Journal of Machine Learning Research 13 (2013) ?-? Submitted 2/13; Published ??/??

Prediction and Clustering in Signed Networks: A Local to Global Perspective

Kai-Yang Chiang KYCHIANG@CS.UTEXAS.EDU
Cho-Jui Hsieh CJHSIEH@CS.UTEXAS.EDU
Nagarajan Natarajan NAGA86@CS.UTEXAS.EDU
Inderjit S. Dhillon INDERJIT@CS.UTEXAS.EDU
Department of Computer Science
University of Texas at Austin
Austin, TX 78701, USA

Ambuj Tewari TEWARIA@UMICH.EDU
Department of Statistics, and
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, MI 48109, USA

Editor: Jure Leskovec

Abstract

The study of social networks is a burgeoning research area. However, most existing work is on networks that simply encode whether relationships exist or not. In contrast, relationships in *signed* networks can be positive ("like", "trust") or negative ("dislike", "distrust"). The theory of social balance shows that signed networks tend to conform to some local patterns that, in turn, induce certain global characteristics. In this paper, we exploit both local as well as global aspects of social balance theory for two fundamental problems in the analysis of signed networks: sign prediction and clustering. Local patterns of social balance have been used in the past for sign prediction. We define more general measures of social imbalance (MOIs) based on ℓ-cycles in the network and give a simple sign prediction rule. Interestingly, by examining measures of social imbalance, we show that the classic Katz measure, which is used widely in unsigned link prediction, also has a balance theoretic interpretation when applied to signed networks. Motivated by the global structure of balanced networks, we propose an effective low rank modeling approach for both sign prediction and clustering. We provide theoretical performance guarantees for our low-rank matrix completion approach via convex relaxations, scale it up to large problem sizes using a matrix factorization based algorithm, and provide extensive experimental validation including comparisons with local approaches. Our experimental results indicate that, by adopting a more global viewpoint of social balance, we get significant performance and computational gains in prediction and clustering tasks on signed networks. Our work therefore highlights the usefulness of the global aspect of balance theory for the analysis of signed networks.

Keywords: signed networks, sign prediction, balance theory, low rank model, matrix completion, graph clustering.

1. Introduction

The study of networks is a highly interdisciplinary field that draws ideas and inspiration from multiple disciplines including biology, computer science, economics, mathematics, physics, sociology, and statistics. In particular, *social network analysis* deals with networks that form between people. With roots in sociology, social network analysis has evolved considerably. Recently, a major force in its evolution has been the growing importance of online social networks that were themselves enabled by the Internet and the World Wide Web. A natural result of the proliferation of online social networks has been the increased involvement in social network analysis of people from computer science, data mining, information studies, and machine learning.

Traditionally, online social networks have been represented as graphs, with nodes representing entities, and edges representing relationships between entities. However, when a network has like/dislike, love/hate, respect/disrespect, or trust/distrust relationships, such a representation is inadequate since it fails to encode the *sign* of a relationship. Recently, online networks where two opposite kinds of relationships can occur have become common. For example, online review websites such as Epinions allow users to either like or dislike other people's reviews. Such networks can be modeled as *signed networks*, where edge weights can be either greater or less than 0, representing positive or negative relationships respectively. The development of theory and algorithms for signed networks is an important research task that cannot be succesfully carried out by merely extending the theory and algorithms for unsigned networks in a straightforward way. First, many notions and algorithms for unsigned networks break down when edge weights are allowed to be negative. Second, there are some interesting theories that are applicable only to signed networks.

Perhaps the most basic theory that is applicable to signed social networks but does not appear in the study of unsigned networks is that of *social balance* (Harary, 1953; Cartwright and Harary, 1956). The theory of social balance states that relationships in friend-enemy networks tend to follow patterns such as "an enemy of my friend is my enemy" and "an enemy of my enemy is my friend". A notion called weak balance (Davis, 1967) further generalizes social balance by arguing that in many cases an enemy of one's enemy can indeed act as an enemy. Both strong and weak balance are defined in terms of *local* structure at the level of triangles. Interestingly, the local structure dictated by balance theory also leads to a special *global* structure of signed networks. We review the connection between local and global structure of balance signed networks in Section 2.

Social balance has been shown to be useful for prediction and clustering tasks for signed networks. For instance, consider the *sign prediction problem* where the task is to predict the (unknown) sign of the relationship between two given entities. Ideas derived from local balance of signed networks can be succesfully used to obtain algorithms for sign prediction (Leskovec et al., 2010a; Chiang et al., 2011). In addition, the *clustering problem* of partitioning the nodes of a graph into tightly knit clusters turns out to be intimately related to weak balance theory. We will see how a clustering into mutually antagonistic groups naturally emerges from weak balance theory (see Theorem 3 for more details).

The goal of this paper is to develop algorithms for prediction and clustering in signed networks by adopting the local to global perspective that is already present in the theory of social balance. What we find particularly interesting is that the local-global interplay

that occurs in the *theory* of social balance also occurs in our *algorithms*. We hope to convince the reader that, even though the local and global definitions of social balance are theoretically equivalent, algorithmic and performance gains occur when a more global approach in algorithm design is adopted.

We mentioned above that a key challenge in designing algorithms for signed networks is that the existing algorithms for unsigned networks may not be easily adapted to the signed case. For example, it has been shown that spectral clustering algorithms for unsigned networks cannot, in general, be directly extended to signed networks (Chiang et al., 2012). However, we do discover interesting connections between methods meant for unsigned networks and those meant for signed networks. For instance, in the context of sign prediction, we see that the *Katz measure*, which is widely used for unsigned link prediction, actually has a justification as a sign prediction method in terms of balance theory. Similarly, methods based on *low rank matrix completion* can be motivated using the global viewpoint of balance theory. Thus, we see that existing methods for unsigned network analysis can reappear in signed network analysis albeit due to different reasons.

Here are the key contributions we make in this paper:

- We provide a local to global perspective of the sign prediction problem, and show that our global methods are superior on synthetic as well as real-world data sets.
- In particular, we propose sign prediction methods based on local structures (triads and higher-order cycles) and low-rank modeling. The methods that use local structures are motivated by a local viewpoint of social balance, whereas the low-rank modeling approach can be viewed as a global approach motivated by a corresponding global viewpoint of social balance.
- We show that the Katz measure used for unsigned networks can be interpreted from a social balance perspective: this immediately yields a sign prediction method.
- We provide theoretical guarantees for sign prediction and signed network clustering of balanced signed networks, under mild conditions on their structure.
- We provide comprehensive experimental results that establish the superiority of global methods over local methods studied in the past and other state-of-the-art approaches.

Parts of this paper have appeared previously in Chiang et al. (2011) and Hsieh et al. (2012). The sign prediction methods based on paths and cycles were first presented in Chiang et al. (2011), and low-rank modeling in Hsieh et al. (2012). In this paper, we provide a more detailed and unified treatment of our previous research; in particular, we provide a local-to-global perspective of the proposed methods, and a much more comprehensive theoretical and experimental treatment.

The organization of this paper is guided by the local versus global aspects of social balance theory. We first review some basics of signed networks and balance theory in Section 2. We recall notions such as (strong) balance and weak balance while emphasizing the connections between local and global structures of balanced signed networks. We will see that local balance structure is revealed by triads (triangles) and cycles, while global balance structure manifests itself as clusterability of the nodes in the network. An understanding

of the local viewpoint of social balance as well as its global implications are reviewed in Section 2 to help the reader better appreciate the methods developed in the paper.

We introduce sign prediction methods motivated from the local viewpoint of social balance in Section 3. In particular, we propose measures of social imbalance (MOIs), and a simple sign prediction method for a given measure of imbalance. The proposed measures of imbalance satisfy the property that they are zero if and only if the network is balanced. While the imbalance measure based on triads has already been studied, the local definition of social balance based on general cycles can be used to obtain a more general measure of imbalance. We also propose a simple and efficient relaxation of the proposed measure. We show the validity of the relaxation in Theorem 5. An infinite-order version of the measure for sign prediction leads to a familiar proximity measure for (unsigned) networks called the Katz measure, as stated in Theorem 7. Other than serving to reinterpret the Katz measure from the perspective of social balance in signed networks, this result is the first connection that we see between link prediction in unsigned networks and sign prediction in signed networks. The sign prediction method recently proposed by Leskovec et al. (2010a) was also motivated by the local definition of social balance. This method, however, is limited to using triangles in the network. In our experiments, we observe that a variant of their method that considers higher-order cycles performs better.

In Section 4, we develop a completely global approach based on the global structure of balanced signed networks. We appeal to the global clusterability of complete weakly balanced networks (stated in Theorem 3) to develop our global approach. Broadly, we show that such networks have low rank adjacency matrices, so that we can solve the sign prediction problem by reducing it to a low rank matrix completion problem. Specifically, we show that the adjacency matrix of a complete k-weakly balanced network, that can be partitioned into $k > 2$ groups such that within-cluster edges are positive and the rest are negative, has rank k (Theorem 8). The result follows by observing that the column space of the signed adjacency matrix is spanned by a set of k linearly independent vectors corresponding to the k groups.

Our approach attempts to fill in the unobserved (missing) edges of a signed network so that the resulting network is weakly balanced. Our result on the low-rank nature of signed adjacency matrices allows us to pose the sign prediction problem as a low-matrix completion problem. The inherent rank constraint in the problem is non-convex. Therefore, we resort to using multiple approximation strategies for solving the problem. First, we look at a standard convex relaxation of rank constraint using the trace norm. The approach comes with recovery guarantees due to Candés and Tao (2009), and it requires the adjacency matrix to be ν-incoherent (see Definition 6). We analytically show that incoherence, in the case of complete k-weakly balanced signed networks, is directly related to the notion of *group imbalance*, which measures how skewed the group sizes are (Theorem 10). We would expect a large group imbalance to make the recovery of the adjacency matrix harder. We rigorously show the recovery guarantee in terms of group imbalance for signed networks in Theorem 11. Unfortunately, solving the aforementioned convex relaxation is computationally prohibitive in practice. We discuss two approaches for approximately solving the low-rank matrix completion problem: one based on Singular Value Projection proposed by Jain et al. (2010) and the other based on matrix factorization, which is both scalable and empirically more accurate.

Furthermore, the low rank modeling approach can also be used for the clustering of a signed network. Our clustering method proceeds as follows. First, we use a low-rank matrix completion algorithm on its adjacency matrix. Then we cluster the top-k eigenvectors of the completed matrix using any feature-based clustering algorithm. By doing so, we show that under the same assumptions that guarantee the recovery of signed networks, the true clusters can be identified from the top-k eigenvectors (Theorem 12).

In Section 5, we show some evidence of local and global balance in real networks. Our experiments on synthetic and real networks show that global methods (based on low rank models) generally perform better, in terms of accuracy of sign prediction, than local methods (based on triads and cycles). Finally, we discuss related work in Section 6, and state our conclusions in Section 7.

2. Signed Networks and Social Balance

In this section, we set up our notation for signed networks, review the basic notions of balance theory, and describe the two main tasks (sign prediction and clustering) addressed in this paper.

2.1 Categories of Signed Networks

The most basic kind of a signed network is a *homogeneous* signed network. Formally, a homogeneous signed network is represented as a graph with the adjacency matrix $A \in \{-1, 0, 1\}^{n \times n}$, which denotes relationships between entities as follows:

$$A_{ij} = \begin{cases} 1, & \text{if } i \ \& \ j \text{ have positive relationship,} \\ -1, & \text{if } i \ \& \ j \text{ have negative relationship,} \\ 0, & \text{if relationship between } i \ \& \ j \text{ is unknown (or missing).} \end{cases}$$

We should note that we treat a zero entry in A as an *unknown* relationship instead of no relationship, since we expect any two entities have some (hidden) positive or negative attitude toward each other even if the relationship itself might not be observed. From an alternative point of view, we can assume there exists an underlying *complete* signed network $A^\star$, which contains relationship information between all pairs of entities. However, we only observe some entries of $A^\star$, denoted by Ω. Thus, the partially observed network A can be represented as:

$$A_{ij} = \begin{cases} A^\star_{ij}, & \text{if } (i,j) \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$$

A signed network can also be *heterogeneous*. In a heterogeneous signed network, there can be more than one kind of entity, and relationships between two, same or different, entities can be positive and negative. For example, in the online video sharing website Youtube, there are two kinds of entities – users and videos, and every user can either *like* or *dislike* a video. Therefore, the Youtube network can be seen as a bipartite signed network, in which the positive and negative links are between users and videos.

In this paper, we will focus our attention on homogeneous signed networks, i.e. networks where relationships are between the same kind of entities. For heterogeneous signed

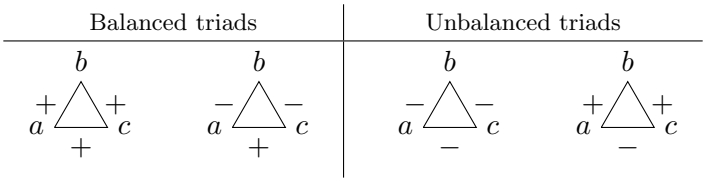


Table 1: Configurations of balanced and unbalanced triads.

networks, it is possible to do some preprocessing to reduce them to homogeneous networks. For instance, in a Youtube network, we could possibly infer the relationships between users based on their taste of videos. These preprocessing tasks, however, are not trivial.

