

## Research Article

# Sign Inference for Dynamic Signed Networks via Dictionary Learning

Yi Cen, Rentao Gu, and Yuefeng Ji

State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Yuefeng Ji; [jyf@bupt.edu.cn](mailto:jyf@bupt.edu.cn)

Received 19 July 2013; Accepted 15 October 2013

Academic Editor: Dewei Li

Copyright © 2013 Yi Cen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mobile online social network (mOSN) is a burgeoning research area. However, most existing works referring to mOSNs deal with static network structures and simply encode whether relationships among entities exist or not. In contrast, relationships in signed mOSNs can be positive or negative and may be changed with time and locations. Applying certain global characteristics of social balance, in this paper, we aim to infer the unknown relationships in dynamic signed mOSNs and formulate this sign inference problem as a low-rank matrix estimation problem. Specifically, motivated by the Singular Value Thresholding (SVT) algorithm, a compact dictionary is selected from the observed dataset. Based on this compact dictionary, the relationships in the dynamic signed mOSNs are estimated via solving the formulated problem. Furthermore, the estimation accuracy is improved by employing a dictionary self-updating mechanism.

## 1. Introduction

Over the past few years, a number of mobile applications that allow users to enjoy networking have emerged. Correspondingly, there has been a proliferation in mobile online social networks (mOSNs). With the ubiquitous use of mobile devices and a rapid shift of technology, it is worthy to investigate the mOSNs from a privacy or security standpoint [1, 2]. The related applications are also extensive such as authentication and recommendation online. In this context, researches about mobile online networks where two opposite kinds of relationships can occur have become common; people not only form links to indicate friendship, support, or approval but signify disapproval or distrust of the opinions of others. It is natural to model such networks as signed networks, where the sign of a link weight can be either positive or negative, representing the status of a relationship. Analogous to traditional social networks analysis, the relationships in signed mOSNs can be represented as a graph, where nodes denote the objects (e.g., people or mobile terminals) and signed edges denote the relationships or links (e.g., a communication made between two people). The link structure of the resulting graph can be exploited to detect

underlying groups of objects, predict missing links, and handle many other tasks [3–17].

One of the most fundamental theories that are applicable to signed social networks is social structural balance [5, 6, 16]. Structural balance corresponds to the possibility of exactly dividing the signed graph into two adversary subcommunities such that all edges within each subcommunity have positive weights while all edges joining agents of different communities have negative weights. Obviously, graphs of nonnegative weights are a special case of structural balance, in which one of the two subcommunities is empty. Since the assumption that structural balance exists in a real signed network might be too extreme, a concept called weak structural balance further generalizes structural balance by discussing the multiadversary-subcommunities partition of the signed graph [7].

Structural balance and weak structural balance have been shown to be valid to analyze signed networks. For instance, the sign inference problem, which aims to infer the unknown relationship between two objects, can be solved by mining balance information of signed networks from local and global perspectives [8–10, 12–17]. With the help of the result inferred, it is possible to predict the relationships so that legitimate

participants can eliminate networking security vulnerabilities. Nevertheless, most of these state-of-the-art methods for sign inference problem are mainly considered from a static point of view, and dynamic scenarios are rarely taken into account. Therefore, it is necessary to establish a rational dynamic network model to infer the sign of relationships.

Actually, there exist several inherent qualities of mOSNs that are challenging to reliably sense the global states of relationships for the large networks in practice [2]. First, in contrast to traditional social networks, the observations of relationships in mOSNs are closely associated with the geographical environment, as well as the relative locations and signal coverage of mobile terminals/network access points. Due to these spatial constraints, such observations, which seem linearly related to the global data of relationships (i.e., linearly sampled from the global data), are bound to miss a significant number of values. Consequently, they are not sufficient to unambiguously infer the true status by the traditional solutions of linear-inverse problem in general. Second, in mOSNs different relations between entities may appear at different times. Accordingly, observations of the networks vary during a time period long enough. These dynamic interactions over time essentially introduce time dimension to the problem of mining, the potential relationship structures. Third, despite maintaining the dynamic performance, the underlying relationships in reality always display some “redundancy” attributed to the gradual/periodic variation [3], the relative stability, and so forth. Owing to the aforementioned characteristics of mOSNs, the mass redundant data generated in variant scenarios will result in resource challenge. Hence, although many observers collect features for at least part of the networks, there are still serious impediments to reliable large-scale or network-wide data processing. After these aspects of mOSNs are learned, it is reasonable to organize the entire relationship dataset in the form of tensor coincident with its spatiotemporal structure. Meanwhile, efficient relationship inference approaches associated with the tensor model are required to overcome the obstacles of this data processing.

The aim of this paper is to develop algorithms for the sign inference in signed mOSNs in global and spatiotemporal evolution perspectives. In particular, we assume that the signed mOSN possesses an underlying dynamic weakly balanced complete network structure. Suppose that we are given an incomplete networking observation tensor (or 3-dimensional array), which consists of the adjacency matrices corresponding to the snapshots of the underlying dynamic weakly balanced complete network at times  $T_0, T_0+1, \dots, T_0+T-1, T_0+T$ . Then the sign inference task is to estimate the sign patterns of all possible links in the dynamic complete network at time  $T_0+T$ . Utilizing the low-rank property of the weak structural balance and the features extracted from the observation tensor, we consider the inference via the incomplete relationship data as an underdetermined linear-inverse problem and develop an approach via a low-rank matrix reconstruction to solve this problem. Moreover, we regard the observation tensor as the training data set and choose a dictionary from it to improve the validity and efficiency of our inference approach. The dictionary

selection method is designed by reducing the size of an overcomplete feature set extracted from the training dataset. Also, a dictionary self-updating mechanism is introduced to improve accuracy of the inference.

Here are the key contributions we make in this paper.

- (i) A dictionary selection approach based on group sparsity has been designed to generate a set containing minimal sizes of features to increase computational efficiency. Specifically, the observation tensor is considered to be the raw materials for feature extraction.
- (ii) The sign inference problem referring to the weakly balanced mOSNs is formulated as a low-rank matrix reconstruction from the selected dictionary. Under certain mild conditions, a low-rank matrix reconstruction algorithm is applied to solve the sign inference problem, and it turns out to be much more accurate and efficient than other inference methods in the literature. A dictionary self-updating mechanism is also introduced to adjust the dynamic characteristics of the network and improve the sensing accuracy.

The rest of this paper is organized as follows. In Section 2, we build the model of the dynamic signed network. Some basics of balance theory are also reviewed for the sake of integrality. In Section 3, we first extract the initial candidate feature pool from the observation tensor and propose a dictionary selection approach. Then we propose our low-rank matrix reconstruction method to solve the sign inference problem. The implementation details of the dictionary self-updating procedure are also proposed. In Section 4, we conduct numerical experiments which demonstrate the validity of our network model for sign inference and justify the performance of our methods as well. Finally, we present our conclusions in Section 5.

