Extended Structural Balance Theory for Modeling Trust in Social Networks

Yi Qian
Department of Computer Science
Rensselaer Polytechnic Institute
Email: qiany3@rpi.edu

Sibel Adalı
Department of Computer Science
Rensselaer Polytechnic Institute
Email: adalis@rpi.edu

***Abstract*—Modeling trust in very large social networks is a hard problem due to the highly noisy nature of these networks that span trust relationships from many different contexts, based on judgments of reliability, dependability and competence and the relationships vary in their level of strength. In this paper, we introduce a new extended balance theory as a foundational theory of trust in networks. Our theory preserves the distinctions between trust and distrust as suggested in the literature, but also incorporates the notion of relationship strength which can be expressed as either discrete categorical values, as pairwise comparisons or as metric distances. Our model is novel, has sound social and psychological basis, and captures the classical balance theory as a special case. We then propose a convergence model, describing how an imbalanced network evolves towards new balance and formulate the convergence problem of a social network as a Metric Multidimensional Scaling (MDS) optimization problem. Finally, we show how the convergence model can be used to predict edge signs in social networks, and justify our theory through experiments on real datasets.**

I. INTRODUCTION

Modeling trust in very large social networks is a hard problem due to the highly noisy nature of these networks. These networks span trust relationships from many different contexts, based on judgments of reliability, dependability and competence [1]. Furthermore, trust relationships vary in their level of strength as participants may or may not know each other well. One specific problem of interest is inferring new relationships and making recommendations based on existing relationships. To this date, many algorithms for such prediction problems are based on machine learning methods. These methods generate multiple structural features from the known relationships, and then use these features to classify the unknown relationships [2][3][4]. In particular, the concept of structural balance is widely applied when developing these algorithms. For example, Leskovec et al. generate a class of "triad features" in their prediction algorithm, i.e. a pair of relationships constraining a third relationship [4]. They also use the structural balance as a touchstone to see the congruence between their practical results and the long studied theory. In essence, the transitivity of trust and distrust within a triad can be easily inferred from the balance theory. Other algorithms [2] make use of similar assumptions without explicit mention of balance theory in which trust and distrust relationships are mapped to metric distances on a continuous range. While these algorithms show strong prediction performance, there is no underlying theory that explains how structural balance can be mapped to metric distances. That is, a theory that stands as the cornerstone of trust prediction is missing. Without such a theory, it is not possible to develop principled algorithms and study to which degree a specific network conforms to the theory, and to illustrate the importance of various balance related properties of networks. Furthermore, in today's networks in which individuals interact with a large number of others, one expects that the relationships and their trust level will vary considerably. As a result, it is not clear how Cartwright-Harary's balance theory [5] can be applied to large scale social networks in which nodes may have different types of trust and distrust relationships, ranging from strong to weak.

In this paper, we address all these concerns and make the following unique contributions. First, we introduce a new extended balance theory that allows arbitrary relationship strengths. We express balance with two simple principles that preserve the meanings of trust and distrust: positive and negative trust relationships. We show how balance can be modeled when strength of relationships are expressed as either discrete categorical values, pairwise comparisons or metric distances using the same principles. Our novel method has sound social and psychological basis, and captures the classical balance theory as a special case. Next, we propose a convergence model, describing how an imbalanced network evolves towards new balance. The assumption behind our model is that in resolving tensions within imbalanced relationships, people tend to avoid the effort of changing relationships if possible. The introduction of extended balance theory allows us to formulate the convergence problem of a social network as a Metric Multidimensional Scaling (MDS) optimization problem. Finally, we show how the convergence model can be used to predict edge signs in social networks, and justify our theory through experiments on real datasets. Stress majorization is applied to solve MDS [6], and our method consistently matches and outperforms the state of the art. In addition, we show promising results towards providing solutions for the harder *link prediction problem* [7].

II. STRUCTURAL BALANCE THEORY & RELATED WORK

When modeling relationships between pairs of individuals, positive relationships are representative of liking, loving,

This research is supported through the authors' participation in the Network Science Collaborative Technology Alliance (NS-CTA) sponsored by the U.S. Army Research Laboratory under Agreement Number W911NF-09-2-0053. The content of this paper does not necessarily reflect the position or policy of the U.S. Government, no official endorsement should be inferred or implied. We also would like to thank T. DuBois for sharing the implementation details of his algorithm.

valuing or approving someone, and negative relationships are representative of disvaluing, disapproving or negatively valuing someone [5]. Suppose we consider trust from the perspective of trustworthiness where one trusts another if they are considered to be truthful, to have integrity and to have positive intentions towards the other. This definition of trustworthiness has a definite affective component that can easily be considered as an extension of the above definition. A positive relationship results in trust because the other person is considered to have positive values from the perspective of the trustor. Similarly, a negative relationship results in distrust because the other person is considered to have specific faults that would prevent them from being trusted [1].

Trust and distrust are not symmetric constructs. A person who is not trusted may eventually become trusted as a result of positive evidence. However, a distrusted person may not be trusted even after many positive experiences. As a result, one has to treat both types of relationships differently. Note that often trust and distrust relationships need not be mutual: A may trust B, but B may distrust A. However, we will treat all trust relationships mutual in this paper as in all the methods we discuss in this section and leave the study of the one-sided or mutual (cooperative) relationships to future work.