In the remaining part of the paper, we will use the term "network" as an abbreviation for "signed network", unless we explicitly specify otherwise. In addition, we will now mainly focus on undirected signed graphs (i.e. A is symmetric) unless we specify otherwise. For a directed signed network, a simple but sub-optimal way to apply our methods is by considering the symmetric network, $\text{sign}(A + A^T)$. Of course, making the network symmetric erases edges with conflicting signs between a pair of nodes. It is important to know how much information is lost in the process. We found that in real networks, the percentage of conflicting edges is extremely small (see Table 4 in Section 5). The observation suggests that making the network undirected preserves the sign structure for the most part, and is sufficient for analysis.

2.2 Social Balance

A key idea behind many methods that estimate a high dimensional complex object from limited data is the exploitation of *structure*. In the case of signed networks, researchers have identified various kinds of non-trivial structure (Harary, 1953; Davis, 1967). In particular, one influential theory, known as social balance theory, states that relationships between entities tend to be *balanced*. Formally, we say a triad (or a triangle) is *balanced* if it contains an even number of negative edges. This is in agreement with beliefs such as "a friend of my friend is more likely to be my friend" and "an enemy of my friend is more likely to be my enemy". The configurations of balanced and unbalanced triads are shown in Table 1.

Though social balance specifies the patterns of triads, one can generalize the balance notion to general ℓ-cycles. An ℓ-cycle is defined as a simple path from some node to itself with length equal to ℓ. The following definition extends social balance to general ℓ-cycles:

Definition (Balanced ℓ-cycles) *An ℓ-cycle is said to be balanced when it contains an even number of negative edges.*

Table 2 shows some instances of balanced and unbalanced cycles based on the above definition. To define balance for general networks, we first define the notion of balance for *complete* networks:

Definition (Balanced complete networks) *A complete network is said to be balanced when all triads in the network are balanced.*

Of course, most real networks are not complete. To define balance for general networks, we adopt the perspective of a missing value estimation problem as follows:

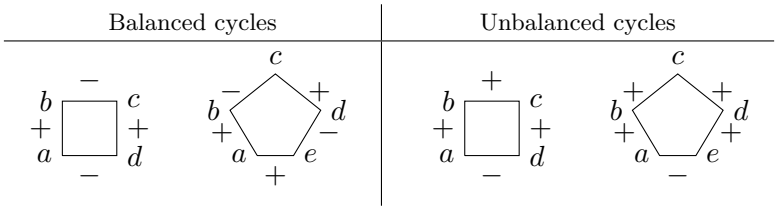


Table 2: Some instances of balanced and unbalanced cycles.

Definition (Balanced networks) *A (possibly incomplete) network is said to be balanced when it is possible to assign* ± 1 *signs to all missing entries in the adjacency matrix, such that the resulting complete network is balanced.*

So far, the notion of balance is defined by specifying patterns of *local* structures in networks (i.e. the patterns of triads). The following result from balance theory shows that balanced networks have a *global* structure.

Theorem 1 (Balance theory, (Cartwright and Harary, 1956)) *A network is balanced iff either (i) all edges are positive, or (ii) we can divide nodes into two clusters (or groups), such that all edges within clusters are positive and all edges between clusters are negative.*

Now we revisit balanced ℓ-cycles defined at the beginning of this subsection. Under that definition, we can verify if a network is balanced or not by looking at all cycles in the network due to the following well-known theorem (whose proof can be found in Easley and Kleinberg (2010, Chapter 5)).

Theorem 2 *A network is balanced iff all its ℓ-cycles are balanced.*

One possible weakness of social balance theory is that the defined balance relationships might be too strict. In particular, researchers have argued that the degree of imbalance in the triad with two positive edges (the fourth triad in Table 1) is much stronger than that in the triad with all negative edges (the third triad in Table 1). By allowing triads with all negative edges, a weaker version of balance notion can be defined (Davis, 1967).

Definition (Weakly balanced complete networks) *A complete network is said to be weakly balanced when all triads in the network are weakly balanced.*

Note that ℓ-cycles with an odd number of negative edges are allowed under weak balance. The definition for general incomplete networks can be obtained by adopting the perspective of a missing value estimation problem:

Definition (Weakly balanced networks) *A (possibly incomplete) network is said to be weakly balanced when it is possible to obtain a weakly balanced complete network by filling the missing edges in its adjacency matrix.*

Though the above definitions define weak balance in terms of patterns of local triads, one can show that weakly balanced networks have a special global structure, analogous to Theorem 1:

Theorem 3 (Weak balance theory, Davis (1967)) *A network is weakly balanced iff either (i) all of its edges are positive, or (ii) we can divide nodes into k clusters, such that all the edges within clusters are positive and all the edges between clusters are negative.*

Note that when $k = 2$, this theorem simply reduces to Theorem 1.

At this juncture, it would be helpful to recall that balance and weak balance are more natural for the analysis of undirected (signed) networks. More recently, Leskovec et al. (2010b) analyzed directed signed networks using the concept of "status", that is characteristic of directed networks. We do not explore the "status structure" of signed networks in this paper, but the theory seems to be promising and worthy of study in the future.

2.3 Key Problems in Signed Network Analysis

As in classical social network analysis, we are interested in what we can infer given a signed network topology. In particular, we will focus on two core problems — sign prediction and clustering.

In the *sign prediction* problem, we intend to infer the unknown relationship between a pair of entities i and j based on partial observations of the entire network of relationships. More specifically, if we assume that we are given a (usually incomplete) network A sampled from some underlying (complete) network $A^\star$, then the sign prediction task is to recover the sign patterns of one or more edges in $A^\star$. This problem bears similarity to the *structural link prediction problem* in *unsigned* networks (Liben-Nowell and Kleinberg, 2007; Menon and Elkan, 2011). Note that the *temporal link prediction problem* has also been studied in the context of an unsigned network evolving in time. The input to the prediction algorithm then consists of a series of networks (snapshots) instead of a single network. We do not consider such temporal problems in this paper.

Clustering is another important problem in network analysis. Recall that according to weak balance theory (Theorem 3), we can find k groups such that they are mutually antagonistic in a weakly balanced network. Motivated by this, the clustering task in a signed network is to identify k antagonistic groups in the network, such that most edges within the same cluster are positive while most edges between different clusters are negative. Notice that since the (weak) balance notion only applies to signed networks, most traditional clustering algorithms for unsigned networks cannot be directly applied.

3. Local Methods for Sign Prediction

The definition of structural balance based on triangles is local in nature. A natural approach for designing sign prediction algorithms proceeds by reasoning locally in terms of unbalanced triangles, which motivates the following *measure of imbalance*:

$$\mu_{\text{tri}}(A) := \sum_{\tilde{\sigma} \in SC_3(A)} \mathbf{1}\left[\tilde{\sigma} \text{ is unbalanced}\right] , \tag{1}$$

where $SC_3(A)$ refers to the set of triangles (simple cycles of length-3) in the network A. In general, we use $SC_\ell(A)$ to denote the set of all simple ℓ-cycles in the network A. A definition essentially similar to the one above appears in the recent work of van de Rijt

(2011, p. 103) who observes that the equivalence

$$\mu_{\text{tri}}(A) = 0 \text{ iff } A \text{ is balanced}$$

holds only for complete graphs. For an incomplete graph, imbalance might manifest itself only if we look at longer simple cycles. Accordingly, we define a higher-order analogue of (1),

$$\mu_\ell^s(A) := \sum_{i=3}^{\ell} \beta_i \sum_{\tilde{\sigma} \in SC_i(A)} \mathbf{1}\left[\tilde{\sigma} \text{ is unbalanced}\right] \ , \tag{2}$$

where $\ell \geq 3$ and β_i's are coefficients weighting the relative contributions of unbalanced simple cycles of different lengths. If we choose a decaying choice of β_i, like $\beta_i = \beta^i$ for some $\beta \in (0,1)$, then we can even define an infinite-order version,

$$\mu_\infty^s(A) := \sum_{i \geq 3} \beta_i \sum_{\tilde{\sigma} \in SC_i(A)} \mathbf{1}\left[\tilde{\sigma} \text{ is unbalanced}\right] \ . \tag{3}$$

It is clear that $\mu_\infty^s(\cdot)$ is a genuine measure of imbalance in the sense formalized by the following corollary of Theorem 2.

Corollary 4 *Fix an observed graph A. Let $\beta_i > 0$ be any sequence such that the infinite sum in* (3) *is well-defined. Then, $\mu_\infty^s(A) > 0$ iff A is unbalanced.*

The basic idea of using a measure of imbalance for predicting the sign of a given query link $\{i,j\}$, such that $i \neq j$ and $A_{i,j} = 0$, is as follows. Given the observed graph A and query $\{i,j\}$, $i \neq j$, we construct two graphs: $A^{+(i,j)}$ and $A^{-(i,j)}$. These are obtained from A by setting A_{ij} and A_{ji} to $+1$ and -1 respectively. Formally, these two augmented graphs are defined as:

$$A_{uv}^{+(i,j)} = \begin{cases} 1, & \text{if } (u,v) = (i,j) \text{ or } (j,i) \\ A_{uv}, & \text{otherwise.} \end{cases} \qquad A_{uv}^{-(i,j)} = \begin{cases} -1, & \text{if } (u,v) = (i,j) \text{ or } (j,i) \\ A_{uv}, & \text{otherwise.} \end{cases}$$

Then, given a measure of imbalance, denoted as $\mu(\cdot)$, the predicted sign of $\{i,j\}$ is simply:

$$\text{sign}\left(\mu\left(A^{-(i,j)}\right) - \mu\left(A^{+(i,j)}\right)\right) \ . \tag{4}$$

Note that, to use the above for prediction, we should use a $\mu(\cdot)$ for which the quantity (4) is efficiently computable. The measure of imbalance based on triads $\mu_{\text{tri}}(A)$ and the more general measure $\mu_\infty^s(A)$ involve counting simple cycles in a graph. However, enumerating simple cycles of a graph is NP-hard [1]. To get around this computational issue, we slightly change the definition of $\mu_\ell(\cdot)$ to the following.

$$\mu_\ell(A) := \sum_{i=3}^{\ell} \beta_i \sum_{\sigma \in C_i(A)} \mathbf{1}\left[\sigma \text{ is unbalanced}\right] \ . \tag{5}$$

1. By straightforward reduction to the Hamiltonian cycle problem (Karp, 1972).

As before, we allow $\ell = \infty$ provided the β_i's decay sufficiently rapidly.

$$\mu_\infty(A) := \sum_{i \geq 3} \beta_i \sum_{\sigma \in C_i(A)} \mathbf{1}\left[\sigma \text{ is unbalanced}\right] \ . \tag{6}$$

The only difference between these definitions and (2),(3) is that here we sum over *all* cycles (denoted by $C_i(A)$), not just simple ones. However, we still get a valid notion of imbalance as stated by the following result.

Theorem 5 *Fix an observed graph A. Let $\beta_i > 0$ be any sequence such that the infinite sum in* (6) *is well-defined. Then, $\mu_\infty(A) > 0$ iff A is unbalanced.*

It turns out, somewhat surprisingly, that computing (4) with $\mu(A) = \mu_{\text{tri}}(A)$ simply reduces to computing (i,j) entry in A^2. The following key lemma gives an efficient way to compute (4) with $\mu(\cdot) = \mu_\ell(\cdot)$ in general. Indeed, it amounts to computing higher powers of the adjacency matrix.

Lemma 6 *Fix A and let $i \neq j$ be such that $(i,j) \notin \Omega$. Let $A^{+(i,j)}$ and $A^{-(i,j)}$ be the augmented graphs. Then, for any $\ell \geq 3$,*

$$\sum_{\sigma \in C_\ell\left(A^{-(i,j)}\right)} \mathbf{1}\left[\sigma\right] \quad - \sum_{\sigma \in C_\ell\left(A^{+(i,j)}\right)} \mathbf{1}\left[\sigma\right] = A^{\ell-1}_{i,j} \ ,$$

where $\mathbf{1}\left[\sigma\right]$ is the abbreviation of $\mathbf{1}\left[\sigma \text{ is unbalanced}\right]$.

Using Lemma 6, it is easy to see that the prediction (4) using (5) reduces to

$$\text{sign}\left(\mu_\ell\left(A^{-(i,j)}\right) - \mu_\ell\left(A^{+(i,j)}\right)\right) = \text{sign}\left(\sum_{t=3}^{\ell} \beta_t A^{t-1}_{i,j}\right) \ ,$$

and the prediction (4) using (6) reduces to

$$\text{sign}\left(\mu_\infty\left(A^{-(i,j)}\right) - \mu_\infty\left(A^{+(i,j)}\right)\right) = \text{sign}\left(\sum_{\ell \geq 3} \beta_\ell A^{\ell-1}_{i,j}\right) \ . \tag{7}$$

Lemma 6 is also key to interpreting a classical proximity measure called the *Katz* measure in the context of sign prediction, discussed next. Proofs of Theorem 5 and Lemma 6 are presented in Appendix A.

3.1 Katz Measure is Valid for Signed Networks

The classic method of Katz (1953) has been used successfully for *unsigned* link prediction (Liben-Nowell and Kleinberg, 2007). Formally, given an unsigned network A with a query node-pair (i,j), the Katz measure for the link (i,j) is defined as:

$$Katz(i,j) := \sum_{\ell=2}^{\infty} \beta^\ell A^\ell_{ij},$$

where $\beta > 0$ is a constant so that the above infinite sum is well defined ($\beta < \frac{1}{\|A\|_2}$ suffices, where $\|A\|_2$ is the spectral norm of A). [2] Intuitively, the Katz measure sums up all possible paths between i and j. The number of length ℓ paths can grow exponentially as ℓ increases (for example, there are $\Omega(n^{\ell-1})$ length ℓ paths between i and j if the network is complete). Therefore, the contributions from length ℓ paths are exponentially damped by the constant β^ℓ. One can also verify the above definition has the following matrix form:

$$Katz(i,j) = ((I - \beta A)^{-1} - I - \beta A)_{ij}.$$

A higher Katz score indicates more proximity between nodes i and j, and the link (i,j) is therefore more likely to exist or to form in the future.

