Low Rank Modeling of Signed Networks

Cho-Jui Hsieh
Dept. of Computer Science
University of Texas at Austin
Austin, TX 78712-1188, USA
cjhsieh@cs.utexas.edu

Kai-Yang Chiang
Dept. of Computer Science
University of Texas at Austin
Austin, TX 78712-1188, USA
kychiang@cs.utexas.edu

Inderjit S. Dhillon
Dept. of Computer Science
University of Texas at Austin
Austin, TX 78712-1188, USA
inderjit@cs.utexas.edu

ABSTRACT

Trust networks, where people leave trust and distrust feedback, are becoming increasingly common. These networks may be regarded as signed graphs, where a positive edge weight captures the degree of trust while a negative edge weight captures the degree of distrust. Analysis of such signed networks has become an increasingly important research topic. One important analysis task is that of sign inference, i.e., infer unknown (or future) trust or distrust relationships given a partially observed signed network. Most state-of-the-art approaches consider the notion of structural balance in signed networks, building inference algorithms based on information about links, triads, and cycles in the network. In this paper, we first show that the notion of weak structural balance in signed networks naturally leads to a global low-rank model for the network. Under such a model, the sign inference problem can be formulated as a low-rank matrix completion problem. We show that we can perfectly recover missing relationships, under certain conditions, using state-of-the-art matrix completion algorithms. We also propose the use of a low-rank matrix factorization approach with generalized loss functions as a practical method for sign inference — this approach yields high accuracy while being scalable to large signed networks, for instance, we show that this analysis can be performed on a synthetic graph with 1.1 million nodes and 120 million edges in 10 minutes. We further show that the low-rank model can be used for other analysis tasks on signed networks, such as user segmentation through signed graph clustering, with theoretical guarantees. Experiments on synthetic as well as real data show that our low rank model substantially improves accuracy of sign inference as well as clustering. As an example, on the largest real dataset available to us (Epinions data with 130K nodes and 840K edges), our matrix factorization approach yields 94.6% accuracy on the sign inference task as compared to 90.8% accuracy using a state-of-the-art cycle-based method — moreover, our method runs in 40 seconds as compared to 10,000 seconds for the cycle-based method.

1. INTRODUCTION

Social network analysis has received a lot of attention recently. Traditionally, online networks such as Facebook or World Wide Web can be viewed as graphs, with nodes representing entities, and edges representing relationships between entities. Recently, trust networks have also become increasingly common where two opposite kinds of relationships exist between entities. For example, online review websites such as Epinions allow users to either like or dislike others' reviews. Such networks can be modeled as *signed networks*, where edge weights are $+1$ or -1, representing positive or negative relationships respectively.

Perhaps the most basic yet significant belief in signed networks is *structural balance* [11, 4]. Structural balance states that people in signed networks tend to follow patterns such as "an enemy of my friend is my enemy" and "an enemy of my enemy is my friend", and so on [4]. It is important to note that since balance notion applies only to signed networks, therefore, algorithms for signed networks can be somewhat different from algorithms for unsigned networks. Structural balance has been shown to be useful for analysis tasks for signed networks. For instance, the *sign inference problem*, which aims to infer the unknown relationship between two entities, can be achieved by learning from balance information of signed networks [17, 5]. Nevertheless, these state-of-the-art methods for sign inference problem mainly consider structural balance, while a more general notion - weak balance [7] - is not taken into account. Therefore, it is natural to ask what can be further inferred from weak balance.

In this paper, we propose a *low rank model* by observing that complete weakly balanced networks have a low-rank structure. Therefore, many analysis tasks such as sign inference and clustering can be posed as low-rank matrix completion. The advantages of taking a matrix completion approach are as follows. First, many matrix completion algorithms provide theoretical recovery guarantees under certain conditions [19, 12]. Moreover, many algorithms such as Alternating Least Squares (ALS) and Stochastic Gradient Descent (SGD) can efficiently find effective solutions for problems with billions of nonzero entries [14]. We will empirically demonstrate that our proposed low rank model is both accurate and scalable in many applications.

We summarize the contributions of this paper:

- We explore a new model for signed networks - the low rank model - from both theoretical and practical points of view. We show that the low-rank structure arises naturally from weak balance theory. In practice, we show that real-life signed networks tend to have low-rank structure, so it is natural to apply the low rank model to real datasets.
- We show that the sign inference problem can be formulated as a low-rank matrix completion problem. Under certain conditions, we prove that matrix completion can perfectly recover missing links. We also apply matrix factorization algorithms to sign inference problems, and these turn out to be much more accurate and efficient than other inference methods.
- We show that our low rank model can be used for clustering. Our algorithm first infers missing values using matrix completion, and then performs the clustering, thus enabling us to get guarantees for clustering under certain conditions. This theoretical guarantee cannot be achieved by other spectral methods [15].

The paper is organized as follows. In Section 2, we review some recent work related to this paper. In Section 3, we propose our low rank model, and show that the sign inference problem can be modeled as a low-rank matrix completion problem. We present two approaches; matrix completion and matrix factorization. In Section 4, we show how to do clustering using the low rank model. In Section 5, we conduct experiments which convincingly demonstrate that our low rank model improves sign inference accuracy as well as clustering results. Finally, we present our conclusions in Section 6.

2. RELATED WORK

Signed network analysis has a rich history dating back to the 1950s — the notion of structural balance was formulated and analyzed by Harary and Carwright [11, 4], who formally defined balanced triads and proved global structural results for signed balanced networks (stated as Theorem 1 in Section 3). Davis [7] further generalized the notion of balance to weak balance, by allowing triads where all edges are negative. In Section 3, we will elaborate on weak balance and show that it naturally leads to our proposed low-rank model.

