

# Relevance Measurements in Online Signed Social Networks

Tyler Derr

Data Science and Engineering Lab  
Michigan State University  
derrtyle@msu.edu

Suhang Wang

Computer Science and Engineering  
Arizona State University  
suhang.wang@asu.edu

Chenxing Wang

Data Science and Engineering Lab  
Michigan State University  
wangch88@msu.edu

Jiliang Tang

Data Science and Engineering Lab  
Michigan State University  
tangjili@msu.edu

## ABSTRACT

Measuring relevance for two nodes is fundamental to social network analysis, which has been proven to benefit many network analysis tasks and applications such as link prediction, node classification, community detection, search and recommendations. The majority of existing relevance measurements focused on unsigned social networks (or networks with only positive links). However, social media provides mechanisms that allow online users to specify negative links in addition to positive ones. For example, Slashdot users can create foe links; users in Epinions can establish distrust relations; while users in Facebook and Twitter can block or unfriend others. Thereby, social networks with both positive and negative links (or signed social networks) have become ubiquitous in social media and attracted increasing attention in recent years. On the one hand, it is evident from recent studies that negative links have added value in a number of analytical tasks. On the other hand, the availability of negative links challenges existing relevance measurements designed for unsigned networks. Hence, we need dedicated relevance measurements for signed social networks. In this paper, we present an initial and comprehensive investigation on signed relevance measurements and design numerous relevance measurements for signed social networks from both local and global perspectives. Empirical experiments on four real-world signed social networks demonstrate the importance of negative links in building signed relevance measurements and their effects on social network analysis tasks.

## KEYWORDS

signed networks, relevance measurements, balance theory

### ACM Reference Format:

Tyler Derr, Chenxing Wang, Suhang Wang, and Jiliang Tang. 2018. Relevance Measurements in Online Signed Social Networks. In *Proceedings of KDD 14th International Workshop on Mining and Learning with Graphs (MLG'18)*, 8 pages. [https://doi.org/10.475/123\\_4](https://doi.org/10.475/123_4)

---

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 bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
MLG'18, August 2018, London, United Kingdom  
© 2018 Copyright held by the owner/author(s).  
ACM ISBN 123-4567-24-567/08/06...\$15.00  
[https://doi.org/10.475/123\\_4](https://doi.org/10.475/123_4)

## 1 INTRODUCTION

Traditionally network analysis has focused on unsigned networks. However, many online social networking services provide mechanisms that allow users to create not only positive links, but also negative relations. These social networks with both positive and negative links are known as signed social networks, where the negative links that users give can denote their foes (e.g., Slashdot), those they distrust (e.g., Epinions), or “unfriended” friends and blocked users (e.g., Facebook and Twitter). It is due to this diverse set of signed networks appearing in today’s social media that has led to their increased attention in the recent years; as well as the increased availability due to the more and more popularity of online social media [2, 15, 23, 43].

Relevance, which measures how relevant two nodes are in a social network, is one of the keystones of social network analysis. This has been shown by their usage in diverse social network analysis tasks and applications such as link prediction [3, 45], node classification [5], community detection [38], search and recommendations [44]. The vast majority of existing node relevance measurements have been designed for unsigned networks (or social networks with only positive links) [1, 4]. However, the availability of negative links in signed networks poses tremendous challenges to unsigned relevance measurements. For instance, most unsigned relevance measurements require all links to be positive [31]. Meanwhile, the fundamental principles and theories of signed networks are substantially different from those of unsigned networks. For example, some social theories such as balance theory [14] are only applicable to signed networks, while social theories for unsigned networks such as homophily may not be applicable to signed networks [37]. Therefore, relevance measurements for signed networks need dedicated efforts since it cannot be executed by simply applying those for unsigned networks.

On the other hand, the existence of negative links also brings about unprecedented opportunities in signed relevance measurements. 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 [13, 22], and they can also boost the performance of recommender systems [27, 40]. Thereby, negative links could offer the potential to help us develop novel relevance measurements for signed networks. There are a few very

recent works in designing similarities for link prediction [17, 33]. However, a general and systematic investigation on signed relevance measurements and their effects on signed network analysis are still desired since it can greatly advance our understandings about signed social networks.

In this paper, we perform the initial and comprehensive study on the problem of measuring relevance on signed social networks. Analogous to relevance research in unsigned networks, we aim to investigate the following: (a) how to make use of both positive and negative links in signed relevance measurements; and (b) what are the effects of these measurements on signed network analysis. The main contributions of the paper are summarized as follows:

- Design numerous relevance measurements for signed social networks from both local and global perspectives;
- Investigate the connection signed relevance measurements have with balance theory and signed network properties; and
- Study the effects of signed relevance measurements with four real-world datasets on two signed network analysis tasks - sign prediction and tie strength prediction.

The rest of this paper is organized as follows. In Section 2, we first formally define the problem of signed relevance. Then, in Section 3, we review related work in relevance measurements and signed networks. We describe the four signed network datasets used in this paper, a preliminary analysis of the data, along with some validation for balance theory in Section 4. Then, in Section 5, we present numerous relevance measurements specific to signed networks. In Section 6 we perform experiments for predicting link signs and tie strength predictions when using the signed relevance algorithms previously discussed in Section 5. Finally, conclusions are given along with our future work in Section 7.