While Katz has been used as an effective proximity measure for link prediction in unsigned networks, it is not obvious what the *signed* Katz score corresponds to in signed networks. Following (7), the connection between Katz measure and $\mu_\infty(\cdot)$ stands out. The following theorem shows that by considering a sign prediction method based on $\mu_\infty(\cdot)$ we obtain an interesting interpretation of the Katz measure on a signed network from a balance theory viewpoint.

Theorem 7 (Balance Theory Interpretation of the Katz Measure) *Consider the sign prediction rule* (4) *using* $\mu_\infty(\cdot)$ *in the reduced form* (7). *In the special case when* $\beta_\ell = \beta^{\ell-1}$ *with* β *small enough* ($\beta < 1/\|A\|_2$), *the rule can be expressed as the Katz prediction rule for edge sign prediction, in closed form:*

$$\operatorname{sign}\left(\left((I - \beta A)^{-1} - I - \beta A\right)_{i,j}\right) .$$

Our sign prediction rule for a given measure of imbalance relies on social balance theory for signed networks. However, real world networks may not conform to the prediction of balance theory or may do so only to a certain extent. Furthermore, balance theory was developed for undirected networks and hence methods based on measures of imbalance can deal only with undirected networks. To partly mitigate the lapse, we can use measures of imbalance to derive *features* that can then be fed to a supervised learning algorithm (like logistic regression) along with the signs of the known edges in the network. When we learn weights for such features, we are weakening our reliance on social balance theory but can naturally deal with directed graphs. By using features constructed from higher-order cycles, we extend the supervised approach used by Leskovec et al. (2010a) that was limited to learning from triads-based features. While the learning approach itself is straightforward, construction of higher-order features for directed signed networks requires some attention. We defer the details to Appendix B. In the experiments (Section 5), we denote the local method (7) corresponding to the measure of imbalance based on cycles of length ℓ by MOI-ℓ. Note that MOI-∞ refers to the signed Katz measure. The supervised learning method discussed in Appendix B, where we use logistic regression to train weights for features constructed from higher-order cycles of length up to ℓ, is referred to as HOC-ℓ in the experiments.

2. Our definition is slightly different from the definition in Liben-Nowell and Kleinberg (2007), where the summation starts from $\ell = 1$. However, the measures are identical for the purpose of link prediction, as the prediction needs to be made only when the query nodes i and j have no existing edge, i.e. $A_{ij} = 0$.

4. Global Methods: Low Rank Modeling

In Section 3, we have seen how to use ℓ-cycles for sign prediction. We have also seen that ℓ-cycles play a major role in how balance structure manifests itself *locally*. By increasing ℓ, the level at which balance structure is considered becomes less localized. Still, it is natural to ask whether we can design algorithms for signed networks by directly making use of their *global* structure. To be more specific, let us revisit the definition of complete weakly balanced networks. In general, weak balance can be defined from either a local or a global point of view. From a local point of view, a given network is weakly balanced if all *triads* are weakly balanced, whereas from a global point of view, a network is weakly balanced if its *global structure* obeys the clusterability property stated in Theorem 3. Therefore, it is natural to ask whether we can directly use this global structure for sign prediction. In the sequel, we show that weakly balanced networks have a "low-rank" structure, so that the sign prediction problem can be formulated as a low rank matrix completion problem.

We begin by showing that given a complete k-weakly balanced network, its adjacency matrix $A^\star$ has rank at most k:

Theorem 8 (Low Rank Structure of Signed Networks) *The adjacency matrix $A^\star$ of a complete k-weakly balanced network has rank 1 if $k \leq 2$, and has rank k for all $k > 2$.*

Proof Since $A^\star$ is k-weakly balanced, the nodes can be divided into k groups, say $S^{(1)}, S^{(2)}, \ldots, S^{(k)}$. Suppose group $S^{(i)}$ contains nodes $s_1^{(i)}, s_2^{(i)}, \ldots, s_{n_i}^{(i)}$, then the corresponding columns vectors of $A^\star$ all have the following form (after suitable reordering of nodes):

$$\mathbf{b}_i = [-1 \quad \cdots \quad -1 \underbrace{1 \quad \cdots \quad 1}_{\text{the } i^{th} \text{ group}} -1 \quad \cdots \quad -1]^T,$$

and so the column space of $A^\star$ is spanned by $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$.

First consider $k \leq 2$, i.e., the network is strongly balanced. If $k = 1$, it is easy to see that $\text{rank}(A^\star) = 1$. If $k = 2$, then $\mathbf{b}_1 = -\mathbf{b}_2$. Therefore, $\text{rank}(A^\star)$ is again 1.

Now consider $k > 2$. In this case, we argue that $\text{rank}(A^\star)$ exactly equals k by showing that $\mathbf{b}_1, \ldots, \mathbf{b}_k$ are linearly independent. We consider the $k \times k$ square matrix M such that $M_{ij} = -1, \forall i \neq j$ and $M_{ii} = 1, \forall i$. It is obvious that $\mathbf{1} = [1\ 1 \cdots 1]^T$ is an eigenvector of M with eigenvalue $-(k-2)$. We can further construct the other $k-1$ linearly independent eigenvectors, each with eigenvalue 2:

$$\mathbf{e}_1 - \mathbf{e}_2,\ \mathbf{e}_1 - \mathbf{e}_3,\ \ldots,\ \mathbf{e}_1 - \mathbf{e}_k,$$

where $\mathbf{e}_i \in \mathbb{R}^k$ is the i^{th} column of the $k \times k$ identity matrix. These $k-1$ eigenvectors are clearly linearly independent. Therefore, $\text{rank}(M) = k$.

From the above we can show that $\text{rank}(A^\star) = k$. Suppose that $\mathbf{b}_1, \ldots, \mathbf{b}_k$ are not linearly independent, then there exists $\alpha_1, \ldots, \alpha_k$, with some $\alpha_i \neq 0$, such that $\sum_{i=1}^k \alpha_i \mathbf{b}_i = \mathbf{0}$. Using this set of α's, it is easy to see that $\sum_{i=1}^k \alpha_i M_i = \mathbf{0}$ (where M_i is the i^{th} column of M), but this contradicts the fact that $\text{rank}(M) = k$. Therefore, $\text{rank}(A^\star) = k$. ■

Figure 1 is an example of a complete 3-weakly balanced network. As shown, we see its adjacency matrix can be expressed as a product of two rank-3 matrices, indicating its rank is no more than three. In fact, by Theorem 8, we can conclude that its rank is exactly 3.

$$\begin{pmatrix} 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -1.045 & 0.265 & -0.402 \\ -1.045 & 0.265 & -0.402 \\ -1.045 & 0.265 & -0.402 \\ 0.319 & -1.260 & -0.830 \\ 0.919 & 0.670 & -0.541 \\ 0.919 & 0.670 & -0.541 \end{pmatrix} \begin{pmatrix} -1.045 & 0.265 & 0.402 \\ -1.045 & 0.265 & 0.402 \\ -1.045 & 0.265 & 0.402 \\ 0.319 & -1.260 & 0.830 \\ 0.919 & 0.670 & 0.541 \\ 0.919 & 0.670 & 0.541 \end{pmatrix}^T$$

Figure 1: An illustrative example of the low-rank structure of a 3-weakly balanced network. The network can be represented as a product of two rank-3 matrices, and so the adjacency matrix has rank no more than 3.

The above reasoning shows that (adjacency matrices of) complete weakly balanced networks have low rank. However, most real networks are not complete graphs. Recall that in order to define balance on incomplete networks, we try to fill in the unobserved or missing edges (relationships) so that balance is obtained. Following this desideratum, we can think of sign prediction in signed networks as a low-rank matrix completion problem. Specifically, suppose we observe entries $(i, j) \in \Omega$ of a complete signed network $A^\star$. We want to find a complete matrix by assigning ± 1 to every unknown entry, such that the resulting complete graph is weakly balanced and hence, the completed matrix is low rank. Thus, our missing value estimation problem can be formulated as:

$$\begin{aligned} \text{minimize} \quad & \text{rank}(X) \\ \text{s.t.} \quad & X_{ij} = A^\star_{ij}, \ \forall \ (i,j) \in \Omega, \\ & X_{ij} \in \{\pm 1\}, \ \forall \ (i,j) \notin \Omega. \end{aligned} \tag{8}$$

Once we obtain the minimizer of (8), which we will denote by $X^\star$, we can infer the missing relationship between i and j by simply looking up the sign of the entry $X^\star_{ij}$. So the question is whether we can solve (8) efficiently. In general, (8) is known to be NP-hard; however, recent research has shown the surprising result that under certain conditions, the low-rank matrix completion problem (8) can be solved by convex optimization to yield a *global* optimum in polynomial time (Candés and Recht, 2008). In the following subsections, we identify such conditions as well as approaches to approximately solve (8) for real-world signed networks.

4.1 Sign Prediction via Convex Relaxation

One possible approximate solution for (8) can be obtained by dropping the discrete constraints and replacing $\text{rank}(X)$ by the trace norm of X (denoted by $\|X\|_*$), which is the tightest convex relaxation of rank (Fazel et al., 2001). Thus, a convex relaxation of (8) is:

$$\begin{aligned} \text{minimize} \quad & \|X\|_* \\ \text{s.t.} \quad & X_{ij} = A^\star_{ij}, \ \forall \ (i,j) \in \Omega. \end{aligned} \tag{9}$$

It turns out that, under certain conditions, by solving (9) we can recover the *exact* missing relationships from the underlying complete signed network. This result is the consequence of recent research (Candés and Recht, 2008; Candés and Tao, 2009) which has

shown that perfect recovery is possible if the observed entries are uniformly sampled and $A^\star$ has high incoherence, which may be defined as follows:

Definition (Incoherence) *An $n \times n$ matrix X with singular value decomposition $X = U\Sigma V^T$ is ν-incoherent if*

$$\max_{i,j} |U_{ij}| \leq \frac{\sqrt{\nu}}{\sqrt{n}} \quad \textit{and} \quad \max_{i,j} |V_{ij}| \leq \frac{\sqrt{\nu}}{\sqrt{n}}. \tag{10}$$

Intuitively, higher incoherence (smaller ν) means that large entries of the matrix are not concentrated in a small part. The following theorem shows that under high incoherence and uniform sampling, solving (9) exactly recovers $A^\star$ with high probability.

Theorem 9 (Recovery Condition (Candés and Tao, 2009)) *Let $A^\star$ be an $n \times n$ matrix with rank k, with singular value decomposition $A^\star = U\Sigma V^T$. In addition, assume $A^\star$ is ν-incoherent. Then there exists some constant C, such that if $C\nu^4 nk^2 \log^2 n$ entries are uniformly sampled, then with probability at least $1 - n^{-3}$, $A^\star$ is the unique optimizer of (9).*

In particular, if the underlying matrix has bounded rank (i.e. $k = O(1)$), the number of sampled entries required for recovery reduces to $O(\nu^4 n \log^2 n)$.

Based on Theorem 9, we now show that the notion of incoherence can be connected to the relative sizes of the clusters in signed networks. As a result, by solving (9), we will show that we can recover the underlying signed network with high probability if there are no "small" groups. To start, we define the *group imbalance* of a signed network as follows:

Definition (Group Imbalance) *Let $A^\star$ be the adjacency matrix of a complete k-weakly balanced network with n nodes, and let $n_1, \ldots, n_k$ be the sizes of the groups. Group imbalance τ of $A^\star$ is defined as*

$$\tau := \max_{i=1,\ldots,k} \frac{n}{n_i}. \tag{11}$$

By definition, $k \leq \tau \leq n$. Larger group imbalance τ indicates the presence of a very small group, which would intuitively make recovery of the underlying network harder (under uniform sampling). For example, consider an extreme scenario that a k-weakly balanced network contains n nodes, with two groups that contain only one node each. Then if nodes $n-1$ and n form these singleton groups, the adjacency matrix of this network has group imbalance $\tau = n$ with $A^\star_{n-1,n-1} = A^\star_{n,n} = 1$ and $A^\star_{n-1,n} = A^\star_{n,n-1} = -1$. However, without observing $A^\star_{n-1,n}$ or $A^\star_{n,n-1}$, it is impossible to determine whether the last two nodes are in the same cluster, or whether each of them belongs to an individual cluster. When n is very large, the probability of observing one of these two entries will be extremely small. Therefore, under uniform sampling of $O(n \log^2 n)$ entries, it is unlikely that any matrix completion algorithm will be able to exactly recover this network.

Motivated by this example, we now analytically show that group imbalance τ determines the possibility of recovery. We first show the connection between τ and incoherence ν.

Theorem 10 (Incoherence of Signed Networks) *Let $A^\star$ be the adjacency matrix of a complete k-weakly balanced network with group imbalance τ. Then $A^\star$ is τ-incoherent.*

Proof Recall that incoherence ν is defined as the maximum absolute value in the (normalized) singular vectors of $A^\star$, which are identical to its eigenvectors (up to signs), since $A^\star$ is symmetric.