An important analysis task on signed networks is the sign inference problem. This problem was first considered by Guha et al.[9]. More recently, Kunegis et al.[15, 16] reconsidered this problem by using varied similarity functions and kernels such as matrix exponential and signed Laplacian, on the signed link structure of the network. Leskovec et al.[17] proposed a machine learning formulation of this problem, arguing that learning from only local triangular structure of edges can achieve high accuracy. Chiang et al.[5] generalized [17] by showing that longer cycles in the signed network reveal balance information in the network — using these additional "features" for learning led to an improvement in inference accuracy. Our modeling approach in this paper is distinct from existing work, as we first show that the global viewpoint of structural balance (as opposed to the local triad structure) naturally leads to a low-rank model for the network, and then show that the sign inference problem may be regarded as a low-rank matrix completion problem. Our approach is much more scalable than previous approaches, and leads to higher inference accuracy as well. Some recent work also considers the link inference problem as network completion [10, 13]. The goal in [10, 13] is to reconstruct the underlying (unsigned) network topology given partially observed links and/or nodes. In contrast, the main goal of our work is to infer the (unobserved) signed relationships between all pairs of entities. Moreover, our low-rank matrix completion approach arises from the notion of weak structural balance, which only applies in signed networks.

Clustering or community detection is another important task in network analysis, and has been well studied for unsigned social networks using many varied approaches [18, 8]. However, extending these algorithms to signed networks is not obvious since it has been shown that clustering on signed networks is highly related to (weak) balance theory [11, 4, 7]. Thus several tailored approaches have been proposed for clustering of signed networks [20]. Recently, Kunegis et al.[16] proposed a spectral approach using the so-called "signed" Laplacian, and showed that partitioning signed networks using the signed Laplacian kernel is analogous to considering ratio cut on unsigned networks. Our approach is somewhat similar to [16] in that we also consider the spectra of signed graphs. However, our clustering algorithm proceeds by first completing the underlying graph using low-rank matrix completion, and then performing the clustering. This important difference makes our clustering results much more reliable, especially on graphs where the observed signed relationships are sparse.

Sign inference using our low rank model is closely related to matrix completion problem. In the last five years, there has been substantial research that has studied exact recovery conditions for this problem [19, 3, 2], and algorithms with theoretical guarantees have also been proposed [1, 12]. Matrix factorization is another approximation technique for matrix completion. Though this approach is notoriously hard to analyze, it is very competitive in practice [14]. While the matrix completion problem is considered mostly in collaborative filtering, our low rank model arises naturally from weak balance of signed networks. We will discuss the details of matrix completion and matrix factorization in Section 3.

3. LOW RANK MODELING

In this section, we investigate the structure of signed networks and show that a low-rank structure intuitively emerges when we consider so-called weakly balanced networks. First, we introduce a few preliminaries before going into details.

We will consider a signed network to be a graph $G = (V, E, A)$, where V is the vertex set of size n, E is the edge set of size m, and $A \in \mathbb{R}^{n \times n}$ is the signed adjacency matrix associated with G. The entries of A are as follows:

$$A_{ij} = \begin{cases} 1, & \text{if } i \ \& \ j \text{ trust each other,} \\ -1, & \text{if } i \ \& \ j \text{ distrust each other,} \\ 0, & \text{if relationship of } (i,j) \text{ is unknown (or missing)} \end{cases}$$

For now we restrict ourselves to symmetric relationships, i.e., A is a symmetric matrix, although asymmetric relationships are possible. Note that for us, $A_{ij} = 0$ does not imply no relationship between i and j; it just means we do not currently observe the relationship. We denote the set of known entries by Ω, i.e. $(i,j) \in \Omega$ iff A_{ij} is observed. A network is complete if all entries in A are observed.

We first review some basics of structural balance. A complete network is called (strongly) *balanced* if all triads in the graph have (i) all positive edges or (ii) only one posi-

$$
\begin{pmatrix}
\boxed{1 \quad 1 \quad 1} & -1 & -1 & -1 \\
1 \quad 1 \quad 1 & -1 & -1 & -1 \\
1 \quad 1 \quad 1 & -1 & -1 & -1 \\
-1 & -1 & -1 & 1 & -1 & -1 \\
-1 & -1 & -1 & -1 & 1 & 1 \\
-1 & -1 & -1 & -1 & 1 & 1
\end{pmatrix}
=
\begin{pmatrix}
-1.045 & 0.265 & -0.402 \\
-1.045 & 0.265 & -0.402 \\
-1.045 & 0.265 & -0.402 \\
0.319 & -1.259 & -0.830 \\
0.919 & 0.607 & -0.541 \\
0.919 & 0.607 & -0.541
\end{pmatrix}
\begin{pmatrix}
-1.045 & -1.045 & -1.045 & 0.319 & 0.919 & 0.919 \\
0.265 & 0.265 & 0.265 & -1.259 & 0.607 & 0.607 \\
-0.402 & -0.402 & -0.402 & -0.830 & -0.541 & -0.541
\end{pmatrix}
$$

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

tive edge. With this local definition, it can be shown that complete balanced networks have a special global structure:

Theorem 1 (Global "Strong Balance" Structure [11])
A complete network is balanced iff all edges are positive, or the vertices can be divided into two different groups such that all edges within the same group are positive, and all edges between the two groups are negative.

The assumption that strong structural balance exists in a real signed network might be too extreme, so a more relaxed notion of balance is the so-called "weak" structural balance theory. The formal definition of weak balance is as follows:

Definition 1 (Weak Balance for Complete Graphs [7])
A complete signed network is weakly balanced iff there is no triad in the network that contains two positive edges and one negative edge.

Similar to Theorem 1, this local definition implies a simple global structure of weakly balanced networks:

Theorem 2 (Global "Weak Balance" Structure [7])
A complete signed network is weakly balanced iff all edges are positive, or the vertices can be divided into several groups such that within-group edges are positive and between-group edges are negative.