## 2 PROBLEM STATEMENT

A signed network  $\mathcal{G}$  is composed of a set of  $N$  nodes (i.e., users)  $\mathcal{U} = \{u_1, u_2, \dots, u_N\}$ , a set of positive links  $\mathcal{E}^+$  and a set of negative links  $\mathcal{E}^-$ . We represent signed links between users in an adjacency matrix,  $\mathbf{A} \in \mathbb{R}^{N \times N}$ , where  $A_{ij} = 1$  if  $u_i$  has a positive link to  $u_j$ ,  $-1$  if  $u_i$  creates a negative link to  $u_j$ , and  $0$  when  $u_i$  has no link to  $u_j$ . Furthermore, we can separate a signed network into two networks, one containing only positive links and the other with only negative links, which we can represent in the adjacency matrices  $\mathbf{A}^+ \in \mathbb{R}^{N \times N}$  and  $\mathbf{A}^- \in \mathbb{R}^{N \times N}$ , respectively. We represent a positive link from  $u_i$  to  $u_j$  with  $A_{ij}^+ = 1$  and  $A_{ij}^- = 0$  otherwise. Similarly, we represent a negative link from  $u_i$  to  $u_j$  with  $A_{ij}^- = 1$  and  $A_{ij}^+ = 0$  otherwise.

We use  $\mathbf{R} \in \mathbb{R}^{N \times N}$  to denote the relevance score matrix, where  $R_{ij}$  represents the relevance from user  $u_i$  to user  $u_j$ . Note that signed relevance values are not necessarily symmetrical. In Table 1 we summarize the major notations used in this work, where  $d_i$  and  $N_i$  denote degree and the set of neighbors of  $u_i$  in an unsigned network, respectively.

With the aforementioned notations and definitions, the problem of signed relevance is formally defined as follows:

*Given the signed network  $\mathcal{G} = (\mathcal{U}, \mathcal{E}^+, \mathcal{E}^-)$ , we aim to develop a measure  $f$  to infer relevance scores between all pairs of users (i.e.,  $\mathcal{U} \times \mathcal{U}$ ).*

**Table 1: Notations.**

Notations	Descriptions
$\mathbf{R}$	Node relevance matrix
$\mathbf{A}$	Adjacency matrix
$\mathbf{A}^+ (\mathbf{A}^-)$	Adjacency matrix of only positive (negative) links
$ \mathbf{A} $	Absolute adjacency matrix
$d_i$	Degree of node $u_i$
$N_i$	Set of neighbors for node $u_i$
$N_i^{in} (N_i^{out})$	Set of incoming (outgoing) neighbors for node $u_i$
$N_i^+ (N_i^-)$	Set of positive (negative) neighbors for node $u_i$
$s_{ij}$	Link sign between users $u_i$ and $u_j$
$\mathbf{B}_l (U_l)$	Matrix holding the number of $l$ length paths having an even (odd) number of negative links
$X_{ij}$	the $(i,j)$ entry of the matrix $\mathbf{X}$

The signed relevance problem is different but related to two signed network analysis tasks including sign prediction [22], and signed tie strength prediction [19] as :

- Sign prediction is to develop a predictor to make use of the signed network  $\mathcal{G}$  to predict the sign of links whose sign has been suppressed [22]. However, the signed relevance problem is to infer a real valued relevance score (typically in the range  $[-1, 1]$ ) for all pairs of users.
- The signed tie strength prediction is defined to infer a strength value in  $[-1, 1]$  for an existing link [19]. However, signed node relevance seeks to define a signed relevance value for all pairs of users.

Note that via relevance measurements, we can infer relevance scores of all pairs of users; hence they can be naturally used for both sign prediction and signed tie strength prediction. Actually, we evaluate the performance of various proposed relevance measurements in this paper via these two tasks.

## 3 RELATED WORK

Our work is related to node relevance measurements and signed network analysis, therefore below we briefly overview them.

**Node Relevance Measurements:** Measuring node relevance is fundamental to social network analysis. Most of existing node relevance measurements have been developed for unsigned social networks. According to the information used by a given measurement we can roughly categorize them into local and global methods. Local methods, commonly referred as structural equivalence [26], use local neighborhood information. Representative local measurements include common neighbors and its variants, Jaccard Index [30] and its variants such as Sorensen Index [32], Adamic-Adar Index [1], and Preferential Attachment Index [4]. Global methods not only utilize the local neighborhoods but also propagate the relevance information through the whole network. Representative global measurements include Katz [18], SimRank [16], ASCOS and ASCOS++ [7, 8], and random walk with restart (RWR) and its variants [39]. One recent work extends RWR for personalized ranking in signed social networks [17] and a few recent works studied node similarities for link prediction [33]. However, to the best of our knowledge, this work is the initial and comprehensive study about node relevance measurements in signed social networks.

**Signed Network Analysis:** With roots in social psychology [6, 14], signed network analysis has attracted increasing attention in

**Table 2: Statistics of four signed social networks.**

Network	# Users	# Positive Links	# Negative Links
Bitcoin-Alpha	3,784	22,651	1,556
Bitcoin-OTC	5,901	32,448	3,526
Slashdot	79,116	392,179	123,218
Epinions	131,828	717,667	123,705

recent years. However, the development of tasks of signed social network analysis is highly imbalanced [36]. Some tasks have been extensively studied such as social balance in signed networks [11, 46], link sign prediction [9, 22], and community detection [10, 21]; some tasks are still in the very early stages of development such as signed network embedding [41] and negative link prediction [35]; while others have not been comprehensively investigated such as node relevance measurements and signed network modeling. A comprehensive overview about signed network analysis can be found in [36].

## 4 DATA ANALYSIS

In this section, we will first introduce the datasets we will use for this study and then perform preliminary analysis with them.

### 4.1 Datasets

In this work, we collect four signed network datasets to study signed relevance measurements, i.e., Bitcoin-Alpha<sup>1</sup>, Bitcoin-OTC<sup>2</sup>, Slashdot<sup>3</sup> and Epinions<sup>4</sup>. Below we describe more details about these datasets.