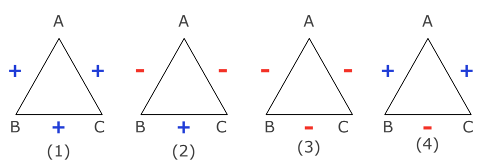

Fig. 1. Classic structural balance with four different types of triads

Structural balance theory (SBT) is based on the assumption that certain types of relationships when viewed from a local perspective are more natural for psychological reasons [8]. The local level is defined as a triangle or triad consisting of relationships between three people. It is natural for three people to be friends: Alice (A) is friends with Bob (B), Bob is friends with Chris (C) and Chris is friends with Alice. This triangle (marked (1) in Figure 1) is considered completely positive. Similarly, a relation where two friends, Bob and Chris have a common enemy Alice, is also natural (triangle (2) in Figure 1). It is considered that in such natural relationships, there is no tension in the interactions. However, it is less natural for Alice and Bob, and Alice and Chris to be friends, but Bob and Chris to be enemies (triangle (4) in Figure 1). This situation is likely to generate tension as now Alice must avoid spending time with Bob when Chris is around. As a result, all definitions of structural balance would consider triangles 1 and 2 as balanced, and 4 as imbalanced.

The last type of relationship is one containing all negative edges. This type of relationship can be considered balanced as there is no specific conflict when three people all dislike each other but spend no time together. As a result, in Davis's weak structural balance theory (WSBT), triangle (3) is considered balanced [8][9]. However, there is an opportunity for one of the pairs in this triangle to become friends, and team up against the common enemy. For this reason, (strong) SBT, considers triangle (3) unbalanced as well [5][8].

A complete network in which all pairs of people are connected to each other satisfies the WSBT if all triangles in it are balanced with respect to WSBT. In this case, the network can be divided into a set of communities $C_1, \ldots, C_k$ such that within each community, nodes are positively connected, and across different communities, all connections are negative. This main result underlies many trust inference algorithms. Algorithms that attempt to solve the *edge sign prediction problem* which involves assessing which links in a network are positive and which links are negative, are often based on SBT/WSBT either implicitly or explicitly.

Guha et. al. [3] introduce one of the earliest methods that addresses the propagation of both trust and distrust. They propose the concepts of direct propagation, co-citation and backwards propagation, and compute trust propagation by repeating matrix operations that combine the three types of propagations. They report an overall 85% prediction accuracy over data samples from Epinion that has equal number of positive and negative edges. Leskovec, Huttenlocher and Kleinberg [4] conduct a series of experiments on three large datasets: Epinions, Slashdot and Wikipedia. In particular, they collect two classes of features, one of which is based on degree and the other is based on triads. These relatively local features form a high dimensional space on which they perform standard machine learning methods and perform edge sign predictions. Similar to our work, they interpret some of their results in terms of the classical balance theory [5], but unlike our work, they do not use balance theory as a starting point.

The recent work by DuBois et. al. [2] is also related to our paper from an algorithmic point of view. This work stands out as it provides very good prediction performance for the edge sign prediction problem; $80-90\%$ on all of the three datasets used in [4] for both positive and negative edges. In this work, two features are computed for each signed edge: the first one is based on path probability (PP, $O(n^2)$) [10] and the second one uses a force directed algorithm (FD, $O(kn)$ at each iteration where k is the average degree of the network) [2]. In the method of FD, the authors map trust and distrust relationships to metric distances: the larger the distance is, the more negative (less positive) the relation is. Then, the edge sign prediction problem is mapped to a graph drawing problem. In this paper, we show that some of the assumptions underlying this algorithm can be formally defined as part of a general structural balance theory that not only works for simple positive and negative edges, but also takes into account trust strength when applicable. Being able to deal with strength also enables us to state the explicit optimization criteria in the metric space for the graph drawing problem. As a result, we are able to compare the prediction performance with respect to an optimal placement of nodes according to our theory. We show our method matches and outperforms the results of this method for different data sets.

Notice that both the force directed algorithm (FD) from [2] and the stress majorization (SM) that we use in this paper have been used in the field of graph drawing [6]. In FD, an attractive force is assigned between endpoints of each positive edge and a repelling force is assigned between endpoints of each negative edge. Nodes are initially randomly laid out, and the system is simulated until a stable equilibrium is reached when the total kinetic energy is below certain threshold. The relation between every pair of nodes is represented by the distance between the two end nodes in the stable layout of the network.

While FD is simple to implement, it operates on a local pairwise level, instead of a global level. This can lead to problems if the local forces end up not being sufficient to hold small groups together. Alternatively, if negative forces are too high, then the network may continuously expand in space and the algorithm may never converge. As a result, FD requires carefully tuned parameters for a specific network. In contrast, SM is a mature approach that guarantees monotonic convergence for drawing graphs. Moreover, in [2], there is no force between pairs of unconnected nodes which can result in unintuitive distances for such pairs. In fact as we show in our results, the FD method maps unconnected nodes to a predominantly positive range. This presents a problem for using this algorithm for solving the *link prediction problem* [7]. Link prediction is a harder problem since networks are often sparse and one needs to find the few edges that are true positive with high probability. We show that our results are very promising on this front.

In addition, to the best of our knowledge, none the existing methods provide a principled way to study the principles underlying trust and distrust relationships in very large networks with varying degrees of relationship strengths. It is unclear to which degree SBT or WSBT balance theory is valid for many large networks in which some or most relationships are simple acquaintances [11] instead close friendship relationships. An acquaintance may not result in the same type of structural constraints. For example, if Alice knows Bob, and Bob knows Chris, but Chris dislikes Alice, this may not cause much stress in the existing relationships if Alice, Bob and Chris rarely spend time with each other, i.e. their relationship is not strong. However, there are still some implications for the network overall when we consider acquaintances as well as friendships. We examine those in the next sections and provide a flexible theory of balance that generalizes WSBT. We show that our theory allows us to formulate convergence as an optimization problem, which can be solved by stress majorization and illustrate that our algorithm achieves better performance than those cited in the literature [2] while also providing a principled way to approach the link prediction problem when signed edges are present.