Thus we can say that a network is k-weakly balanced iff it can be perfectly divided into k groups, $k \in \mathbb{N}$. Note that Theorem 1 can be regarded as a special case of Theorem 2 with $k \leq 2$. We now show that the adjacency matrix A of a complete k-weakly balanced network is low rank. With a suitable reordering of nodes, A can be represented as a block-diagonal matrix where all entries within the diagonal blocks are $+1$'s, and all entries within off-diagonal blocks are all -1's. The following theorem proves that the adjacency matrix of a complete k-weakly balanced network has rank up to k.

Theorem 3 (Low Rank Structure of Signed Networks)
The adjacency matrix A 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 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 column vectors $A_{:,s_1^{(i)}}, \ldots, A_{:,s_{n_i}^{(i)}}$ are all identical to 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 is spanned by $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$.

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

consider $k > 2$. In this case, we argue that $\text{rank}(A) = k$ by showing that $\mathbf{b}_1, \ldots, \mathbf{b}_k$ are linearly independent. We consider the following $k \times k$ square matrix:

$$
M =
\begin{bmatrix}
1 & -1 & \cdots & -1 & -1 \\
-1 & 1 & \cdots & -1 & -1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
-1 & -1 & \cdots & 1 & -1 \\
-1 & -1 & \cdots & -1 & 1
\end{bmatrix}
$$

It is obvious that $\mathbf{1} = [1 \ 1 \cdots 1]^T$ is an eigenvector of M with eigenvalue $-(k-2)$. We can further construct $k-1$ linearly independent eigenvectors all 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 ith 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 $\text{rank}(A) = k$. 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 $\sum_{i=1}^{k} \alpha_i \mathbf{b}_i = 0$. Using this set of α's, it is easy to see that $\sum_{i=1}^{k} \alpha_i M_{:,i} = 0$, but this contradicts the fact that $\text{rank}(M) = k$. Therefore, $\text{rank}(A) = k$. $\square$

Figure 1 is an example of a complete 3-weakly balanced network. As shown, its adjacency matrix can be expressed as a product of two rank-3 matrices, indicating its rank is no more than three. By Theorem 3, we can conclude that $\text{rank}(A) = 3$.

The above reasoning shows that complete weakly balanced graphs are low rank, however, most real networks are not complete graphs. One way to define balance on graphs that are not complete is to try to fill in the unobserved or missing edges(relationships) so that balance is obtained:

Definition 2 (Weak Balance for General Graphs)
A signed network is weakly balanced iff it is possible to add missing edges to the network, with appropriate sign, so that the resulting complete graph is weakly balanced.

Hence, sign inference in trust networks can be thought of as a low-rank matrix completion problem. Specifically, given a signed network with observed edges $A_{ij}, (i,j) \in \Omega$, we want to find a complete matrix by assigning ± 1 to every unknown entry, such that the resulting complete graph is (nearly) weakly balanced and hence, the completed matrix is low rank. Thus, our missing value estimation problem can be formulated as:

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

Once we obtain the minimizer of (1), which we will denote by X^*, we can infer the missing relationship between i and j by simply looking up the sign of the entry X^*_{ij}. However, it is known that solving (1) is NP-hard in general. Recent research on low-rank matrix completion has shown the surprising result that in many cases, problem (1) can be solved to yield the *global* optimal in polynomial time [3]. In the following subsections, we identify such conditions as well as approaches to approximately solve (1) for real-world signed networks.

3.1 Inference via Matrix Completion

One possible approximate solution for (1) can be obtained by dropping the discrete constraints and replacing $\text{rank}(X)$ by $\|X\|_*$, where $\|X\|_*$ denotes the trace norm of X, which is the tightest convex relaxation of rank. Thus, a convex relaxation of (1) is:

$$\text{minimize} \ \ \|X\|_* \ \text{s.t.} \ \ X_{ij} = A_{ij}, \ \forall \ (i,j) \in \Omega. \quad (2)$$

It turns out that, under certain condition, by solving (2) we can recover the *exact* missing relationships from the underlying complete signed network. This surprising result is the consequence of recent research [3, 2] which has shown that perfect recovery from the observations is possible if the observed entries are uniformly sampled and X^* has high incoherence, which may be defined as follows:

Definition 3 (Incoherence)
An $m \times n$ matrix X^* with singular value decomposition $X^* = USV^T$ is μ-incoherent if

$$\max_{i,j} |U_{ij}| \leq \sqrt{\mu}/\sqrt{m} \ \text{ and } \ \max_{i,j} |V_{ij}| \leq \sqrt{\mu}/\sqrt{n}. \quad (3)$$

Intuitively, higher incoherence (smaller μ) means that large entries of X^* are not concentrated in a small part of the matrix, and so uniform sampling is sufficient to recover X^*. The following theorem summarizes the exact recovery condition that we will use in this paper:

Theorem 4 (Recovery Condition [2])
Let X^* be a matrix of bounded rank ($k = O(1)$) with singular value decomposition $X^* = USV^T$. Assume X^* is μ-incoherent and more than $C\mu^4 n \log^2 n$ entries are uniformly sampled, then with probability at least $1 - n^{-3}$, X^* is the unique optimizer of (2).