Let $\mathbf{u}$ be any unit eigenvector of $A^\star$ ($\|\mathbf{u}\|_2 = 1$) with eigenvalue λ. Suppose i and j are in the same group, then the i^{th} and j^{th} rows of $A^\star$ are identical, i.e., $A^\star_{i,:} = A^\star_{j,:}$. As a result, the i^{th} and j^{th} elements of all eigenvectors will be identical (since $u_i = A^\star_{i,:}\mathbf{u}/\lambda = A^\star_{j,:}\mathbf{u}/\lambda = u_j$). Thus, $\mathbf{u}$ has the following form:

$$\mathbf{u} = [\underbrace{\alpha_1, \alpha_1, \ldots, \alpha_1}_{n_1}, \underbrace{\alpha_2, \ldots, \alpha_2}_{n_2}, \ldots, \underbrace{\alpha_k, \ldots, \alpha_k}_{n_k}]^T. \tag{12}$$

Because $\|\mathbf{u}\|_2 = 1$, $\sum_{i=1}^k n_i \alpha_i^2 = 1$, and so $n_i \alpha_i^2 \leq 1$, $\forall i$, which implies $|\alpha_i| \leq 1/\sqrt{n_i}$, $\forall i$. Thus,

$$\max_i |u_i| = \max_i |\alpha_i| \leq \max_i \frac{1}{\sqrt{n_i}} = \max_i \frac{\sqrt{n/n_i}}{\sqrt{n}} \leq \frac{\sqrt{\tau}}{\sqrt{n}}.$$

Therefore, $A^\star$ is τ-incoherent. ■

Putting together Theorems 9 and 10, we now have the main theorem of this subsection:

Theorem 11 (Recovery Condition for Signed Networks) *Suppose we observe edges A_{ij}, $(i,j) \in \Omega$, from an underlying k-weakly balanced signed network $A^\star$ with n nodes, and suppose that the following assumptions hold:*

A. k is bounded ($k = O(1)$),

B. the set of observed entries Ω is uniformly sampled, and

C. number of samples is sufficiently large, i.e. $|\Omega| \geq C\tau^4 n \log^2 n$, where τ is the group imbalance of the underlying complete network $A^\star$.

Then $A^\star$ can be perfectly recovered by solving (9), with probability at least $1 - n^{-3}$.

In particular, if n/n_i is upper bounded so that τ is a constant, then we only need $O(n \log^2 n)$ observed entries to *exactly* recover the complete k-weakly balanced network.

4.2 Sign Prediction via Singular Value Projection

Though the convex optimization problem (9) can be solved to yield the global optimum, the computational cost of solving it might be too prohibitive in practice. Therefore, recent research provides more efficient algorithms to approximately solve (8) (Cai et al., 2010; Jain et al., 2010). In particular, we consider the Singular Value Projection (SVP) algorithm proposed by Jain et al. (2010) which attempts to solve the low-rank matrix completion problem in an efficient manner. The SVP algorithm considers a robust formulation of (8) as follows:

$$\begin{aligned} \text{minimize} \quad & \|\mathcal{P}(X) - A\|_F^2 \\ \text{s.t.} \quad & \text{rank}(X) \leq k, \end{aligned} \tag{13}$$

where the projection operator $\mathcal{P}$ is defined as:

$$(\mathcal{P}(X))_{ij} = \begin{cases} X_{ij}, & \text{if } (i,j) \in \Omega \\ 0, & \text{otherwise.} \end{cases}$$

Note that the objective (13) recognizes that there might be some violations of weak balance in the observations A, and minimizes the squared-error instead of trying to enforce exact equality as in (9). In an attempt to optimize (13), the SVP algorithm iteratively calculates the gradient descent update $\hat{X}^{(t)}$ of the current solution $X^{(t)}$, and projects $\hat{X}^{(t)}$ onto the non-convex set of matrices whose rank are at most k using SVD. After the SVP algorithm terminates and outputs $\bar{X}$, one can take the sign of each entry of $\bar{X}$ to obtain an approximate solution of (8). The SVP procedure for sign prediction is summarized in Algorithm 1.

Algorithm 1: Sign Prediction via Singular Value Projection (SVP)

Input: Adjacency matrix A, rank k, tolerance ϵ, max iteration $t_{\max}$, step size η
Output: $\bar{X}$, the completed low-rank matrix that approximately solves (8)

1. Initialize $X^{(0)} \leftarrow 0$ and $t \leftarrow 0$.
2. Do
   - $\hat{X}^{(t)} \leftarrow X^{(t)} - \eta(\mathcal{P}(X^{(t)}) - A)$
   - $[U_k, \Sigma_k, V_k] \leftarrow$ Top k singular vectors and singular values of $\hat{X}^{(t)}$
   - $X^{(t+1)} \leftarrow U_k \Sigma_k {V_k}^T$
   - $t \leftarrow t + 1$

   while $\|\mathcal{P}(X^{(t)}) - A\|_F^2 > \epsilon$ and $t < t_{\max}$
3. $\bar{X} \leftarrow \text{sign}(X^{(t)})$

In addition to its efficiency, experimental evidence provided by Jain et al. (2010) suggests that if observations are uniformly distributed, then all iterates of the SVP algorithm are ν-incoherent, and if this occurs, then it can be shown that the matrix completion problem (8) can be exactly solved by SVP. In Section 5, we will see that SVP performs well in recovering weakly balanced networks.

4.3 Sign Prediction via Matrix Factorization

A limitation of both convex relaxation and SVP is that they require uniform sampling to ensure good performance. However, this assumption is violated in most real-life applications, and so these approaches do not work very well in practice. In addition, both the methods cannot scale to very large datasets, as they require computation of the SVD. Thus, we use a gradient based matrix factorization approach as an approximation to the signed network completion problem. In Section 5, we will see that this matrix factorization approach can boost the accuracy of estimation as well as scale to large real networks.

In the matrix factorization approach, we consider the following problem:

$$\min_{W,H \in \mathbb{R}^{n \times k}} \sum_{(i,j) \in \Omega} (A_{ij} - (WH^T)_{ij})^2 + \lambda \|W\|_F^2 + \lambda \|H\|_F^2. \tag{14}$$

Problem (14) is non-convex, and an alternating minimization algorithm is commonly used to solve it. Recent theoretical results show that the alternating minimization algorithm provably solves (14) under similar conditions as trace-norm minimization (Jain et al., 2013). The matrix factorization approach is widely used in practical collaborative filtering applications as its performance is competitive to or better than trace-norm minimization, while scalability is much better. For example, to solve the Netflix problem, (14) has been

applied with a fair amount of success to factorize datasets with 100 million ratings (Koren et al., 2009).

Nevertheless, there is an issue when modeling signed networks using (14): the squared loss in the first term of (14) tends to force entries of WH^T to be either $+1$ or -1. However, what we care about in this completion task is the consistency between sign$((WH^T)_{ij})$ and sign(A_{ij}) rather than their difference. For example, $(WH^T)_{ij} = 10$ should have zero loss when $A_{ij} = +1$ if only the signs are important.

To resolve this issue, instead of using the squared loss, we use a loss function that only penalizes the inconsistency in *sign*. More precisely, objective (14) can be generalized as:

$$\min_{W,H\in\mathbb{R}^{n\times k}} \sum_{(i,j)\in\Omega} \ell oss\ (A_{ij}, (WH^T)_{ij}) + \lambda\|W\|_F^2 + \lambda\|H\|_F^2. \tag{15}$$

In order to penalize inconsistency of sign, we can change the loss function to be the sigmoid or squared-hinge loss:

$$\begin{aligned} \ell oss_{\text{ sigmoid}}(x, y) &= 1/(1 + \exp(xy)), \\ \ell oss_{\text{ square-hinge}}(x, y) &= (\max(0, 1 - xy))^2. \end{aligned} \tag{16}$$

In Section 5, we will see that applying sigmoid or squared-hinge loss functions slightly improves prediction accuracy.

4.4 Time Complexity of Sign Prediction Methods

There are two main optimization techniques for solving (14) for large-scale data: Alternating Least Squares (ALS) and Stochastic Gradient Descent (SGD) (Koren et al., 2009). ALS solves the squared loss problem (14) by alternately minimizing W and H. When one of W or H is fixed, the optimization problem becomes a least squares problem with respect to the other variable, so that we can use well developed least squares solvers to solve each subproblem. Given an $n \times n$ observed matrix with m observations, it requires $O(mk^2)$ operations to form the Hessian matrices, and $O(nk^3)$ operations to solve each least squares subproblem. Therefore, the time complexity of ALS is $O(t_1(mk^2 + nk^3))$ where t_1 is the number of iterations.

However, ALS can only be used when the loss function is the squared loss. To solve the general form (15) with various loss functions, we use stochastic gradient descent (SGD). In SGD, for each iteration, we pick an observed entry (i, j) at random, and only update the i^{th} row $\mathbf{w}_i^T$ of W and the j^{th} row $\mathbf{h}_j^T$ of H. The update rule for $\mathbf{w}_i^T$ and $\mathbf{h}_j^T$ is given by:

$$\begin{aligned} \mathbf{w}_i^T &\leftarrow \mathbf{w}_i^T - \eta \left(\frac{\partial \ell oss\ (A_{ij}, (WH^T)_{ij})}{\partial \mathbf{w}_i^T} + \lambda \mathbf{w}_i^T \right), \\ \mathbf{h}_j^T &\leftarrow \mathbf{h}_j^T - \eta \left(\frac{\partial \ell oss\ (A_{ij}, (WH^T)_{ij})}{\partial \mathbf{h}_j^T} + \lambda \mathbf{h}_j^T \right), \end{aligned} \tag{17}$$

where η is a small step size. Each SGD update costs $O(k)$ time, and the total cost of sweeping through all the entries is $O(mk)$. Therefore, the time complexity for SGD is $O(t_2 mk)$, where t_2 is the number of iterations taken by SGD to converge. Notice that

HOC	LR-ALS	LR-SGD
$O(2^\ell nm)$	$O(t_1(nk^3 + mk^2))$	$O(t_2 km)$

Table 3: Time complexity of cycle-based method (HOC) and low rank modeling methods (LR-ALS, LR-SGD). The HOC time only considers feature computation time. The time for low rank modeling consists of total model construction time.

Algorithm 2: Clustering with Matrix Completion

Input: Adjacency matrix A, number of clusters k

Output: Cluster indicators

1. $X^\star \leftarrow$ Completion(A) with any matrix completion algorithm.
2. $U \leftarrow$ Top k eigenvectors of $X^\star$.
3. Run any feature-based clustering algorithm on U.

although the complexity of SGD is linear in k, it usually takes many more iterations to converge compared with ALS, i.e., $t_2 > t_1$. For other approaches to solve (15) for large scale data, please see Yu et al. (2013).

On the other hand, all cycle-based algorithms introduced in Section 3 require time at least $O(nm)$, because they involve multiplication of m-sparse $n \times n$ matrices in model construction. In particular, the time complexity for HOC-ℓ methods is $O(2^\ell nm)$ (see Appendix B for details), which is much more expensive than both ALS and SGD as shown in Table 3 (note that in real large-scale social networks, $m > n \gg t_1, t_2, k$).

4.5 Clustering Signed Networks

In this section, we see how to take advantage of the low-rank structure of signed networks to find clusters. Based on weak balance theory, the general goal of clustering for signed graphs is to find a k-way partition such that most within-group edges are positive and most between-group edges are negative. One of the state-of-the-art algorithms for clustering signed networks, proposed by Kunegis et al. (2010), extends spectral clustering by using the signed Laplacian matrix. Given a partially observed signed network A, the signed Laplacian $\bar{L}$ is defined as $\bar{D} - A$, where $\bar{D}$ is a diagonal matrix such that $\bar{D}_{ii} = \sum_{j \neq i} |A_{ij}|$. By this definition, the clustering of signed networks can be derived by computing the top k eigenvectors of $\bar{L}$, say $U \in \mathbb{R}^{n \times k}$, and subsequently running the k-means algorithm on U to get the clusters. This procedure is analogous to the standard spectral clustering algorithm on unsigned graphs; the only difference being that the usual graph Laplacian is replaced by the signed Laplacian.

However, there is no theoretical guarantee that the use of the signed Laplacian can recover the true groups in a weakly-balanced signed network. To overcome this theoretical defect, we now give an algorithm which, under certain conditions, is able to recover the real structure even with partial observations. The key idea is to first use a low-rank matrix completion algorithm before performing the clustering. The following theorem shows that the eigenvectors of the completed matrix possess a desirable property.

Theorem 12 *Let A_{ij}, $(i,j) \in \Omega$, be entries observed from a complete k-weakly balanced network $A^\star$ with n nodes, and assume that the solution of (9) is $X^\star$ with eigenvectors $U = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k]$. If the assumptions in Theorem 11 are all satisfied, then the rows of U, $U_{i,:} = U_{j,:}$ iff i and j are in the same cluster in $A^\star$ with probability at least $1 - n^{-3}$.*

Proof From Theorem 11, we know the recovered matrix $X^\star$ will be $A^\star$ with probability $\geq 1 - n^{-3}$ if the assumptions hold. Suppose $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are the k eigenvectors of $X^\star$. From the proof of Theorem 10, the eigenvectors will have the form in (12), which means that the i^{th} and j^{th} rows of U, $U_{i,:} = U_{j,:}$ if i and j are in the same cluster. Furthermore, when i and j are in different clusters, $A^\star_{i,:} \neq A^\star_{j,:}$, so $U_{i,:}$ cannot equal $U_{j,:}$. This proves the theorem. ■

Following this theorem, the true clusters can be identified from the eigenvectors of $X^\star$ when the assumptions in Theorem 11 hold. Therefore, perfect clustering is guaranteed in this scenario.

More generally, we can use any matrix completion method to complete A. For example, if we take SVP as the matrix completion approach, we can obtain a perfect clustering result if all iterates of the algorithm are ν-incoherent. Under the latter condition, SVP can recover $A^\star$ exactly, so the property of eigenvectors in Theorem 12 can again be used. Our clustering algorithm that uses matrix completion is summarized in Algorithm 2.