III. Extended Structural Balance Theory (ESBT)

In this section, we introduce a new and more fine-tuned way to define structural balance that we will call extended structural balance theory, or ESBT for short. The psychological explanation for weak structural balance relies on the concept of stress. Certain situations cause stress in interactions and as a result, are not considered natural. So, in balanced situations such stress must not exist. We define this stress more precisely as a function of the strength of relationships.

In classical balance theory, relations are restricted to binary values (+/−), which we interpret as trust and distrust. When Cartwright and Harrary first formalized the theory of structural balance, they also suggested that relationships of interest exist in varying degrees, and that their theory is built on the incomplete representation of strengths of relations [5]. Tie strength is a well-studied concept in social psychology. A person may have close friends and acquaintances (strong/weak ties), with different trust expectations. A strong tie may represent a trust relation corresponding constructs relevant to high risk situations, while a weak tie may be trusted for low risk situations like providing private information [11].

To model this distinction, we consider a scenario where relationships have varying strengths. A strong positive link represents a close friendship or family tie, i.e. strong trust constructs, and a strong negative link represents hatred, i.e. strong distrust constructs. However, many other types of trust relationships may exist in between the spectrum of (strong) trust and (strong) distrust. For example, a negative bias may be considered a weak distrust and a weak tie may represent weak (positive) trust.

A complete balance theory should be able to deal with relationhips with strengths. As a first step, we need to have a measurement of relations with various strengths. While it is arguable whether such strengths can be expressed by numerical values, it is fairly clear that the strength of any two relations can be compared. For positive relations such as liking, valuing or approving, two relationships are comparable in terms of which one is stronger than the other. Similar argument applies to two negative relations. Finally, a positive relation and a negative relation are comparable by their signs. Hence, relations with strengths by nature inherit a total ordering.

Let the collection of relations with strengths be E. An edge (A, B) and a relation with associated strength e will be used interchangeably in later discussion. We pick the ordering $\preceq$ such that, $e_1 \preceq e_2$ denotes e_1 is positively equivalent to or stronger than (or negatively equivalent to or weaker than resp.) e_2. That is, if e_1, e_2 are both positive relations, $e_1 \preceq e_2$ if e_1 is at least as positive as e_2 in terms of strengths; if e_1, e_2 are both negative relations, $e_1 \preceq e_2$ if e_1 is at most as negative as e_2 in terms of strengths; if e_1, e_2 are of different signs, $e_1 \preceq e_2$ if e_1 is positive and e_2 is negative. In the simplest case where we have only positive and negative relations, we have that $+ \preceq -$.

We also consider a **neutral relationship** as one that is unbiased, which will be denoted as O. Basically a neutral relationship is a non-negative and non-positive relationship, corresponding to no opinion and no bias. In the case of incomplete networks, classic balance theory implicitly considered two types of triads with neutral relations balanced: "$+, +, O$", "$+, -, O$" [8]. With the introduction of neutral relations, E can be partitioned into three subsets: positive relationships P, negative relations N and neutral relations O. Following the definition of ordering $\preceq$, it is clear that for any $e_+ \in P$, $e_O \in O$, $e_- \in N$, $e_+ \preceq e_O \preceq e_-$ holds. We use $\langle e_1, e_2 \rangle$ to denote the set of relations $\langle e_1, e_2 \rangle = \{e \mid e_1 \preceq e \preceq e2\}$. Hence, given $\langle e_1, e_2 \rangle$, the lower bound e_1 represents the strongest possible relationship and the upper bound represents e_2 represents the weakest possible relationship in this range.

A. *Principles of Structural Balance*

A triad is the smallest unit in balance theory, within which two of its relations cause influence over the third one. Such an influence will limit the range of comfortable relations of the third relation in a balanced state; and if it goes out of range, tension occurs and participants will suffer from **stress**. Participants will seek relationship changes to resolve this type of stress. We call such range of relations **tolerance**, with which we interpret structural balance at a finer level.

Given a network of nodes $G = (V, E)$, there exists a tolerance for each pair (A, B) of nodes of the form $\langle e_1, e_2 \rangle$ which is constrained by the triads (A, B) is part of. When there is no constraint, i.e. stress, on a specific relationship, the tolerance includes any relationship in E. We propose two principles regarding tolerance: transitivity and heterophily.

Principle 1 (Transitivity of positive relationships): Let individuals A, B, C in a network form a traid, and (A, B), $(A, B)^{'}$, (B, C) be positive. Suppose $T = \langle e_1, e_2 \rangle$ denotes the tolerance of (A, C) based on relations $(A, B), (B, C)$, and $T' = \langle e'_1, e'_2 \rangle$ denotes the tolerance based on $(A, B)^{'}, (B, C)$. If $(A, B)' \preceq (A, B)$ then we have that $e'_2 \preceq e_2$.

Furthermore, there exists a $e_{sp} \in P$ such that if $(A, B) \preceq e_{sp}$ and $(B, C) \preceq e_{sp}$, then $e_2 \preceq e_O$ holds for all $e_O \in O$. That is, T will only have positive relations beyond a certain threshold e_{sp}.

In other words, the fact that B are friends with both A and C provides the freedom for A and C to become friends; and there is stress on A and C to get close. The stronger the relation between (A, B) and (B, C), the higher the stress between (A, C) to be connected (more) positively, and the resulting tolerance is restricted to be more positive.

The stress that is based on positive relations has been frequently defined by SBT and WSBT. Positive relations in a triad cause stress for the remaining relations to be positive. As a result, both in SBT and WSBT, a balanced network consists of communities that are connected to each other with positive ties. When we consider the strength of relations, we generalize this by saying that the more positive two of the relations are in a tie, there is lesser tolerance for negative trust values.