Based on Theorem 4, 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 (2), we can recover the underlying signed network with high probability if there are no extremely small groups. More precisely, we define the *group imbalance* of a signed network as follows:

Definition 4 (Group Imbalance)
Let X^* 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 X^* is defined as

$$\tau \equiv \max_{i=1,\ldots,k} n/n_i. \quad (4)$$

By definition, $k \leq \tau \leq n$. Intuitively, larger group imbalance τ indicates the presence of a very small group, which would 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 containing only one node. Then the adjacency matrix of this network has group imbalance $\tau = n$ with the following form:

$$X^* = \begin{bmatrix} 1 & \cdots & \cdots & -1 & -1 \\ \vdots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & -1 & -1 \\ -1 & \cdots & -1 & 1 & -1 \\ -1 & \cdots & -1 & -1 & 1 \end{bmatrix}$$

However, without observing $X^*_{n,n-1}$ or $X^*_{n-1,n}$, it is impossible to determine whether the last two nodes are in the same cluster, or 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, no matrix completion algorithm can exactly recover this network under uniform sampling.

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 μ.

Theorem 5 (Incoherence of Signed Networks)
Any complete k-weakly balanced network is τ-incoherent where τ is the group imbalance measurement.

PROOF. Let X^* be the adjacency matrix of a k-weakly balanced complete network. Recall from Definition 3 that μ is defined as the maximum absolute value in the (normalized) singular vectors of X^*, which are the same as eigenvectors of X^* since the adjacency matrix is symmetric.

Let $\mathbf{u}$ be any eigenvector of X^* ($\|\mathbf{u}\|_2 = 1$) with eigenvalue λ. Suppose i and j are in the same group, namely $X^*_{i,:} = X^*_{j,:}$, we have $u_i = X^*_{i,:}\mathbf{u}/\lambda = X^*_{j,:}\mathbf{u}/\lambda = u_j$. Thus, $\mathbf{u}$ has the following form:

$$\mathbf{u} = [\underbrace{v_1, v_1, \ldots, v_1}_{n_1}, \underbrace{v_2, \ldots, v_2}_{n_2}, \ldots, \underbrace{v_k, \ldots, v_k}_{n_k}]^T. \quad (5)$$

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

$$\max_i |u_i| = \max_j |v_j| \leq \max_j \frac{1}{\sqrt{n_j}} = \max_j \frac{\sqrt{n/n_j}}{\sqrt{n}} \leq \frac{\sqrt{\tau}}{\sqrt{n}}.$$

Therefore, X^* is τ-incoherent. $\square$

Putting together Theorems 4 and 5, we now have the main theorem of this subsection:

Theorem 6 (Recovery Condition for Signed Networks)
Suppose we observe edges $A_{ij}, (i,j) \in \Omega$, from an underlying k-weakly balanced signed network X^*, 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 X^*.
Then X^* can be perfectly recovered by solving (2), with probability at least $1 - n^{-3}$.

In particular, if n_i/n is lower 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.

It is known that the convex problem (2) can be exactly solved by an SDP. However, the computational cost of SDP

might be too prohibitive in practice. Recent research provides more efficient algorithms to approximately solve (1) [1, 12]. In our experiment, we use the SVP algorithm proposed by Jain et al.[12] which attempts to solve matrix completion problem in an efficient manner. Experimental evidence in [12] shows that all iterates of the SVP algorithm are μ-incoherent, in which case the matrix completion problem (1) can be exactly solved by SVP. In Section 5, we will see that SVP performs well in recovering weakly balanced networks.

3.2 Inference via Matrix Factorization

Though matrix completion algorithms can guarantee recovery for weakly balanced networks under certain conditions, convex relaxation (2) does not work very well in real-life applications, where observed values are not uniformly distributed, which violates one of the assumptions in Theorem 6. In addition, the methods for solving (2) cannot scale to very large datasets. 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 a matrix factorization approach can not only boost the accuracy of estimation but also scale to large real networks.

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

$$\min_{W,H \in \mathbb{R}^{k \times n}} \sum_{(i,j) \in \Omega} (A_{ij} - (W^T H)_{ij})^2 + \lambda \|W\|_F^2 + \lambda \|H\|_F^2. \quad (6)$$

Although problem (6) is non-convex, it is widely used in practical collaborative filtering applications as the performance is competitive or better as compared to trace-norm minimization, while scalability is much better. For example, to solve the Netflix problem, (6) has been applied with a fair amount of success to factorize the dataset with 100 million ratings [14].

Nevertheless, there is an issue when modeling signed networks using (6): the square loss in the first term of (6) tends to force entries of $W^T H$ to be either $+1$ or -1. However, what we care about in this completion task is the consistency between $\text{sign}((W^T H)_{ij})$ and $\text{sign}(A_{ij})$ rather than their difference. For example, $(W^T H)_{ij} = 10$ 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 (6) can be generalized as:

$$\min_{W,H \in \mathbb{R}^{k \times n}} \sum_{(i,j) \in \Omega} \ell(A_{ij}, (W^T H)_{ij}) + \lambda \|W\|_F^2 + \lambda \|H\|_F^2. \quad (7)$$

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

$$\ell_{\text{sigmoid}}(x, y) = 1/(1 + \exp(xy)),$$
$$\ell_{\text{square-hinge}}(x, y) = (\max(0, 1 - xy))^2. \quad (8)$$

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

Time complexity. There are two main optimization techniques for solving (7) for large-scale data: Alternating Least Squares (ALS) and Stochastic Gradient Descent (SGD) [14]. ALS solves the squared loss problem (6) 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

Table 1: Network Statistics

	# nodes	# edges	+ edges	- edges
Wikipedia	7,065	103,561	78.7%	21.2%
Slashdot	82,144	549,202	77.4%	22.6%
Epinions	131,828	840,799	85.0%	15.0%

an $n \times n$ observed matrix with m observations, the time complexity for each subproblem requires $O(mk^2)$ operations to form the Hessian matrices, and $O(nk^3)$ to solve the least squares problem. 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 square loss. To solve the general form (7) 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 ith column of W and the jth column of H, denoted by $\mathbf{w}_i$ and $\mathbf{h}_j$, respectively. The update rule for $\mathbf{w}_i$ is given by:

$$\mathbf{w}_i \leftarrow \mathbf{w}_i - \eta \left(\frac{\partial \ell(A_{ij}, (W^T H)_{ij})}{\partial \mathbf{w}_i} + \lambda \mathbf{w}_i \right), \quad (9)$$

where η is a small step size. The update rule for $\mathbf{h}_j$ is similar to (9). Since each SGD update (9) costs $O(k)$ time, after a sweep through all known entries it will take $O(mk)$ time. Therefore, the time complexity for SGD is $O(t_2mk)$, where t_2 is the number of iterations taken by SGD to converge. Notice that although the complexity of SGD is linear in k, it usually takes many more iterations to converge compared with ALS, i.e., $t_2 > t_1$.