It should not be surprising that our clustering algorithm is superior to (signed) spectral clustering. In some sense, our approach can be viewed as a spectral method, except that it first fills in the missing links from the training data by doing matrix completion. This step is simple yet crucial in signed networks as it overcomes the sparsity of the network. We will see that our clustering algorithm outperforms the (signed) spectral clustering method in Section 5.

5. Experimental Results

We now present experimental results for sign prediction and clustering using our proposed methods. For sign prediction, we show that local methods, such as MOI and HOC (see Section 3), yield better predictive accuracy when longer cycles are considered. In addition, if we consider the global low-rank structure of the network, prediction via matrix factorization further outperforms local methods in terms of both accuracy and running time. For clustering, we show that clustering via low rank model gives us significantly better results than clustering via signed Laplacian. These results suggest the usefulness of the global perspective on social balance.

5.1 Description of Data Sets

In our experiments, we consider both synthetic and real-life datasets. To construct synthetic networks, we first consider a complete k-weakly balanced network $A^\star$, and sample entries from $A^\star$ to form the partially observed network A, with three controlling parameters: *sparsity* s, *noise level* ϵ and *sampling process* $\mathcal{D}$. The sparsity s controls the percentage of edges we sample from $A^\star$. The noise level ϵ specifies the probability that the sign of a sampled edge is flipped. The sampling process $\mathcal{D}$ specifies how the sampled entries are distributed. In particular, we will focus on two sampling distributions: uniform and power-law distribu-

Table 4: Network Statistics

	# nodes	# edges	+ edges	- edges	edges with conflicting signs
Wikipedia	7,065	103,561	78.8%	21.2%	0.71%
Slashdot	82,144	549,202	77.4%	22.6%	0.64%
Epinions	131,828	840,799	85.0%	15.0%	0.57%

tion, denoted as $\mathcal{D}_{\text{uni}}$ and $\mathcal{D}_{\text{pow}}$ respectively. Thus, a partially observed network A can be described as $A = A^\star(s, \epsilon, \mathcal{D})$.

We also consider three real-life signed networks: Epinions, Slashdot, Wikipedia. Epinions is a consumer review network in which users can either trust or distrust other consumer's reviews. Slashdot is a discussion web site in which users can recognize others as friends or foes. Wikipedia is a who-votes-for-whom network in which users can vote for or against others to be administrators in Wikipedia. These three datasets have previously been used as benchmarks for sign prediction (Leskovec et al., 2010a; Chiang et al., 2011). Table 4 shows the statistics of the three networks.

5.2 Evidence of Local and Global Patterns in Real Signed Networks

We have seen that cycles in signed networks exhibit structural balance according to balance theory, and that we can make use of cycles for predictions (see Section 3). Indeed, cycles tend to be balanced in real-life networks. In all three real networks we consider, Leskovec et al. (2010b) found that balanced triads (i.e. 3-cycles) are much more likely to be observed than unbalanced triads. Our study also shows that the local patterns (i.e. ℓ-cycles) of the three networks tend to be balanced. For each network A, we consider all patterns of 3-cycles and 4-cycles in the symmetric network $\text{sign}(A + A^T)$. For convenience, we use $C_{\ell i}$ to denote the i^{th} pattern (of signs) of an ℓ-cycle. The patterns of these cycles are shown in Table 5. We first calculate the probability that the configuration of a given ℓ-cycle is $C_{\ell i}$, denoted $P(C_{\ell i})$. We then randomly shuffle the signs of edges in the network and calculate the same probability on the shuffled network, which is denoted $P_0(C_{\ell i})$. Thus $P_0(C_{\ell i})$ can be viewed as the (expected) probability that $C_{\ell i}$ is observed if the signs of edges have no particular pattern. With the two probabilities, we calculate the "surprise" of observing $C_{\ell i}$ as follows:

$$S(C_{\ell i}) := \frac{\Delta_\ell P(C_{\ell i}) - \Delta_\ell P_0(C_{\ell i})}{\sqrt{\Delta_\ell P_0(C_{\ell i})(1 - P_0(C_{\ell i}))}},$$

where Δ_ℓ is the number of ℓ-cycles in the network. The above quantity is basically the number of standard deviations that the observed value of $C_{\ell i}$ differs from the expected value of $C_{\ell i}$ in the shuffled network. See Leskovec et al. (2010b) for more discussion.

Table 5 shows the observed probability, the expected probability, and the surprise value of each $C_{\ell i}$ in three networks. Although it is not true that *each* of the balanced cycles is much more likely to appear, the last two rows in Table 5 show that differences between $P(C)$ and $P_0(C)$ and the surprise values of *overall* balanced 3- and 4-cycles are quite large. This implies that given an arbitrary 3- or 4-cycle in these networks, it is very likely to be balanced. Overall, we find that local balanced patterns are somewhat significant.

Table 5: Statistics of balanced and unbalanced ℓ-cycles, $\ell = 3, 4$ (note that $\sum_i P(C_{\ell i}) = \sum_i P_0(C_{\ell i}) = 1$). The first 6 cycles in the table are balanced while the last 4 cycles are unbalanced. The last two rows show that overall balanced 3-cycles and 4-cycles are much more than expected.

	Epinions			Slashdot			Wikipedia		
Type of cycle	$P(C_{\ell i})$	$P_0(C_{\ell i})$	$S(C_{\ell i})$	$P(C_{\ell i})$	$P_0(C_{\ell i})$	$S(C_{\ell i})$	$P(C_{\ell i})$	$P_0(C_{\ell i})$	$S(C_{\ell i})$
$C_{31} : +++$	0.8259	0.5754	1107.0	0.7301	0.4502	425.2	0.6996	0.4806	335.4
$C_{33} : +--$	0.0791	0.0706	72.3	0.1364	0.1260	23.5	0.0840	0.1105	-64.7
$C_{41} : ++++$	0.7538	0.4777	14464.7	0.6723	0.3435	5120.8	0.6080	0.3757	3557.6
$C_{43} : ++--$	0.0911	0.0787	1210.6	0.1127	0.1286	-352.1	0.1007	0.1155	-344.1
$C_{44} : +-+-$	0.0065	0.0393	-4418.5	0.0138	0.0645	-1528.0	0.0139	0.0578	-1396.4
$C_{46} : ----$	0.0103	0.0008	8722.8	0.0263	0.0030	3147.7	0.0054	0.0022	505.4
$C_{32} : ++-$	0.0834	0.3493	-1218.4	0.1125	0.4111	-458.7	0.2052	0.3987	-302.5
$C_{34} : ---$	0.0117	0.0047	220.9	0.0211	0.0127	56.9	0.0013	0.0102	8.5
$C_{42} : +++-$	0.1174	0.3875	-14508.8	0.1413	0.4211	-4191.5	0.2473	0.4167	-2548.5
$C_{45} : +---$	0.0208	0.0160	1017.7	0.0337	0.0392	-212.0	0.0247	0.0320	-309.3
Balanced 3-cycles	0.9050	0.6459	1182.9	0.8665	0.5763	443.9	0.7835	0.5911	299.6
Balanced 4-cycles	0.8617	0.5965	14147.8	0.8250	0.5397	4234.7	0.7280	0.5513	2635.6

On the other hand, in Section 4, we have seen that low rank structure emerges when we theoretically examine weakly balanced networks. We now show that real networks tend to exhibit low-rank structure to a much greater extent compared to random networks. As a baseline, for each real network we create two corresponding random networks for comparison: the first one is the (symmetric) ER network generated from the Erdös-Rényi model (Erdös and Rényi, 1960) that preserves the sparsity and the ratio of positive to negative edges of the compared real network. The second one is the shuffled network with the same network structure as the compared real network, except that we randomly shuffle the signs of edges.

The experiment is conducted as follows. We first derive the low-rank complete matrix $X^\star$ by running matrix completion on the observed entries A_{ij}. Then, we look at the relative error on the observed set Ω:

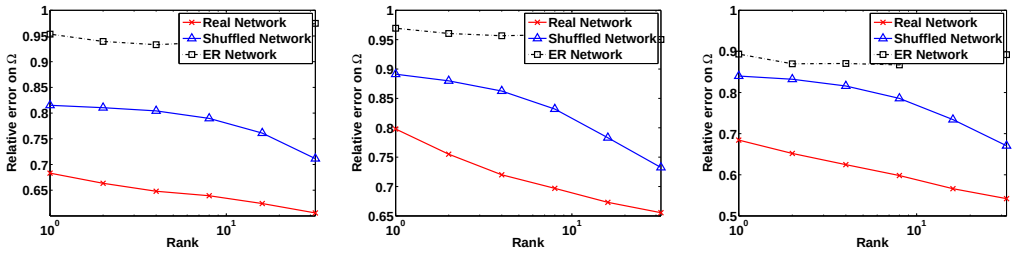
$$\text{err}_\Omega = \frac{\|W \circ (X^\star - A)\|_F}{\|A\|_F}, \tag{18}$$

where $W_{ij} = 1$ if $(i, j) \in \Omega$ and $W_{ij} = 0$ otherwise, and $\circ$ denotes element-wise multiplication. Clearly, smaller err_Ω indicates better approximation for the observed entries.

In our experiment, we use matrix factorization for matrix completion, with ranks $k = 1, 2, 4, 8, 16$ and 32. For each network (real networks and their corresponding random networks), we complete the network with different k values and compute err_Ω. The result is shown in Figure 2. Compared to the two random networks, the three real-life networks achieve much smaller err_Ω for each small k. This suggests that low-rank matrices provide a better approximation of the observed entries for real-life networks.

5.3 Sign Prediction

We now compare the performance of our proposed methods for sign prediction. As introduced in Section 3, there are two families of cycle-based methods: one based on measures of



(a) Epinions (b) Slashdot (c) Wikipedia

Figure 2: Relative error on Ω, the observed entries, between adjacency matrix and completed matrix, for real-life networks versus random networks. Real-life networks achieve much smaller relative error for every k as compared with random networks.

imbalance (MOI), and the other based on the supervised learning using higher order cycles (HOC). Both families depend on a parameter $\ell \geq 3$ that denotes the order of the cycles that the method is based on. For MOI, we consider all ℓ less than 10 as well as ∞ (recall that in this case, MOI becomes the Katz measure), and for HOC we consider $\ell = 3, 4, 5$. Note that the set of features used by HOC-$(\ell + 1)$ is a strict superset of the features used by HOC-ℓ.

We also consider two global approaches for low rank matrix completion – Singular Value Projection and matrix factorization from Sections 4.2 and 4.3 respectively. The SVP approach (denoted as LR-SVP) is chosen to demonstrate that perfect recovery can be achieved if the observations are uniformly distributed. For matrix factorization, we consider the ALS method that solves problem (14), as well as SGD methods that solve the general problem (15) with sigmoid loss and square-hinge loss, defined in (16). We denote these methods as LR-ALS, LR-SIG and LR-SH, respectively.

5.3.1 Results on Synthetic Datasets

We first compare all categories of approaches on synthetic datasets. We choose LR-SVP, LR-ALS, MOI-∞ and HOC-3 as representatives of the two approaches of low rank matrix completion, MOI-based, and HOC-based methods respectively. We consider the underlying network $A^\star$ to be a complete 5-weakly balanced network, where the five clusters have sizes 100, 200, 300, 400 and 500. Instead of observing all of $A^\star$, we assume that we only observe a partial network A using three procedures: uniform sampling, uniform sampling with noise, and sampling with power-law distribution. For each algorithm, we input the observed entries as training data and calculate the sign prediction accuracy on the rest of the entries.

Uniform sampling: In this scenario, we generate several observed networks $A = A^\star(s, 0, \mathcal{D}_{\text{uni}})$. We vary s from 0.001 to 0.1 and plot the prediction accuracy in Figure 3a. Under this setting, LR-SVP and LR-ALS outperform the cycle-based methods. We observe that MOI-∞ performs the worst with accuracy only 50%-70%. However, if we repeat the same experiment substituting $A^\star$ with $A_2^\star$, where $A_2^\star$ is a complete strongly balanced network whose two groups have size 1000, we observe that MOI and global methods perform alike as shown in Figure 3b. This is because MOI uses cycle-based measurements to make more

cycles become *balanced*. This prediction policy is most appropriate when $k = 2$ (that is, the underlying network $A^\star$ has strong balance), but performs poorly when the underlying network is weakly balanced (i.e. more than two groups). HOC-3 works much better than MOI-∞ since it learns a classifier from cycle-based features rather than simply making cycles balanced, but its accuracy drops dramatically when s is less than 0.05. On the other hand, both LR-SVP and LR-ALS show high accuracy for all $s \geq 0.01$. In particular, LR-SVP can achieve 100% accuracy when $s > 0.07$, which reconfirms the theoretical recovery guarantee stated in Theorem 11. Moreover, LR-ALS can also recover the ground truth, an observation that is consistent with previous results.

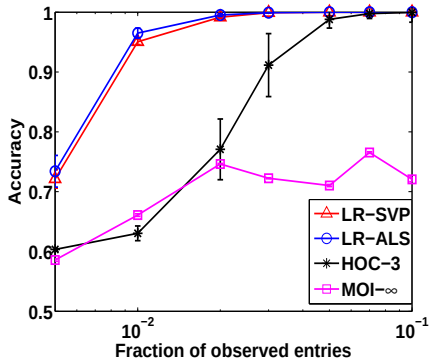
Uniform sampling with noise: To make the synthetic data more similar to real data, we further add noise to the observations. We generate observed networks $A = A^\star(0.1, \epsilon, \mathcal{D}_{\text{uni}})$, where ϵ varies from 0.01 to 0.25. The result is shown in Figure 3c. We can see that global methods are still clearly better than cycle-based methods when noise level is higher. Moreover, LR-SVP perfectly recovers $A^\star$ when the noise level $\epsilon < 0.05$, and LR-ALS also achieves perfect recovery with a smaller ϵ.

