbf{P} \in \mathbb{R}^{N \times M}$ denotes the user-post authorship matrix where $\mathbf{P}_{ij} = 1$ if u_i creates p_j and 0 otherwise; $\mathbf{X} \in \mathbb{R}^{M \times F}$ denotes the attribute-value representation of \mathcal{P} and $\mathbf{Y} \in \mathbb{R}^{M \times c}$ is the label indicator matrix where $\mathbf{Y}_{ij} = 1$ if p_i is labeled as the j th class and 0 otherwise. *The problem of data classification with signed networks is that of training classifiers to predict class labels for unseen posts by utilizing data instances $\{\mathbf{X}, \mathbf{Y}\}$ and their contextual information from signed networks $\{\mathbf{P}, \mathbf{A}\}$.*

Research on data classification with unsigned networks found that class labels of posts from the same user are likely to be consistent and that users with links are likely to generate posts with similar class labels [Tang and Liu 2012]. There are two popular ways of exploiting contextual information from unsigned networks for data classification based on these two findings. One way is to convert contextual information into correlation links between posts. This boils down to the problem of combining content and correlation links for data classification [Qi and Davison 2009]. The other way is that we first extract constraints from contextual information for posts and extend traditional classifiers to model these constraints such as LapRLS from Least Squares in Belkin et al. [2006] and LapSVM from Support Vector Machines [Sindhwani et al. 2005]. To address the problem of data classification with signed networks, we may need to understand the structure of positive and negative links in signed networks in relation to attributes and labels of posts. For example, what are the properties of posts from users with negative links in terms of attributes and labels? If users have both positive and negative links, then what are the differences in terms of their posts? If users with positive links are more likely to generate similar posts to users with negative links, then the problem boils down to that of classification with relative comparisons [Schultz and Joachims 2004].

5.3.2. Data Clustering with Signed Networks. Differing from data classification, data clustering is unsupervised learning, that is, the label information \mathbf{Y} is not available. The problem of data clustering with signed networks is to find f that identifies k post clusters so posts in the same cluster are more similar to each other than to those in other

clusters by using information in $\{\mathbf{X}, \mathbf{P}, \mathbf{A}\}$ and can be formulated as follows:

$$f : \{\mathbf{X}, \mathbf{P}, \mathbf{A}\} \rightarrow \{C_1, C_2, \dots, C_k\}, \quad (41)$$

where C_i is the i th clusters identified by a clustering function f .

By introducing the concept of pseudo-labels, unsupervised learning problems can be transformed into supervised learning problems [Masaeli et al. 2010; Cai et al. 2010]. Hence, an intuitive research direction is to transform clustering with signed networks into classification with signed networks with pseudo-labels. It is likely that posts from users with negative links may be from different clusters and negative links may serve as additional constraints when we cluster posts. Therefore, another possible direction for data clustering with signed networks is to transform data clustering algorithms with unsigned networks by considering negative links as constraints and these constraints force posts from users with negatives links to different clusters, which behaviors similarly to traditional constraint clustering problem [Wagstaff et al. 2001]. Recent research investigates how to embed signed networks into a latent space where nodes sit closer to their “friends” than their “enemies” [Cygan et al. 2012; Pardo et al. 2013; Kermarrec and Thraves 2014]. Similarly, we can develop algorithms to embed the combination of signed networks and posts to learn representations for users and posts simultaneously.

6. CONCLUSIONS

The availability of large-scale signed networks in social media has encouraged increasing attention on mining signed networks. Signed networks are unique in terms of basic concepts, principles, and properties of specific computational tasks. This survey article provides a comprehensive overview about mining signed networks in social media. We first introduce basic concepts, principles, and properties of signed networks, including signed network representations (Section 2.1), properties of positive and negative links, and social theories for signed networks. Then, we classify various tasks into node-oriented, link-oriented, and application-oriented groups. Some of these tasks are well studied, whereas others need further investigation. For each group, we review well-studied tasks with representative algorithms and also discuss some tasks that are not sufficiently studied with formal definitions together with promising research directions.

In reviewing representative algorithms of well-studied tasks, for the methodology perspective, we notice that social theories such as balance theory and status theory are widely used in mining signed networks and we summarize three major ways in applying social theories in mining signed networks, that is, feature engineering, constraint generating, and objective defining.

