Examining spectral space of complex networks with positive and negative links

Leting Wu, Xiaowei Ying and Xintao Wu*

Software and Information Systems Department,
University of North Carolina at Charlotte,
9201 University City Blvd., Charlotte, NC 28223, USA
E-mail: lwu8@uncc.edu
E-mail: xying@uncc.edu
E-mail: xwu@uncc.edu
*Corresponding author

Aidong Lu

Computer Science Department,
University of North Carolina,
9201 University City Blvd., Charlotte, NC 28223, USA
E-mail: alu1@uncc.edu

Zhi-Hua Zhou

National Key Laboratory for Novel Software Technology,
Nanjing University,
163 Xianlin Avenue, Nanjing 210046, China
E-mail: zhouzh@lamda.nju.edu.cn

Abstract: Previous studies on social networks are often focused on networks with only positive relations between individual nodes. As a significant extension, we conduct the spectral analysis on graphs with both positive and negative edges. Specifically, we investigate the impacts of introducing negative edges and examine patterns in the spectral space of the graph's adjacency matrix. Our theoretical results show that communities in a *k*-balanced signed graph are distinguishable in the spectral space of its signed adjacency matrix even if connections between communities are dense. This is quite different from recent findings on unsigned graphs, where communities tend to mix together in the spectral space when connections between communities become dense. We further conduct theoretical studies based on graph perturbation to examine spectral patterns of general unbalanced signed graphs. We illustrate our theoretical findings with various empirical evaluations on both synthetic data and real world Correlates of War data.

Keywords: spectral analysis; balanced graph; matrix perturbation; social network.

Reference to this paper should be made as follows: Wu, L., Ying, X., Wu, X., Lu, A. and Zhou, Z-H. (xxxx) 'Examining spectral space of complex networks with positive and negative links', *Int. J. Social Network Mining*, Vol. X, No. Y, pp.000–000.

Biographical notes: Leting Wu is a PhD student in Information Technology at the University of North Carolina at Charlotte since 2009. She earned her BS in Applied Mathematics from Fudan University of China in 2006 and her MA in Mathematics from State University of New York at Buffalo in 2008. Her major research interests include spectral analysis of social networks.

Xiaowei Ying received his PhD in Information Technology at the University of North Carolina at Charlotte in 2011. He received his BA in Mathematics from Fudan University of China in 2006. His major research interests include privacy-preserving data mining and social network analysis.

Xintao Wu is an Associate Professor in the Department of Software and Information Systems at the University of North Carolina at Charlotte, USA. He received his PhD in Information Technology from George Mason University in 2001, BS in Information Science from the University of Science and Technology of China in 1994, and ME in Computer Engineering from the Chinese Academy of Space Technology in 1997. His major research interests include data mining and knowledge discovery, data privacy and security.

Aidong Lu is an Associate Professor in the Department of Computer Science at the University of North Carolina at Charlotte, USA. She has received her PhD from Purdue University in 2005 and MS and BS from Tsinghua University in 2001 and 1999 respectively. Her research interests include visualisation and human computer interaction.

Zhi-Hua Zhou is a Professor in the Department of Computer Science and Technology at the National Key Laboratory for Novel Software Technology at Nanjing University, China. He received his BSc, MSc and PhD in Computer Science from Nanjing University in 1996, 1998 and 2000, respectively, all with the highest honour. His major research interests include machine learning, data mining, pattern recognition and image retrieval.

1 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 contrast to the extensive studies on social networks that restrict to only positive relationships between individuals, in this paper we study *signed* networks with both positive and negative relationships.

Signed networks were originally used in anthropology and sociology to model friendship and enmity (Davis, 1967; Hage and Harary, 1983). The motivation for signed networks arose from the fact that psychologists use –1, 0, and 1 to represent disliking, indifference, and liking, respectively. Graph topology of signed networks can then be expressed as an adjacency matrix where an entry is 1 (or –1) if the relationship is positive (or negative) and 0 if the relationship is absent.

Spectral analysis techniques for 0–1 matrices corresponding to a given network have been well developed. As a significant extension, in this paper we investigate the influence of introducing negative edges in the graph topology and examine community patterns in the spectral space of its signed adjacency matrix. We start from k-balanced signed graphs which have been extensively examined in social psychology, especially from the stability of sentiments perspective (Inohara, 2002). Our theoretical results show that communities in a *k*-balanced signed graph are distinguishable in the spectral space of its signed adjacency matrix even when connections between communities are dense. This is very different from recent findings on unsigned graphs (e.g., Ying and Wu, 2009; Prakash et al., 2010), where communities tend to mix together when connections between communities become dense. We provide a theoretical explanation through treating the *k*-balanced signed graph as a perturbed variant of a disconnected *k*-block network. We further conduct theoretical studies based on graph perturbation to examine spectral patterns of general unbalanced signed graphs. We illustrate our theoretical findings with various empirical evaluations.

