On Spectral Analysis of Signed and Dispute Graphs

Leting Wu*, Xintao Wu†, Aidong Lu* and Yuemeng Li*
* University of North Carolina at Charlotte, Email: {lwu8,alu1,yli60}@uncc.edu
† University of Arkansas, Email: xintaowu@uark.edu

Abstract—This paper presents a study of signed networks from both theoretical and practical aspects. On the theoretical aspect, we conduct theoretical study based on matrix perturbation theorem for analyzing community structures of complex signed networks and show how the negative edges affect distributions and patterns of node spectral coordinates in the spectral space. We prove and demonstrate cluster orthogonality for two types of signed networks: graph with dense inter-community mixed sign edges and $k$-dispute graph. We show why the line orthogonality pattern does not hold in the spectral space for these two types of networks. On the practical aspect, we have developed a clustering method to study signed networks and $k$-dispute networks. Empirical evaluations on both synthetic and real networks show our algorithm outperforms existing clustering methods on signed networks in terms of accuracy.

I. INTRODUCTION

Most social network analysis approaches focused on unsigned graphs, where an edge between two nodes represents a presence of a relationship (e.g., trust or friendship) between two individuals. However, relationships could be inherently negative to express distrust or dislike. In the adjacency matrix $A$ of a signed graph, $a_{ij} = 1$ if the two individuals like each other, $a_{ij} = -1$ if the two dislike each other, and $a_{ij} = 0$ if the two are indifferent.

In order to analyze signed graph, researchers extended some of those existing measures and algorithms for unsigned graph to fit with signed graphs. In [1], the authors extended the definition of modularity on signed graph. Another work [2] showed spectral analysis of signed graphs by the extended definition of Laplacian matrix. However, they failed to clearly relate the structures in signed graph with patterns in the spectral space. In [3], the authors conducted the spectral analysis of approximate $k$-balanced signed graphs by applying matrix perturbation. In the ideal $k$-balanced graph case, edges within one community (inner-community edges) are all positive and edges across two communities (inter-community edges) are all negative. Such a graph is called a $k$-balanced graph where $k$ denotes the number of communities.

The community structure in a signed graph and its spectral space is the combined effect of both positive edges and negative edges. The work [3] only dealt with a special type of signed networks, i.e., $k$-balanced networks where negative connections exist across communities and positive connections exist inside communities. In this paper, we present a spectral analysis framework for analyzing community structures of complex signed networks and show how the negative edges affect distributions and patterns of node spectral coordinates in the spectral space. Specifically, we study the community structure of the following two types of signed networks.

- Type I: inner-community connections are dense with positive edges whereas inter-community connections are mixed with both positive and negative edges.
- Type II: inner-community connections are absent or very sparse whereas inter-community connections are dense with negative edges. A $k$-partite graph with disputing communities (termed as $k$-dispute) is an ideal case of such graph where the relationship such as untrust and dispute mainly exist between communities and rarely inside communities.

Our contributions are two-fold. First, we theoretically prove the clusterability of community structure in signed adjacency spaces of the above two types of signed networks. We have found that both types of signed networks demonstrate significant Orthogonal cluster patterns in the $k$-dimensional spectral subspace spanned by the eigenvectors corresponding to the largest absolutes of eigenvalues. We give out theoretical explanation of the patterns and show them with illustration examples to verify the findings. Second, the discovered orthogonality property in the adjacency eigenspace has potential for a series of applications. In this paper, we present an effective graph partition algorithm, UniAdjCluster, which utilizes the cluster orthogonality property in the adjacency eigenspace. The idea is to project node coordinates to the unit sphere in the spectral space and then apply the classic $k$-means to find clusters. Our empirical evaluations on synthetic data and real-world social networks validate our theoretical findings and show the effectiveness of our graph partition algorithm.

II. RELATED WORK

There is also a large literature on examining the eigenvectors of the graph Laplacian or normal matrix for unsigned networks with various applications such as spectral clustering and graph visualization. In spectral analysis of the Laplacian matrix or the normal matrix, the coordinates are arranged to make the sum of all the distance between two nodes smallest. In their projection spaces, closely related nodes are pulled together to form clusters. Many spectral clustering methods, which exploit this cluster property, have been developed including ratio cut, normalized cut, and min-max cut (refer to a recent tutorial [4]).