- Feature Engineering*: It helps extract features for computational models according to social theories. For example, in link prediction, triangle-based features are extracted based on balance theory to improve link predilection [Leskovec et al. 2010a], while triad features are extracted based on status theory to predict signs of links in Tang et al. [2012].
- Constraint Generating*: It generates constraints from social theories for computational models. Regularization is one of the most popular ways to implement constraint generating. For example, a regularization term is added to capture signed networks for recommendation based on generalized balance theory [Forsati et al. 2014], and balance regularization is defined in Tang et al. [2015] to apply balance theory for negative link prediction.
- Objective Defining*: It uses social theories to define the objectives of the computational models. For example, in Amelio and Pizzuti [2013], based on balance theory, two

objectives are developed for community detection, and balance theory and status theory are explicitly captured in the objective functions for sign prediction [Yang et al. 2012].

While from the technique perspective, we find that similar techniques such as random walk, low-rank approximation, and spectral clustering are adopted by various tasks of mining signed networks:

- Random Walk: Given a network and a starting node, we select one of its neighbors randomly and move to the neighbor. Then we choose a neighbor of this node at random and walk to it and so on. The (random) sequence of nodes selected this way is a random walk on the network [Lovász 1993]. The techniques of random walk are used in various tasks of mining signed networks such as node ranking [Traag et al. 2010] and community detection [Yang et al. 2007].
- Low-Rank Approximation: Low-rank approximation aims to find a low-rank matrix such that the cost function, which measures the fit between the low-rank matrix and a given matrix, is optimized. It captures the low-rank structure of signed networks for link prediction [Hsieh et al. 2012] and it is one of the major techniques to build recommender systems with signed networks.
- Spectral Clustering: Spectral clustering is derived from the graph partition problem, which aims to find a partition such that the cut (the number of links between two disjoint sets of nodes) is minimized. Spectral clustering is one of the most popular approaches for community detection [Kunegis et al. 2010]. Meanwhile, it can naturally generate vector representations for nodes thus it is also widely used in other tasks such as link prediction [Chiang et al. 2013].

ACKNOWLEDGMENTS

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

- Mohammad Ali Abbasi, Jiliang Tang, and Huan Liu. 2014. Scalable learning of users preferences using networked data. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*. ACM, 4–12.
- Charu Aggarwal and Karthik Subbian. 2014. Evolutionary network analysis: A survey. *ACM Comput. Surv.* 47, 1 (2014), 10.
- Charu C. Aggarwal. 2011. *An Introduction to Social Network Data Analytics*. Springer.
- Priyanka Agrawal, Vikas K. Garg, and Ramasuri Narayanam. 2013. Link label prediction in signed social networks. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*. AAAI Press, 2591–2597.
- Nir Ailon, Yudong Chen, and Xu Huan. 2013. Breaking the small cluster barrier of graph clustering. *arXiv preprint arXiv:1302.4549* (2013).
- Alessia Amelio and Clara Pizzuti. 2013. Community mining in signed networks: A multiobjective approach. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ACM, 95–99.
- Pranay Anchuri and Malik Magdon-Ismael. 2012. Communities and balance in signed networks: A spectral approach. In *Proceedings of the 2010 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'12)*. IEEE, 235–242.
- T. Antal, P. L. Krapivsky, and S. Redner. 2005. Dynamics of social balance on networks. *Phys. Rev. E* 72, 3 (2005), 036121.
- Robert Axelrod and D. Scott Bennett. 1993. A landscape theory of aggregation. *Br. J. Pol. Sci.* 23, 02 (1993), 211–233.
- Nikhil Bansal, Avrim Blum, and Shuchi Chawla. 2004. Correlation clustering. *Mach. Learn.* 56, 1–3 (2004), 89–113.