There is a point when the strength of the two positive relations are strong enough such that it is imbalanced for (A, C) to remain unfriended, i.e. neutral. This observation is inspired by the "strong triadic closure" in [8]. In trust literature, it is often referred as the "transitivity of trust" though transitivity is also used in other contexts.

Principle 2 (Heterophily in relationships): Let individuals A, B, C in a network form a triad. Suppose $T = \langle e_1, e_2 \rangle$ denotes the tolerance of (A, C) based on relations $(A, B), (B, C)$, and $T' = \langle e'_1, e'_2 \rangle$ denotes the tolerance based on $(A, B)^{'}, (B, C)$. We have that if $(A, B)' \preceq (A, B) \preceq (B, C)$, or $(B, C) \preceq (A, B) \preceq (A, B)'$, then $e_1 \preceq e'_1$.

Furthermore, suppose there is a well-defined concept of difference between two relations with strengths. Then, the larger difference between (A, B) and (B, C) is, less positive e_1 (the lower bound) is.

Given individuals A, B, C in a network, if the relationship between (A, B) and the relation between (B, C) differs to some extent, then the tolerance is geared towards the negative. Furthermore, the more different the strength of the relationships are, the tolerance is geared towards more negative values.

The second type of stress is an interpretation of homophily. We note that homophily, i.e. having common friends or enemies, may sometimes cause stress (in $+, +, +$) but sometimes it does not (in $-, -, -$ for WSBT). However, lack of homophily, which we call heterophily does cause stress. For example, consider the case $+, +, -$ for (A, B), (B, C) and (A, C). There is stress on (A, C) to be positive due to transitivity. But, there is also stress on (A, B) and (B, C) to be negative. Either way, the result will be more desirable: either all being friends, or having two friends with a common enemy. We call the second type of stress the principle of heterophily. The more different the ties are (the strongest difference is between distrust and trust, and the weakest difference is between two identical trust ties), the more pressure there is for the tie to be negative. At a point when the difference between (A, B) and (B, C) is significant enough, we argue that a positive value for (A, C) will cause imbalance. This is inspired by the observation that two people who have severely conflicted relationships with a common neighbor, e.g., one is the other's close friend's enemy, are not likely to be friends.

(A, B)	(A, C)	Tolerance for (B, C)	
$+$	$+$	$\langle +, O \rangle$	Transitivity
$+$	$-$	$\langle O, - \rangle$	Heterophily
$-$	$-$	$\langle +, - \rangle$	No stress

TABLE I. TOLERANCE RULES IN STRUCTURAL BALANCE THEORY

The two principles help interpret balance in terms of relations with strengths precisely. A triad is balanced if all relationship strengths are within the tolerance implied by the other relationships in the triad.

Definition 1 (Balance): A triad A, B, C is balanced if for all pairs (A, B), given the tolerance $\langle e_1, e_2 \rangle$ of (A, B) with respect to $(B, C), (A, C)$, we have that $(A, B) \in \langle e_1, e_2 \rangle$. Given a network G of relationships, G is said to be balanced if for all triads in the network are balanced.

The tolerance rules for the Davis's balance theory are given in Table I. Notice that neutral relations "O" are also added. This is because triangles "$+, +, O$" and "$+, -, O$" are allowed implicitly in their theory in the general case of incomplete graphs. According to the table, triads (1), (2) and (3) from Figure 1 are balanced as each relation strength is within the tolerance, but triad (4) is not balanced. Hence, our theory generalizes WSBT in the classical balance theory.

B. Balance theorems with weak and strong ties.

In this section, we show how our reasoning can be applied to a network with multiple types of relationships. We consider a set of discrete labels (shown in Table II) that have been discussed in previous literature and show that we can reason about balance in such a network using our two principles.

Strong positive ties, s+ (trust) are similar to a close friendship. There is a strong expectation of reciprocity, similarity of tastes (homophily), common intentions and benevolence towards each other [12]. The traditional definition of SBT is based on these types of positive relationships.

Strong negative ties, s- (distrust) are generally explained as having negative experiences with someone which is indicative of their negative intentions, unreliability and overall belonging to groups that are not considered trustworthy [13]. Even when a distrusted person behaves in a trustworthy way, this could be considered a trick to get one to trust them.

However, these are not all the different classes of relationships that one might consider in a network. While one might consider a continuum of tie strength, we summarize some

Relation Type	Interpretation
Strongly positive (s+)	close friendship, trust
Weakly positive (w+)	aquiantance
Neutral (O)	unbiased relation, no relation
Weakly negative (w-)	minor disagreement, negative bias
Strongly negative (s-)	hatred, distrust

TABLE II. STRENGTH OF RELATIONS: $s+ \preceq w+ \preceq O \preceq w- \preceq s-$.

additional discrete classes. **Weak positive ties, w+ (weak trust)** can be considered a utilitarian type of trust. Interacting with someone who is only trusted partially is more risky, but can be acceptable in certain situations. For example, Uzzi [14] uses the term embedded vs. arm's length ties to distinguish between the two types of trust. While a close friend is highly trusted, they may not have access to the resources a more risky contact may provide. Granovetter [11] uses the term weak tie to talk about a relationship that is an acquaintance, not a close friend. Weak ties give access to less privileged information than strong ties, but come from outside of one's close network. In both cases, there is a trust relationship between two people, but this does not imply a continuous interaction or a strong affective component as in trust.

We also introduce **weak negative ties, w- (weak distrust)** to model cases in which there is a certain amount of distrust as a result of biases stemming from social groups people belong to or heresay that may not be as strong as distrust [15]. In essence, the burden of proof of one's trustworthiness is much higher in distrust than in weak distrust, but in both cases, positive evidence is not evaluated in the same way as in trusting relations. These five types of relationships are summarized in Table II.