## 2. Background and Preliminaries

*2.1. Dynamic Signed Network Structure.* Formally, a dynamic undirected signed network is represented as a dynamic graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  is the vertex set of size  $n$  and  $\mathcal{E}$  is the edge set varying over time. A network snapshot denoted by  $\mathcal{S}_t = (\mathcal{V}, \mathcal{E}_t, \mathbf{A}^{(t)})$  presents the connections of  $\mathcal{G}$  observed at time  $t$ . Here,  $\mathcal{E}_t$  is the subset of  $\mathcal{E}$  and  $\mathbf{A}^{(t)} \in \{-1, 0, 1\}^{n \times n}$  is the adjacency matrix of  $\mathcal{S}_t$  with the signed weights

$$a_{ij}^{(t)} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ have positive relationship,} \\ -1 & \text{if } i \text{ and } j \text{ have negative relationship,} \\ 0 & \text{if relationship between } i \text{ and } j \text{ is unknown.} \end{cases} \quad (1)$$

Particularly, for each time  $t$ , a zero entry in  $\mathbf{A}^{(t)}$  is treated as an unknown relationship based on the acknowledgement that some potential attitudes exist between any two entities, even if the relationship itself is not observed. From this viewpoint, we can assume that there exists an underlying dynamics complete signed network  $\overline{\mathcal{G}}$ , in which only some partial relationships are observed at times

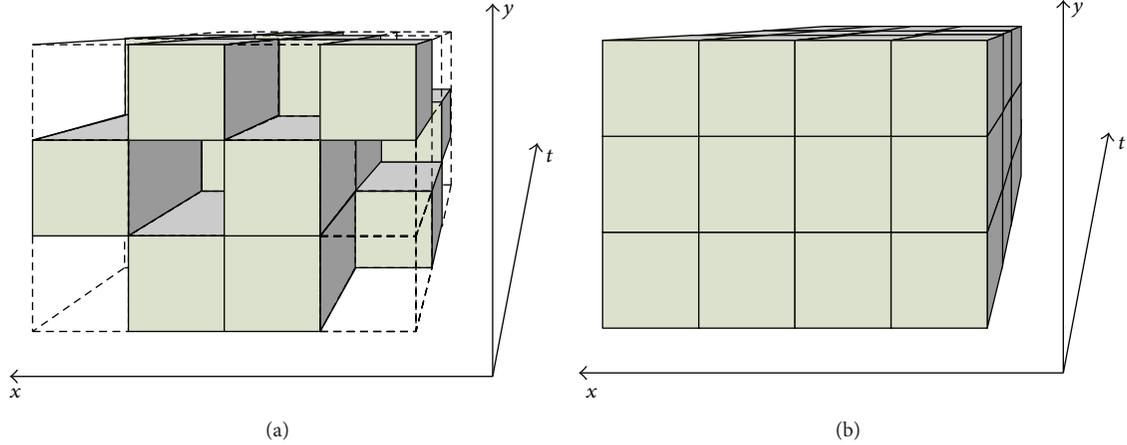


FIGURE 1: Illustration of the adjacency tensor, the cube units symbolize the data of relationships: (a) the adjacency tensor of the observed network and (b) the adjacency tensor of the underlying complete network.

$T_0, T_0 + 1, \dots, T_0 + T - 1$ , respectively. Correspondingly, we let  $\overline{\mathcal{A}} \in \{-1, 1\}^{n \times n \times T}$  denote the three-dimensional tensor that contains relationship information between all pairs of entities in  $\overline{\mathcal{E}}$ . Thus, the observation tensor  $\mathcal{A}$  consisting of a series of network snapshots can be represented as

$$\mathcal{A}_{i,j,t} = a_{ij}^{(t)} = \begin{cases} \overline{\mathcal{A}}_{i,j,t}, & (i, j, t) \in \Omega, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where  $\Omega$  is the index set of the observed entries. Let  $\mathcal{P}_\Omega$  be the orthogonal projection operator onto the span of tensors vanishing outside  $\Omega$  so that the  $(i, j, t)$ th component of  $\mathcal{P}_\Omega(\mathcal{X})$  is equal to  $\mathcal{X}_{i,j,t}$  when  $(i, j, t) \in \Omega$  and zero otherwise. Then we have  $\mathcal{P}_\Omega(\overline{\mathcal{A}}) = \mathcal{A}$  (shown in Figure 1) and  $\mathcal{P}_{\Omega_t}(\overline{\mathbf{A}}^{(t)}) = \mathbf{A}^{(t)}$  for each time slice  $t$ , where  $\bigcap_t \Omega_t = \emptyset$  and  $\bigcup_t \Omega_t = \Omega$ .

While the above kind of signed networks is called homogeneous, that is, relationships of the networks are between the same kinds of entities, a signed network can also be heterogeneous. In a heterogeneous signed network, there can be more than one kind of entities, and relationships between same or different entities can be positive and negative, such as YouTube with two kinds of entities—users and videos. Moreover, this three-dimensional network adjacency tensor can increase dimensions (e.g., spatial dimension, etc.) to adapt to a wider range of scenarios. In this paper, we mainly focus our attention on three-dimensional homogeneous signed networks.

**2.2. Weak Structural Balance.** Structural balance theory was first formulated by Heider [18] in order to understand the structure in a network of individuals whose mutual relationships are characterized in terms of friendship and hostility. Formally, a triad is considered balanced if the product of the signs in the triad is positive; that is, it contains an even number of negative edges. This is in agreement with principles 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” [6]. The configurations of balanced and unbalanced triads are shown in Figure 2. One possible weakness of this theory is that the defined balance relationships might be too strict. In this perspective, by extending the fundamental beliefs in real networks, weak structural balance is proposed as a way of eliminating the assumption that “the enemy of my enemy is my friend” [7]. Equivalently, the case that “the enemy of my enemy is my enemy” is permitted. Therefore, the local structure of weak balance posits that only triads with exactly two positive edges are implausible and that all other kinds of triads should be permissible (also illustrated in Figure 2).

The formal definition of weakly balanced networks is as follows.

**Definition 1** (weakly balanced networks [7]). A (possibly incomplete) network is weakly balanced if and only if it is possible to obtain a weakly balanced complete network by filling the missing edges in its adjacency matrix. Furthermore, in terms of patterns of global structure, a complete network is weakly balanced if and only if the vertex set can be divided into  $r$  clusters,  $r \geq 1$ , such that all the edges within clusters are positive and all the edges between clusters are negative.

There exists the literature discussing the approaches of clustering and sign prediction with respect to signed networks. Ideas derived from local balance of signed networks can be successfully used to yield algorithms for sign inference [9, 10]. Meanwhile, several works analyze the social interrelations from global perspective of structural balance [8, 13–15, 17]. In particular, it is shown in [8] that the adjacency matrix of weakly balanced networks has a “low-rank” structure, and the sign prediction methods based on low-rank modeling were proposed as well.

**Theorem 2** (low-rank structure of signed networks [8]). *The adjacency matrix  $\overline{\mathbf{A}} \in \{1, -1\}^{n \times n}$  of a complete  $r$ -weakly balanced network has rank 1, if  $r \leq 2$ , and has rank  $r$  for all  $r > 2$ .*

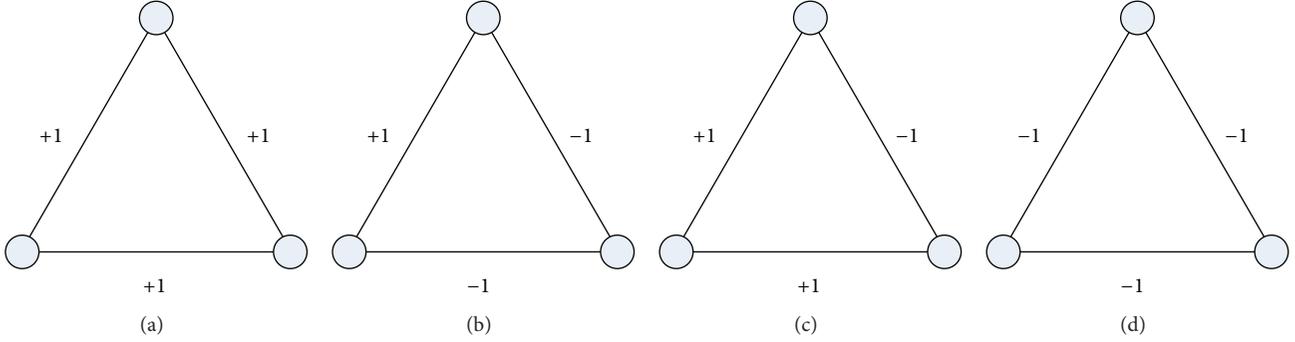


FIGURE 2: Signed undirected connectivity configurations mentioned in Section 2.2: (i) (a) and (b) are balanced triads, (c) and (d) are unbalanced triads, and (ii) (a), (b), and (d) are weakly balanced triads.