Sampling with power-law distribution: As Sections 4.1 and 4.2 pointed out, the exact recovery guarantees of convex relaxation and SVP for matrix completion crucially rely on the assumption that observed entries are uniformly sampled. However, in most real networks (for example, Slashdot in Kunegis et al. (2009)), the degree distribution of observed entries follows a power law. Therefore, we examine how the approaches perform on power-law distributed networks. The power-law distributed networks are generated using the Chung-Lu-Vu (CLV) model proposed by Chung et al. (2004), which allows one to generate random graphs with arbitrary expected degree sequence. Similar to the uniform sampling case, we perform the sign prediction task on $A = A^\star(s, 0, \mathcal{D}_{\text{pow}})$ varying s from 0.001 to 0.1, and plot the prediction accuracy in Figure 3d. We can see that MOI-∞ still has poor performance for weakly balanced graphs. However, unlike the uniform sampling case, LR-SVP has lower accuracy rate compared to HOC-3 when $s < 0.1$. On the other hand, LR-ALS still performs better than all other methods on power-law distributed graphs.

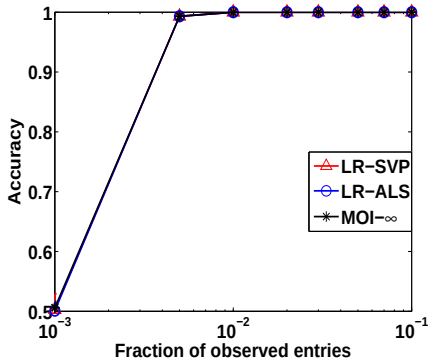
From results on synthetic data shown in Figure 3, we can conclude that global methods generally do better than local methods, and the low rank model with matrix factorization (LR-ALS) performs the best in most cases, even when observed entries are not uniformly distributed.

5.3.2 Results on Real-life Datasets

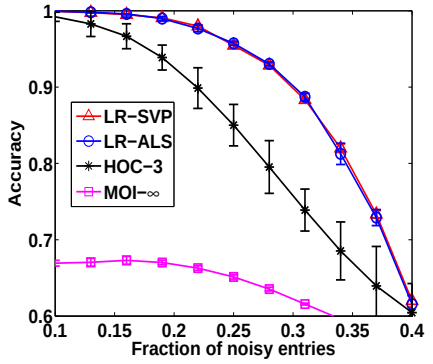
Now we further evaluate our sign prediction methods on three real-life networks. To begin with, we evaluate and compare MOI methods using a *leave-one-out* type methodology: each edge in the network is successively removed and the method tries to predict the sign of that edge using the rest of the network. Figure 4 shows the accuracy of MOI based methods. Note that the accuracy is shown for edges with embeddedness under a certain threshold. First, we see that the accuracy is a non-decreasing function of the embeddedness threshold. Next, it is clear that higher-order methods perform significantly better than the MOI-3 (triangle-based) method. Finally, the performance boost is larger for edges with low embeddedness. This is expected as edges of low embeddedness by definition do not have many common neighbors for their end-points, and higher-order cycles have relatively better



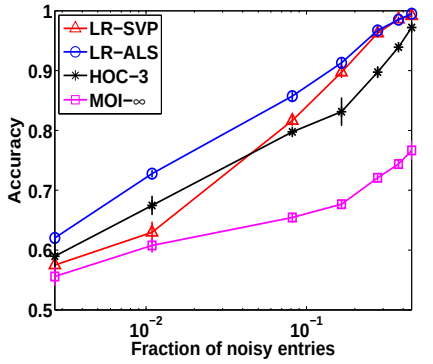
(a) Uniformly sampled without noise ($k = 5$)



(b) Uniformly sampled without noise on balanced networks ($k = 2$)



(c) Uniformly sampled with noise ($k = 5$)

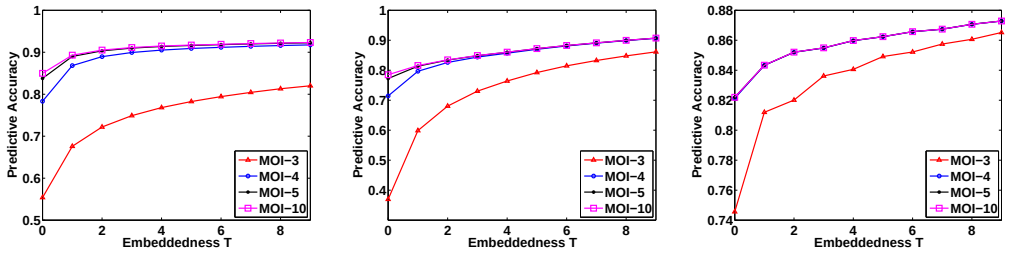


(d) Power-law distributed networks ($k = 5$)

Figure 3: Sign prediction accuracy of local and global methods on synthetic datasets. On (strongly) balanced networks (3b), MOI-∞ is seen to perform as well as low rank modeling methods (LR-SVP and LR-ALS). However, in weakly balanced networks, global methods LR-SVP and LR-ALS outperform cycle-based methods such as MOI-∞ and HOC-3 (supervision on high order cycles). In addition, low rank modeling with matrix factorization (LR-ALS) is more robust than singular value projection (LR-SVP) when the observations are sampled from a power-law distribution.

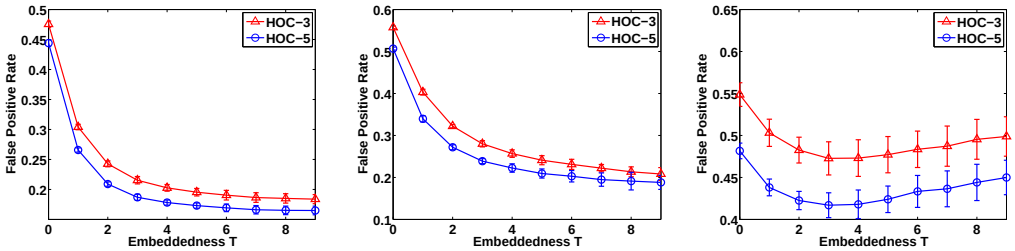
information for such edges. We also observe from our experiments that beyond $\ell = 5$, the performance gain is not very significant.

Next, we compare HOC methods with various ℓ to see how much high order cycles can benefit us in supervision. We resort to 10-*fold cross-validation*: we (randomly) created ten disjoint test folds each consisting of 10% of the total number of edges in the network. For each test fold, the remaining 90% of the edges serve as the training set. (For a given test fold, the feature extraction and logistic model training are done on a graph with the test edges removed, not just the signs.) To evaluate HOC methods, we consider not only prediction accuracies but also false-positive rates. We report accuracies as well as false-positive rates by averaging them over the ten folds. As shown in Table 6, in all the datasets, there is a



(a) Epinions (b) Slashdot (c) Wikipedia

Figure 4: Accuracy of MOI-based methods for cycle lengths $\ell = 3, 4, 5, 10$. These plots show the accuracy of MOI-ℓ methods for edges with embeddedness *at least* T for various thresholds T. We see that the difference in the performance of MOI-3 and higher order methods is larger for edges with lower embeddedness. We also see that the improvement obtained by going beyond order 5 is not very significant.



(a) Epinions (b) Slashdot (c) Wikipedia

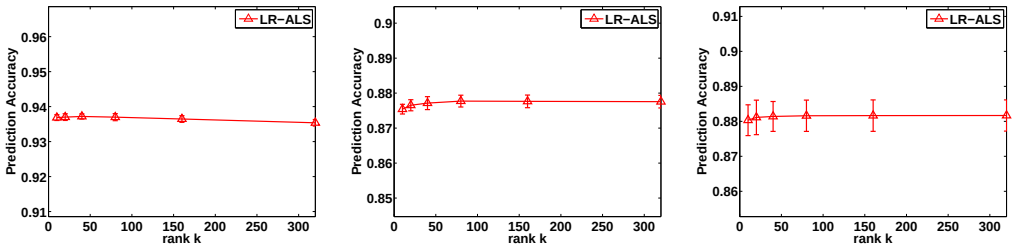
Figure 5: False positive rates of higher order cycle (HOC) Methods for $\ell = 3, 5$. These plots show the false positive rate of HOC-ℓ methods for edges with embeddedness *at least* T for various thresholds T. We see that considering higher order cycles has the benefit of significantly reducing false-positives while simultaneously achieving slightly better overall accuracy (refer to Table 6). However, unlike what we see for MOI methods, the improvement here does not seem to depend strongly on edge embeddedness. The false positive rates for HOC-4 are very similar to that of HOC-5 and hence are not shown for clarity.

small improvement in accuracy by using higher order cycles (HOC-5). The false positive rate, however, reveals a more interesting phenomenon in Figure 5. Indeed, higher order methods (such as HOC-5) significantly reduce the false positive rate as compared to HOC-3. However Figure 5 shows that, unlike MOI based methods, edge embeddedness does not seem to affect the decrease in false positive rate for HOC methods. We see this trend across all the datasets.

At this point, we see that for cycle-based methods, considering higher order cycles benefits the accuracy of sign prediction and lowers the false positive rate. Furthermore, the results are consistent across the three diverse networks. These results confirm the intuition that getting more global information improves quality of prediction, and motivate us to consider the global structure of networks.

Table 6: The sign prediction accuracy for cycle-based methods (MOI and HOC) and low rank modeling methods (LR-ALS, LR-SIG and LR-SH) along with standard deviation if the accuracy is averaged by 10-fold cross validation. Note that for MOI methods we report leave-one-out accuracy. We can see that the low rank modeling approaches are better than cycle-based methods.

	Epinions	Slashdot	Wikipedia
MOI-3	0.5539	0.3697	0.7456
MOI-10	0.8497	0.7850	0.8220
HOC-3	0.9014 ± 0.0013	0.8303 ± 0.0018	0.8424 ± 0.0063
HOC-5	0.9080 ± 0.0012	0.8469 ± 0.0015	0.8605 ± 0.0049
LR-ALS	0.9437 ± 0.0007	0.8774 ± 0.0018	0.8814 ± 0.0043
LR-SIG	$\mathbf{0.9448} \pm 0.0009$	0.8806 ± 0.0017	$\mathbf{0.8839} \pm 0.0049$
LR-SH	0.9437 ± 0.0015	$\mathbf{0.8835} \pm 0.0015$	0.8810 ± 0.0042



(a) Epinions (b) Slashdot (c) Wikipedia

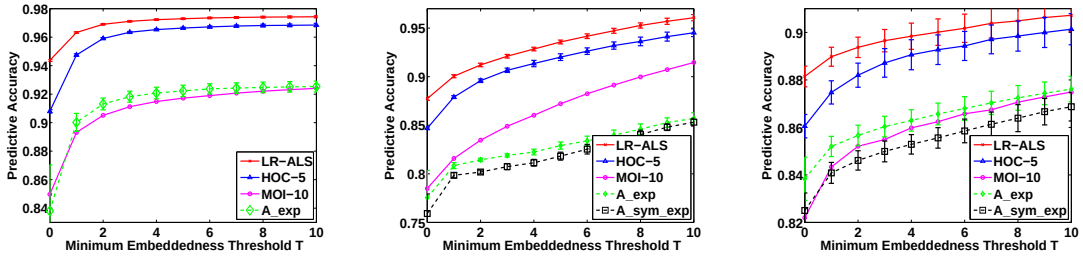
Figure 6: Sign prediction accuracy for low rank modeling with matrix factorization (LR-ALS) with different ranks. We see that LR-ALS is quite robust to the rank.

Now we turn our attention to low rank modeling approaches. We have seen that LR-SVP does not perform well under power-law distributions of observed relationships in synthetic networks (see Figure 3d), so we consider the more robust matrix factorization approach for solving the matrix completion problem, including LR-ALS, LR-SIG and LR-SH, for experiments on real datasets. Again, we use 10-fold cross validation setting, and report the average prediction accuracy for each dataset in Table 6. From the table, we observe that global methods clearly outperform cycle-based methods. In particular, we observe that HOC-5 only improves HOC-3 by less than 1.5%, while global methods consistently improve the accuracy of HOC-5 by more than 2% over all datasets. In addition, LR-SIG and LR-SH further improve the accuracy of LR-ALS. This shows that the sigmoid and square-hinge losses are more suitable for sign prediction, which supports the discussion in Section 4.3. On real datasets, we do not have prior information about the target rank k. However, Figure 6 shows that the performance of LR-based methods is not sensitive to k.

In Figure 7, we further select a representative of each category, MOI-10, HOC-5 and LR-ALS, and show their performances with different levels of edge embeddedness (LR-SIG and LR-SH perform similar to LR-ALS in all datasets). In addition, we also compare our methods with the methods A_sym_exp and A_exp proposed by Kunegis et al. (2009), which

predicts the sign of edges using matrix exponential with low rank approximation[3] [4]. For LR-ALS, A_sym_exp and A_exp we choose the rank $k = 40$.

From Figure 7 we see that matrix exponential and MOI methods perform alike as one would expect. HOC learns the weights carefully to determine which configurations of cycles are more important, and therefore performs better than the former two methods that use fixed weights. Also, one might expect that cycle-based approaches should perform better on edges with higher embeddedness because of the relatively richer cycle information available. However, LR-ALS achieves higher prediction accuracy regardless of the embeddedness. All the above results show that global methods are more effective than local methods.