Different from the Laplacian matrix or normal matrix, the properties of the adjacency eigenspace have only received attentions in some recent work. The authors in [5] found an EigenSpoke pattern wherein, when plotted the the singular
vectors of graphs against each other, they have clear, separate lines that neatly align along specific axes. In [6], the authors demonstrated the line orthogonality pattern in the adjacency eigenspace of approximate $k$-block graphs. In [7], the authors further gave explicit formulae (as well as its conditions) to quantify how much orthogonal lines rotate from the canonical axes and how far spectral coordinates of nodes with direct links to other communities deviate from the line of their own community. The authors developed spectrum based fraud detection methods that exploit the spectral space of the adjacency matrix of the underlying network topology to identify random link attacks [8].

III. BACKGROUND

A. Notation

A signed network or graph $G$ is a set of $n$ nodes connected by a set of $m$ links which could be positive or negative. The network considered here is symmetric, connected, and without self-loops. Let $\lambda_i$ be the $i$-th eigenvalue of the symmetric adjacency matrix $A$ and $x_i$ the corresponding eigenvector. The eigenvalues are arranged in certain sequence according to the type of the graphs. $x_{ij}$ denotes the $j$-th entry of $x_i$. The spectral decomposition of $A$ is $A = \sum_i \lambda_i x_i x_i^T$.

Formula (1) illustrates the notation. The eigenvector $x_i$ is represented as a column vector. In most social networks, there exist $k$ leading eigenvalues that have value or magnitude significantly greater than the rest ones. The corresponding eigenvectors $x_i (i = 1, \ldots, k)$ contain most topological information of the original graph in the spectral space. The $k$-dimensional spectral space is spanned by $(x_1, \ldots, x_k)$. When we project node $u$ in this $k$-dimensional subspace with $x_i$ as the basis, the row vector $\alpha_u = (x_{1u}, x_{2u}, \ldots, x_{ku})$ is its coordinate of in this subspace. We call $\alpha_u$ the spectral coordinate of node $u$.

$$\begin{pmatrix} x_1 \\ x_i \\ x_k \\ x_n \end{pmatrix} \downarrow \\
\begin{pmatrix} x_{11} & \cdots & x_{i1} & \cdots & x_{k1} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{1n} & \cdots & x_{in} & \cdots & x_{kn} & \cdots & x_{nn} \end{pmatrix} \alpha_u = \begin{pmatrix} x_{1u} \\ x_{iu} \\ x_{ku} \\ x_{nu} \end{pmatrix}$$

(1)

B. Spectral Perturbation

For a symmetric $n \times n$ matrix $A$ with a symmetric perturbation $E$, the matrix after perturbation can be written as $\tilde{A} = A + E$. Let $\lambda_i$ be the $i$-th largest eigenvalue of $A$ with its eigenvector $x_i$. Similarly, $\tilde{\lambda}_i$ and $\tilde{x}_i$ denote the eigenvalue and eigenvector of $\tilde{A}$. It has been shown that the perturbed eigenvector $\tilde{x}_i$ can be approximated by a linear function involving all original eigenvectors (refer to Theorem V.2.8 in [9]). We quote it as below.

Lemma 1: Let $U = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, $S = \text{diag}(\lambda_1, \ldots, \lambda_{i-1}, \lambda_{i+1}, \ldots, \lambda_n)$, and $\beta_{ij} = x_i^T E x_j$. The eigenvector $\tilde{x}_i (i = 1, \ldots, k)$ can be approximated as:

$$\tilde{x}_i \approx x_i + U(\lambda_i - S)^{-1} U^T E x_i$$

(2)

when the following conditions hold:

1) $\delta = |\lambda_i - \lambda_{i+1}| - \|x_i^T E x_i\|_2 - \|U^T E U\|_2 > 0$;
2) $\gamma = \|U^T E x_i\|_2 < \frac{1}{2}\delta$.