The Bitcoin-Alpha network is a signed network we collected from Bitcoin Alpha. Similarly we collected Bitcoin-OTC from Bitcoin OTC. Both of these datasets were collected from publicly available data from their respective websites. The two Bitcoin sites are open market websites that allow users to buy and sell things. Due to the anonymity behind users' Bitcoin accounts, users of these websites form trust networks to prevent against scammers (e.g., fake users who are just attempting to have another user send them bitcoins, but never deliver their end of the deal, which is usually the delivery of some other monetary good). In addition to the signed networks, users in both websites can specify scores in the range [1,10] (or [-10,-1]) to indicate the positive (or negative) tie strength. Note that negative links in both websites are visible to the public.

The Slashdot dataset was obtained from [20]. Slashdot focuses on providing technology news since 1997. One of the unique features is that since 2002 the website has allowed users to explicitly mark other users as their friends (positive links) or foes (negative links). Note that negative links in Slashdot are only visible to users who login to the system.

We have also collected a dataset from the product review site Epinions where users can establish trust (positive) and distrust (negative) links. In addition, users can write reviews for items from certain pre-defined categories. Note that negative links in Epinions are totally invisible to the public, but in this dataset, negative links were given by Epinions staff for research purposes.

<sup>1</sup><http://www.btcalpha.com>

<sup>2</sup><https://www.bitcoin-otc.com>

<sup>3</sup><http://www.slashdot.org>

<sup>4</sup><http://www.epinions.com>

**Table 3: Reciprocal Links in Signed Social Networks.**

Datasets	Positive Links	Negative Links
Bitcoin-Alpha	85.4%	18.0%
Bitcoin-OTC	83.8%	17.8%
Slashdot	30.7%	7.4%
Epinions	34.8%	3.8%

Some statistics are demonstrated in Table 2. We note from the table that in all datasets, negative links are sparser than positive links, thus negative links could have different properties from positive links. Meanwhile, previous studies suggest that balance theory is helpful to explain social phenomena in signed networks [22]. Thus, in the following subsections, we study properties of negative links analogous to positive links and validate balance theory in four real-world signed networks.

### 4.2 Degree Distributions

As we know, the distributions of in- or out-degrees of positive links in unsigned networks follow power-law distributions – most nodes with small degrees while a few nodes with large degrees [4]. In this subsection, we examine whether similar distributions can be observed for positive and negative links in signed social networks.

For each user, we calculate the numbers of in- and out-degrees for positive and negative links, separately. We observed that the distributions of in- and out-degrees of positive and negative links in four signed networks followed a power-law degree distributions. For instance, a few nodes give a large number of negative links; while many nodes only give few negative links.

### 4.3 Reciprocal Links in Signed Social Networks

Links in directed social networks can be generally categorized into reciprocal (two-way) and parasocial (one-way) links [31]. Reciprocal links among nodes in unsigned networks are usually treated as the basis to create stable social ties and play an important role in the formation and evolution of networks [24]. In this subsection, we study reciprocal links in signed social networks.

For a pair of users  $(u_i, u_j)$ , there are four types of reciprocal links –  $(u_i + u_j, u_j + u_i)$ ,  $(u_i + u_j, u_j - u_i)$ ,  $(u_i - u_j, u_j - u_i)$  and  $(u_i - u_j, u_j + u_i)$ , where  $u_i + u_j$  (or  $u_i - u_j$ ) denotes that there is a positive link (or a negative link) from  $u_i$  to  $u_j$ . We analyzed our four signed networks and found that among four types of reciprocal links, there are few  $(u_i + u_j, u_j - u_i)$  and  $(u_i - u_j, u_j + u_i)$ . Therefore, our analysis on reciprocal links focuses on  $(u_i + u_j, u_j + u_i)$  and  $(u_i - u_j, u_j - u_i)$ . We calculate if  $u_i$  has a positive link (or a negative link) to  $u_j$ , how likely  $u_j$  also has a positive link (or a negative link) to  $u_i$ . The results on four signed networks are shown in Table 3.

From the table, we make some key observations. The percent of reciprocal positive links is much higher than that of reciprocal negative links in all four signed social networks. Though in all four websites, positive links are always visible to the public, the percent of reciprocal positive links in Bitcoin-Alpha and Bitcoin-OTC is much higher than that in Slashdot and Epinions. Users in Bitcoin Alpha and OTC exchange bitcoins with others; while users share free content (news or reviews) with others in Slashdot and Epinions. Thus, Bitcoin Alpha and OTC users need much stronger social

ties for bitcoin trading in the online worlds than users in Slashdot and Epinions who consume online free content. The percent of reciprocal negative links in Bitcoin-Alpha and Bitcoin-OTC is much higher than that in Slashdot, where the percent of reciprocal negative links in Slashdot is much higher than that in Epinions. These four websites have different access controls to negative links. In Bitcoin Alpha and OTC, negative links are totally visible to the public; only users who login to the Slashdot can see negative links; while negative links are totally private in Epinions. Exposing negative links may cause revenges that consequently could lead to more reciprocal negative links [34].

#### 4.4 Balance Theory in Signed Networks

Social theories such as homophily [28] play an important role in building node relevance measurements for unsigned social networks [25]. In this subsection, we investigate one of the most fundamental social theories related to signed social networks, i.e., balance theory [6], that could be helpful in building node relevance measurements in signed social networks.

Generally, balance theory is based on the intuition that “the friend of my friend is my friend” and “the enemy of my enemy is my friend” [6]. We adopt  $s_{ij}$  to denote the link sign between two users  $u_i$  and  $u_j$  where  $s_{ij} = 1$  (or  $s_{ij} = -1$ ) if there is a positive (or negative) link between  $u_i$  and  $u_j$ . Balance theory suggests that a cycle is balanced if there are an even number of negative links, but we typically only focus on triads (or 3-cycles) [22]. A triad of three users  $(u_i, u_j, u_k)$  is balanced if  $s_{ij} = 1$  and  $s_{jk} = 1$ , then  $s_{ik} = 1$ ; or  $s_{ij} = -1$  and  $s_{jk} = -1$ , then  $s_{ik} = 1$ . Therefore, for a triad, there are four possible sign combinations  $(+, +, +)$ ,  $(+, +, -)$ ,  $(+, -, -)$  and  $(-, -, -)$ , while only  $(+, +, +)$  and  $(+, -, -)$  are balanced. Note that balance theory is only applicable to undirected signed network, and thus we ignore the link directions when applying it to directed signed networks following the discussions in [22]. We count each of the four sign combinations and find that 92.0%, 91.5%, 94.5% and 92.4% of triads in Bitcoin-Alpha, Bitcoin-OTC, Slashdot and Epinions are balanced, respectively.