On the other hand, all previous link or cycle-based sign inference algorithms [17, 5] require time at least $O(nm)$ because all of them contain some $n \times n$ sparse matrix multiplication steps in model construction. Moreover, for all length-l paths, the number of features is exponential in l. Therefore, assuming the number of features in consideration is d, the time complexity for various methods will be $O(dnm)$. The time complexity is summarized in the following table:

HOC	ALS	SGD
$O(dnm)$	$O(t_1(nk^3 + mk^2))$	$O(t_2km)$

Since in real large-scale social networks, $m > n \gg t_1, t_2, k$, this shows our alternative minimization approach is much more efficient for sign inference.

3.3 Low Rank Structure in Real Datasets

We now show that real networks tend to exhibit low-rank structure to a much greater extent than random networks. We consider three large-scale online social networks – Wikipedia, Epinions[17], and Slashdot[15][1]. Table 1 shows the statistics of these datasets. We compare these real networks with random networks as the baseline. The random network is created using the Erdös-Rényi model with sparsity equal to the Wikipedia network.

To measure the closeness of observed entries between the original network and the completed matrix, we first derive the low-rank complete matrix A^* by conducting matrix completion using the observed entries A_{ij}. Then, we look at the relative error on the observed set Ω: $\text{err}_\Omega = \|W \circ (A^* - A)\|_F / \|A\|_F$, where $W_{ij} = 1$ if $(i, j) \in \Omega$ and $W_{ij} = 0$ otherwise, and $\circ$ denotes element-wise multiplica-

---

[1]All the three data sets can be downloaded from SNAP (http://snap.stanford.edu)

tion. Clearly, smaller err_Ω indicates better approximation for the observed entries.

In our experiment, we choose matrix factorization approach for matrix completion, with ranks $k = 1, 2, 4, 8, 16$ and 32. For each network (three real datasets and the random network), we complete the network with different k and compute err_Ω. The result is shown in Figure 2. Compared with the purely random network, 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 each real-life network, as compared to random Erdös-Rényi graphs.
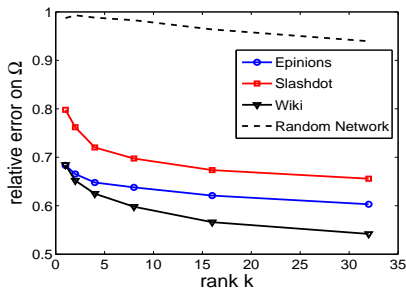

Figure 2: Relative error between adjacency matrix and completed matrix with respect to observed entries, for real-life networks versus a random network. Real-life networks achieve much smaller relative error for every k as compared with the random network.

4. CLUSTERING

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 clustering algorithms [16] extends the notion of Laplacian to signed networks, and proposes a spectral clustering algorithm based on a signed Laplacian matrix. Given a partially observed signed network A, the signed Laplacian is defined as $\bar{D} - A$, where $\bar{D}$ is a diagonal matrix in which $\bar{D}_{ii} = \sum_{j \neq i} |A_{ij}|$. By this definition, the ratio cut 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 that since in Theorem 3 we proved that the k-weakly balanced graphs have rank up to k, we can obtain good clustering by first running a matrix completion algorithm, say trace-norm minimization, on A. The following theorem shows that the eigenvectors of the completed matrix possess a desirable property:

Theorem 7
Let A_{ij}, $(i, j) \in \Omega$, be entries observed from a complete k-weakly balanced network X^*, and assume that the solution of (2) is X with eigenvectors $U = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k]$. If the

---

Algorithm 1: Clustering with Matrix Completion
Input: Adjacency matrix A, number of clusters k
Output: Cluster indicators
1. $X \leftarrow \text{Completion}(A)$ with any matrix completion algorithm.
2. $U \leftarrow$ Top k eigenvectors of X.
3. Run any feature-based clustering algorithm on U.

assumptions in Theorem 6 are all satisfied, then with high probability $U_{i,:} = U_{j,:}$ iff i and j are in the same cluster in X^*.

PROOF. From Theorem 6, we know the recovered matrix X will be X^* with high probability. Suppose $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are the k eigenvectors of X^*. From the proof of Theorem 5, the eigenvectors will have the form in (5), which means $U_{i,:} = U_{j,:}$ if i and j are in the same cluster. Furthermore, when i and j are in different clusters, $X^*_{i,:} \neq X^*_{j,:}$, so $U_{i,:}$ cannot equal to $U_{j,:}$. This proves the theorem. $\square$

Following this theorem, the true clusters can be identified from the eigenvectors of X when the assumptions in Theorem 6 hold. Therefore, perfect clustering is guaranteed in this scenario.

More generally, we can use any matrix completion method discussed in Section 3 to complete A. For example, if we take SVP as the matrix completion approach, we can derive perfect clustering result if all iterates of the algorithm are μ-incoherent. This is because under this condition, SVP can recover X^* exactly, so the property of eigenvectors in Theorem 7 can again be used. Our clustering algorithm that uses matrix completion is summarized in Algorithm 1.