C. Spectral Properties of Unsigned Graph with Approximate $k$-block Community Structure

In [7], the authors conducted theoretical studies to demonstrate why node points in an unsigned graph with $k$ communities exhibit the $k$ orthogonal line patterns in the projected spectral subspace of the graph’s adjacency matrix. We quote their main results below thus we can compare and clearly show the differences from our theoretical results on signed networks.

For a graph with $k$ disconnected communities $C_1, \ldots, C_k$ of size $n_1, \ldots, n_k$ respectively ($\sum n_i = n$), its adjacency matrix $A$ can be written as a block-wise diagonal matrix:

$$A = \begin{pmatrix} A_1 & 0 \\ & \ddots \\ & 0 & A_k \end{pmatrix}$$

(3)

where $A_i$ is the $n_i \times n_i$ adjacency matrix of $C_i$. Let $\lambda_{C_i}$ be the largest eigenvalue of $A_i$ in magnitude with eigenvector $x_{C_i} \in \mathbb{R}^{n_i}$. When the sizes of the $k$ disconnected communities are comparable, $\lambda_i = \lambda_{C_i}, i = 1, \ldots, k$.

Lemma 2: For a graph with $k$ disconnected comparable communities as shown in 3, for all $i = 1, \ldots, k$ and $j = 1, \ldots, n$, $\lambda_i \gg |\lambda_j|$. The first $k$ eigenvectors of $A$ have the following form:

$$(x_1, x_2, \ldots, x_k) = \begin{pmatrix} x_{C_1} \\ 0 \\ x_{C_2} \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 0 \end{pmatrix},$$

and all the entries of $x_i$ are non-negative.

Proposition 1: For a graph with $k$ disconnected comparable communities as shown in 3, spectral coordinates of all nodes locate on the $k$ axes $\xi_1, \ldots, \xi_k$ where $\xi_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ is the canonical basis and the $i$-th entry of $\xi_i$ is 1. Specifically, for any node $u \in C_i$, its spectral coordinate has the form

$$\alpha_u = (0, \ldots, 0, x_{iu}, 0, \ldots, 0).$$

(4)

Theorem 1: Denote a graph as $\tilde{A} = A + E$ where $A$ is as shown in 3 and $E$ denotes the edges across communities. The leading $k$ eigenvectors of $A$ can be approximated as

$$\tilde{x}_i \approx x_i + \sum_{j=1,j \neq i}^{k} \beta_{ij} x_j + \frac{1}{\lambda_i} E x_i.$$

Its matrix formula is shown as

$$(\tilde{x}_1, \ldots, \tilde{x}_k) \approx (x_1, \ldots, x_k)R + E(x_1, \ldots, x_k),$$

(5)
For a node $u \in C_i$, let $\Gamma^i_u$ denote its neighbors in $C_j$ for $j \neq i$, and $\Gamma^i_u = \emptyset$. The spectral coordinate of $u$ can be approximated as

$$\alpha_u \approx x_{iu}r_i + \left( \sum_{v \in \Gamma^i_u} e_{uv}x_{1v} \left/ \lambda_1 \right. + \ldots + \sum_{v \in \Gamma^i_u} e_{uv}x_{kv} \left/ \lambda_k \right. \right)$$

(6)

where scalar $x_{iu}$ is the only non-zero entry in its original spectral coordinate shown in 4, $e_{uv}$ is the $(u, v)$ entry of $E$, and $r_i$ is the $i$-th row of the following matrix

$$R = \begin{pmatrix} 1 & \frac{\beta_{i1}}{\lambda_1 - \lambda_2} & \ldots & \frac{\beta_{ik}}{\lambda_1 - \lambda_k} \\ \frac{\beta_{i1}}{\lambda_2 - \lambda_1} & 1 & \ldots & \frac{\beta_{ik}}{\lambda_2 - \lambda_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\beta_{i1}}{\lambda_k - \lambda_1} & \frac{\beta_{i2}}{\lambda_k - \lambda_2} & \ldots & 1 \end{pmatrix}.$$  

(7)

**Proposition 2:** For a graph $\tilde{A} = A + E$, spectral coordinates form $k$ approximately orthogonal clusters. Specifically, for any node $u \in C_i$, if it is not directly connected with other communities, $\alpha_u$ lies on the line $r_i$; otherwise, $\alpha_u$ deviates from lines $r_j$ ($j = 1, \ldots, k$), where $r_j$ is the $j$-th row of matrix $R$ shown in Equation 7.