Actually, since the global viewpoint of weak balance stated in Definition 1 obeys clustering characteristics presented in Theorem 2, for  $\bar{\mathbf{A}}$ , there exists an invertible matrix  $\mathbf{P}$  such that

$$\mathbf{P}\bar{\mathbf{A}}\mathbf{P}^T = \begin{pmatrix} \mathbf{1}_{n_1} & & & \\ & \mathbf{1}_{n_2} & & \\ & & \mathbf{1}_{n_3} & \\ & & & \ddots \\ & & & & \mathbf{1}_{n_r} \end{pmatrix}, \quad (3)$$

where  $\mathbf{1}_{n_i}$  on the primary diagonal is an  $n_i$ -order square matrix whose entries are all 1 ( $\sum_{i=1}^r n_i = n$ ) and the other entries of  $\mathbf{P}\bar{\mathbf{A}}\mathbf{P}^T$  are all  $-1$ . The  $n_i$ -order square matrix indicates the  $i$ th cluster.

*Notation.* For  $\mathbf{X} \in R^{m \times n}$ , let the mixed norm  $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^m \|\mathbf{X}_i\|_2$ ; the soft-thresholding operator  $\mathcal{D}_\tau(\cdot) : \mathbf{X} \in R^{m \times n} \mapsto \mathbf{Y} \in R^{m \times n}$  is also defined obeying

$$\mathbf{Y}_i = \begin{cases} 0 & \text{if } \|\mathbf{X}_i\|_2 \leq \tau, \\ \left(1 - \frac{\tau}{\|\mathbf{X}_i\|_2}\right) \mathbf{X}_i & \text{otherwise,} \end{cases} \quad (4)$$

where  $\mathbf{X}_i$  and  $\mathbf{Y}_i$  denote the  $i$ th row of  $\mathbf{X}$  and  $\mathbf{Y}$ , respectively [15]. The invertible vectorization is denoted by  $\text{vec}(\cdot) : R^{m \times n} \mapsto R^{mn}$ .

Let  $\mathcal{S}_{\mu,L}^{1,1}(R^{m \times n})$  be the class of convex functions with Lipschitz gradient [19]. A continuous differentiable function  $f(\mathbf{Y})$  belongs to  $\mathcal{S}_{\mu,L}^{1,1}(R^{m \times n})$  for some  $0 \leq \mu \leq L$  if for any  $\mathbf{X}, \mathbf{Y} \in R^{m \times n}$  we have both of the following:

$$\begin{aligned} \|\nabla f(\mathbf{X}) - \nabla f(\mathbf{Y})\|_F &\leq L\|\mathbf{X} - \mathbf{Y}\|_F, \\ \langle \nabla f(\mathbf{X}) - \nabla f(\mathbf{Y}), \mathbf{X} - \mathbf{Y} \rangle &\geq \mu\|\mathbf{X} - \mathbf{Y}\|_F^2. \end{aligned} \quad (5)$$

### 3. Sign Inference via Dictionary Learning

In this section, we focus on a solution of the sign inference to estimate connection statuses via dictionary learning. As the preparation, we propose a large-scale dictionary selection method to generate the dictionary for inferring. Assume that

we are given a (usually incomplete) network observation tensor  $\mathcal{A}$  sampled from an underlying dynamic weakly balanced complete network  $\bar{\mathcal{G}}$  with the adjacency tensor  $\bar{\mathcal{A}}$ . As the description in Section 1, it is reasonable to suppose that most relationships between entities have their own stability in a long period of time in practice and subsequently the change in the scale of each subcommunity is limited. Apparently, this implies the strong dependence retained among the observed data. Combining these assumptions with the low-rank characteristic of weakly balanced complete networks, we extract an initial feature pool from the observation tensor  $\mathcal{A}$  and propose a dictionary selection method to compress the scale of the feature pool in Section 3.1. The corresponding algorithm is presented, respectively, in Section 3.2. With the trained dictionary, we propose our sign inference approach and dictionary updating mechanism in Section 3.3, which are also inspired by the low-rank characteristic of weakly balanced complete networks.

The method we propose to handle the dictionary selection is motivated by the Singular Value Thresholding (SVT) algorithm, which is a simple and efficient algorithm for nuclear norm minimization problems proposed by Cai et al. [20]. Our basic idea is to obtain the optimal solution of the trace norm minimization problem by solving its dual problem whose objective function can be shown to be continuously differentiable with Lipschitz continuous gradient. Specifically, we prove that the optimal solution of the primary problem can be readily obtained from the optimal solution of the dual problem. We first provide a brief review of the standard SVT algorithm.

Considering the problem

$$\begin{aligned} \min_{\mathbf{X}} \quad & \tau\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X}\|_F^2 \\ \text{subject to} \quad & \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{M}), \end{aligned} \quad (6)$$

Cai et al. [20] give a theoretical analysis that, when  $\tau \rightarrow \infty$ , the optimal solution of problem (6) converges to that of the standard problem:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathbf{X}\|_* \\ \text{subject to} \quad & \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{M}). \end{aligned} \quad (7)$$

Given that  $\tau > 0$ , the SVT algorithm operates as a linear Bregman iteration scheme. Furthermore, by defining the Lagrangian function of problem (6) as

$$\mathcal{L}(\mathbf{X}, \mathbf{Y}) = \tau \|\mathbf{X}\|_* + \frac{1}{2} \|\mathbf{X}\|_F^2 + \langle \mathbf{Y}, \mathcal{P}_\Omega(\mathbf{M} - \mathbf{X}) \rangle, \quad (8)$$

where  $\mathbf{Y}$  is the Lagrangian dual variable, we can derive its dual function as

$$f(\mathbf{Y}) = \inf_{\mathbf{X}} \mathcal{L}(\mathbf{X}, \mathbf{Y}). \quad (9)$$

Cai et al. show that SVT indeed optimizes the dual function  $f(\mathbf{Y})$  via the gradient ascent method.

**3.1. Large-Scale Dictionary Selection.** We address how to select the dictionary given an initial candidate feature pool in this subsection. To this end, we first extract an initial candidate feature pool from  $\mathcal{A}$ , which is sampled from  $\overline{\mathcal{A}}$ . Since  $\overline{\mathcal{A}}$  consists of the adjacency matrices  $\overline{\mathbf{A}}^{(t)}$  ( $t = T_0, T_0 + 1, \dots, T_0 + T - 1$ ), the matrix  $\mathbf{A}^{(t)}$  in  $\mathcal{A}$  can retain the information of  $\overline{\mathbf{A}}^{(t)}$  more or less. Thus, we reserve the group of  $\mathbf{A}^{(t)}$  with relatively higher sample rate to extract features. We use singular value decomposition (SVD) to express each  $\mathbf{A}^{(t)}$  as a series of orthogonal bases in Hilbert space; that is,

$$\mathbf{A}^{(t)} = \sum_{r=1}^{r_t} \sigma_r^{(t)} \mathbf{u}_r^{(t)} (\mathbf{v}_r^{(t)})^T = \sum_{r=1}^{r_t} \sigma_r^{(t)} \mathbf{U}_r^{(t)}, \quad (10)$$

where  $\mathbf{u}_r^{(t)}$  and  $\mathbf{v}_r^{(t)}$  are singular vectors of  $\mathbf{A}^{(t)}$  with eigenvalue  $\sigma_r^{(t)}$ ,  $1 \leq r \leq r_t$ . Without loss of generality, we sort  $\sigma_1^{(t)} \geq \sigma_2^{(t)} \geq \dots \geq \sigma_{r_t}^{(t)}$  the singular values of  $\mathbf{A}^{(t)}$  in descending order, and set

$$L_t = r_t - \arg \max_{r=1, \dots, r_t-1} (\sigma_r^{(t)} \geq \sigma_{r+1}^{(t)}). \quad (11)$$