(a) Epinions (b) Slashdot (c) Wikipedia

Figure 7: Sign prediction accuracy of various methods with different levels of embeddedness, along with standard deviation if the accuracy is averaged by 10-fold cross validation. For MOI methods we report leave-one-out accuracy. These plots show the accuracy for edges with embeddedness at least T. The A_sym_exp method in Epinions cannot achieve accuracy 80% for all embeddedness levels so it is not shown in the plot. We can see that LR-ALS consistently achieves the highest accuracy for all thresholds T.

5.3.3 Running Time Comparison

As discussed in Section 4.3, low rank modeling with matrix factorization is more efficient than cycle-based algorithms in terms of time complexity, which we now confirm. The running times are summarized in Table 7. To show the scalability of matrix factorization methods, we construct a large-scale dataset Cluster10, which contains 1.1 millon nodes and 120 million edges (about 100 times more than Epinions). Cluster10 is constructed by uniformly sampling edges from a 10-weakly balanced network, in which clusters have sizes 20000, 40000, ..., 200000. For matrix factorization approach, we report the time needed to solve the model by ALS (with square loss) and SGD (with sigmoid and square-hinge losses). The time LR-SGD is thus the average time to solve LR-SIG and LR-SH models. For HOC methods the training time is dominated by the feature construction step, thus we only report the time for computation of features. Therefore, the reported time for HOC is an underestimate of the time required to construct the HOC model; even then we can see that the time required by LR-ALS, LR-SIG and LR-SH is much lower than HOC methods.

3. The method A_sym_exp considers the symmetric matrix A_sym $= A + A^T$ and its eigen-decomposition $U\Sigma U^T$, and computes the matrix exponential of A_sym with rank-k approximation, $U_k \exp(\Sigma_k) U_k^T$.
4. They incorrectly refer to A_exp as the exponential of A as they in fact compute A_exp as $U_k \exp(\Sigma_k) V_k^T$, where $U_k \Sigma_k V_k^T$ is the best rank-k approximation of A.

Table 7: Running time (in seconds) for low rank modeling methods (LR-ALS and LR-SGD) and supervision on high order cycles (HOC) on real datasets and a 1.1 million node synthetic data Cluster10. We see that LR methods with matrix factorization are clearly more efficient than cycle-based algorithms.

	HOC-3	HOC-4	HOC-5	LR-ALS	LR-SGD
Wikipedia	18.08	74.52	462.92	**2.26**	2.41
Slashdot	133.4	1,936.0	$> 10,000$	**17.4**	24.7
Epinions	560.64	6,156.8	$> 10,000$	**28.67**	37.2
Cluster10	$> 10,000$	$> 10,000$	$> 10,000$	**455.1**	1,152

In conclusion, for the sign prediction problem, the low rank model with matrix factorization is clearly the best method in terms of accuracy and scalability.

5.4 Clustering

We now show that our clustering approach, which completes the low-rank structure of signed networks before performing clustering, outperforms spectral clustering based on the signed Laplacian (Kunegis et al., 2010). We conduct experiments on synthetic data generated from weakly balanced networks (note that we do not have ground truth for clustering in the real-life datasets). We consider a 10-weakly balanced network $A^\star$ where size of each group is 100, and sample entries from $A^\star$ using uniform sampling and uniform sampling with noise.

To measure the performance of clustering, we calculate the number of edges that satisfy the ground-truth clustering, which is defined by

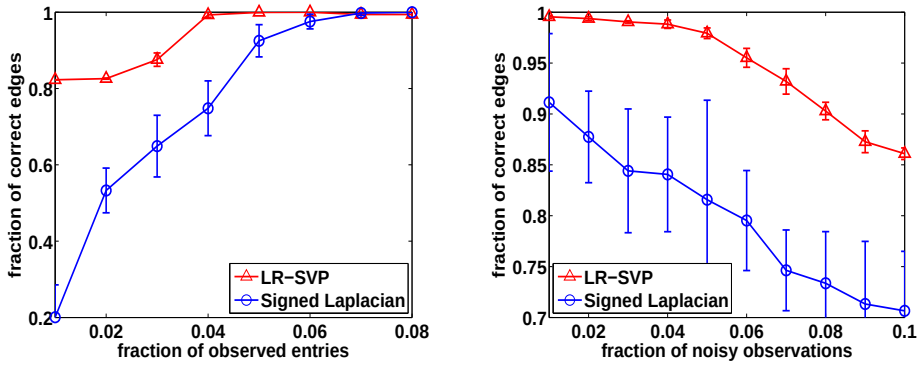
$$\sum_{i,j:s_i=s_j} I(\bar{s_i} = \bar{s_j}) + \sum_{i,j:s_i \neq s_j} I(\bar{s_i} \neq \bar{s_j}). \tag{19}$$

where $s_1, \ldots, s_n$ denote the ground-truth clustering assignment for each node, and $\bar{s}_1, \ldots, \bar{s}_n$ are the clustering results given by the clustering algorithm.

Following the procedure outlined in Section 5.3, in the uniform sampling case, we consider the networks $A = A^\star(s, 0, \mathcal{D}_{\text{uni}})$ with $s \in [0.01, 0.06]$, while in sampling with noise case we consider networks $A = A^\star(0.05, \epsilon, \mathcal{D}_{\text{uni}})$ with $\epsilon \in [0.01, 0.08]$. For each observed network, we apply Algorithm 2 (see Section 4.5) and clustering via the signed Laplacian, and evaluate clustering results using (19). The results of these two scenarios are shown in Figure 8. In both the scenarios, our proposed clustering approach is significantly better than clustering based on the signed Laplacian, and shows that recovering the low-rank structure of signed networks leads to improved clustering results.

6. Related Work

Signed networks have been studied since the early 1950s. Harary and Cartwright were the first to mathematically study structural balance. They defined balanced triads and proved the global structure of balanced signed networks (Harary, 1953; Cartwright and Harary, 1956). Davis (1967) further generalized the balance notion to weak balance by allowing



(a) Data without noise (b) Data with noise

Figure 8: Clustering partially observed synthetic data - clustering with matrix completion using SVP (LR-SVP) performs significantly better than clustering with the signed Laplacian.

triads with all negative edges, and showed that weakly balanced graphs have the global structure of mutual antagonistic groups.

Though theoretical studies of signed networks have been conducted for a long time, it was not until this decade that analysis of real signed networks could be done at a large scale as large real networks have become more accessible recently. For example, Kunegis et al. (2009) performed several analysis tasks on Slashdot, and Leskovec et al. (2010a,b) studied the local and global structure of three real signed networks. They designed several computational experiments to justify that the structure of these signed networks match balance theory.

In this paper, we focused on problems in signed networks. However, these problems have their counterparts in unsigned networks. For instance, structural link prediction in unsigned networks corresponds to the sign prediction problem. Structural link prediction has been well explored, and it is usually solved by computing a similarity measure between nodes (Liben-Nowell and Kleinberg, 2007), such as those proposed by Katz (1953) and Adamic and Adar (2003). The sign prediction problem, however, was not formally considered until the work by Guha et al. (2004), in which they develop a trust propagation framework to predict trust or distrust between entities. More recently, Kunegis et al. (2009, 2010) reconsidered this problem by using various similarity functions and kernels such as matrix exponential and signed Laplacian. Leskovec et al. (2010a) proposed a machine learning formulation of this problem, arguing that learning from only local triangular structure of edges can achieve reasonable accuracy.

Sign prediction using our global method is closely related to the low-rank matrix completion problem. In the last five years, there has been substantial research studying exact recovery conditions for this problem (Candés and Recht, 2008; Candés and Tao, 2009), and algorithms with theoretical guarantees have also been proposed, such as SVT (Cai et al., 2010) and SVP (Jain et al., 2010). While the matrix completion problem has been consid-

ered mostly in collaborative filtering, our low rank model arises naturally from the weak balance of signed networks.

Clustering is another fundamental problem in network analysis. For unsigned networks, there are several proposed algorithms that have been shown to be effective, such as clustering via graph Laplacian (Ng et al., 2001), modularity (Newman, 2006) and multilevel approaches (Dhillon et al., 2007). However, most of these approaches cannot be directly extended to signed networks since weak balance theory does not apply to unsigned networks. As a result, researchers have tried to tailor unsigned network clustering algorithms in order to make them applicable to signed networks. For instance, Doreian and Mrvar (1996) proposed a local search strategy which is similar to the Kernighan-Lin algorithm (Kernighan and Lin, 1970). Starting with an initial clustering assignment, it tries to move nodes one by one to get a more preferable clustering. Yang et al. (2007) proposed an agent-based method which basically conducts a random walk on the graph. Kunegis et al. (2010) generalized spectral algorithms to signed networks. They proposed a spectral approach using the signed Laplacian, and showed that partitioning signed networks into two groups using the signed Laplacian kernel is analogous to considering ratio cut on unsigned networks. Anchuri and Magdon-Ismail (2012) proposed hierarchical iterative methods that solve 2-way signed modularity objectives using spectral relaxation at each hierarchy. Chiang et al. (2012) proposed graph kernels for signed network clustering, and showed that the multilevel framework can be extended to this problem.

Another line of research related to signed graph clustering problem is correlation clustering. The goal of correlation clustering is the following: given n objects where certain pairs of objects are labelled as similar and certain pairs as dissimilar, find a clustering that maximizes the number of similar pairs within clusters, plus the number of dissimilar pairs between clusters. The problem was first considered by Bansal et al. (2004), who proved that finding the optimal correlation clustering is NP-hard, and proposed two approximation algorithms to maximize the number of edges that satisfy "agreement" (a positive within-cluster edge or a negative between cluster-edge) and to minimize the number of edges that do not, under the special case that all pairwise label information is given. Bounds for general correlation clustering setting have been obtained by Demaine et al. (2006). On the other hand, some researchers have also considered the correlation clustering problem from the statistical learning theory viewpoint. For example, Joachims and Hopcroft (2005) give error bounds for the problem if only partial pairs are observed. Recently, Cesa-Bianchi et al. (2012) proposed a method for sign prediction by learning a correlation clustering index. They consider three types of learning models: batch, online and active learning, and provide theoretical bounds for prediction mistakes under each setting. Though there is no social balance notion in the correlation clustering problem, it can be viewed as finding a clustering of signed graph where nodes correspond to objects, and positive/negative edges correspond to similar/dissimilar pairs. Therefore, our proposed method can also be applied to the problem of correlation clustering.

7. Conclusions and Future Work

In this paper, we studied the usefulness of social balance in signed networks, with two fundamental applications: sign prediction and clustering. Starting from a local view of social

balance, we developed sign prediction methods based on length-ℓ cycles. The predictive accuracies are improved if longer cycles are taken into consideration, suggesting that a broader view of local patterns helps in sign prediction. We then considered the global perspective on social balance, and showed that the adjacency matrices of (weakly) balanced networks are low rank. Based on this observation, we modeled the sign prediction problem as a low-rank matrix completion problem. We discussed three approaches to matrix completion: convex relaxation, singular value projection, and matrix factorization. In addition, we applied this low rank modeling technique to the clustering problem. In experiments, we observe that sign prediction via matrix factorization not only outperforms local methods (MOI and HOC), but requires much less running time. Clustering results are also more favorable via the matrix completion approach in comparison with the existing signed Laplacian approach. All of these results consistently demonstrate the effectiveness of the global viewpoint of social balance.

For future work, one possible direction is to explore analysis tasks for heterogeneous signed networks. Since there are different types of entities in heterogeneous networks, currently there are no clear answers to questions such as: do balance relationships exist on such networks? How do we quantitatively measure balance if balance patterns exist? How is balance at a local level related to the global structure? Furthermore, another possible direction is to examine other theories for directed signed networks. Leskovec et al. (2010a,b) has found evidence that status theory holds in real signed networks, but that the patterns conforming to status theory are quite different from those conforming to balance theory. Thus, it is natural to ask how to design algorithms by pursuing global patterns conforming to status theory. These interesting directions are worth exploring in future research.

Acknowledgments

We gratefully acknowledge the support of NSF grants CCF-0916309, CCF-1117055, and DOD Army grant W911NF-10-1-0529. Most of the contribution of Ambuj Tewari to this work occurred while he was a postdoctoral fellow at the University of Texas at Austin.

References

Lada A. Adamic and Eytan Adar. Friends and neighbors on the web. *Social Networks*, 25 (3):211–230, 2003.

Pranay Anchuri and Magdon-Ismail. Communities and balance in signed networks: A spectral approach. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining*, pages 235–242, 2012.

Nikihil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. *Machine Learning*, 56:89–113, 2004.

Jian-Feng Cai, Emmanuel J. Candés, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *Society for Industrial and Applied Mathematics (SIAM)*, 20(4):1956–1982, 2010.

Emmanuel J. Candés and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9:712–772, 2008.

Emmanuel J. Candés and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transaction of Information Theory*, 56(5):2053–2080, 2009.

Dorwin Cartwright and Frank Harary. Structure balance: A generalization of Heider's theory. *Psychological Review*, 63(5):277–293, 1956.

Nicolo Cesa-Bianchi, Claudio Gentile, Fabio Vitale, , and Giovanni Zappella. A correlation clustering approach to link classification in signed networks. In *The 25th Annual Conference on Learning Theory*, pages 34.1–34.20, 2012.

Kai-Yang Chiang, Nagarajan Natarajan, Ambuj Tewari, and Inderjit S. Dhillon. Exploiting longer cycles for link prediction in signed networks. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pages 1157–1162, 2011.