### IV. Spectral Analysis of Signed Graphs

In this section, we study the effect of positive and negative edges and prove clusterability of two types of signed networks. The key idea is to apply matrix perturbation and show the effect of positive and negative edges between two communities can cancel each other to certain extent so that the $k$ communities still appear to be $k$ orthogonal clusters in the spectral space spanned by the principal eigenvectors.

#### A. Type I: Graphs with Mixed Positive/Negative Inter-Community Edges

Denote the adjacency matrix of a Type I signed graph as $\tilde{A} = A + E$ where $A$ is as shown in Equation 3 and $E$ denotes the signed edges across communities. Because $E$ is not sparse, we cannot apply Lemma 1 directly to get the approximation of the first few eigenvectors of $A$ (conditions in Lemma 1 are not satisfied).

Our following theoretical results show that communities are still distinguishable when the difference between the number of positive inter-community edges and the number of negative inter-community edges is small even though their sum is large. We further show that line pattern is lost when too many cross community edges are present. Nodes with no connection with other communities or with equal weight of positive and negative connection with other communities form a central line for the community and other nodes scatter around lines. Central lines for different communities are orthogonal to each other thus we observe $k$ clusters orthogonal to each other.

**Corollary 1:** Denote the adjacency matrix of a signed graph as $\tilde{A} = A + E$ where $A$ is as shown in Equation 3 and $E$ denotes the signed edges across communities. If $E$ can be decomposed into $E = E_r + E_t$ satisfying:

- $E_r$ contains similar number of positive and negative edges so that $x_i^T E_r x_i \approx 0$;
- $E_t$ contains mostly one kind of edges and $E_t$ satisfies the condition in Lemma 1;

we conclude that the leading $k$ eigenvectors of $\tilde{A}$ can be approximated as

$$\tilde{x}_i \approx x_{iu} + \sum_{j=1,j\neq i}^{k} \frac{\tilde{\beta}_{ji}}{\lambda_i - \lambda_j} x_j$$

(8)

$$+ \frac{1}{\lambda_i} E_r x_i + \frac{1}{\lambda_i^2} E_t E_r x_i + \frac{1}{\lambda_j} E_t E_r x_i,$$

where

$$\tilde{\beta}_{ji} \approx x_j^T E_r x_i + \frac{1}{\lambda_i} x_i^T E_t E_r x_i + \frac{1}{\lambda_j} x_j^T E_t E_r x_i.$$  

We include all our detailed proofs in [10]. The key idea is to decompose $E$ into a series of $E_s$ for $s = 1, \ldots, t - 1$ where each $E_s$ contains a very small number of mixed signed edges. According to Lemma 2, all entries of $x_i$ are non-negative. So the part of $\tilde{\beta}_{ji}$ decided by the positive edges can cancel with that decided by the negative edges. We have $x_i^T E_r x_i \approx 0$. Hence we can apply Lemma 1 to get the approximated eigenvectors iteratively over a series of small perturbations.

Similar as in Theorem 1, we can derive the form of spectral coordinates. We thus have the spectral coordinate of node $u$ as following:

**Proposition 3:** A graph as $\tilde{A} = A + E$ that satisfies the condition of Corollary 1, the spectral coordinate of $u$ can be approximated as

$$\alpha_u \approx x_{iu}r_i + E_r \left( \frac{x_1}{\lambda_1} + \ldots + \frac{x_k}{\lambda_k} \right)$$

(9)

$$(E_t E_r) u \left( \frac{x_1}{\lambda_1} + \ldots + \frac{x_k}{\lambda_k} \right) + (E_t) u \left( \frac{x_1}{\lambda_1} + \ldots + \frac{x_k}{\lambda_k} \right) (\tilde{R} - I)$$

where $x_{iu}$ is the only non-zero entry in its original spectral coordinate shown in 4, $(\cdot)_u$ means the $u$-th row of the corresponding matrices.