Then, due to the low-rank property of the weakly balanced complete adjacency matrix, we keep the group of  $\mathbf{U}_r^{(t)}$  corresponding to the  $L_t$  largest  $\sigma_r^{(t)}$  as the features. By this procedure, we extract an initial candidate feature pool as  $\{\mathbf{U}_r^{(t)} : T_0 \leq t \leq T_0 + T - 1, 1 \leq r \leq L_t\}$ , where each matrix  $\mathbf{U}_r^{(t)} \in R^{n \times n}$  denotes a feature. Equivalently, we can discuss  $\mathcal{Q} = \{\text{vec}(\mathbf{U}_r^{(t)}) : T_0 \leq t \leq T_0 + T - 1, 1 \leq r \leq L_t\}$  and form the matrix  $\Phi = [\text{vec}(\mathbf{U}_1), \text{vec}(\mathbf{U}_2), \dots, \text{vec}(\mathbf{U}_S)]$  for convenience, where  $\text{vec}(\mathbf{U}_s) = \text{vec}(\mathbf{U}_r^{(t)})$ ,  $1 \leq r \leq L_t, 1 \leq s \leq S = \sum_t L_t$ , and  $T_0 \leq t \leq T_0 + T - 1$ .

Due to massive data of the initial feature pool  $\Phi$ , we hope to find an optimal subset to form the dictionary  $\Psi = [\text{vec}(\mathbf{U}_1), \text{vec}(\mathbf{U}_2), \dots, \text{vec}(\mathbf{U}_K)]$  such that the set  $\Phi$  can be well reconstructed by  $\Psi$  and the size of  $\Psi$  is as small as possible. To achieve this goal, we select  $\Psi$  such that the rest of the features in  $\Phi$  can be well reconstructed using it. Analogous to the optimization problem in [21], the basic problem is formulated as follows:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathbf{X}\|_{2,1} \\ \text{subject to} \quad & \Phi \mathbf{X} = \Phi, \end{aligned} \quad (12)$$

where  $\Phi \in R^{N \times S}$  ( $N = n^2$ ),  $\mathbf{X} \in R^{S \times S}$ , and  $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^S \|\mathbf{X}_i\|_2$ . Apparently,  $\|\mathbf{X}\|_{2,1}$  enforces the group sparsity on the variable  $\mathbf{X}$  and the optimal solution usually contains zero rows. This means that not all features in  $\Phi$  are necessary to be selected to reconstruct any data sample.

Motivated by SVT, we have the equivalent problem of (12) as follows:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathbf{X}\|_{2,1} + \frac{1}{2} \|\mathbf{X}\|_F^2 \\ \text{subject to} \quad & \Phi \mathbf{X} = \Phi. \end{aligned} \quad (13)$$

The Lagrangian function of problem (13) is defined as

$$\mathcal{L}(\mathbf{X}, \mathbf{Y}) = \tau \|\mathbf{X}\|_{2,1} + \frac{1}{2} \|\mathbf{X}\|_F^2 + \langle \mathbf{Y}, \Phi - \Phi \mathbf{X} \rangle, \quad (14)$$

and its dual function is

$$f(\mathbf{Y}) = \inf_{\mathbf{X}} \mathcal{L}(\mathbf{X}, \mathbf{Y}). \quad (15)$$

We first examine the properties of the dual function  $f(\mathbf{Y})$  and then show how to achieve the optimal solution of the problem (13) from its dual optimum directly. As the mixed norm  $\|\mathbf{X}\|_{2,1}$  is not differentiable, it is difficult to optimize the dual function  $f(\mathbf{Y})$  directly. However, we can obtain a useful property of the dual function  $f(\mathbf{Y})$  as follows.

**Theorem 3.** For all  $\tau \geq 0$ , the dual function  $f(\mathbf{Y})$  is continuously differentiable with Lipschitz continuous gradient at most  $M$ . Furthermore, the primal optimal  $\widehat{\mathbf{X}}$  of the problem (13) is given by

$$\widehat{\mathbf{X}} = \mathcal{D}_\tau(\Phi \widehat{\mathbf{Y}}), \quad (16)$$

when the dual optimal  $\widehat{\mathbf{Y}}$  of the problem (13) is obtained.

The proof of Theorem 3 is based on the following results.

**Lemma 4.** For each  $\tau \geq 0$  and  $\mathbf{Y} \in R^{m \times n}$ , one has

$$\mathcal{D}_\tau(\mathbf{Y}) = \arg \min_{\mathbf{X}} \tau \|\mathbf{X}\|_{2,1} + \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2. \quad (17)$$

As a matter of fact, considering the following optimization problem:

$$\min_{x \in R} \tau |x| + \frac{1}{2} (x - y)^2, \quad (18)$$

it is easy to show that the unique solution admits a closed form called the soft-thresholding operator, following a terminology introduced by Donoho and Johnstone [22]; it can be written that

$$\widehat{y} = \begin{cases} 0 & \text{if } |x| \leq \tau, \\ \left(1 - \frac{\tau}{|x|}\right) x & \text{otherwise.} \end{cases} \quad (19)$$

Thus, from a generalized view, one has Lemma 4.

Also, the following result can be deduced based on the properties of Moreau-Yosida regularization [23].

**Lemma 5.** For any  $\mathbf{X}, \mathbf{Y} \in R^{m \times n}$ , one has

$$\|\mathcal{D}_\tau(\mathbf{X}) - \mathcal{D}_\tau(\mathbf{Y})\|_F^2 \leq \langle \mathcal{D}_\tau(\mathbf{X}) - \mathcal{D}_\tau(\mathbf{Y}), \mathbf{X} - \mathbf{Y} \rangle. \quad (20)$$

It follows that  $\mathcal{D}_\tau(\mathbf{Y})$  is globally Lipschitz continuous with modulus 1.

*Proof of Theorem 3.* Since

$$\begin{aligned} f(\mathbf{Y}) &= \inf_{\mathbf{X}} \mathcal{L}(\mathbf{X}, \mathbf{Y}) \\ &= \inf_{\mathbf{X}} \left( \tau \|\mathbf{X}\|_{2,1} + \frac{1}{2} \|\mathbf{X}\|_F^2 + \langle \mathbf{Y}, \Phi - \Phi \mathbf{X} \rangle \right) \\ &= \inf_{\mathbf{X}} \left( \tau \|\mathbf{X}\|_{2,1} + \frac{1}{2} \|\mathbf{X}\|_F^2 + \langle \mathbf{Y}, \Phi \rangle - \langle \Phi^T \mathbf{Y}, \mathbf{X} \rangle \right) \\ &= \inf_{\mathbf{X}} \left( \tau \|\mathbf{X}\|_{2,1} + \frac{1}{2} \|\mathbf{X} - \Phi^T \mathbf{Y}\|_F^2 \right) \\ &\quad + \langle \mathbf{Y}, \Phi \rangle - \frac{1}{2} \|\Phi^T \mathbf{Y}\|_F^2 \\ &= g(\mathbf{Y}) + \langle \mathbf{Y}, \Phi \rangle - \frac{1}{2} \|\Phi^T \mathbf{Y}\|_F^2 \end{aligned} \quad (21)$$