(A,B)	(A,C)	(B,C)'s tolerance	(A,B)	(A,C)	(B,C)'s tolerance
$s+$	$s+$	$\langle s+, w+\rangle$	$w+$	$w-$	$\langle w+, s-\rangle$
$s+$	$w+$	$\langle s+, O\rangle$	$w+$	$s-$	$\langle O, s-\rangle$
$s+$	O	$\langle s+, w-\rangle$	O	O	$\langle s+, s-\rangle$
$s+$	$w-$	$\langle O, s-\rangle$	O	$w-$	$\langle s+, s-\rangle$
$s+$	$s-$	$\langle w-, s-\rangle$	O	$s-$	$\langle w+, s-\rangle$
$w+$	$w+$	$\langle s+, w-\rangle$	$w-$	$w-$	$\langle s+, s-\rangle$
$w+$	O	$\langle s+, s-\rangle$	$w-$	$s-$	$\langle s+, s-\rangle$
			$s-$	$s-$	$\langle s+, s-\rangle$

TABLE III. TOLERANCE OF STRONG, WEAK & NEUTRAL RELATIONS

Given these relationship types, we now describe the tolerance for different triads in Table III and the resulting imbalanced triads or structures in Table IV. Notice that the triads with two positive relations and one negative relation are imbalanced as they are in classic balance theory, except for the cases in which all three relations as weak. In fact, the types of triads that consist of weak relations and neutral relations only are not considered to be imbalanced structures. The argument here is that when all relations are weak or neutral, the influence inside the triad is not significant enough to draw tension. Also, triads of type "$s+$ $s+$ O" and "$s+$ $s-$ O" are considered to be imbalanced structures. The arguments against each type of imbalanced structure is listed in Table IV.

C. Relation Distance and General Expression of Balance

The concept of extended balance is meaningful only if the tolerance rules can be explicitly defined, so that whether a triad and a network is balanced or not can be determined. Whenever relations are drawn from a finite and small set, this is easy to do. However, it is considerably more complex in the general case when the strengths of relations are drawn from arbitrary numerical values.

Triads and the argument for stress
$(s{+}s{+}s{-})$ $(s{+}s{+}w{-})$ $(s{+}w{+}s{-})$ $(s{+}w{+}s{-})$ $(w{+}w{+}s{-})$: my two friends cannot get along with each other
$(s+s+O)$: my two close friends do not friend each other
$(s+s-O)$: my enemy's close friend does not pick a side

TABLE IV. IMBALANCED TRIADIC STRUCTURES IN THE PRESENCE OF STRONG AND WEAK TIES.

To handle such cases, we refine the measurement of relations with strengths from a total ordering to positive real values. In particular, we define function $\psi : E \rightarrow R^+$ such that, for two relations $e_1, e_2 \in E$, $\psi(e_1) \leq \psi(e_2)$ if and only if $e_1 \preceq e_2$. Since positive values can be seen as metric distances, we call $\psi(e)$ the relation distance of e. In other words, relations with varying strengths are represented by distances with different lengths. More negative strengths are represented by larger distances and more positive strengths are represented by smaller distances. We propose the following general rule of tolerance with the concept of relation distance.

Definition 2: Given adjacent relations (A,B) and (B,C), the tolerance of (A,C) is given by $[|\psi(A,B) - \psi(B,C)|, \psi(A,B) + \psi(B,C)]$.

It can be easily checked that the general tolerance rule agrees with Principle 1 and 2 by substituting $\leq$ for $\preceq$. Immediately, we have the following theorem.

Theorem 1: Given a triad (A,B,C), if $\psi(A,B)$, $\psi(B,C)$, $\psi(A,C)$ satisfies the metric triangle inequality, then (A,B,C) is balanced.

To illustrate how the distances can be used to represent a given set of strength values, we revisit WSBT. Consider two thresholds: $b_+ < b_-$ such that if $\psi(e) \geq b_-$ then the relationship e is negative. Similarly, if $\psi(e) \leq b_+$, then e is a positive relationship. For any value $b_+ < \psi(e) < b_-$, the relationship is considered neutral. We can see that as long as $b_- > 2b_+$, Table I is equivalent to Definition 2.

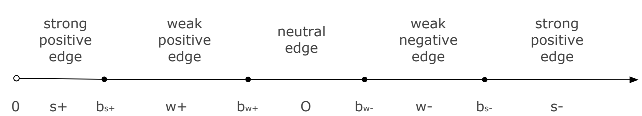

Fig. 2. Partitioning of distances: $(0, b_{s+}] : s+$, $(b_{s+}, b_{w+}) : w-$, $[b_{w+}, b_{w-}] : O$, $(b_{w-}, b_{s-}) : w-$, and $[b_{s-}, \infty) : w+$.

Similarly, to capture the example from Section III-B, we consider the partitioning of the distance domain given in Figure 2. We can see that if the following conditions are satisfied for the boundary parameters:

$$b_{w+} > 2b_{s+} \qquad b_{s-} > b_{w-} + b_{s+}$$
$$b_{s-} > 2b_{w+} \qquad b_{w-} > b_{w+} + b_{s+}$$

the tolerance given in Table III are equivalent to Definition 2, and all triads shown in Table IV are imbalanced according to metric triangle inequality.

IV. Convergence Model

Researchers have long argued that every social network has a tendency towards a balanced state [16]. The next question of interest is if an imbalance rises, in what way will a social network change towards a new balance. It is noted in social psychology literature that people are reluctant to make changes in relations as they tend to avoid the effort needed to make such changes. In a balanced triadic relation, participants are likely to do nothing and keep their pairwise relations as what they were. In an imbalanced triadic relation, participants are likely to make the smallest effort possible to regain triadic balance. We define the concept of relation cost as the effort one needs to take to accomplish a certain relation change. Our convergence model is established based on a unified assumption: every social network converges in a way that requires as little total change in relations as possible to reach a balanced state.