- Ghazaleh Beigi, Jiliang Tang, and Huan Liu. 2016. Signed link analysis in social media networks. In *Proceedings of the 10th International AAAI Conference on Web and Social Media*.
- Mikhail Belkin and Partha Niyogi. 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*, Vol. 14. 585–591.
- Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. 2006. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.* 7 (2006), 2399–2434.
- Smriti Bhagat, Graham Cormode, and S. Muthukrishnan. 2011. Node classification in social networks. In *Social Network Data Analytics*. Springer, 115–148.
- Mustafa Bilgic, Lilyana Mihalkova, and Lise Getoor. 2010. Active learning for networked data. In *Proceedings of the 27th International Conference on Machine Learning (ICML'10)*. 79–86.
- Petko Bogdanov, Nicholas D. Larusso, and Ambuj Singh. 2010. Towards community discovery in signed collaborative interaction networks. In *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops (ICDMW'10)*. IEEE, 288–295.
- Phillip Bonacich and Paulette Lloyd. 2004. Calculating status with negative relations. *Soc. Netw.* 26, 4 (2004), 331–338.
- Christian Borgs, Jennifer Chayes, Adam Kalai, Azarakhsh Malekian, and Moshe Tennenholtz. 2010. A novel approach to propagating distrust. *Internet Netw. Econ.* (2010), 87–105.
- Piotr Borzysmek and Marcin Sydow. 2010. Trust and distrust prediction in social network with combined graphical and review-based attributes. In *Agent and Multi-Agent Systems: Technologies and Applications*. Springer, 122–131.
- Deng Cai, Chiyuan Zhang, and Xiaofei He. 2010. Unsupervised feature selection for multi-cluster data. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 333–342.
- Dorwin Cartwright and Terry C. Gleason. 1966. The number of paths and cycles in a digraph. *Psychometrika* 31, 2 (1966), 179–199.
- Dorwin Cartwright and Frank Harary. 1956. Structural balance: A generalization of Heider's theory. *Psychol. Rev.* 63, 5 (1956), 277.
- Yi Cen, Rentao Gu, and Yuefeng Ji. 2013. Sign inference for dynamic signed networks via dictionary learning. *J. Appl. Math.* 2013 (2013).
- Chien Chin Chen, Yu-Hao Wan, Meng-Chieh Chung, and Yu-Chun Sun. 2013b. An effective recommendation method for cold start new users using trust and distrust networks. *Inform. Sci.* 224 (2013), 19–36.
- Wei Chen, Laks V. S. Lakshmanan, and Carlos Castillo. 2013a. Information and influence propagation in social networks. *Synth. Lect. Data Manag.* 5, 4 (2013), 1–177.
- Wei Chen, Yajun Wang, and Siyu Yang. 2009. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 199–208.
- Yi Chen, Xiao-long Wang, and Bo Yuan. 2013. Overlapping community detection in signed networks. *arXiv preprint arXiv:1310.4023* (2013).
- Kai-Yang Chiang, Cho-Jui Hsieh, Nagarajan Natarajan, Ambuj Tewari, and Inderjit S. Dhillon. 2013. Prediction and clustering in signed networks: A local to global perspective. *arXiv preprint arXiv:1302.5145* (2013).
- Kai-Yang Chiang, Nagarajan Natarajan, Ambuj Tewari, and Inderjit S. Dhillon. 2011. Exploiting longer cycles for link prediction in signed networks. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. ACM, 1157–1162.
- Kai-Yang Chiang, Joyce Jiyoung Whang, and Inderjit S. Dhillon. 2012. Scalable clustering of signed networks using balance normalized cut. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. ACM, 615–624.
- Jinsook Cho. 2006. The mechanism of trust and distrust formation and their relational outcomes. *J. Retail.* 82, 1 (2006), 25–35.
- Fan Chung, Alexander Tsias, and Wensong Xu. 2013. Dirichlet pagerank and ranking algorithms based on trust and distrust. *Internet Math.* 9, 1 (2013), 113–134.
- Peter Clifford and Aidan Sudbury. 1973. A model for spatial conflict. *Biometrika* 60, 3 (1973), 581–588.
- David Cohn and Huan Chang. 2000. Learning to probabilistically identify authoritative documents. In *ICML*. Citeseer, 167–174.
- James S. Coleman. 1988. Social capital in the creation of human capital. *Am. J. Sociol.* (1988), S95–S120.
- Marek Cygan, Marcin Pilipczuk, Michał Pilipczuk, and Jakub Onufry Wojtaszczyk. 2012. Sitting closer to friends than enemies, revisited. In *Mathematical Foundations of Computer Science 2012*. Springer, 296–307.