#### 4.5 Discussions

We summarize the observations from the above preliminary data analysis as – (1) properties of negative links could be different from positive links, which makes signed social networks be distinct from unsigned social networks; hence, though node relevance measurements have been extensively studied, we still need dedicated efforts to systematically investigate signed relevance measurements; and (2) most of triads in signed social networks satisfy balance theory; thus it can guide us to build advanced and novel signed relevance measurements.

### 5 SIGNED RELEVANCE MEASUREMENTS

Node relevance measurements have been extensively studied in unsigned networks. According to our preliminary data analysis in the last section, the availability of negative links makes signed networks unique in many aspects such as properties and balance theory. In this section, analogous to unsigned networks, we develop node relevance measurements for signed networks.

#### 5.1 Definitions

Many node relevance measurements have been proposed for unsigned networks. According to the information used, we can mostly divide them to local and global measurements. Local measurements only use local neighborhood information such as common neighbors; while global measurements utilize the whole structural information such as the Katz relevance. Meanwhile, node relevance measurements can be undirected and directed, corresponding to undirected and directed networks. Note that we could use any method that requires a directed network for an undirected network, since undirected networks are simply directed networks where each edge has both directions. In this work, we will group signed relevance measurements as local and global methods.

With node relevance measurements for unsigned networks, there are three strategies to design signed ones. The first is to only use  $A^+$  in the calculation of node relevance scores. This strategy completely ignores the negative links and could result in over-estimation of the impact of positive links [36]. The second strategy would be to convert negative links in the signed network into positive links, thus converting the signed network into an unsigned network. Such a network can be represented by the matrix  $\tilde{A}$  where  $\tilde{A}_{ij} = |A_{ij}|$ . Ignoring signs of links not only overlooks the differences between negative and positive links; but also eliminates the applicability of balance theory to signed networks. Our third strategy is to take advantage of negative links and balance theory to develop signed relevance measurements based on unsigned methods. In the following subsections, we will detail how to apply the third strategy to representative unsigned node relevance measurements.

#### 5.2 Local Methods

In this subsection, we build local signed relevance measurements based on representative local methods for unsigned networks including common neighbors and Jaccard Index [26, 29]. For each unsigned measurement, we will first briefly introduce it, then detail how to design the signed one and finally discuss its connection to balance theory.

##### 5.2.1 Common neighbors. Unsigned Common neighbors (UCN):

If two nodes share a lot of common friends, they are likely to be relevant to one another. Based on this intuition, UCN defines the relevance score between  $u_i$  and  $u_j$  as the number of common neighbors, which is formally defined as:

$$R_{ij} = |N_i \cap N_j|$$

where  $|x|$  denotes the size of the set  $x$ .

**Signed Common neighbors (SCN):** UCN cannot be directly applied to include negative links. Therefore, we define SCN as follows:

$$R_{ij} = (|N_i^+ \cap N_j^+| + |N_i^- \cap N_j^-|) - (|N_i^+ \cap N_j^-| + |N_i^- \cap N_j^+|)$$

We can interpret SCN as number of common neighbors of  $u_i$  and  $u_j$  where they agree on the polarity of the sign  $(|N_i^+ \cap N_j^+| + |N_i^- \cap N_j^-|)$  and then subtracting the number of neighbors that they disagree on the sign  $(|N_i^+ \cap N_j^-| + |N_i^- \cap N_j^+|)$ .

*Connection to Balance Theory:* If  $u_i$  and  $u_j$  agree with the majority of the signs of their neighbors, i.e.,  $(|N_i^+ \cap N_j^+| + |N_i^- \cap N_j^-|) >$

$(|N_i^+ \cap N_j^-| + |N_i^- \cap N_j^+|)$ , then  $R_{ij}$  is positive which will lead to more balanced triads. Otherwise, they have more disagreements on the signs, i.e.,  $(|N_i^+ \cap N_j^-| + |N_i^- \cap N_j^+|) > (|N_i^+ \cap N_j^+| + |N_i^- \cap N_j^-|)$ , then  $R_{ij}$  is negative, which will also result in more balanced triads. Therefore, SCN aims to force more triads with  $u_i$  and  $u_j$  to be balanced.

**5.2.2 Jaccard Index. Unsigned Jaccard Index (UJI):** UCN only considers the number of common neighbors of  $u_i$  and  $u_j$ , but it ignores the number of unique neighbors these two users have. Therefore, UCN is likely to give users with large numbers of neighbors high relevance scores. To mitigate such effect, UJI standardizes the UCN scores by the number of unique neighbors the two users have and defined as:

$$R_{ij} = \frac{|N_i \cap N_j|}{|N_i \cup N_j|}$$

**Signed Jaccard Index (SJI):** Similar to how UJI is defined from UCN, SJI is defined as SCN divided by the total number of unique neighbors  $u_i$  and  $u_j$  have:

$$R_{ij} = \frac{SCN_{ij}}{|N_i^+ \cup N_i^- \cup N_j^+ \cup N_j^-|}$$

*Connection to Balance Theory:* Similar to SCN, SJI targets to force more triads to be balanced.