and  $g(\mathbf{Y})$  is the Moreau-Yosida regularization of the mixed norm  $\|\cdot\|_{2,1}$ , using the well-known properties of Moreau-Yosida regularization [23], we get the results that  $g(\mathbf{Y})$  is a globally continuously differentiable convex function. Moreover,  $\nabla g(\mathbf{Y}) = \Phi(\Phi^T \mathbf{Y} - \mathcal{D}_\tau(\Phi^T \mathbf{Y}))$  and  $\nabla g(\mathbf{Y})$  is continuously differentiable with Lipschitz continuous gradient  $\rho$ ; that is, for any  $\mathbf{Y}_1, \mathbf{Y}_2 \in R^{N \times S}$ ,

$$\begin{aligned} \|\nabla g(\mathbf{Y}_1) - \nabla g(\mathbf{Y}_2)\|_F &\leq \|\Phi^T(\mathbf{Y}_1 - \mathbf{Y}_2)\|_F \\ &\leq \rho \|\mathbf{Y}_1 - \mathbf{Y}_2\|_F, \end{aligned} \quad (22)$$

where  $\rho = \sup_{\|\mathbf{Z}\|_F=1, \mathbf{Z} \in R^{N \times S}} \|\Phi^T \mathbf{Z}\|_F$ . Then the gradient of  $f(\mathbf{Y})$  can be obtained as follows:

$$\begin{aligned} \nabla f(\mathbf{Y}) &= \nabla g(\mathbf{Y}) + \Phi - \Phi \Phi^T \mathbf{Y} \\ &= \Phi(\Phi^T \mathbf{Y} - \mathcal{D}_\tau(\Phi^T \mathbf{Y})) + \Phi - \Phi \Phi^T \mathbf{Y} \\ &= \Phi - \Phi \mathcal{D}_\tau(\Phi^T \mathbf{Y}). \end{aligned} \quad (23)$$

It follows that, for any  $\mathbf{Y}_1, \mathbf{Y}_2 \in R^{N \times S}$ ,

$$\begin{aligned} \|\nabla f_2(\mathbf{Y}_1) - \nabla f_2(\mathbf{Y}_2)\|_F &= \|\Phi - \Phi \mathcal{D}_\tau(\Phi^T \mathbf{Y}_1) - \Phi + \Phi \mathcal{D}_\tau(\Phi^T \mathbf{Y}_2)\|_F \\ &= \|\Phi(\mathcal{D}_\tau(\Phi^T \mathbf{Y}_1) - \mathcal{D}_\tau(\Phi^T \mathbf{Y}_2))\|_F \\ &\leq \|\Phi \Phi^T(\mathbf{Y}_1 - \mathbf{Y}_2)\|_F \\ &\leq M \|\mathbf{Y}_1 - \mathbf{Y}_2\|_F, \end{aligned} \quad (24)$$

where the first inequality follows from (20) and  $M = \sup_{\|\mathbf{Z}\|_F=1, \mathbf{Z} \in R^{N \times S}} \|\Phi \Phi^T \mathbf{Z}\|_F$ . When the dual optimal  $\hat{\mathbf{Y}}$  is obtained, by using the result of (21), we can get

$$\begin{aligned} \hat{\mathbf{X}} &= \arg \min_{\mathbf{X}} \mathcal{L}(\mathbf{X}, \mathbf{Y}) \\ &= \arg \min_{\mathbf{X}} \left( \tau \|\mathbf{X}\|_{2,1} + \frac{1}{2} \|\mathbf{X} - \Phi^T \mathbf{Y}\|_F^2 \right) \\ &= \mathcal{D}_\tau(\Phi \hat{\mathbf{Y}}). \end{aligned} \quad (25)$$

This concludes the proof.  $\square$

Since  $f(\mathbf{Y})$  is the dual function of the objective function (13),  $f(\mathbf{Y})$  is concave. Let

$$\begin{aligned} q(\mathbf{Y}) &= -f(\mathbf{Y}) \\ &= - \left( \tau \|\mathcal{D}_\tau(\Phi \mathbf{Y})\|_{2,1} \right. \\ &\quad \left. + \frac{1}{2} \|\mathcal{D}_\tau(\Phi \mathbf{Y})\|_F^2 + \langle \mathbf{Y}, \Phi - \Phi \mathcal{D}_\tau(\Phi \mathbf{Y}) \rangle \right), \end{aligned} \quad (26)$$

which is convex. Thus, the following holds for any  $\mathbf{Y}_1, \mathbf{Y}_2 \in R^{N \times S}$ :

$$\langle q(\mathbf{Y}_1) - q(\mathbf{Y}_2), \mathbf{Y}_1 - \mathbf{Y}_2 \rangle \geq 0. \quad (27)$$

It is also easy to show that  $q(\mathbf{Y})$  belongs to the class  $\mathcal{S}_{0,M}^{1,1}(R^{N \times S})$  and

$$\nabla q(\mathbf{Y}) = -\Phi(\mathbf{I} - \mathcal{D}_\tau(\Phi^T \hat{\mathbf{Y}})), \quad (28)$$

where  $\mathbf{I} \in R^{S \times S}$  is the identity matrix. Therefore, we can solve problem (13) by minimizing the objective function  $q(\mathbf{Y})$ ; that is,

$$\min_{\mathbf{Y}} q(\mathbf{Y}). \quad (29)$$

Therefore, the dictionary  $\Psi$  is selected by the optimal solution  $\hat{\mathbf{Y}}$ ; that is, the  $i$ th column of  $\Phi$  is chosen to be the atom of  $\Psi$  if  $\|\hat{\mathbf{Y}}_i\|_2 \neq 0$ . The optimization algorithm is presented in the next subsection.

**3.2. Optimization Methods.** In this subsection, we develop an efficient optimization algorithm to solve the dual problem (29). Because the objective function  $q(\mathbf{Y})$  is continuously differentiable with Lipschitz continuous gradient, it is feasible to utilize gradient-based optimization methods to achieve the optimal solution for their simplicity and low complexity within each iteration. However, classical gradient-based methods for functions with Lipschitz continuous gradient converge at a rate of  $O(1/N)$ , where  $N$  is the number of iterations during optimization [19]. In fact, this is too slow especially when dealing with large-scale datasets. Note that Nesterov showed in his work [24] that an accelerated gradient algorithm can be constructed such that  $O(1/N^2)$ , the lower bound on the convergence rate for gradient-based methods [25], is achieved when minimizing unconstrained

smooth functions. With this consideration, in the following we propose an accelerated thresholding algorithm to solve these smooth convex optimization problems using Nesterov's method with an adaptive line search scheme [19, 26].