With the concept of relation distance, we are able to express the structure of a social network by drawing it in the Euclidean space. The strength of each relation is expressed by the distance between their locations. Notice that every layout in the Euclidean space automatically satisfies the metric triangle inequality, and hence corresponds to a balanced state of the network. For an imbalanced social network, it is not possible to draw it using its initial relation distances. Hence, our convergence model produces a layout of the social network with minimum total relation cost from the original one.

Let $G = (V, E)$ denote an arbitrary social network, and $G^* = (V, E^*)$ denote a balanced state of G. Let $n * n$ matrix X denote the layout of G^*, with each row vector x_i denoting node i's location in m-dimensional space. For each pair (i, j), $\psi(i, j)$ denotes its relation distance in G, and $d_{i,j}(X)$ denotes distance between i and j in X, i.e., its relation distance in G^*. Given an edge $(i, j) \in E$, the relation cost on (i, j) is given by:

$$c_{i,j}(X) = w_{\psi(i,j)} * (d_{i,j}(X) - \psi(i, j))^2$$

where the weight value is a function of the original distance. The weight function can take into account the difficulty of changing a relation. For example, it is generally easier to change a neutral relation than a positive or a negative relation that incorporate an initial bias. The study of optimal weights is beyond the scope of this paper. However, we consider three main classes of weights:

$$w_{\psi(i,j)} = \begin{cases} w_+ & \text{if } \psi(i,j) \text{ is a positive edge} \\ w_O & \text{if } \psi(i,j) \text{ is a neutral edge} \\ w_- & \text{if } \psi(i,j) \text{ is a negative edge} \end{cases}$$

If $w_O << w_+$ and $w_O << w_-$, then neutral edges would have very little influence on the already established positive/negative relations.

Definition 3: Let $G = (V, E)$ be a social network where E is a set of weighted edges. Its converged network $G^* = (V, E^*)$ is given by layout matrix X with $d_{i,j}(X)$ as the relation distance between every pair (i, j), such that the total relation cost $\sigma(X)$ is minimized:

$$\sigma(X) = \min_X \sum_{i<j\leq n} w_{\psi(i,j)} * (d_{i,j}(X) - \psi(i, j))^2$$

The optimization of relation cost is in fact a Metric Multidimensional Scaling problem (MDS) by assigning nodes a location in metric space. The total cost function is called stress in MDS, and is often minimized through an optimization strategy called *stress majorization* [6]. Stress majorization is an iterative method that guarantees monotonically decreasing stress in each iteration, and returns a locally minimum solution. It is recognized as a principled technique in the field of graph drawing. The algorithm, however, requires $O(n^3)$ time and $O(n^2)$ space. Due to its complexity, stress majorization is applicable on graphs with limited size when missing edges are explicitly represented as neutral ties. We are in the process of developing an approximation method for this problem as a result. In this paper, however, we will investigate the performance of the exact solution to stress majorization.

V. Experimental Results

In this section, we first focus on the *edge sign prediction problem*. Suppose we are given a social network with signs, but a small fraction of the edge signs are "hidden". How can we predict these signs with the information provided by the rest of network? The convergence model is able to predict these "hidden" signs. Let's denote the original social network with all signed edges as G, the network consisting of hidden edges as G_h, and the network consisting of the remaining edges as G_r. The edges (relations) between each pair of nodes is measured by $\{+, -, O\}$. We run the convergence model on G_r, and denote the network after convergence as $G_r^{'}$. We expect that the signs of the hidden edges in $G_r^{'}$ largely agree with the true signs.

By the assumption that every social network has a tendency towards balance, it can be inferred that G is largely balanced at any moment. Hence, the majority of G_r is balanced. The only exceptions are the components with hidden edges, which are of sign O in G_r. By the principle of total relation cost minimization, the changes mostly occur on the O-sign hidden edges during the convergence. We expect the hidden edges in G_h to have their true signs in $G_r^{'}$ if G is largely balanced.

We compare our algorithm to force directed algorithm (FD) in [2]. Note that we have tuned our implementation of FD to provide similar performance reported in this work. Even though this work combines two algorithms, in our comparison experiments, we find that FD alone gives equally good prediction performance on all three datasets as the combination. Due to space limitations, we exclude the detailed study of PP and FD/PP combination from this paper.

We use the same three datasets as it is in [2] [4] to conduct our experiments, all provided by the Stanford Large Network Dataset Collection. (1) *Epinions* is a product review website where users give reviews and ratings on product articles. Users can choose to trust or distrust others. The network contains more than 100,000 users and over 700,000 trust/distrust edges. (2) *Slashdot* is a technology news website where users rate each other as friends or foes. The dataset released in February 2009 contains over 77,000 users and over 900,000 friend/foe edges. (3) *Wikipedia* elections collects the votes by Wikipedia users in elections for promoting candidates as administrators. Each user can give a supporting (positive) or opposing (negative) vote on the promotion of another. The dataset has about 7,000 users and around 100,000 votes (edges).