### 5.3 Global Methods

The global methods make use of not only the local neighborhoods, but also allow for the propagation of relevance information to pass through the whole network. Most of the global methods for unsigned networks assume that two users  $u_i$  and  $u_j$  should have high relevance if they have neighbors that have high relevance. In this subsection, we detail how to design global signed relevance measurements based on representative unsigned ones and then connect them to balance theory.

**5.3.1 Katz. Unsigned Katz (UK):** This method sums over the collection of all paths from  $u_i$  to  $u_j$  and has an exponential decay on the weight associated with the count of paths as the length increases [18]:

$$R_{ij} = \sum_{l=1}^{\infty} \beta^l \cdot |\text{paths}_{i,j}^l| = \sum_{l=1}^{\infty} \beta^l [A^l]_{ij} \quad (1)$$

where  $|\text{paths}_{i,j}^l|$  is the count of paths of length  $l$  from  $u_i$  to  $u_j$ . Note that we should have  $\beta < 1$  so that longer paths will be assigned less weight than shorter paths. This can be formulated recursively to handle the counting of paths of varying length as follows:

$$R_{ij} = \frac{\beta}{d_i} \sum_{k=1}^N A_{ik} R_{kj} + \delta_{ij}$$

Note that  $\delta_{ij}$  is used to ensure that every node in the network has a high relevance to themselves (i.e., "self-similarity") and is a diagonal term defined as  $\delta = I$ . The relevance scores are also normalized from each user  $u_i$  based on their degree  $d_i$ .

**Signed Katz (SK):** Balance theory states that a  $k$ -cycle in a signed social network is balanced if it contains an even number of negative edges and unbalanced if it contains an odd number of

negative edges. With relevance scores from SK, we expect more balanced  $k$ -cycles than unbalanced ones involving users  $u_i$  and  $u_j$ . To achieve this, we would therefore need to choose the sign of the node relevance  $R_{ij}$  to be either positive or negative, such that it optimizes over all the cycles involving  $u_i$  and  $u_j$  (i.e., all the paths between  $u_i$  and  $u_j$ ). We let  $B_l$  and  $U_l$  denote the matrices that hold the number of paths with an even and odd number of negative links in paths of length  $l$ , respectively. As done in UK, we also can similarly allow the decay of importance on the longer paths. Our formulation with its recurrence relation for the calculation of paths of length  $l$  having an even or odd number of negative edges is the following:

$$R = \sum_{l=1}^Y \beta^l f(B_l, U_l) \quad (2)$$

where

$$\begin{aligned} B_l &= B_{l-1}A^+ + U_{l-1}A^-, \\ U_l &= B_{l-1}A^- + U_{l-1}A^+, \\ B_1 &= A^+, \quad U_1 = A^- \end{aligned}$$

and  $f(B_l, U_l)$  is a function to combine the counts of paths with even and odd number of negative links. Next we will discuss the inner working of SK. When counting paths of length 1 (i.e., a direct edge connecting the two nodes), we set  $B_1$  to be  $A^+$  since having a positive edge is trivially having an even number of negative links in a path of length 1, and similarly reasoned for initializing  $U_1$  to be  $A^-$ . We assume that  $B_{l-1}$  and  $U_{l-1}$  represent the paths of length  $l-1$  having an even and odd number of negative edges, respectively, between the path of nodes. Adding one positive link ( $A^+$ ) to a path in  $B_{l-1}$  or adding a negative link ( $A^-$ ) to a path in  $U_{l-1}$  will result in a path of length  $l$  with an even number of negative links. This then leads to the recursively definition of  $B_l = B_{l-1}A^+ + U_{l-1}A^-$ . Similarly, we can obtain the definition of  $U_l = B_{l-1}A^- + U_{l-1}A^+$ .

**THEOREM 5.1.** *When we choose  $f(B_l, U_l) = (B_l - U_l)$  and  $A \in \mathbb{R}^{N \times N}$ , where  $A_{ij} = 1$  if  $u_i$  has a positive link to  $u_j$ ,  $-1$  if  $u_i$  creates a negative link to  $u_j$ , and 0 when  $u_i$  has no link to  $u_j$ , signed Katz in Eq. 2 is equivalent to applying unsigned Katz in Eq. 1 on the signed network adjacency matrix defined as  $A$ .*

**PROOF.** To prove the theorem, we only need to show that:  $B_l - U_l = A^l$ . We use mathematical induction as follows:

**Basis:** Let  $l = 1$ , based on our definition of  $B_1$  and  $U_1$ , we have  $(B_1 - U_1) = (A^+ - A^-) = A^1$ .

**Inductive Hypothesis:** Suppose the theorem holds for  $l = k$ . In other words,  $(B_k - U_k) = A^k$ .

**Inductive Step:** Let  $l = k + 1$ . Then our left side is  $(B_{k+1} - U_{k+1}) = ((B_k A^+ + U_k A^-) - (B_k A^- + U_k A^+)) = (B_k - U_k)(A^+ - A^-) = A^k(A) = A^{k+1}$ , which completes the proof.  $\square$

*Connection to Balance Theory:* SK is built based on balance theory. SCN and SJI are designed to force more balanced triads (or 3-cycles), while SK further pushes for more  $l$ -cycle to be balanced. If the majority of paths between  $u_i$  and  $u_j$  (taking into account the exponential decay based on path length) have an even number of

negative links, according to balance theory, we should have a positive node relevance between them. Similarly, when having an odd number of negative edges, we want to have a negative relevance. Therefore, if we count the number of paths between  $u_i$  and  $u_j$  with an even or odd number of negative edges, then we can subtract the number with an odd number of negative links from the number of paths having an even number of links, since this will give us the optimal choice of sign between  $u_i$  and  $u_j$  as mentioned above. More specifically, if the resulting value is positive, the node relevance between  $u_i$  and  $u_j$  is positive, otherwise negative.