If $E_t = 0$, we have $\tilde{\beta}_{ji} \approx 0$ so that $\alpha_u \approx (0, \ldots, 0, x_{iu}, 0, \ldots, 0) + (E_r) u \left( \frac{x_1}{\lambda_1} + \ldots + \frac{x_k}{\lambda_k} \right)$. $E_r$ is the summation of impact of node $u$’s neighbors in $C_i$ through the balanced part $E_r$. Though we assume that we have equally positive and negative edges between communities in $E_r$, individual nodes could still have more positive inter-community edges than negative ones or vice versa. When inter-community connection is dense, many nodes have a large number of inter-community edges that are not equally positive and negative and thus have large jump from the central line $(0, \ldots, 0, x_{iu}, 0, \ldots, 0)$. The line pattern is no longer kept.

When $E_r = 0$ and $E_t$ satisfies the condition in Lemma 1, the graph has very sparse inter-community connection. The result is similar to that of unsigned graph. Line pattern is still kept. However, nodes can scatter on both side of the central line instead of just one side. With neither $E_r$ nor $E_t$ is zero, there are two extra terms in Equation 9 compared with
Equation 6 in Theorem 1. The two extra terms capture further derivation of the nodes from their central lines. We summarize our result below.

**Proposition 4:** The spectral coordinates of $\tilde{A}$ form $k$ approximately orthogonal clusters. For node $u \in C_i$ with no connection with other communities or with similar numbers positive and negative edges with the same community, $\alpha_u$ lies close to $\tilde{r}_z$. Other nodes scatter around $\tilde{r}_z$.

**B. Type II: Signed Graph with k Disputing Communities**

In disputing networks, negative connections exist mostly in between two different communities and the graph.

**Definition 1:** A $k$-dispute graph represents a graph with $k$ communities under two constraints. First, there are no links inside the communities. Second, nodes from different communities are densely connected with negative edges. The adjacency matrix $A_p$ can be written in the following form with proper permutation of the nodes:

$$A_p = \begin{pmatrix}
0 & B_{12} & \cdots & B_{1k} \\
B_{21} & 0 & \cdots & B_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
B_{k1} & B_{k2} & \cdots & 0
\end{pmatrix},$$

where $B_{ij}$ is the $n_i \times n_j$ matrix to represent the relationships between community $i$ and community $j$. We call $A_p$ as a $k$-partite matrix.

We assume all nonzero entries in $A_p$ negative. Our following results can naturally be applied to regular $k$-partite graphs where all nonzero entries in $A_p$ are positive. Let $|\zeta_1| > \cdots > |\zeta_n|$ be the eigenvalues of $A_p$ and $z_i$ be the corresponding eigenvectors. We similarly define spectral coordinates $\alpha_u$ as in Equation 11.

$$\alpha_u \rightarrow \begin{pmatrix}
z_1 \\
z_2 \\
z_k \\
z_{n}
\end{pmatrix} = \begin{pmatrix}
z_{1u} & z_{1n} \\
z_{2u} & z_{2n} \\
z_{ku} & z_{kn} \\
z_{nu}
\end{pmatrix} \begin{pmatrix}
z_1 \\
z_2 \\
z_k \\
z_{n}
\end{pmatrix}$$

Since there are only negative edges, the largest eigenvalue in magnitude is negative according to Perron-Frobenius theorem. The corresponding eigenvector has all the entries nonnegative. The challenge is to characterize the spectral patterns of $k$-dispute graph.

Our idea is to construct a new block matrix $Q$ by subtracting the first eigenpairs from the graph $A_p$. We find that $Q$ has blocks with positive entries on the diagonal and the blocks with positive and negative entries off the diagonal, which fits with part of conditions in Corollary 1. Orthogonalizing the first $k$ eigenvectors of $Q$ with the removed eigenvector, we can get the approximated eigenvectors of the graph with only negative relations and show it clusterable in the spectral space spanned by these eigenvectors.