We recall Nesterov's method with an adaptive line search scheme as follows. Take the unconstrained smooth convex minimization problem  $\min_{\mathbf{y} \in \mathbb{R}^n} q(\mathbf{y})$ , for instance, where  $q(\mathbf{y})$  belongs to  $\mathcal{S}_{\mu, L}^{1,1}(\mathbb{R}^n)$ ,  $\mu \geq 0$ , and  $L < +\infty$ . Nesterov's method for this problem utilizes two sequences:  $\{\mathbf{y}_l\}$  and  $\{\mathbf{s}_l\}$ ,  $\mathbf{y}_l, \mathbf{s}_l \in \mathbb{R}^n$ . The searching point  $\mathbf{s}_l$  satisfies

$$\mathbf{s}_l = \mathbf{y}_l + \beta_l (\mathbf{y}_l - \mathbf{y}_{l-1}), \quad (30)$$

where  $\beta_l$  is a tuning parameter. The approximate solution  $\mathbf{y}_{l+1}$  can be computed as a gradient step of  $\mathbf{s}_l$  as

$$\mathbf{y}_{l+1} = \mathbf{s}_l - \frac{1}{L_l} \nabla q(\mathbf{s}_l), \quad (31)$$

where  $1/L_l$  is the step size. Starting from an initial point  $\mathbf{y}_0$ ,  $\mathbf{s}_l$  and  $\mathbf{y}_{l+1}$  can be computed recursively according to (30) and (31) and can arrive at the optimal solution  $\hat{\mathbf{y}}$ . Although it has been shown that Nesterov's method is a very powerful optimization technique for class  $\mathcal{S}_{\mu, L}^{1,1}(\mathbb{R}^n)$  [19], how to choose  $\beta_l$  and  $1/L_l$  in each iteration is a critical issue in Nesterov's method. When they are set properly, the sequence  $\{\mathbf{y}_l\}$  can converge to the optimal  $\hat{\mathbf{y}}$  at a certain convergence rate. As a well-known scheme for setting  $\beta_l$  and  $L_l$ , Nesterov's constant scheme assumes  $\beta_l$  and  $L_l$  to be constant [19], while Nemirovski's line search scheme requires  $L_l$  to monotonically increase, and  $\beta_l$  is independent of  $L_l$  [27]. Both of the settings result in slow convergence.

To overcome this drawback, an adaptive line search scheme for Nesterov's method is proposed in [26]. Under the assumption that  $\bar{\mu}$ , the low bound of  $\mu$ , is known in advance, this scheme is built upon the estimate sequence [19] defined as follows.

**Definition 6** (estimate sequence [19]). A pair of sequences  $\{\phi_l(\mathbf{y})\}$  and  $\{\lambda_l \geq 0\}$  is called an estimate sequence of function  $q(\mathbf{y})$  if  $\lim_{k \rightarrow \infty} \lambda_k = 0$  and  $\phi_l(\mathbf{y}) \leq (1 - \lambda_l)q(\mathbf{y}) + \lambda_l \phi_0(\mathbf{y})$ , for all  $\mathbf{y} \in \mathbb{R}^n$ .

The estimate sequence defined in Definition 6 has the following important property.

**Theorem 7** (see [19]). Let  $\{\phi_l(\mathbf{y})\}$  and  $\{\lambda_k \geq 0\}$  be an estimate sequence. For any sequence  $\{\mathbf{y}_l\}$ ,  $q(\mathbf{y}_l) - \hat{q} \leq \lambda_k (\phi_0(\hat{\mathbf{y}}) - \hat{q}) \rightarrow 0$  if  $q(\mathbf{y}_l) \leq \hat{\phi}_k \equiv \min_{\mathbf{y} \in \mathbb{R}^n} \phi_k(\mathbf{y})$ , where  $\hat{q}$  is the optimal objective function value.

We further specify the estimation sequence in [19]:

$$\phi_l(\mathbf{y}) = \hat{\phi}_l + \frac{\gamma_l}{2} \|\mathbf{y} - \mathbf{v}_l\|^2, \quad (32)$$

where the sequences  $\{\gamma_l\}$ ,  $\{\mathbf{v}_l\}$ , and  $\{\hat{\phi}_l\}$  satisfy

$$\begin{aligned} \mathbf{v}_{l+1} &= \frac{1}{\gamma_{l+1}} \left( (1 - \alpha_l) \gamma_l \mathbf{v}_l + \bar{\mu} \alpha_l \mathbf{s}_l - \alpha_l \nabla q(\mathbf{s}_l) \right), \\ \gamma_{l+1} &= (1 - \alpha_l) \gamma_l + \bar{\mu} \alpha_l, \\ \hat{\phi}_{l+1} &= (1 - \alpha_l) \hat{\phi}_l + \alpha_l q(\mathbf{s}_l) \\ &\quad - \frac{\alpha_l^2}{2\gamma_{l+1}} \|\nabla q(\mathbf{s}_l)\|^2 + \frac{\alpha_l (1 - \alpha_l) \gamma_l}{\gamma_{l+1}} \\ &\quad \times \left( \frac{\bar{\mu}}{2} \|\mathbf{v}_l - \mathbf{s}_l\|^2 + \langle \nabla q(\mathbf{s}_l), \mathbf{v}_l - \mathbf{s}_l \rangle \right). \end{aligned} \quad (33)$$

Then Algorithm 2 in [26] is proposed by modifying Nemirovski's line search scheme with the adaptive parameters of this sequence, which satisfy Theorem 7.

Note that Theorem 3 indicates that the objective function  $q(\mathbf{Y})$  satisfies the conditions of using Nesterov's method with an adaptive line search scheme. Therefore we directly extend Algorithm 2 in [26] to the high-dimensional scenarios to solve (29). The complete procedures are summarized in Algorithm 1.

In Algorithm 1, the while loop from Step 4 to Step 13 is designed to choose a proper step size to satisfy Step 8. As the Lipschitz gradient of  $q(\mathbf{Y})$  is  $M$ ,  $L_l$  is upper bounded by  $2M$  since Step 8 always holds when  $L_l \geq M$  [27]. In Step 14, we initialize  $L_{l+1} = h(\theta)L_l$ , where

$$h(\theta) = \begin{cases} 1, & 1 \leq \theta \leq 5, \\ 0.8, & \theta > 5, \end{cases} \quad (34)$$

and  $\theta > 1$  due to the condition in Step 8 [26]. Apparently, when  $\theta$  is large,  $L_{l+1}$  can be adjusted to avoid the step size  $1/L_l$  becoming too small, which may slow down the convergence rate.

**3.3. Sign Inference and Dictionary Update Mechanism.** This subsection details how to use the dictionary to solve the sign inference problem. Actually, this problem bears similarity to the sign prediction problem in the static signed networks or the unsigned networks varying periodically [3, 8, 11, 12]. In this paper, we intend to infer the unknown relationship between a pair of entities  $i$  and  $j$  based on partial relationship observations of the entire dynamic network at time  $T_0 + T$ . We expect to accomplish this task with the help of the dictionary constructed by the relationship data for times  $T_0$  through  $T_0 + T - 1$ . As aforementioned, there exists strong dependence between the connection status at time  $T_0 + T$  and the history relationship dataset in the dynamic network. We formulate the sign inference problem as follows:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \Psi \mathbf{x}\|_2^2 + \|\mathbf{x}\|_1, \quad (35)$$

where  $\Psi$  is the dictionary and  $\mathbf{y}$  is the invertible vectorization of the matrix  $\mathbf{A}^{(T_0+T)}$  observed at time  $T_0 + T$ ; that is,  $\mathbf{y} = \text{vec}(\mathbf{A}^{(T_0+T)})$ . Because  $\bar{\mathbf{A}}^{(T_0+T)} = \sum_r \sigma_r^{(T_0+T)} \mathbf{U}_r^{(T_0+T)}$  by using SVD and subsequently

```

(1) Input:  $\tilde{\mu}, \alpha_{-1} = 0.5, \mathbf{Y}_{-1} = \mathbf{Y}_0, L_{-1} = L_0, \gamma_0 \geq \tilde{\mu}, \lambda_0 = 1$ 
(2) Output:  $\mathbf{Y}_N$ 
(3) for  $l = 1, 2, \dots, N$  do
(4)   while true do
(5)     compute  $\alpha_k \in (0, 1)$  as the root of  $L_l \alpha_l^2 = (1 - \alpha_l) \gamma_l + \alpha_l \tilde{\mu}$ ,
            $\gamma_{l+1} = (1 - \alpha_l) \gamma_l + \alpha_l \tilde{\mu}, \beta_l = \frac{(1 - \alpha_{l-1}) \gamma_l}{(\gamma_l + L_l \alpha_l) \alpha_{l-1}};$ 
(6)     compute  $\mathbf{S}_l = \mathbf{Y}_l + \beta_l (\mathbf{Y}_l - \mathbf{Y}_{l-1})$ 
(7)     compute  $\mathbf{Y}_{l+1} = \mathbf{S}_l - (1/L_l) \nabla q(\mathbf{S}_l)$ 
(8)     if  $q(\mathbf{Y}_{l+1}) \leq q(\mathbf{S}_l) - (1/2L_l) \|\nabla q(\mathbf{S}_l)\|_F^2$  then
(9)       goto Step 14
(10)    else
(11)       $L_l = 2L_l$ 
(12)    end if
(13)  end while
(14)  set  $\theta = 2L_l \frac{q(\mathbf{S}_l) - q(\mathbf{Y}_{l+1})}{\|\nabla q(\mathbf{S}_l)\|_F^2}, L_{l+1} = h(\theta) L_l$ 
(15)  set  $\lambda_{l+1} = (1 - \alpha_l) \lambda_l$ 
(16) end for