- James A. Davis. 1967. Clustering and structural balance in graphs. *Human Relations* 20, 2 (1967), 181–187.
- Martine De Cock and Paulo Pinheiro Da Silva. 2006. A many valued representation and propagation of trust and distrust. In *Fuzzy Logic and Applications*. Springer, 114–120.
- Cristobald de Kerchove and others. 2009. *Ranking Large Networks: Leadership, Optimization and Distrust*. Ph.D. Dissertation. UCL.
- Cristobald De Kerchove and Paul Van Dooren. 2008. The PageTrust algorithm: How to rank web pages when negative links are allowed? In *SDM*. SIAM, 346–352.
- Patrick Doreian and Andrej Mrvar. 1996. A partitioning approach to structural balance. *Soc. Netw.* 18, 2 (1996), 149–168.
- Thomas DuBois, Jennifer Golbeck, and Aravind Srinivasan. 2011. Predicting trust and distrust in social networks. In *Privacy, Security, Risk and Trust (Passat), 2011 IEEE 3rd International Conference on and 2011 IEEE 3rd International Conference on Social Computing (SocialCom)*. IEEE, 418–424.
- Giuseppe Facchetti, Giovanni Iacono, and Claudio Altafini. 2011. Computing global structural balance in large-scale signed social networks. *Proc. Natl. Acad. Sci. U.S.A.* 108, 52 (2011), 20953–20958.
- Pengyi Fan, Hui Wang, Pei Li, Wei Li, and Zhihong Jiang. 2012. Analysis of opinion spreading in homogeneous networks with signed relationships. *J. Stat. Mech.: Theor. Exp.* 2012, 08 (2012), P08003.
- Rana Forsati, Mehrdad Mahdavi, Mehrnosh Shamsfard, and Mohamed Sarwat. 2014. Matrix factorization with explicit trust and distrust side information for improved social recommendation. *ACM Trans. Inform. Syst.* 32, 4 (2014), 17.
- Ove Frank and Frank Harary. 1980. Balance in stochastic signed graphs. *Soc. Netw.* 2, 2 (1980), 155–163.
- Lise Getoor and Christopher P. Diehl. 2005. Link mining: A survey. *ACM SIGKDD Explorations Newsletter* 7, 2 (2005), 3–12.
- Jacob Goldenberg, Barak Libai, and Eitan Muller. 2001a. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Market. Lett.* 12, 3 (2001), 211–223.
- Jacob Goldenberg, Barak Libai, and Eitan Muller. 2001b. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Acad. Market. Sci. Rev.* 9, 3 (2001), 1–18.
- Sergio Gómez, Pablo Jensen, and Alex Arenas. 2009. Analysis of community structure in networks of correlated data. *Phys. Rev. E* 80, 1 (2009), 016114.
- Mark Granovetter. 1978. Threshold models of collective behavior. *Am. J. Sociol.* (1978), 1420–1443.
- Ramathan Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. 2004. Propagation of trust and distrust. In *Proceedings of the 13th International Conference on World Wide Web*. ACM, 403–412.
- Frank Harary. 1959. On the measurement of structural balance. *Behav. Sci.* 4, 4 (1959), 316–323.
- Frank Harary and others. 1953. On the notion of balance of a signed graph. *Mich. Math. J.* 2, 2 (1953), 143–146.
- Frank Harary and Jerald A. Kabell. 1980. A simple algorithm to detect balance in signed graphs. *Math. Soc. Sci.* 1, 1 (1980), 131–136.
- Frank Harary and Helene J. Kimmel. 1979. Matrix measures for transitivity and balance*. *J. Math. Sociol.* 6, 2 (1979), 199–210.
- Russell Hardin. 2004. Distrust: Manifestations and management. *Distrust* 8 (2004), 3–33.
- Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev. 2012a. Detecting subgroups in online discussions by modeling positive and negative relations among participants. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, 59–70.
- Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev. 2012b. Extracting signed social networks from text. In *Workshop Proceedings of TextGraphs-7 on Graph-based Methods for Natural Language Processing*. Association for Computational Linguistics, 6–14.
- Taher H. Haveliwala. 2002. Topic-sensitive pagerank. In *Proceedings of the 11th International Conference on World Wide Web*. ACM, 517–526.
- Fritz Heider. 1946. Attitudes and cognitive organization. *J. Psychol.* 21, 1 (1946), 107–112.
- Nancy M. Henley, Robert B. Horsfall, and Clinton B. De Soto. 1969. Goodness of figure and social structure. *Psychol. Rev.* 76, 2 (1969), 194.
- Yao Ping Hou. 2005. Bounds for the least Laplacian eigenvalue of a signed graph. *Acta Math. Sin.* 21, 4 (2005), 955–960.
- Cho-Jui Hsieh, Kai-Yang Chiang, and Inderjit S. Dhillon. 2012. Low rank modeling of signed networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 507–515.