By the eigen-decomposition of a matrix, $A_p = \sum_{i=1}^{n} \zeta_i z_i z_i^T$. Since there are only negative edges in $A_p$, $\zeta_1$ is negative according to Perron-Frobenius theorem. We remove the effect of $z_1$ from $A_p$ by introducing $Q = A_p - \zeta_1 z_1 z_1^T$.

Write $z_1^T = (z_{C_1}, \ldots, z_{C_k})^T$ where $z_{C_i}$ are the entries of $z_1$ corresponding to community $C_i$. $Q$ contains a block-wise diagonal matrix:

$$A = \begin{pmatrix}
|\zeta_1| z_{C_1} z_{C_1}^T & 0 & \cdots & 0 \\
0 & |\zeta_2| z_{C_2} z_{C_2}^T & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & |\zeta_k| z_{C_k} z_{C_k}^T
\end{pmatrix},$$

and $E = Q - A$ has the corresponding blocks on the diagonal zero and the off diagonal blocks representing the connection between $C_i$ and $C_j$ as $B_{ij} = -\zeta_i z_{C_i}, z_{C_j}^T$.

A is the same with Equation 3 except for the entries relaxed to real nonnegative numbers. $E$ contains both positive and negative real number entries. The proof of Corollary 1 does not require 0-1 entries so that we can apply it to $Q$. In order to satisfy the conditions of Corollary 1, we need to properly decompose $E$ into $E_r$ and $E_l$. The first $k$ normalized eigenvectors of $A$ are:

$$(x_1, \ldots, x_k) = \begin{pmatrix}
\frac{z_{C_1}}{||z_{C_1}||_2} & 0 & \cdots & 0 \\
0 & \frac{z_{C_2}}{||z_{C_2}||_2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{z_{C_k}}{||z_{C_k}||_2}
\end{pmatrix}$$

and their corresponding eigenvalues are $||z_{C_1}||_2 \zeta_1, \ldots, ||z_{C_k}||_2 \zeta_k$. Without loss of generality, we assume $||z_{C_1}||_2 \geq \cdots \geq ||z_{C_k}||_2$. We also notice that the rest of $A$ are all zeros which give a gap between $k$ and $k + 1$ eigenvalues as big as $||z_{C_k}||_2 \zeta_k$.

Let

$$q_{ij} = 1 - \frac{|\zeta_1| ||z_{C_i}||_2 ||z_{C_j}||_2}{z_{C_i}^T B_{ij} z_{C_j}}$$

and

$$E_l = \begin{pmatrix}
0 & q_{12} B_{12} & \cdots & q_{1k} B_{1k} \\
q_{21} B_{21} & 0 & \cdots & q_{2k} B_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
q_{k1} B_{k1} & q_{k2} B_{k2} & \cdots & 0
\end{pmatrix}.$$

Then $x_j^T E_i x_i = x_j^T (E - E_l) x_i = |\zeta_1| (1 - q_{jj}) ||z_{C_i}||_2 ||z_{C_j}||_2 = 0$. Suppose $E_l$ now is dominated by edges of one kind of sign. We can now apply Corollary 1 to get the approximated eigenvectors of $Q$. With proper transformation, we can derive the approximated eigenvectors for $A_p$.

**Proposition 5:** Define $q_{ij}$ and $E_l$ as in Equation 12 and 13. If $E_l$ satisfies the conditions of Lemma 1, the eigenvectors of $A_p$ can be approximated by the following form:

$$(z_1, \tilde{z}_2, \ldots, \tilde{z}_k) \approx (x_1, \cdots, x_k) TK + E \left(0, \frac{x_1}{||z_{C_1}||_2 |\zeta_1|}, \ldots, \frac{x_{k-1}}{||z_{C_{k-1}}||_2 |\zeta_1|} \right) K$$
When $k$ is an upper triangle matrix decided by Gram-Schmidt process to have $TK$ orthogonal. The second to $k$-th eigenvalues are of the different sign of $\zeta_1$.