```

ALGORITHM 1: Adaptive line search scheme for dictionary selection.

$\text{vec}(\overline{\mathbf{A}}^{(T_0+T)}) = \sum_r \sigma_r^{(T_0+T)} \text{vec}(\mathbf{U}_r^{(T_0+T)})$ , we will estimate  $\overline{\mathbf{A}}^{(T_0+T)}$  in the form of vector and transform the low-rank matrix reconstruction problem into a traditional  $l_1$ -norm minimization problem in compressive sensing. We solve (35) by applying backtracking-based adaptive orthogonal matching pursuit (BAOMP) method, which incorporates a simple backtracking technique to detect the previously chosen atoms' reliability and then deletes the unreliable atoms at each iteration [28]. Then we force that  $a_{ij}^{(T_0+T)}$ , the element of the resulting matrix, is equal to 1 if  $a_{ij}^{(T_0+T)} > 1$  or equal to -1 if  $a_{ij}^{(T_0+T)} \leq 0$ , to ensure the elements coinciding with the value setting of relationships.

Furthermore, assume that we are given a sequence input samples  $\mathbf{Y} = [\mathbf{y}^{(T_0+T)}, \mathbf{y}^{(T_0+T+1)}, \dots, \mathbf{y}^{(T)}]$ , where  $\mathbf{y}^{(t)} = \text{vec}(\mathbf{A}^{(t)})$ ,  $T_0 + T \leq t \leq \tilde{T}$ , the task of the sign inference becomes to reconstruct the complete adjacency matrices  $\overline{\mathbf{A}}^{(t)}$  one by one. Since the  $\overline{\mathbf{A}}^{(t)}$  may contain some features which are not included in dictionary, it is necessary to add these features into the dictionary to increase the accuracy of the inference. However, the inferred matrix is not the original matrix exactly and consequently the unobserved relationships are not really known. In contrast, the observed adjacency matrix  $\mathbf{A}^{(t)}$  retains all existing relationships. For this reason, we only use  $\mathbf{A}^{(t)}$  to extract the features rather than the optimal solution of (35). We apply the extracting approach in Section 3.2 and add the complementary features into the dictionary. Note that this operation will continuously increase the scale of the dictionary while the samples keep inputting for inference; the dictionary selection approach proposed in Section 3.2 will be applied to compact the dictionary once the size of the dictionary exceeds a predetermined bound.

## 4. Numerical Experiment

In this section, we perform experiments on synthetic networks and show that our low-rank model and dictionary learning method outperform other methods on the task of the sign inference for dynamic signed networks. To ensure that our results are reliable, we conduct all experiments 20 times and average out the results from all of the trials.

To construct synthetic networks, we first consider a weakly balanced complete network  $\overline{\mathcal{G}}$  whose adjacency tensor is  $\overline{\mathcal{A}}$ . The slide of  $\overline{\mathcal{A}}$  at time  $t$  is an adjacency matrix  $\overline{\mathbf{A}}^{(t)}$  in the form of (3). In addition, only a few patterns of  $\overline{\mathbf{A}}^{(t)}$  exist in  $\overline{\mathcal{A}}$ . The observation tensor  $\mathcal{A}$  is formed by sampling some entries from  $\overline{\mathcal{A}}$ . Concretely, we let the adjacency tensor  $\overline{\mathcal{A}}$  of  $\overline{\mathcal{G}}$  consist of 50  $250 \times 250$  matrices of complete 4-weakly balanced structure. For the network  $\overline{\mathcal{G}}$ , four clusters are generated randomly. The size of each cluster is larger than 20 and the sum of the sizes is 250. We further assume that only a part of network relationships is observed by uniform sampling with probability  $p \in (0, 1)$ . It results in  $n^2 p$  entries being randomly sampled from  $\overline{\mathbf{A}}^{(t)}$ , where  $p$  is the fraction of observed entries. We choose a set of matrices whose lost rates are from 0.05 to 0.55 and apply the approach proposed in Section 3.2 to select the dictionary  $\Psi$ .

With the dictionary  $\Psi$  and the given observed matrix  $\mathbf{A}^{(t)}$  at time  $t \geq T_0 + T$ , the task of the sign inference is achieved by solving (35). We use BAOMP to estimate the complete matrix  $\overline{\mathbf{A}}^{(t)}$  and compare the performance of our approach to two state-of-the-art methods, alternating least square (ALS) [29] and singular value projection (SVP) [30], for the sign inference problem. Different from accuracy defined by the relative error on the observed set in [8], we utilize the

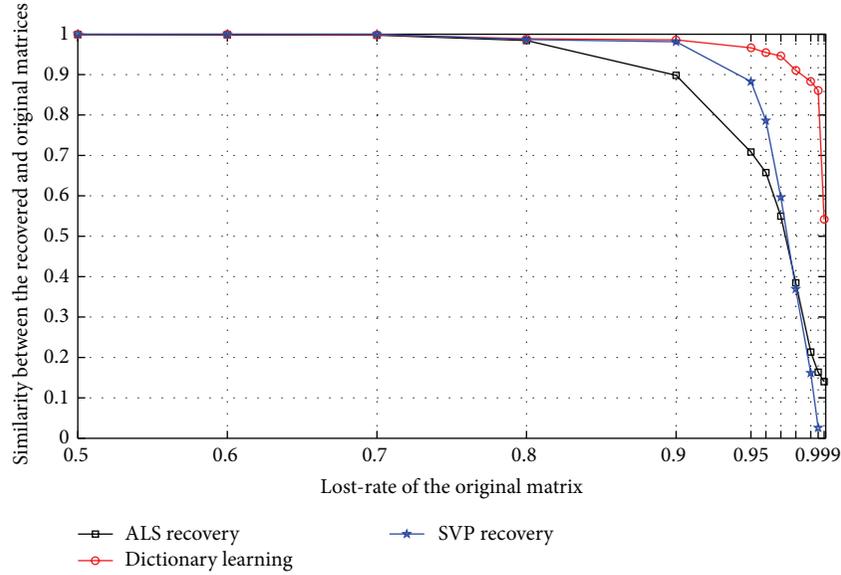


FIGURE 3: Accuracy of sign inference algorithms on synthetic datasets. In general, we can see that dictionary learning outperforms ALS and SVP.