- Xia Hu, Lei Tang, Jiliang Tang, and Huan Liu. 2013. Exploiting social relations for sentiment analysis in microblogging. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*. ACM, 537–546.
- Bernardo A. Huberman, Daniel M. Romero, and Fang Wu. 2008. Social networks that matter: Twitter under the microscope. *arXiv preprint arXiv:0812.1045* (2008).
- Amin Javari and Mahdi Jalili. 2014. Cluster-based collaborative filtering for sign prediction in social networks with positive and negative links. *ACM Trans. Intell. Syst. Technol.* 5, 2 (2014), 24.
- Jonathan Q. Jiang. 2015. Stochastic blockmodel and exploratory analysis in signed networks. *arXiv preprint arXiv:1501.00594* (2015).
- Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. 2003. The eigentrust algorithm for reputation management in P2P networks. In *Proceedings of the 12th International Conference on World Wide Web*. ACM, 640–651.
- David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 137–146.
- Anne-Marie Kermarrec and Christopher Thraves. 2014. Signed graph embedding: When everybody can sit closer to friends than enemies. *arXiv preprint arXiv:1405.5023* (2014).
- Jon M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *J. ACM* 46, 5 (1999), 604–632.
- S. J. Knapskog. 1998. A metric for trusted systems. In *Proceedings of the 21st National Security Conference*. Citeseer, 16–29.
- David Knoke and Song Yang. 2008. *Social Network Analysis*. Vol. 154. Sage.
- Ling-Qi Kong and Meng-Long Yang. 2011. Improvement of clustering algorithm FEC for signed networks. *J. Comput. Appl.* 31, 5 (2011), 1395–1399.
- Jérôme Kunegis. 2014. Applications of structural balance in signed social networks. *arXiv preprint arXiv:1402.6865* (2014).
- Jérôme Kunegis, Andreas Lommatzsch, and Christian Bauchhage. 2009. The slashdot zoo: Mining a social network with negative edges. In *Proceedings of the 18th International Conference on World Wide Web*. ACM, 741–750.
- Jérôme Kunegis, Julia Preusse, and Felix Schwagereit. 2013. What is the added value of negative links in online social networks? In *Proceedings of the 22nd International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 727–736.
- Jérôme Kunegis, Stephan Schmidt, Andreas Lommatzsch, Jürgen Lerner, Ernesto William De Luca, and Sahin Albayrak. 2010. Spectral analysis of signed graphs for clustering, prediction and visualization. In *SDM*, Vol. 10. SIAM, 559–559.
- Yann Lecun, Bengio Yoshua, and Hinton Geoffrey. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.
- Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010a. Predicting positive and negative links in online social networks. In *Proceedings of the 19th International Conference on World Wide Web*. ACM, 641–650.
- Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010b. Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1361–1370.
- Dong Li, Zhi-Ming Xu, Nilanjan Chakraborty, Anika Gupta, Katia Sycara, and Sheng Li. 2014b. Polarity related influence maximization in signed social networks. *PloS One* 9, 7 (2014), e102199.
- Wei Li, Pengyi Fan, Pei Li, Hui Wang, and Yiguang Pan. 2013. An opinion spreading model in signed networks. *Mod. Phys. Lett. B* 27, 12 (2013).
- Yanhua Li, Wei Chen, Yajun Wang, and Zhi-Li Zhang. 2013. Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*. ACM, 657–666.
- Yanhua Li, Wei Chen, Yajun Wang, and Zhi-Li Zhang. 2014. Voter model on signed social networks. *Internet Math* 11, 2 (2015), 93–133.
- Yadong Li, Jing Liu, and Chenlong Liu. 2014a. A comparative analysis of evolutionary and memetic algorithms for community detection from signed social networks. *Soft Comput.* 18, 2 (2014), 329–348.
- David Liben-Nowell and Jon Kleinberg. 2007. The link-prediction problem for social networks. *J. Am. Soc. Inform. Sci. Technol.* 58, 7 (2007), 1019–1031.
- Bo Long, Zhongfei Mark Zhang, Xiaoyun Wu, and Philip S. Yu. 2006. Spectral clustering for multi-type relational data. In *Proceedings of the 23rd International Conference on Machine Learning*. ACM, 585–592.
- László Lovász. 1993. Random walks on graphs: A survey. *Combinatorics* 2, 1 (1993), 1–46.