Kai-Yang Chiang, Joyce Whang, and Inderjit S. Dhillon. Scalable clustering of signed networks using balance normalized cut. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pages 615–624, 2012.

Fan Chung, Linyuan Lu, and Van Vu. Spectra of random graphs with given expected degrees. *Internet Mathematics*, 1:6313–6318, 2004.

James A. Davis. Clustering and structural balance in graphs. *Human Relations*, 20(2): 181–187, 1967.

Erik D. Demaine, Dotan Emanuel, Amos Fiat, and Nicole Immorlica. Correlation clustering in general weighted graphs. *Theoretical Computer Science*, 361(2-3):172–187, 2006.

Inderjit S. Dhillon, Yuqiang Guan, and Brian Kulis. Weighted graph cuts without eigenvectors: A multilevel approach. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 29(11):1944–1957, 2007.

Patrick Doreian and Andrej Mrvar. A partitioning approach to structural balance. *Social Networks*, 18(21):149–168, 1996.

David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. 2010.

P. Erdös and A. Rényi. On the evolution of random graphs. In *Publication of the mathematical institute of the Hungarian academy of sciences*, pages 17–61, 1960.

Maryam Fazel, Haitham Hindi, and Stephen P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *American Control Conference*, volume 6, pages 4734–4739, 2001.

R. V. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th International Conference on World Wide Web*, pages 403–412, 2004.

Frank Harary. On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 2(2):143–146, 1953.

Cho-Jui Hsieh, Kai-Yang Chiang, and Inderjit S. Dhillon. Low rank modeling of signed networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 507–515, 2012.

Prateek Jain, Reghu Meka, and Inderjit S. Dhillon. Guaranteed rank minimization via singular value projection. In *Advances in Neural Information Processing Systems*, pages 934–945, 2010.

Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, pages 665–674, 2013.

Thorsten Joachims and John Hopcroft. Error bounds for correlation clustering. In *Proceedings of the 22nd international conference on Machine learning*, ICML '05, New York, NY, USA, 2005. ACM.

Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103, 1972.

Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1): 39–43, 1953.

Brian W. Kernighan and Shen Lin. An efficient heuristic procedure for partitioning graphs. *Bell system technical journal*, 49(1):291–307, 1970.

Yehuda Koren, Robert M. Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42:30–37, 2009.

Jérôme Kunegis, Andreas Lommatzsch, and Christian Bauckhage. The slashdot zoo: Mining a social network with negative edges. In *Proceedings of the 18th International Conference on World Wide Web*, pages 741–750, 2009.

Jérôme Kunegis, Stephan Schmidt, Andreas Lommatzsch, Jürgen Lerner, Ernesto W. DeLuca, and Sahin Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proceedings of the SIAM International Conference on Data Mining*, pages 559–570, 2010.

Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Predicting positive and negative links in online social networks. In *Proceedings of the 19th International Conference on World Wide Web*, pages 641–650, 2010a.

Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1361–1370, 2010b.

David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031, 2007.

Aditya Krishna Menon and Charles Elkan. Link prediction via matrix factorization. *Proceedings of the 2011 European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 437–452, 2011.

M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of USA*, 103(23):8577–8582, 2006.

Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: analysis and an algorithm. In *advances in neural information processing systems*, pages 849–856. MIT Press, 2001.

Arnout van de Rijt. The micro-macro link for the theory of structural balance. *Journal of Mathematical Sociology*, 35(1):94–113, 2011.

Bo Yang, William Cheung, and Jiming Liu. Community mining from signed social networks. *IEEE Transaction on Knowledge and Data Engineering*, 19(10):1333–1348, 2007.

Hsiang-Fu Yu, Cho-Jui Hsieh, Si Si, and Inderjit S. Dhillon. Parallel matrix factorization for recommender systems. *Knowledge and Information Systems(KAIS)*, 2013.

Appendix A: Proofs

Proof of Theorem 5 One direction is trivial. If A is unbalanced then there is an unbalanced simple cycle. However, any simple cycle is obviously a cycle and hence the sum in (6) will be strictly positive.

For the other direction, suppose $\mu_\infty(A) > 0$. This implies there is an unbalanced cycle σ in the graph. Decompose the unbalanced cycle into finitely many simple cycles. We will be done if we could show that one of these simple cycles has to be unbalanced. It is easy to see why this is true: if all of these simple cycles were balanced, they all would have had an even number of negative edges, but then the total number of negative edges in σ could not have been odd. ■

Proof of Lemma 6 Define the sets of ℓ-cycles,

$$C_\ell^+(i,j) := \{\sigma \in C_\ell\Big(A^{+(i,j)}\Big) : \sigma \text{ includes } (i,j)\}$$
$$C_\ell^-(i,j) := \{\sigma \in C_\ell\Big(A^{-(i,j)}\Big) : \sigma \text{ includes } (i,j)\} \ ,$$

that include the edge (i,j). Note that, since $A^{+(i,j)}$ and $A^{-(i,j)}$ only differ in the sign of the edge (i,j), we have,

$$C_\ell\Big(A^{+(i,j)}\Big)\backslash C_\ell^+(i,j) = C_\ell\Big(A^{-(i,j)}\Big)\backslash C_\ell^-(i,j) \ .$$

Thus, we have,

$$\begin{aligned}
&\sum_{\sigma\in C_\ell\left(A^{-(i,j)}\right)} \mathbf{1}\left[\sigma\right] - \sum_{\sigma\in C_\ell\left(A^{+(i,j)}\right)} \mathbf{1}\left[\sigma\right] \\
&= \sum_{\sigma\in C_\ell^-(i,j)} \mathbf{1}\left[\sigma\right] + \sum_{\sigma\in C_\ell\left(A^{-(i,j)}\right)\backslash C_\ell^-(i,j)} \mathbf{1}\left[\sigma\right] - \sum_{\sigma\in C_\ell^+(i,j)} \mathbf{1}\left[\sigma\right] - \sum_{\sigma\in C_\ell\left(A^{+(i,j)}\right)\backslash C_\ell^+(i,j)} \mathbf{1}\left[\sigma\right] \\
&= \sum_{\sigma\in C_\ell^-(i,j)} \mathbf{1}\left[\sigma\right] - \sum_{\sigma\in C_\ell^+(i,j)} \mathbf{1}\left[\sigma\right] \ . \qquad (20)
\end{aligned}$$

Now cycles in $C_\ell^-(i,j)$ are in 1-1 correspondence with paths π in $\mathcal{P}_{\ell-1}(i,j)$ of length $\ell-1$, in the original graph A, that go from i to j. Moreover, $\sigma \in C_\ell^-(i,j)$ is unbalanced iff the corresponding path in $\mathcal{P}_{\ell-1}(i,j)$ has an *even* number of -1's. Similarly, $\sigma \in C_\ell^+(i,j)$ is unbalanced iff the corresponding path in $\mathcal{P}_{\ell-1}(i,j)$ has an *odd* number of -1's. Thus, continuing from (20):

$$\begin{aligned}
&\sum_{\sigma\in C_\ell^-(i,j)} \mathbf{1}\left[\sigma\right] - \sum_{\sigma\in C_\ell^+(i,j)} \mathbf{1}\left[\sigma\right] \\
&= \sum_{\pi\in\mathcal{P}_{\ell-1}(i,j)} \mathbf{1}\left[\pi \text{ has even no. of } -1\text{'s}\right] - \sum_{\pi\in\mathcal{P}_{\ell-1}(i,j)} \mathbf{1}\left[\pi \text{ has odd no. of } -1\text{'s}\right] \\
&= \sum_{i_1,i_2,\ldots,i_{\ell-2}} A_{i,i_1}\cdot A_{i_1,i_2}\cdot\ldots\cdot A_{i_{\ell-2},j} = \left(A^{\ell-1}\right)_{i,j} \ ,
\end{aligned}$$

where the second equality is true because A only has $\pm1, 0$ entries. ■

Appendix B: Supervised Higher-Order Cycle (HOC) methods

We begin by describing the approach used by Leskovec et al. (2010a). The features for an edge are based on the sign configurations of triads it is a part of. Fix an edge $e = (i, j)$. Consider an arbitary common neighbor (in an undirected sense) k of i and j. Links between i and k can be in 4 possible configurations:

$$i \xrightarrow{+} k \qquad i \xleftarrow{+} k$$
$$i \xrightarrow{-} k \qquad i \xleftarrow{-} k \ .$$

Similarly, there are 4 possible configurations for links between k and j. Thus, we can get a total of 16 features for the edge e by considering the number of common neighbors k in each of the $4 \times 4 = 16$ configurations. Though we draw 4 configurations separately, links with different directions can simultaneously exist between i and j, possibly with different sign.

These configurations corresponds to features for a supervised variant of the k-cycle method for $k = 3$. Let A^+ and A^- be the matrices of positive and negative edges such that $A = A^+ + A^-$. In terms of matrix powers, these sixteen features are nothing but the (i, j) entries in the sixteen matrices:

$$A^{b_1} \cdot A^{b_2} \qquad A^{b_1} \cdot \left(A^{b_2}\right)^T \qquad \left(A^{b_1}\right)^T \cdot A^{b_2} \qquad \left(A^{b_1}\right)^T \cdot \left(A^{b_2}\right)^T \ , \tag{21}$$

where $b_1, b_2 \in \{\pm\}$, and $(A^{b_1})^T$ denotes the transpose of A^{b_1}. Note that we have described the features of a *directed* edge $e = (i, j)$.

Using higher-order cycles

A criticism against using only these triangle-based features is that there could be many people in the social network who do not share friends. In fact, this is the case in most of the networks used by Leskovec et al. (2010a). The reason their method is able to predict well on such pairs is that they additionally use seven other "degree-type" features like in-degree and out-degree (and their signed variants). Thus, the prediction for an edge with zero emdeddedness (embeddedness refers to the number of common neighbors of the vertices of an edge) relies completely on the degree-based features. We can additionally incorporate features from higher-order cycles. Generalizing the construction (21), we can define 64 fourth-order features (corresponding to 4-cycles in the graph) of an edge (i, j) as the (i, j) entries in the matrices:

$$\left(A^{b_1}\right)^{t_1} \cdot \left(A^{b_2}\right)^{t_2} \cdot \left(A^{b_3}\right)^{t_3} \ , \tag{22}$$

where $b_i \in \{\pm\}$ indicates whether we look at the positive or negative part of A and $t_i \in \{T, 1\}$ indicates whether or not we transpose it. There are 4 possibilities for each b_i, t_i pair, resulting in a total of $4 \times 4 \times 4 = 64$ possibilities.

By now the reader can guess the construction of features of a general order $\ell \geq 3$. For the edge (i, j), they will be the (i, j) entries in the $4^{\ell-1}$ matrices

$$\left(A^{b_1}\right)^{t_1} \cdot \left(A^{b_2}\right)^{t_2} \cdots \cdot \left(A^{b_{\ell-1}}\right)^{t_{\ell-1}} \ , \tag{23}$$

with $b_i \in \{\pm\}, t_i \in \{T, 1\}$.

Note that the number of features is exponential in ℓ, and therefore it is not feasible to obtain features from arbitrarily long cycles. We use $\ell \leq 5$ for supervised HOC methods in our experiments that are presented in Section 5.

Reducing the Number of Features

The number of features can quickly become unmanageable, and computationally infeasible, as soon as ℓ is beyond 5. While dimensionality of the feature space may be the primary concern, the combinatorial nature of the features also raises the following intuitive concern: the interpretability of features rendered by high-order cycles, say when $\ell = 6$, composed of different signs and directions, is a challenge. For example, it is intuitively hard to appreciate the difference between the two walks $i \xrightarrow{+} k_1 \xrightarrow{+} k_2 \xrightarrow{-} k_3 \xrightarrow{+} k_4 \xrightarrow{+} j$ and $i \xrightarrow{+} k_1 \xrightarrow{+} k_2 \xleftarrow{-} k_3 \xrightarrow{+} k_4 \xrightarrow{+} j$.

With this realization, one way to quickly reduce the number of features, yet retain the information in longer cycles, is to consider the underlying undirected graph, ignoring the directions. In particular, the ℓ^{th} order features will be from the matrices

$$A^{b_1} \cdot A^{b_2} \ldots \cdot A^{b_{k-1}}, \tag{24}$$

with $b_i \in \{\pm\}$. Note that since we are considering the undirected graph, we ensure that the features are symmetric by summing features of the form $A^{b_1}A^{b_2}$ and $A^{b_2}A^{b_1}$. Thus the number of ℓ^{th} order features to compute is reduced to $O(2^\ell)$ from $O(4^\ell)$. Though the number of features is still exponential in ℓ, the construction of features becomes easier for small values of ℓ.

We note that another way to avoid dealing with too many features is to use a *kernel* instead. A kernel computes inner products in feature space without explicitly constructing the feature map. One can then use off-the-shelf SVM classifiers to perform the classification. We leave this promising approach of directly defining a kernel on *pairs of nodes* of a graph and using it for link prediction to future work.

Classifier

We use a simple logistic regression where the imbalance of an edge is modeled as a linear combination of the features, which are imbalances in cycles of various lengths and characteristics themselves. Let V be the set of vertices in the network and $\Phi : V \times V \to \mathbb{R}^p$ denote the feature map. Then,

$$P(A_{ij} = +1) = \frac{1}{1 + \exp\left(-w_0 - \langle \mathbf{w}, \Phi(i,j) \rangle\right)},$$

using which logistic regression is used to learn w_0 and the weight vector $\mathbf{w} = [w_1 \cdots w_p]^T \in \mathbb{R}^p$. The prediction of any query (i, j) is then given by $\text{sign}(P(A_{ij} = +1) - 0.5)$.