### 5.3.2 Asymmetric Similarity Measure for Weighted Networks.

**Unsigned Asymmetric Similarity Measure for Weighted Networks (UASCOS++):** This method is an enrichment of the ASCOS [7] (that itself is an improved variant of SimRank [16] to correctly include both even and odd length paths) to handle weighted networks. The formulation of ASCOS is the following:

$$\mathbf{R}_{ij} = \begin{cases} \frac{c}{|N_i^{in}|} \sum_{k \in N_i^{in}} \mathbf{R}_{kj} & i \neq j \\ 1 & i = j \end{cases}$$

Let  $\overline{\mathbf{P}}_{ij} = \frac{A_{ij}}{d_i^{in}}$  and we can rewrite the formulation as:

$$\mathbf{R} = c\mathbf{P}^T\mathbf{R} + (1 - c)\mathbf{I}$$

It defines the node relevance as the summation of normalized relevance from the incoming neighbors  $u_k$  of  $u_i$  to  $u_j$ . The modifications for ASCOS++ were performed to handle weights on the edges and the formulation is shown below:

$$\mathbf{R}_{ij} = \begin{cases} c \sum_{k \in N_i^{in}} \frac{A_{ik}}{\sum_{q \in N_i^{in}} A_{iq}} (1 - e^{-A_{ik}}) \mathbf{R}_{kj} & i \neq j \\ 1 & i = j \end{cases} \quad (3)$$

The adjustment is that they now normalize each of the edge weights coming into  $i$  by the summation of all the incoming weights into  $i$ . The term  $(1 - e^{-A_{ik}})$  maps the weights to be close to 1 when edge weights are large, and when the weights are small, it maps them close to 0.

**Signed ASCOS++ (SASCOS++):** ASCOS++ has difficulties to directly adapt to signed networks. Assume that a user  $u_i$  has an even number of incoming edges, where half the edges are positive, while the other half are negative. Therefore, this would lead to an undefined value as the summation over all incoming edges to  $u_i$   $\sum_{q \in N_i^{in}} A_{iq}$  is zero.

Another issue with directly applying ASCOS++ (i.e., Eq. 3) is that the resulting relevance scores could contradict with balance theory. To ease our analysis in the following case, let  $\kappa = \sum_{q \in N_i^{in}} A_{iq}$ ,  $\lambda = \frac{A_{ik}}{\kappa}$  and  $\mu = (1 - e^{-A_{ik}})$ . If  $A_{ik} = 1$  and  $\kappa$  is negative, then  $\lambda$  is negative and  $\mu$  is positive. Thus, if  $\mathbf{R}_{kj}$  is also positive, then the product of these three terms (i.e.,  $\lambda, \mu, \mathbf{R}_{kj}$ ) results in  $\mathbf{R}_{ij}$  being negative. Thus the resulting triad between users  $u_i, u_k$ , and  $u_j$  would be  $(+, +, -)$  and does not follow balance theory. Similarly, when  $\mathbf{R}_{kj}$  is negative, the product is positive and the resulting triad  $(+, -, +)$  is also not balanced.

Due to the fact using ASCOS++ with signed networks could inherently disagree with balance theory, this motivates us to build

SASCOS++. We note that when using ASCOS++ with signed networks,  $\mu$  is equal to approximately 0.63 and -1.72 when  $A_{ik}$  is positive or negative, respectively. Thus, it is providing a stronger push in the similarity (by roughly three times) when seeing a negative link. Due to the imbalance of positive and negative links in signed networks, we leave the  $\mu$  term unchanged, but make a change to the normalization (i.e.,  $\kappa$ ). The formulation for SASCOS++ is shown below:

$$\mathbf{R}_{ij} = \begin{cases} c \sum_{k \in N_i^{in}} \frac{A_{ik}}{\sum_{q \in N_i^{in}} |A_{iq}|} (1 - e^{-A_{ik}}) \mathbf{R}_{kj} & i \neq j \\ 1 & i = j \end{cases}$$

*Connection to Balance Theory:* It is easy to verify that SASCOS++ defines a signed relevance measurements aligning with balance theory. In other words, it will push more balanced triads between user relevance scores.

## 6 EXPERIMENTS

We can evaluate the quality of relevance measurements in two ways – (1) explicitly applying them in signed network analysis tasks and (2) implicitly using them to boost signed network analysis tasks. In this work, we assess relevance measurements explicitly by applying them in sign prediction and tie strength prediction. *Note that since the purpose of this paper is not to develop state-of-the-art sign prediction and tie strength prediction algorithms; hence we do not compare relevance measurements with existing sign prediction and tie strength algorithms.* However, to give context to the predictions, we include a baseline to each task that predicts at random. As mentioned in the last section, we can have three strategies to adapt unsigned measurements for signed networks – (1) removing negative links; (2) ignoring signs; and (3) building advanced signed versions based on signed network properties and balance theory. In the following subsections, given an unsigned measurement “X”, we use “X-R” and “X-I” to denote the corresponding measurements applicable to signed networks by removing negative links and ignoring signs, respectively. For example, “UCN-R” and “UCN-I” denote the strategies of adapting “UCN” to signed networks by removing negative links and ignoring signs, separately. The first question we want to answer is – which strategy leads to better measurements. We have built numerous local and global measurements. The second question is – how they perform in terms of different tasks.

For each of the parameterized measurements, we performed cross validation for the parameter tuning for each of the tasks. Note that we have filtered the Epinions and Slashdot datasets to being 15,108 and 16,070 users, 78,851 and 80,440 positive edges, 10,951 and 26,600 negative edges, respectively. Among measurements discussed in the last section, common neighbor (CN), Jaccard Index (JI), and Katz (K) measurements are designed for undirected networks; while ASCOS++ is for directed networks. As mentioned before directed measurements can be naturally applied to undirected ones by considering one undirected link as two directed links. Therefore, we conduct experiments in the undirected setting.