- Hao Ma, Michael R. Lyu, and Irwin King. 2009. Learning to recommend with trust and distrust relationships. In *Proceedings of the 3rd ACM Conference on Recommender Systems*. ACM, 189–196.
- Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. 2011. Recommender systems with social regularization. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*. ACM, 287–296.
- Peter V. Marsden and Noah E. Friedkin. 1993. Network studies of social influence. *Sociol. Methods Res.* 22, 1 (1993), 127–151.
- Seth A. Marvel, Jon Kleinberg, Robert D. Kleinberg, and Steven H. Strogatz. 2011. Continuous-time model of structural balance. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5 (2011), 1771–1776.
- Mahdokht Masaeli, Jennifer G. Dy, and Glenn M. Fung. 2010. From transformation-based dimensionality reduction to feature selection. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 751–758.
- Paolo Massa and Paolo Avesani. 2005. Controversial users demand local trust metrics: An experimental study on epinions.com community. In *Proceedings of the National Conference on Artificial Intelligence*, Vol. 20. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 121.
- Robert M. May and Alun L. Lloyd. 2001. Infection dynamics on scale-free networks. *Phys. Rev. E* 64, 6 (2001), 066112.
- Miller McPherson, Lynn Smith-Lovin, and James M. Cook. 2001. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* (2001), 415–444.
- Abhinav Mishra and Arnab Bhattacharya. 2011. Finding the bias and prestige of nodes in networks based on trust scores. In *Proceedings of the 20th International Conference on World Wide Web*. ACM, 567–576.
- Michael Moore. 1978. An international application of Heider’s balance theory. *Eur. J. Soc. Psychol.* 8, 3 (1978), 401–405.
- Michael Moore. 1979. Structural balance and international relations. *Eur. J. Soc. Psychol.* 9, 3 (1979), 323–326.
- Uma Nalluri. 2014. Utility of distrust in online recommender systems. *Capstone Project Report*.
- Mark E. J. Newman and Michelle Girvan. 2004. Finding and evaluating community structure in networks. *Phys. Rev. E* 69, 2 (2004), 026113.
- Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank citation ranking: Bringing order to the web. Stanford InfoLab.
- Symeon Papadopoulos, Yiannis Kompatsiaris, Athena Vakali, and Ploutarchos Spyridonos. 2012. Community detection in social media. *Data Min. Knowl. Discov.* 24, 3 (2012), 515–554.
- Athanasios Papaioikonomou, Magdalini Kardara, Konstantinos Tserpes, and Dora Varvarigou. 2014. Edge sign prediction in social networks via frequent subgraph discovery. *IEEE Internet Comput.* (2014).
- Eduardo G. Pardo, Mauricio Soto, and Christopher Thraves. 2013. Embedding signed graphs in the line. *J. Combin. Optim.* (2013), 1–21.
- Arti Patidar, Vinti Agarwal, and K. K. Bharadwaj. 2012. Predicting friends and foes in signed networks using inductive inference and social balance theory. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM’12)*. IEEE Computer Society, 384–388.
- Clara Pizzuti. 2009. A multi-objective genetic algorithm for community detection in networks. In *Proceedings of the 21st International Conference on Tools with Artificial Intelligence (ICTAI’09)*. IEEE, 379–386.
- Xiaoguang Qi and Brian D. Davison. 2009. Web page classification: Features and algorithms. *ACM Comput. Surv.* 41, 2 (2009), 12.
- Filippo Radicchi, Daniele Vilone, and Hildegard Meyer-Ortmanns. 2007a. Universality class of triad dynamics on a triangular lattice. *Phys. Rev. E* 75, 2 (2007), 021118.
- Filippo Radicchi, Daniele Vilone, Sooyeon Yoon, and Hildegard Meyer-Ortmanns. 2007b. Social balance as a satisfiability problem of computer science. *Phys. Rev. E* 75, 2 (2007), 026106.
- Matthew Richardson, Rakesh Agrawal, and Pedro Domingos. 2003. Trust management for the semantic web. In *The Semantic Web-ISWC 2003*. Springer, 351–368.
- Thomas C. Schelling. 2006. *Micromotives and Macrobehavior*. W. W. Norton & Company.
- Matthew Schultz and Thorsten Joachims. 2004. Learning a distance metric from relative comparisons. *Advances in Neural Information Processing Systems (NIPS)* (2004), 41.
- Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. 2008. Collective classification in network data. *AI Mag.* 29, 3 (2008), 93.
- Mahsa Shafaei and Mahdi Jalili. 2014. Community structure and information cascade in signed networks. *New Gen. Comput.* 32, 3–4 (2014), 257–269.