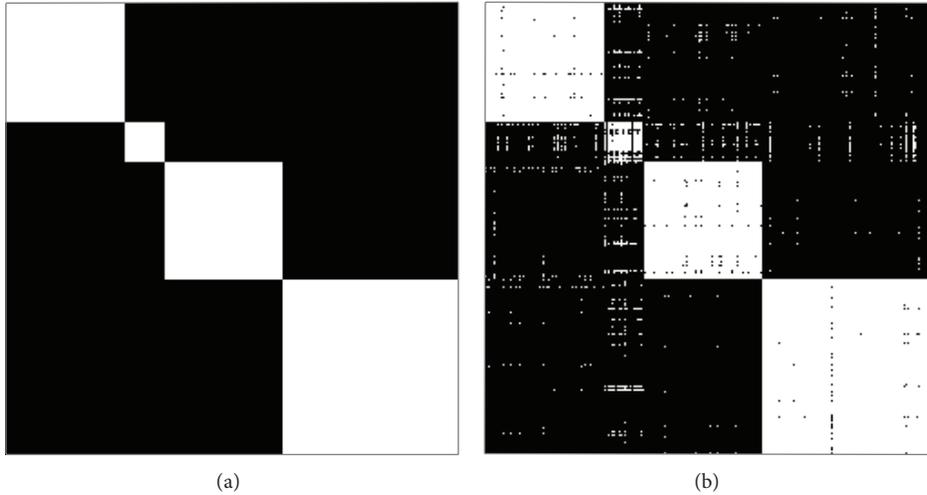


FIGURE 4: An example of the sign inference. (a) illustrates the original matrix. Given the matrix with 98% lost-rate, (b) is the result inferred by dictionary learning method. The similarity of inferred matrix is 0.9347.

similarity between the inferred matrix and the original one to indicate the accuracy of estimation. The definition of the similarity is  $|\langle \bar{\mathbf{A}}^{(t)}, \hat{\mathbf{A}}^{(t)} \rangle| / \|\bar{\mathbf{A}}^{(t)}\|_F \|\hat{\mathbf{A}}^{(t)}\|_F$ . We vary the lost-rate of the original matrix  $\bar{\mathbf{A}}^{(t)}$  from 0.5 to 0.999 and plot the inference accuracy in Figure 3 (lost-rate: 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, and 0.999). Apparently, dictionary learning outperforms ALS and SVP. To present our result more clearly, we also use a visual expression in which the white pixels represent 1 and the black pixels represent -1. Figure 4 shows one example of the sign inference and we find that relationships and the clusters can almost be accurately estimated by our inference approach.

### 5. Conclusion

In this paper, we establish a low-rank tensor model for the dynamic weakly balanced signed networks. With this model, we first extract the feature pool and propose an approach to extract the compact dictionary from pool. To improve the performance of the selection approach, we derive the corresponding dual problem and introduce an accelerated thresholding algorithm to solve the dual problem. Consequently, the optimal solution of the primary problem can be readily obtained from optimizing the dual problem. In addition, combined with the compact dictionary generation method, the sign inference approach is provided for estimating

missing relationships of the dynamic weakly balanced signed networks at a certain time slice. Also, the approach is endowed with the function of the dictionary updating if relationship statuses change.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research was jointly supported by National 973 Program of China (no. 2013CB329204), National 863 Program of China (no. 2011AA01A104), National Natural Science Foundation of China (no. 61100206), Research Fund for Doctoral Program of Higher Education of China (no. 20120005130001), China, and Fund of State Key Laboratory of Information Photonics and Optical Communications (Beijing University of Posts and Telecommunications), China.

## References

- [1] C. Zhang, J. Sun, X. Zhu, and Y. Fang, "Privacy and security for online social networks: challenges and opportunities," *IEEE Network*, vol. 24, no. 4, pp. 13–18, 2010.
- [2] N. Jabeur, S. Zeadally, and B. Sayed, "Mobile social networking applications," *Communications of the ACM*, vol. 56, no. 3, pp. 71–79, 2013.
- [3] D. M. Dunlavy, T. G. Kolda, and E. Acar, "Temporal link prediction using matrix and tensor factorizations," *ACM Transactions on Knowledge Discovery from Data*, vol. 5, no. 2, article 10, 2011.
- [4] A. Clauset, C. Moore, and M. E. J. Newman, "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, no. 7191, pp. 98–101, 2008.
- [5] F. Harary, "On the notion of balance of a signed graph," *The Michigan Mathematical Journal*, vol. 2, no. 2, pp. 143–146, 1953.
- [6] D. Cartwright and F. Harary, "Structural balance: a generalization of Heider's theory," *Psychological Review*, vol. 63, no. 5, pp. 277–293, 1956.
- [7] J. A. Davis, "Clustering and structural balance in graphs," *Human Relations*, vol. 20, no. 2, pp. 181–187, 1967.
- [8] C. J. Hsieh, K. Y. Chiang, and I. S. Dhillon, "Low rank modeling of signed networks," in *Proceedings of the Knowledge Discovery and Data Mining Conference (KDD '12)*, 2012.
- [9] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in *Proceedings of the 19th International World Wide Web Conference (WWW '10)*, pp. 641–650, April 2010.
- [10] K.-Y. Chiang, N. Natarajan, A. Tewari, and I. S. Dhillon, "Exploiting longer cycles for link prediction in signed networks," in *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM '11)*, pp. 1157–1162, October 2011.
- [11] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031, 2007.
- [12] L. Lü and T. Zhou, "Link prediction in complex networks: a survey," *Physica A*, vol. 390, no. 6, pp. 1150–1170, 2011.
- [13] M. Szell, R. Lambiotte, and S. Thurner, "Multirelational organization of large-scale social networks in an online world," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 31, pp. 13636–13641, 2010.
- [14] J. Kunegis, A. Lommatzsch, and C. Bauckhage, "The slashdot zoo: mining a social network with negative edges," in *Proceedings of the 18th International World Wide Web Conference (WWW '09)*, pp. 741–750, ACM, 2009.
- [15] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, pp. 1361–1370, ACM, April 2010.
- [16] J. A. Davis, "Structural balance, mechanical solidarity, and interpersonal relations," *The American Journal of Sociology*, vol. 68, no. 4, pp. 444–462, 1963.
- [17] A. Srinivasan, "Local balancing influences global structure in social networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 5, pp. 1751–1752, 2011.
- [18] F. Heider, "Attitudes and cognitive organization," *The Journal of Psychology*, vol. 21, no. 1, pp. 107–112, 1946.
- [19] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic, Boston, Mass, USA, 2003.
- [20] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [21] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 3449–3456, June 2011.
- [22] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.
- [23] J. B. Hiriart-Urruty and C. Lemarechal, *Convex Analysis and Minimization Algorithms*, vol. 2, Springer, Heidelberg, Germany, 2nd edition, 1996.
- [24] Y. Nesterov, "A method of solving a convex programming problem with convergence rate  $O(1/k^2)$ ," *Soviet Mathematics Doklady*, vol. 27, no. 2, pp. 372–376, 1983.
- [25] A. Nemirovsky and D. Yudin, *Informational Complexity and Efficient Methods for Solution of Convex Extremal Problems*, John Wiley & Sons, New York, NY, USA, 1983.
- [26] J. Liu, J. Chen, and J. Ye, "Large-scale sparse logistic regression," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 547–555, July 2009.
- [27] A. Nemirovski, "Efficient Methods in Convex Programming," Lecture Notes, 1994.
- [28] H. Huang and A. Makur, "Backtracking-based matching pursuit method for sparse signal reconstruction," *IEEE Signal Processing Letters*, vol. 18, no. 7, pp. 391–394, 2011.
- [29] Y. Koren, R. M. Bell, and C. Volinsky, "Matrixfactorization techniques for recommender systems," *IEEE Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [30] P. Jain, R. Meka, and I. Dhillon, "Guaranteed rank minimization via singular value projection," in *Proceedings of the 24th Annual Conference on Neural Information Processing Systems (NIPS '10)*, pp. 934–945, December 2010.