### 6.1 Sign Prediction

For each dataset, we randomly choose 80% as training, and the remaining as testing. We perform relevance measurements on the

**Table 4: Relevance performance in terms of sign prediction.**

Measurements	Bitcoin-Alpha	Bitcoin-OTC	Slashdot	Epinions
UCN-R	0.500	0.523	0.520	0.520
UCN-I	0.501	0.497	0.508	0.508
SCN	0.671	0.716	0.549	0.629
UJI-R	0.499	0.524	0.513	0.522
UJI-I	0.497	0.489	0.503	0.512
SJI	0.669	0.725	0.550	0.630
UK-R	0.517	0.587	0.542	0.560
UK-I	0.488	0.482	0.498	0.538
SK	0.730	0.766	<b>0.693</b>	0.702
UASCOS++-R	0.530	0.603	0.554	0.573
UASCOS+++-I	0.496	0.484	0.497	0.537
SASCOS++	<b>0.765</b>	<b>0.774</b>	0.663	<b>0.705</b>

**Table 5: Relevance performance in terms of tie strength prediction.**

Measurements	Bitcoin-Alpha	Bitcoin-OTC
UCN-R	0.286	0.324
UCN-I	0.291	0.332
SCN	<b>0.277</b>	<b>0.308</b>
UJI-R	0.286	0.324
UJI-I	0.291	0.332
SJI	<b>0.277</b>	<b>0.308</b>
UK-R	0.290	0.326
UK-I	0.295	0.333
SK	0.284	0.320
UASCOS++-R	0.292	0.328
UASCOS+++-I	0.302	0.345
SASCOS++	0.299	0.334
Random	0.648	0.664

training set to get the relevance scores for each pair of users. The signed specific measurements can obtain a relevance score from  $[-1, 1]$ ; hence we directly use the sign of the relevance score to indicate the sign of links. For “X-R” and “X-I”, the relevance score is in  $[0, 1]$ . From the training data, we search an optimal threshold, and then if the relevance score is less than the threshold, we predict a negative link and positive otherwise. Since positive and negative links are usually imbalanced in real-world signed networks, we use Area Under the receiver operating characteristics Curve (AUC) as the metric to assess the performance of link sign prediction. We note that the baseline of random guessing edge signs would result in an AUC of 0.5.

The sign prediction comparison results are shown in Table 4. We note that signed specific relevance measurements perform much better than these that (1) remove negative links and (2) ignore signs. These results suggest the importance of negative links in building node relevance measurements for signed networks. Meanwhile, global signed measurements consistently obtain better sign prediction performance than local signed measurements. We note that global methods consider long circles; while local methods only consider triads. This observation is consistent with that in [9] – long circles contain rich information in helping predict the signs of links.

## 6.2 Tie Strength Prediction

We have only used the two Bitcoin datasets (Bitcoin-Alpha and Bitcoin-OTC) for this task as they are the only two of the four datasets that have a ground truth strength associated with each edge in the network. Note that we have normalized the two datasets to have their strength in the range  $[-1, 1]$  to ensure easy mappings from our presented relevance measurements to the tie strengths associated with these datasets edges.

We directly use the relevance scores of signed specific measurements as the predicted tie strength. While for “X-R” and “X-I”, we use the similar strategy as sign prediction for tie strength prediction – we search an optimal threshold from the training data to map the relevance scores to  $[-1, 1]$ . We provide the entire binary network as input and then attempt to predict the tie strength associated with each edge of the network. Therefore, we use root-mean-square error (RMSE) as the metric to evaluate the performance of tie strength prediction.

The tie strength prediction performance is demonstrated in Table 5. The first observation is that the random tie strength prediction of picking values uniformly in the range  $[-1, 1]$  results in the worst performance. Now, given the context of the random baseline performance, we further discuss the results of the relevance measurements. We note that most of the time, signed specific measurements outperform these that (1) remove negative links or (2) ignore signs for tie strength prediction. The overall best measurement in each dataset was a signed specific measurement. This further supports the importance of negative links in signed relevance measurements. Meanwhile, local signed measurements obtain comparable or even better performance than global signed measurements in tie strength prediction. This observation is different from that of sign prediction. To achieve better sign prediction performance, we only need to predict the sign accurately. However, for tie strength prediction, in addition to signs of links, we also need to predict the strength of the relevance correctly. Thus, local information could be good at predicting relevance strength. In fact, most existing tie strength prediction algorithms for unsigned networks only use local information [12, 42].

## 7 CONCLUSION

Relevance measurements have been extensively studied for unsigned social networks. In recent years, signed network analysis has attracted increasing attention. However, as a fundamental task, relevance measurements are rather limited. In this paper, we offer an initial and comprehensive study on signed relevance measurements. We build multiple local and global measurements guided by signed network properties and balance theory. We further study the impact of signed relevance measurements on two signed network analysis tasks, i.e., sign prediction and tie strength prediction. Experimental results demonstrate that dedicated efforts are necessary to build signed relevance measurements with negative links. We will further investigate the following directions. First, we would like to study other social theories for signed networks and build novel relevance measurements based on them. Second, we will study the impact of signed relevance measurements on more signed network analysis tasks.