- Moshen Shahriari and Mahdi Jalili. 2014. Ranking nodes in signed social networks. *Soc. Network Anal. Min.* 4, 1 (2014), 1–12.
- Tushar Sharma, Ankit Charls, and PramodKumar Singh. 2009. Community mining in signed social networks-an automated approach. In *ICCEA09, Manila 9* (2009).
- Hua-Wei Shen. 2013. *Community Structure: An Introduction*. Springer.
- Vikas Sindhwani, Partha Niyogi, and Mikhail Belkin. 2005. Beyond the point cloud: From transductive to semi-supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning*. ACM, 824–831.
- Michael Speriosu, Nikita Sudan, Sid Upadhyay, and Jason Baldridge. 2011. Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proceedings of the 1st Workshop on Unsupervised Learning in NLP*. Association for Computational Linguistics, 53–63.
- Nidamarthi Srinivas and Kalyanmoy Deb. 1994. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evol. Comput.* 2, 3 (1994), 221–248.
- Panagiotis Symeonidis and Nikolaos Mantas. 2013. Spectral clustering for link prediction in social networks with positive and negative links. *Soc. Network Anal. Min.* 3, 4 (2013), 1433–1447.
- Panagiotis Symeonidis and Eleftherios Tiakas. 2013. Transitive node similarity: Predicting and recommending links in signed social networks. *World Wide Web* (2013), 1–34.
- Michael Szell, Renaud Lambiotte, and Stefan Thurner. 2010. Multirelational organization of large-scale social networks in an online world. *Proc. Natl. Acad. Sci.* 107, 31 (2010), 13636–13641.
- Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. 2011. User-level sentiment analysis incorporating social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1397–1405.
- Jiliang Tang, Charu Aggarwal, and Huan Liu. 2016a. Node classification in signed social networks. In *Proceedings of the SIAM International Conference on Data Mining*.
- Jiliang Tang, Charu Aggarwal, and Huan Liu. 2016b. Recommendations in signed social networks. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 31–40.
- Jiliang Tang, Shiyu Chang, Charu Aggarwal, and Huan Liu. 2015. Negative link prediction in social media. In *Proceedings of the ACM International Conference on Web Search and Data Mining*.
- Jiliang Tang, Xia Hu, Yi Chang, and Huan Liu. 2014b. Predictability of distrust with interaction data. In *Proceedings of the ACM International Conference on Information and Knowledge Management*.
- Jiliang Tang, Xia Hu, and Huan Liu. 2013. Social recommendation: A review. *Soc. Network Anal. Min.* 3, 4 (2013), 1113–1133.
- Jiliang Tang, Xia Hu, and Huan Liu. 2014a. Is distrust the negation of trust? The value of distrust in social media. In *Proceedings of the ACM Hypertext Conference*.
- Jiliang Tang and Huan Liu. 2012. Feature selection with linked data in social media. In *SDM*. SIAM, 118–128.
- Jie Tang, Tiancheng Lou, and Jon Kleinberg. 2012. Inferring social ties across heterogenous networks. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*. ACM, 743–752.
- Jiliang Tang, Jie Tang, and Huan Liu. 2014. Recommendation in social media: Recent advances and new frontiers. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1977–1977.
- Lei Tang and Huan Liu. 2010. Community detection and mining in social media. *Synth. Lect. Data Min. Knowl. Discov.* 2, 1 (2010), 1–137.
- Howard Francis Taylor. 1970. *Balance in Small Groups*. Van Nostrand Reinhold Co.
- Evmaria Terzi and Marco Winkler. 2011. A spectral algorithm for computing social balance. In *Algorithms and Models for the Web Graph*. Springer, 1–13.
- V. A. Traag and Jeroen Bruggeman. 2009. Community detection in networks with positive and negative links. *Phys. Rev. E* 80, 3 (2009), 036115.
- Vincent Traag, Yurii Nesterov, and Paul Van Dooren. 2010. Exponential ranking: Taking into account negative links. *Soc. Inform.* (2010), 192–202.
- Patricia Victor, Chris Cornelis, Martine De Cock, and P. Pinheiro da Silva. 2006. Towards a provenance-preserving trust model in agent networks. In *WWW2006 Conference Proceedings, Special Interest Tracks, Posters and Workshops*.
- Patricia Victor, Chris Cornelis, Martine De Cock, and Ankur Teredesai. 2009. Trust-and distrust-based recommendations for controversial reviews. In *Web Science Conference (WebSci'09: Society On-Line)*.