Proposition 5 converts the clusterability problem of a $k$-dispute graph $A_p$ to whether $E_t$ is a perturbation small enough. When $q_{ij}$’s are close to zeros, $\|E_t\|$ is also close to zero and $A_p$ is thus clusterable. Notice $|\zeta_1|$ is the approximated global density and $\frac{z_i^T E_t z_j}{\|z_i\| \cdot \|z_j\|}$ is the approximated inter-community connection density between $C_i$ and $C_j$. Small $q_{ij}$’s means that the inter-community connections have similar density. Extract a row and we have the spectral coordinates for node $u$ belonging to $C_i$.

Proposition 6: For a $k$-dispute graph $A_p$, spectral coordinate of a node $u$ in community $i$ can be approximated as

$$\alpha_u = \frac{z_{1u}}{\|z_i\|} t_i K + E_u \left( 0, \frac{x_1}{\|z_i\| \|z_C\|}, \ldots, \frac{x_{k-1}}{\|z_{C-1-i}\| \|z_i\|} \right) K,$$

where $t_i$ is the $i$th row of $T$. $t_i K$ and $t_j K$ are approximately orthogonal to each other for $i \neq j$. Two nodes in different communities $C_i$ and $C_j$ stay close to the corresponding vectors so the spectral space has $k$ separable clusters that are almost orthogonal to each other.

C. Unified Adjacency Eigenspace based Clustering

In this section, we present a graph partition algorithm, UniAdjCluster, which utilizes the clusterable pattern in the spectral space of the adjacency matrix.

As shown in Algorithm 1, we calculate the eigenvectors with the largest eigenvalues in magnitude (line 1). If those eigenvalues are positive, it usually indicates that the graph has some communities either with higher density of positive edges inside them or higher density of negative edges outside them. If the largest one of those eigenvalues is positive while the others are negative, it usually indicates that the graph has $k$-partite structure. We then project each spectral coordinate $\alpha_u$ to the unit sphere in the $k$-dimensional subspace by normalizing $\alpha_u$ to its unit length (line 3). With the right value of $k$, we expect to observe that nodes from one community to form a cluster on the unit sphere. Hence there will be $k$ well separated clusters on the unit sphere. We apply the $k$-means clustering algorithm on the unit sphere to produce a partition of the graph (line 4).

We use the Davies-Bouldin Index ($DBI$) as statistics to determine the best fitting. The low $DBI$ indicates output clusters with low intra-cluster distances and high inter-cluster distances.

### Algorithm 1 UniAdjCluster: Unified Adjacency Eigenspace based Clustering

**Input:** $A,K$

**Output:** Clustering results

1. Compute $x_1, \ldots, x_K$ by the eigen-decomposition of $A$ with $|\lambda_1| > \cdots > |\lambda_K|$
2. for $k = 2, \ldots, K$
3. $\alpha_u = (x_{1u}, \ldots, x_{Ku})$ and $\alpha_u = \frac{\alpha_u}{\|\alpha_u\|}$
4. Apply $k$-means algorithm on $\{\alpha_u\}_{u=1,\ldots,n}$
5. Compute fitting statistics from $k$-means algorithm
6. end for

7. Output partitions under $k$ with the best fitting statistics and corresponding eigenvalues

V. Evaluation

A. Datasets

We generate several synthetic graphs for the purpose of illustrating our theoretical results. In all synthetic graphs, we have 5 communities with the number of nodes 200, 180, 170, 150, and 140. For Syn-1 to Syn-3, the inner-community edges of each community, $C_i$, are generated separately with power law degree distribution with the parameter 2.3, whereas the inter-community edges are added with different numbers. Denote ratio of inter-community edges over inner-community edges between $C_i$ and $C_j$ as $p_{ij}$, Syn-1 has $p_{ij} = 20\%$ while Syn-2 increases $p_{ij}$ to 80% and keeps the others same as Syn-1. Syn-3 has $p_{ij} = 80\%$. Note that Syn-1 and Syn-2 are the exactly same as in [7] where the orthogonal line property was illustrated.