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 generates 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

In this section, we perform experiments on synthetic and real networks, and show that our proposed low-rank model for signed networks outperforms other methods on the tasks of sign inference and clustering. To ensure that our results are reliable, we conduct all experiments 10 times, and average the result from all of the trials.

5.1 Sign Inference

Recall that given a partially observed signed graph A, the sign inference task is to predict the signs of the missing links. Although in Section 3 we focused on undirected graphs when introducing our low rank model, we can easily extend our sign inference algorithms to directed graphs since matrix completion and matrix factorization algorithms are easily adapted to the directed case. Therefore, our sign inference experiments are all on networks that are directed.

5.1.1 Algorithms and Parameter Setting

We consider two low-rank modeling approaches proposed in Section 3: Matrix Completion (MC) and Matrix Factor-

ization (MF). For MC methods, we use Singular Value Projection (MC-SVP) since it is efficient and effective in practice [12]. For MF methods, we mainly consider Alternating Least Square (MF-ALS) which uses the squared loss in (7). In real datasets, we also use Stochastic Gradient Descent (SGD) to solve (7) with sigmoid and square-hinge losses, denoted as MF-SGDSIG and MF-SGDSH respectively.

We compare the performance of our low rank model to state-of-the-art approaches, such as cycle-based methods, for the sign inference problem [17, 5]. The first cycle-based methods are the so-called measure of social imbalance (MOI), which predict the sign of an edge so that more cycles become balanced [5]. If we consider cycles of arbitrary length, and exponentially damp their importance based on length, we get a measure we call (MOI-∞), which infers the sign of (i, j) as $\text{sign}\big(((I-\beta A)^{-1}-I-\beta A)_{ij}\big)$. The second cycle-based approach we consider is a supervised learning approach based on high order cycles (HOC), with features derived from cycles of length 3 (i.e. triangles) [17], length 4 and length 5 (see [5] for more details). As in [5], we use HOC-3, HOC-4 and HOC-5 to denote these methods.

Some of these models require parameter setting beforehand, such as the regularization parameter λ in MF (see (7)) and β in MOI-∞. To select these parameters, we conduct a 3-fold cross validation on the training set. With the selected parameters, we then construct the model based on the whole training set and conduct sign inference on the testing data.

5.1.2   Synthetic Datasets

We first compare all categories of approaches on synthetic datasets. We choose MC-SVP, MC-ALS, MOI-∞ and HOC-3 as representatives of MC, MF, MOI-based and HOC-based methods respectively. We fix the underlying signed network X^* 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 X^*, we assume that we only observe a partial network by sampling some entries from X^* using three sampling procedures: uniform sampling, uniform sampling with noise, and sampling with power-law distribution. For each inference algorithm, we input the observed entries as training data and calculate the sign inference accuracy on the rest of elements.

Uniform sampling: In this scenario, we randomly sample $n^2 p$ entries from X^*, where $p \in (0, 1)$ is the fraction of observed entries. We vary p from 0.001 to 0.1 and plot the inference accuracy in Figure 3(a). Clearly, MC-SVP and MF-ALS outperform the cycle-based methods. MOI-∞ performs the worst with accuracy only 50%-70%. This is because MOI uses cycle-based measurements to make more cycles become *balanced*. This inference policy can work only when $k = 2$ (that is, the underlying network has strong balance), but performs poorly when the underlying network is weakly balanced. 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 p is less than 0.05. On the other hand, both MC-SVP and MF-ALS show high accuracy for all $p \geq 0.01$. In particular, MC-SVP can achieve 100% accuracy when $p > 0.07$, which reconfirms the theoretical recovery guarantee stated in Theorem 6. Moreover, although MF-ALS has no theoretical guarantee, it can still 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 into observations. Specifically, in this scenario, each observed entry A_{ij} has sign that is opposite to the true value X^*_{ij} with probability r. For clarity, we fix the fraction of observed entries $p = 0.1$, and increase r from 0.01 to 0.25. The result is shown in Figure 3(b). We can see that our low-rank modeling approaches are still clearly better than cycle-based methods when noise level becomes higher. Moreover, MC-SVP can still perfectly recover X^* when the noise level $r < 0.05$, and MF-ALS can also achieve perfect recovery with a smaller r.

Sampling with power-law distribution: As Section 3 mentioned, the good performance of matrix completion crucially relies on the assumption that observed entries are uniformly sampled. However, in most real networks (for example, Slashdot in [15]), the degree distribution follows power law. Therefore, we examine how all approaches perform on power-law distributed networks. We generate power-law distributed networks using the Chung-Lu-Vu (CLV) model proposed in [6], which allows one to generate random graphs with arbitrary expected degree sequence. Similar to the uniform sampling case, we vary the fraction of observed entries and plot the inference accuracy in Figure 3(c). We can see MOI-∞ still has poor performance for weakly balanced graphs. However, distinct from uniform sampling case, MC-SVP has lower accuracy rate compared to HOC-3 when $p < 0.1$. This shows that MC-SVP cannot work well given non-uniform distributed observations. On the other hand, MF-ALS still performs better than all other methods in power-law distributed graphs.

From all experiments on synthetic data shown in Figure 3, we can conclude that low rank modeling approaches generally do better than cycle-based methods, and the matrix factorization approach (MF-ALS) performs the best in most cases, even in non-uniform distributed networks. This indicates MF approach should be superior than others in real networks. This will be confirmed in the following subsection.

5.1.3   Real-life Datasets

Next we demonstrate that our low rank model is more effective than existing methods in real datasets also. Already in Figure 3(c) we have seen that MC-SVP fails to perform well under power-law distributed networks, so we consider the more robust MF approaches, including MF-ALS, MF-SGDSIG and MF-SGDSH, for experiments on real datasets. We compare these proposed methods with the best cycle-based methods, HOC-3, HOC-4 and HOC-5. Again we use Wikipedia, Slashdot and Epinions to examine sign inference algorithms on real networks. These three datasets have previously been used as benchmarks on sign inference [17, 5].