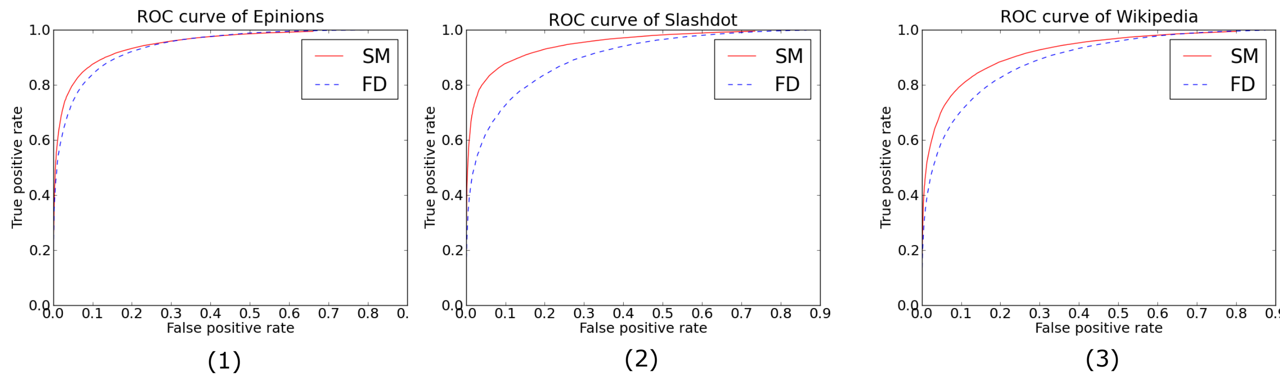

Fig. 3. The ROC curves are drawn upon distances of hidden edges, generated by SM and FD for (1) Epinions, (2) Slashdot and (3) Wikipedia datasets.

Data: M, k, deg, G
Result: Pt, Nt
Get G' = generate-subgraph(M, k, deg, G);
Partition G' into 10 groups of test and training samples;
Create two empty sets Pt, Nt;
for *each of the groups* **do**
  run SM on the training sample and get the layout;
  for *each edge e in the testing sample* **do**
    compute its distance in the layout;
    if *e is a positive edge* **then**
      add its distance to Pt;
    else
      add its distance to Nt;
    end
  end
end

Algorithm 1: SM Prediction

All edges are treated as undirected. Running SM on the entire dataset is infeasible due to both the memory and computational cost. As a result, we generate random samples of our datasets using snowball sampling method in which a small number k of seeds with degree greater than a given threshold deg are selected at random, then all nodes that are adjacent to the seed node are selected iteratively until the desired network size is reached. In our practice, the size of the resulting graph is in the range 3,000-5,000 nodes, k is chosen from 2-10 randomly and deg is chosen from 7-20 randomly. For each dataset, we generate 10 sub-networks and perform 10-fold cross validation. The number of edges in a sub-network of Epinions is around 180,000, for Slashdot 65,000 and for Wikipedia 160,000. In the implementation of SM, the partitioning of the distance domain satisfies $b_+ < 1/2b_-$, conforming to our theory. The weight of each type of edge satisfies $w_O << w_+ < 1/2w_-$. The first inequality has been argued in the previous section. The second one is chosen empirically, indicating that a negative edge has larger influence than a positive one. We use the same setting for all the networks and do not employ any other adjustable parameters.

Edge Sign Prediction. The distances of testing edges are computed by the layout of the training data. Given a distance threshold, the sign of each edge is predicted as positive if and only if its distance is smaller than the threshold. In the previous work, such threshold is computed from the (distance,sign) pairs of the training samples using standard machine learning techniques [2][4]. In this paper, however, we do not concentrate on the learning process. The issue of interest is how good the convergence model performs in separating hidden positive edges from negative ones in terms of distance. Instead of making predictions based on a particular threshold, we draw ROC curves for evaluation which capture the performance of sign prediction for both positive and negative edges across all thresholds and compute the false and true positive rates based on the computed Pt (Nt) values returned by the Algorithm 1. The ROC curves in Figure 3 are drawn upon the Pt (Nt) values from the accumulation of all testing samples.

For all three datasets, we find the ROC curve of SM is on the “northwest” side of the one of FD, which indicates SM is consistently better than FD in separating hidden positive edges from negative ones. Notice that the improvement for Slashdot is the most significant one among the three, possibly due to the fact that Slashdot edges represent “friends” or “foes”, which is by nature a more clear identification of trust/distrust compared to votes in Wikipedia or distrust for reviews in Epinions. As a result, our convergence model produces a very good prediction performance. On the Epinions and Slashdot datasets, the best thresholds on ROC curve give $88 - 90\%$ accuracy on both positive and negative hidden edges. For Wikipedia, SM achieves $83 - 85\%$ at the best threshold. The accuracy rates of Epinions and Wikipidea match the best results from previous work, and Slashdot appears to be the best so far.

General Link Prediction. The *edge sign prediction* only deals with the cases in which we already know that an edge exists in the original network. A more general and harder problem is to predict whether there is a positive or a negative edge between a pair of nodes (link prediction [7]). The difficulty of these problems stem from the fact that social networks are usually sparse with a lot more neutral relations than biased relations. Our convergence model should be able to make general edge predictions based on distances. If larger distances represent more negativeness (less positiveness), then the distance of a neutral relation should be smaller than a negative one and larger than a positive one. As a consequence, the distribution of neutral edges in terms of distance should concentrate in the middle range. We study this distribution, as a preliminary step towards solving the general edge prediction.