Syn-4 and Syn-5 are two signed graphs by flipping 20% and 50% of outer-community edges in Syn-3 to negative. Syn-6, Syn-7, and Syn-8 are three ideal $k$-dispute graphs by removing all the inner-community edges and then randomly adding negative edges between two communities with probability 0.2, 0.4 and 0.6, respectively.

We further generate four perturbed $k$-dispute graphs, Syn-9 to Syn-12, based on Syn-6. Syn-9 is generated by adding inner community edges with the same sign. It has $p_{ij} = (20\%)^{-1}$, which means the ratio of inner-community edges over inter-community edges is 20%. In Syn-10, we further add inner community edges in $C_4$ and $C_5$. $p_{145}$ is now $(80\%)^{-1}$ for these two communities. In Syn-11, all $p_{ij}$ are $(80\%)^{-1}$.

The three real datasets include Facebook, Correlates of War (COW), and Epinions. Refer to [10] for descriptions of these three datasets. We include the statistics of all datasets in Block 1 of Table 1 where “$n$” denotes the number of nodes and “$n(+/−)$” denotes the numbers of positive/negative edges respectively.

We run our algorithm UniAdjCluster on both the synthetic data and real network datasets and show graph partition results in Block 2 of Table 1 where “$k$” denotes the number of output clusters, “$DBI$” is the Davies-Bouldin Index, “Angle” is the average angle between centroids, and “$Q$” is the signed modularity [1]. The high modularity indicates a stronger com-
munity structure that contains more positive inner-community edges and more negative outer-community edges. Similarly, we expect all the angles between centroids of the output clusters are close to 90° since spectral coordinates form k quasi-orthogonal clusters in the determined k-dimensional spectral space.

### B. Accuracy Comparison with Existing Methods

In our experiment, we compare our algorithm UniAdjCluster with two algorithms for signed graphs. The accuracy is defined as $\sum_{i} |C_i \cap \hat{C}_i|$ where $C_i$ denotes the i-th community produced by different algorithms. The first algorithm [2] extends NormalCut to signed graph by modifying the definition of degree matrix. The second algorithm [1] applies an extended definition of modularity and runs various clustering methods such as simulate annealing to maximize the modularity. We show results in Block 3 of Table I where “SNCut” denotes the resolution-limit-free community detection algorithm. Note that we cannot report accuracy values for the three real datasets, Facebook, COW and Epinion since we do not know about their true node-community relations and we did not report accuracy values for Syn-1 to Syn-3 since they are unsigned graphs.

For Type I signed network, Syn-4 and Syn-5, our algorithm produces higher accuracy values than the other two algorithms as shown in Table I. For ideal k-dispute graphs, Syn-6 to Syn-8, all three algorithms achieve 100% accuracy. When lower inner-community connections are added to k-dispute graphs, e.g., as shown for Syn-9, we can still get a 100% accuracy. However, when moderate or large inner-community connections are added, e.g., as shown for Syn-10 and Syn-11, we see decreased accuracy values. The quality of the partitioning produced by our algorithm UniAdjCluster is better than (or comparable with) that produced by the other two algorithms.

### VI. Future Work

In our future work, we plan to conduct spectral analysis of weighted graphs. Our theoretical results do not put restriction on the entries of $A$ or $E$ so we expect they are applicable to weighted graphs. However, the situation in weighted graphs could be more complicated. We also plan to conduct spectral analysis of directed graphs. Many real-world networks have directionality on the edges, making the adjacency matrix non-symmetric. Spectral analysis of directed graphs often transforms the directed network to undirected using various symmetrization approaches or leverages the generalization of the Laplacian matrix for directed graphs. We will study the adjacency eigenspace projection patterns for directed graphs based on the graph perturbation theory and develop spectral clustering based graph partition algorithms for directed graphs.

### ACKNOWLEDGMENTS

This work was supported in part by U.S. National Science Foundation (CCF-1047621) and U.S. National Institute of Health (1R01GM103309).

### REFERENCES