To make the comparison fair, we conduct a 10-fold cross validation and report the average inference accuracy for each dataset in Table 2. We observe that MF-based algorithms clearly outperform cycle-based methods. In particular, we observe that HOC-5 only improves HOC-3 by less than 1.5%, while MF-based algorithms consistently improve the accuracy of HOC-5 by more than 2% over all datasets. In addition, MF-SGDSIG and MF-SGDSH further improve the accuracy of MF-ALS slightly. This shows that the sigmoid loss and square-hinge are more suitable for sign inference, which supports the discussion in Section 3.2.

In Figure 4, we further examine the performance of these algorithms with different levels of edge embeddedness. Embeddedness of edge (i, j) is defined as the number of common neighbors of the nodes i and j, and can be thought as a mea-

(a) Uniformly sample without noise   (b) Uniformly sample with noise   (c) Power-law distributed Networks
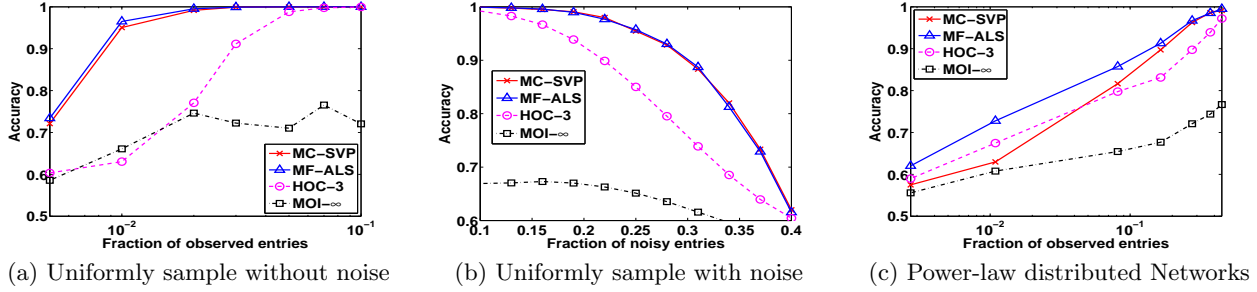
Figure 3: Accuracy of sign inference algorithms on synthetic datasets. In general, we can see that MC-SVP and MF-ALS outperform cycle-based methods such as MOI-∞ and HOC-3. In addition, MF-ALS is more robust than MC-SVP when the observations are sampled from a power-law distribution.

Table 2: The sign inference accuracy for MF-based algorithms and cycle-based algorithms. We can see that the MF-based algorithms are better than cycle-based algorithms.

	Epinions	Slashdot	Wikipedia
HOC-3	0.9014	0.8303	0.8424
HOC-5	0.9080	0.8469	0.8605
MF-ALS	0.9374	0.8774	0.8814
MF-SGDSIG	**0.9465**	0.8789	**0.8830**
MF-SGDSH	0.9437	**0.8835**	0.8810

sure of the proximity between i and j. One might expect that cycle-based approaches should perform better on edges with higher embeddedness because more cycle information is available. However, surprisingly MF-ALS achieves higher inference accuracy regardless of the embeddedness. The performance of MF-SGDSIG and MF-SGDSH is similar to MF-ALS so they are not shown in Figure 4 for clarity.

5.1.4 Computation Time

In addition to inference accuracy, we now compare the running time required by the different methods. As discussed in Section 3.2, matrix factorization methods are more efficient than cycle-based algorithms in terms of time complexity. Here, we further show that MF-based methods are empirically much faster than cycle-based algorithms. The running times are summarized in Table 3. To conduct timing tests on a large signed network, in addition to the three real datasets as described in Table 1, we further construct a large-scale synthetic dataset called Cluster10 where number of edges is 100 times more than Epinions. Cluster10 is generated from a 10-weakly balanced network, in which clusters have sizes 20000, 40000,..., 200000 respectively. There are totally 1.1 million nodes and 120 million edges uniformly observed from the complete graph. We construct this synthetic data to show that our matrix factorization approach can easily scale up to massive graphs compared to HOC-3 and HOC-5. For matrix factorization approach, we report the time needed to solve the model by SGD (with sigmoid and square-hinge) and ALS (with square loss). For HOC methods which build classifiers from cycle-based features, since the time for training phase depends on the classifier, we only report the time for computation of features. Thus the reported time for HOC is an underestimation for constructing the HOC model; even then we can see that the time required by MF-based algorithms is much lower than HOC methods.

In conclusion, for the sign inference problem, we can see

Table 3: Running time (in seconds) for our MF approach and HOC on real datasets and a 1.1 million node synthetic data Cluster10. For HOC methods, we only consider the time for feature computation before the model training, while for MF-based methods we report the total time for constructing the model. We can see that MF-based methods are clearly more efficient than cycle-based algorithms.

	HOC-3	HOC-4	HOC-5	ALS	SGD
Wiki	18.08	74.52	462.92	**2.26**	2.41
Slashdot	133.4	1936.0	>10,000	**17.4**	24.7
Epinions	560.64	6156.8	>10,000	**28.67**	37.2
Cluster10	>10,000	>10,000	>10,000	**455.1**	1152

that our low rank model outperforms other traditional inference methods. In particular, the matrix factorization approach is clearly the most competitive method in terms of accuracy and scalability.