For each dataset, we generate samples based on random source nodes as before, except that we exclude the edges

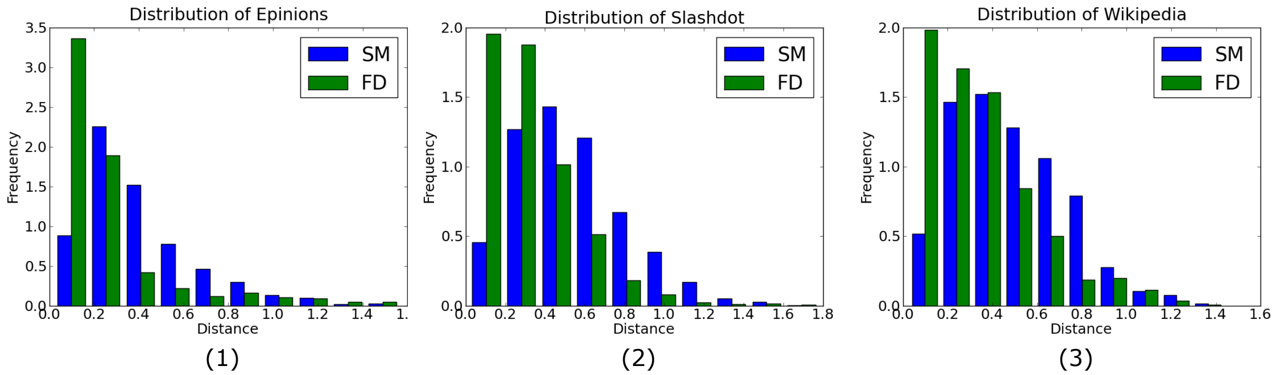

Fig. 4. The histograms are drawn upon distances of neutral testing edges, generated by SM and FD for (1) Epinions, (2) Slashdot and (3) Wikipedia datasets.

between the k source nodes. Instead of cross validation, we use the entire sub-network for training, and use the $k(k-1)/2$ edges between the k source nodes as testing data, whose signs are available in the original dataset (positive, negative or neutral, i.e. no link). After convergence the distances of these edges should be representative of their true signs. We repeat the experiments 50 times over all three datasets, and collect the distances for only the neutral testing edges. Figure 4 shows that the distances of neutral testing edges generated by SM do relatively concentrate in the middle-range of values following an almost Gaussian distribution. In contrast, the majority of neutral testing edges's distances by FD have small values, implying a positive prediction is much more likely for FD than for SM. However, SM provides more flexibility as the distances are distributed over a larger range with an almost Gaussian distribution, allowing us to test different tunable algorithms. As a result, our model is a good starting point for developing algorithms for solving the general *link prediction problem*.

VI. Conclusions

In this paper, we introduced a general model for structural balance theory that can handle relation strengths and generalizes the classical balance theory. We showed that our notion of balance can be mapped to triangular inequality over metric distances and the issue of convergence can be modeled as the metric multidimensional scaling problem for which stress majorization provides exact solutions. We have shown that our theory can be used to effectively solve the edge sign prediction problem and its performance matches and exceeds state of the art for this problem. This is due to the fact that positive and negative edges are mapped to a continuous range of strengths based on the constraints provided by the other nodes. However, in contrast with previous work, our method is aware of global constraints based on balance which results in better results overall. Furthermore, the solutions provided by our method can also be used to solve the harder *link prediction problem*.

We are investigating various avenues of future work. An approximation algorithm of stress majorization has been developed in the context of social networks. Furthermore, we are currently testing how the inclusion of relation strength improves performance by considering other actions of the users that imply the existence of a social tie. Our method also has applications to many related problems like clustering and link prediction, which we are currently investigating. We are studying the various properties of our theory in general networks and how it can be extended to an asymmetric interpretation of links. Our method also allows us to study and compare the characteristics existing networks towards balance such as the ratio between positive and negative distances, distribution of neutral edges. These measures can help us develop new insights into the nature of adversarial relationships in different networks.

References

[1] S. Adalı, *Modeling Trust Context in Networks*. Springer Briefs, 2013.

[2] T. DuBois, J. Golbeck, and A. Srinavasan, “Predicting trust and distrust in social networks,” in *IEEE Int. Conf. on Social Computing*, 2011.

[3] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, “Propagation of trust and distrust,” in *WWW*, 2004, pp. 403–412.

[4] J. Leskovec, D. Huttenlocher, and J. Kleinberg, “Predicting positive and negative links in online social networks,” in *WWW*, 2010, pp. 641–650.

[5] D. Cartwright and F. Harary, “Structural balance: a generalization of heider's theory,” *Psychological Review*, vol. 63, no. 5, pp. 277–293, 1956.

[6] E. R. Gansner, Y. Koren, and S. North, “Graph drawing by stress majorization,” *J. Pach (Ed.), Graph Drawing, Lecture Notes in Computer Science*, vol. 3383, pp. 239–250, 2005.

[7] D. Liben-Nowell and J. Kleinberg, “The link prediction problem for social networks,” in *CIKM*, 2003, pp. 556–559.

[8] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[9] J. Davis, “Clustering and structural balance in graphs,” *Human Relations*, pp. 181–187, 1967.

[10] T. DuBois, J. Golbeck, and A. Srinivasan, “Rigorous probabilistic trust-inference with applications to clustering,” in *Proc. of the 2009 WI-IAT Conf.*, vol. 01. IEEE Computer Society, 2009, p. 655658.

[11] M. Granovetter, “The strength of weak ties,” *American Journal of Sociology*, vol. 78, pp. 1–22, 1973.

[12] M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll, “Understanding and sharing intentions: The origins of cultural cognition,” *Behavioral and Brain Sciences*, vol. 28, no. 5, pp. 675–91, 2005.

[13] S. T. Fiske, A. J. Cuddy, and P. Glick, “Universal dimensions of social cognition: warmth and competence,” *Trends in Cognitive Sciences*, vol. 11, no. 2, pp. 77 – 83, 2007.

[14] B. Uzzi, “The sources and consequences of embeddedness for the economic performance of organizations: The network effect.” *American sociological review*, vol. 61, pp. 674–698, 1996.

[15] D. Ames, S. Fiske, and A. Todorov, “Impression formation: A focus on others intents,” in *The Oxford Handbook of Social Neuroscience*, J. Decety and J. Cacioppo, Eds. Oxford University Press, 2011, pp. 419–433.

[16] P. Doreian, “Event sequences as generators of social network evolution,” *Social Networks*, vol. 24, pp. 93–119, 2002.