- Patricia Victor, Nele Verbiest, Chris Cornelis, and Martine De Cock. 2013. Enhancing the trust-based recommendation process with explicit distrust. *ACM Trans. Web* 7, 2 (2013), 6.
- Erik Volz and Lauren Ancel Meyers. 2007. Susceptible–infected–recovered epidemics in dynamic contact networks. *Proc. Roy. Soc. B: Biol. Sc.* 274, 1628 (2007), 2925–2934.
- Kiri Wagstaff, Claire Cardie, Seth Rogers, Stefan Schrödl, and others. 2001. Constrained k-means clustering with background knowledge. In *ICML*, Vol. 1. 577–584.
- Jun Wang, Fuji Ren, and Lei Li. 2014. Recognizing sentiment of relations between entities in text. *IEEE Trans. Electr. Electron. Eng.* 9, 6 (2014), 614–620.
- Zhaoming Wu, Charu C. Aggarwal, and Jimeng Sun. 2016. The troll-trust model for ranking in signed networks. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, 447–456.
- Rongjing Xiang, Jennifer Neville, and Monica Rogati. 2010. Modeling relationship strength in online social networks. In *Proceedings of the 19th International Conference on World Wide Web*. ACM, 981–990.
- Bo Yang, William K. Cheung, and Jiming Liu. 2007. Community mining from signed social networks. *IEEE Trans. Knowl. Data Eng.* 19, 10 (2007), 1333–1348.
- Shuang-Hong Yang, Alexander J. Smola, Bo Long, Hongyuan Zha, and Yi Chang. 2012. Friend or frenemy? Predicting signed ties in social networks. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 555–564.
- Jihang Ye, Hong Cheng, Zhe Zhu, and Minghua Chen. 2013. Predicting positive and negative links in signed social networks by transfer learning. In *Proceedings of the 22nd International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1477–1488.
- Xiaofeng Yu and Junqing Xie. 2014a. Learning interactions for social prediction in large-scale networks. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 161–170.
- Xiaofeng Yu and Jun Qing Xie. 2014b. Modeling mutual influence between social actions and social ties. (2014).
- Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. 2014. *Social Media Mining: An Introduction*. Cambridge University Press.
- Jiawei Zhang, Qianyi Zhan, Lifang He, Charu Aggarwal, and Philip Yu. 2016. Trust hole identification in signed networks. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery*.
- Tongda Zhang, Haomiao Jiang, Zhouxiao Bao, and Yingfeng Zhang. 2013. Characterization and edge sign prediction in signed networks. *J. Industr. Intell. Inform.* 1, 1 (2013).
- Elena Zheleva and Lise Getoor. 2009. To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles. In *Proceedings of the 18th International Conference on World Wide Web*. ACM, 531–540.
- Q. Zheng and D. B. Skillicorn. 2015. Spectral embedding of signed networks. In *SDM*. SIAM.
- Xiaolong Zheng, Daniel Zeng, and Fei-Yue Wang. 2014. Social balance in signed networks. *Inform. Syst. Front.* (2014), 1–19.
- Shenghuo Zhu, Kai Yu, Yun Chi, and Yihong Gong. 2007. Combining content and link for classification using matrix factorization. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 487–494.
- Cai-Nicolas Ziegler and Georg Lausen. 2005. Propagation models for trust and distrust in social networks. *Inform. Syst. Front.* 7, 4–5 (2005), 337–358.
- Kiyana Zolfaghar and Abdollah Aghaie. 2010. Mining trust and distrust relationships in social web applications. In *Proceedings of the 2010 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*. IEEE, 73–80.

Received January 2015; revised April 2016; accepted June 2016