5.2 Clustering

In this subsection, we show that our proposed clustering approach, which completes the low-rank structure of signed networks before performing clustering, outperforms spectral clustering based on the signed Laplacian [16]. Similar to Section 5.1.2, 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 X^* where size of each group is 100. We then observe entries from X^* with two sampling procedures: 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

$$\sum_{i,j:s_i=s_j} I(\bar{s}_i = \bar{s}_j) + \sum_{i,j:s_i \neq s_j} I(\bar{s}_i \neq \bar{s}_j).$$

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 the previous subsection, in the uniform sampling case, we draw pn^2 i.i.d. samples from all the n^2 edges. Similarly, in sampling with noise, we flip the sign of each observed edge with probability r. The results of these two scenarios are shown in Figure 5. In both scenarios, our proposed clustering approach is significantly better than clustering based on the signed Laplacian. This shows that recovering the low-rank structure of signed networks leads to improved clustering results.
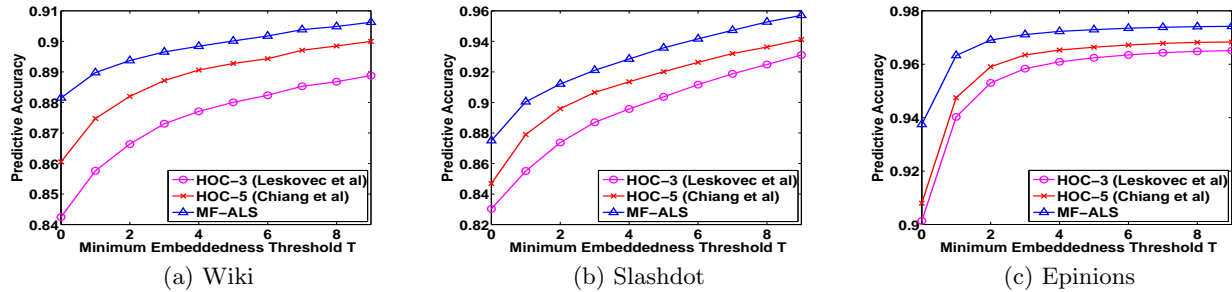
(a) Wiki   (b) Slashdot   (c) Epinions

Figure 4: Accuracy of sign inference algorithms with different levels of embeddedness. These plots show the accuracy for edges whose embeddedness is at least T. We can see that MF-ALS consistently gives us higher accuracy for all thresholds T.
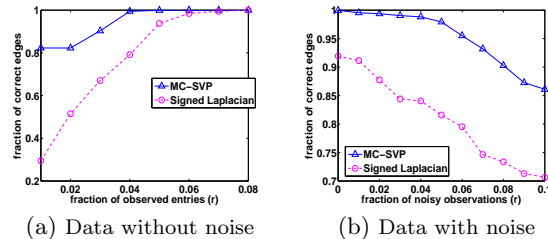

(a) Data without noise   (b) Data with noise

Figure 5: Clustering partially observed synthetic data. Figure 5(a) is the result without noise and Figure 5(b) is the result with noise. In both cases, clustering with MC-SVP performs significantly better than using signed Laplacian.

6. CONCLUSIONS

In this paper, we have proposed a low rank modeling approach for signed network analysis. We have shown that the low-rank structure of signed networks naturally emerges from weak balance theory. The sign inference problem in such networks can thus be modeled as a low-rank matrix completion problem. We first showed that missing links in a signed network can be exactly recovered by matrix completion algorithms under certain conditions, and then introduced a more efficient matrix factorization approach for sign inference. Furthermore, we showed that the low rank model can also be used for clustering. Experiments conducted on both synthetic data and real networks show that our low rank model improves sign inference significantly, in terms of both accuracy and speed. Clustering results also become more favorable by making use of the low rank model.

7. ACKNOWLEDGEMENT

8. REFERENCES

[1] J.-F. Cai, E. J. candés, and Z. Shen. A singular value thresholding algorithm for matrix completion. *Society for Industrial and Applied Mathematics (SIAM)*, 20(4):1956–1982, 2010.

[2] E. J. Candés and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56(5):2053–2080, 2009.

[3] E. J. Canés and B. Recht. Exact matrix completion via convex optimization. *Found. of Comput. Math.*, 9:712–772, 2008.

[4] D. Cartwright and F. Harary. Structure balance: A generalization of Heider's theory. *Psychological Review*, 63(5):277–293, 1956.

[5] K.-Y. Chiang, N. Natarajan, A. Tewari, and I. S. Dhillon. Exploiting longer cycles for link prediction in signed networks. In *CIKM*, 2011.

[6] F. Chung, L. Lu, and V. Vu. Spectra of random graphs with given expected degrees. *Internet Mathematics*, 1, 2004.

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

[8] I. S. Dhillon, Y. Guan, and B. Kulis. Weighted graph cuts without eigenvectors: A multilevel approach. *IEEE Trans. on pattern analysis and machine intelligence*, 29(11):1944–1957, 2007.

[9] R. V. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *WWW*, 2004.

[10] S. Hanneke and E. P. Xing. Network completion and survey sampling. In *AISTATS*, 2009.

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

[12] P. Jain, R. Meka, and I. Dhillon. Guaranteed rank minimization via singular value projection. In *NIPS*, pages 934–945, 2010.

[13] M. Kim and J. Leskovec. The network completion problem: infering missing nodes and edges in networks. In *SDM*, 2011.

[14] Y. Koren, R. M. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42:30–37, 2009.

[15] J. Kunegis, A. Lommatzsch, and C. Bauckhage. The slashdot zoo: Mining a social network with negative edges. In *WWW*, pages 741–750, 2009.

[16] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. DeLuca, and S. Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *SDM*, pages 559–570, 2010.

[17] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. In *WWW*, 2010.

[18] A. Y. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *NIPS*, 2001.

[19] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[20] B. Yang, W. Cheung, and J. Liu. Community mining from signed social networks. *IEEE Trans. on knowledge and data engineering*, 19(10):1333–1348, 2007.