## REFERENCES

- [1] Lada A Adamic and Eytan Adar. 2003. Friends and neighbors on the web. *Social networks* 25, 3 (2003), 211–230.
- [2] Pranay Anchuri and Malik Magdon-Ismael. 2012. Communities and balance in signed networks: A spectral approach. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*. IEEE, 235–242.
- [3] Lars Backstrom and Jure Leskovec. 2011. Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 635–644.
- [4] Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *science* 286, 5439 (1999), 509–512.
- [5] Smriti Bhagat, Graham Cormode, and S Muthukrishnan. 2011. Node classification in social networks. In *Social network data analytics*. Springer, 115–148.
- [6] Dorwin Cartwright and Frank Harary. 1956. Structural balance: a generalization of Heider's theory. *Psychological review* 63, 5 (1956), 277.
- [7] Hung-Hsuan Chen and C Lee Giles. 2013. ASCOS: an asymmetric network structure context similarity measure. In *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*. IEEE, 442–449.
- [8] Hung-Hsuan Chen and C Lee Giles. 2015. Ascosp++: An asymmetric similarity measure for weighted networks to address the problem of simrank. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 10, 2 (2015), 15.
- [9] 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.
- [10] 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.
- [11] Giuseppe Facchetti, Giovanni Iacono, and Claudio Altafini. 2011. Computing global structural balance in large-scale signed social networks. *Proceedings of the National Academy of Sciences* 108, 52 (2011), 20953–20958.
- [12] Eric Gilbert and Karrie Karahalios. 2009. Predicting tie strength with social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 211–220.
- [13] Ramanathan 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.
- [14] Fritz Heider. 1946. Attitudes and cognitive organization. *The Journal of psychology* 21, 1 (1946), 107–112.
- [15] 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.
- [16] Glen Jeh and Jennifer Widom. 2002. SimRank: a measure of structural-context similarity. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 538–543.
- [17] Jinhong Jung, Woojeong Jin, Lee Sael, and U. Kang. 2016. Personalized Ranking in Signed Networks Using Signed Random Walk with Restart. In *IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain*. 973–978. <https://doi.org/10.1109/ICDM.2016.0122>
- [18] Leo Katz. 1953. A new status index derived from sociometric analysis. *Psychometrika* 18, 1 (1953), 39–43.
- [19] Srijan Kumar, Francesca Spezzano, VS Subrahmanian, and Christos Faloutsos. 2016. Edge weight prediction in weighted signed networks. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*. IEEE, 221–230.
- [20] Jérôme Kunegis, Andreas Lommatzsch, and Christian Bauckhage. 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.
- [21] Jérôme Kunegis, Stephan Schmidt, Andreas Lommatzsch, Jürgen Lerner, Ernesto W De Luca, and Sahin Albayrak. 2010. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*. SIAM, 559–570.
- [22] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Predicting positive and negative links in online social networks. In *Proceedings of the 19th international conference on World wide web*. ACM, 641–650.
- [23] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Signed networks in social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1361–1370.
- [24] Jundong Li, Jiliang Tang, Yilin Wang, Yali Wan, Yi Chang, and Huan Liu. 2017. Understanding and Predicting Delay in Reciprocal Relations. *arXiv preprint arXiv:1703.01393* (2017).
- [25] David Liben-Nowell and Jon Kleinberg. 2007. The link-prediction problem for social networks. *journal of the Association for Information Science and Technology* 58, 7 (2007), 1019–1031.
- [26] Francois Lorrain and Harrison C White. 1971. Structural equivalence of individuals in social networks. *The Journal of mathematical sociology* 1, 1 (1971), 49–80.
- [27] Hao Ma, Michael R Lyu, and Irwin King. 2009. Learning to recommend with trust and distrust relationships. In *Proceedings of the third ACM conference on Recommender systems*. ACM, 189–196.
- [28] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* 27, 1 (2001), 415–444.
- [29] Mark Newman. 2010. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA.
- [30] Gerard Salton and Michael J McGill. 1986. *Introduction to modern information retrieval*. (1986).
- [31] John Scott. 2012. *Social network analysis*. Sage.
- [32] Thorvald Sorensen. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr.* 5 (1948), 1–34.
- [33] Panagiotis Symeonidis and Eleftherios Tiakas. 2014. Transitive node similarity: predicting and recommending links in signed social networks. *World Wide Web* 17, 4 (2014), 743–776.
- [34] Michael Szell, Renaud Lambiotte, and Stefan Thurner. 2010. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences* 107, 31 (2010), 13636–13641.
- [35] Jiliang Tang, Shiyu Chang, Charu Aggarwal, and Huan Liu. 2015. Negative link prediction in social media. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 87–96.
- [36] Jiliang Tang, Yi Chang, Charu Aggarwal, and Huan Liu. 2016. A survey of signed network mining in social media. *ACM Computing Surveys (CSUR)* 49, 3 (2016), 42.
- [37] Jiliang Tang, Xia Hu, and Huan Liu. 2014. Is distrust the negation of trust?: the value of distrust in social media. In *Proceedings of the 25th ACM conference on Hypertext and social media*. ACM, 148–157.
- [38] Lei Tang and Huan Liu. 2010. Community detection and mining in social media. *Synthesis Lectures on Data Mining and Knowledge Discovery* 2, 1 (2010), 1–137.
- [39] Hanghang Tong, Christos Faloutsos, and Jia-yu Pan. 2006. Fast Random Walk with Restart and Its Applications. In *Data Mining, 2006. ICDM'06. Sixth International Conference on*. IEEE, 613–622.
- [40] 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)*.
- [41] Suhang Wang, Jiliang Tang, Charu Aggarwal, Yi Chang, and Huan Liu. 2017. Signed network embedding in social media. SDM.
- [42] 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.
- [43] Bo Yang, William Cheung, and Jiming Liu. 2007. Community mining from signed social networks. *IEEE transactions on knowledge and data engineering* 19, 10 (2007).
- [44] Hongzhi Yin, Bin Cui, Jing Li, Junjie Yao, and Chen Chen. 2012. Challenging the Long Tail Recommendation. *Proc. VLDB Endow.* 5, 9 (May 2012), 896–907. <https://doi.org/10.14778/2311906.2311916>
- [45] Zhijun Yin, Manish Gupta, Tim Weninger, and Jiawei Han. 2010. A unified framework for link recommendation using random walks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on*. IEEE, 152–159.
- [46] Xiaolong Zheng, Daniel Zeng, and Fei-Yue Wang. 2015. Social balance in signed networks. *Information Systems Frontiers* 17, 5 (2015), 1077–1095.