2 Notation

A signed graph G can be represented as a symmetric adjacency matrix $A_{n\times n}$ with $a_{ij} = 1$ if there is a positive edge between the nodes i and j, $a_{ij} = -1$ if there is a negative edge between the nodes i and j, and $a_{ij} = 0$ otherwise. A has n real eigenvalues. Let λ_i be the i^{th} largest eigenvalue of A with the eigenvector x_i, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Let x_{ij} denote the j^{th} entry of x_i. The spectral decomposition of A is $A = \sum_i \lambda_i x_i x_i^T$.

$$\begin{array}{c} \\ \\ \\ \alpha_u \rightarrow \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccc} x_1 & & x_i & & x_k & & x_n \\ & & \downarrow & & & & \end{array} \\ \left(\begin{array}{cc|c|cccc} x_{11} & \cdots & x_{i1} & \cdots & x_k & \cdots & x_{n1} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \hline x_{1u} & \cdots & x_{iu} & \cdots & x_{ku} & \cdots & x_{nu} \\ \hline \vdots & & \vdots & & \vdots & & \vdots \\ x_{1n} & \cdots & x_{in} & \cdots & x_{kn} & \cdots & x_{nn} \end{array}\right) \end{array} \tag{1}$$

Formula (1) illustrates our notions. The eigenvector x_i is represented as a column vector. There usually exist k leading eigenvalues that are significantly greater than the remaining ones for networks with k well-separated communities. We call the row vector $\alpha_u = (x_{1u}, x_{2u}, \cdots, x_{ku})$ the spectral coordinate of node u in the k-dimensional subspace spanned by $(x_1, \cdots, x_k)$. This subspace contains most topological information of the original graph. We denote the i^{th} canonical basis as $\xi_i = (0,\ldots,0, 1, 0,\ldots,0)$, where the i^{th} entry of ξ_i is 1 and all other entries are zero.

Let E be a symmetric perturbation matrix, and B be the adjacency matrix after perturbation, i.e., $B = A + E$. Similarly, let μ_i be the i^{th} largest eigenvalue of B with eigenvector y_i, and y_{ij} is the j^{th} entry of y_i. Row vector $\tilde{\alpha}_u = (y_{1u}, \cdots, y_{ku})$ is the spectral coordinate of node u after perturbation.

3 The spectral property of *k*-balanced graph

The *k*-balanced graph is one type of signed graphs that have received extensive examinations in social psychology. It was shown that the stability of sentiments is equivalent to k-balanced (clusterable). A necessary and sufficient condition for a signed graph to be k-balanced is that the signed graph does not contain a cycle with exactly one negative edge (Davis, 1967).

Definition 1: Graph G is a k-balanced graph if the node set V can be divided into k non-trivial disjoint subsets such that $V_1,\ldots,V_k$, edges connecting any two nodes from the same subset are all positive, and edges connecting any two nodes from different subsets are all negative.

The k node sets, $V_1,\ldots,V_k$, naturally form k communities denoted by $C_1,\ldots,C_k$ respectively. Let $n_i = |V_i| \left(\sum_i n_i = n\right)$, and A_i be the $n_i \times n_i$ adjacency matrix of community C_i. After ordering the nodes properly, the adjacency matrix B of a k-balanced graph can be written as:

$$B = A + E, \text{ where } A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_k \end{pmatrix}, \tag{2}$$

and E represents the negative edges across communities. More generally, $e_{uv} = 1(-1)$ if a positive(negative) edge is added between the nodes u and v, and $e_{uv} = 0$ otherwise.

3.1 Non-negative block-wise diagonal matrix

For a graph with k disconnected communities, its adjacency matrix A is shown in (2). Let ν_i be the largest eigenvalue of A_i with eigenvector z_i of dimension n_i. Without loss of generality, we assume $\nu_1 > \cdots > \nu_k$. Since the entries of A_i are all non-negative, with Perron-Frobenius theorem (Stewart and Sun, 1990), ν_i is positive and all the entries of its eigenvector z_i are non-negative. When the k communities are comparable in size, ν_i is the i^{th} largest eigenvalues of A (i.e., $\lambda_i = \nu_i$), and the eigenvectors of A_i can be naturally expanded to the eigenvalues of A as follows:

$$\left(x_1, x_2, \cdots, x_k\right) = \begin{pmatrix} z_1 & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_k \end{pmatrix} \tag{3}$$

Now, consider the node u in community C_i. Note that all the entries in x_i are non-negative, and the spectral coordinate of node u is just the u^{th} row of the matrix in (3). Then, we have

$$\alpha_u = \left(0, \cdots, 0, x_{iu}, 0, \cdots, 0\right), \tag{4}$$

where $x_{iu} > 0$ is the only non-zero entry of α_u. In other words, for a graph with k disconnected comparable communities, spectral coordinates of all nodes locate on k positive half-axes of canonical basis $\xi_1,\cdots,\xi_k$ and nodes from the same community locate on the same half axis.

3.2 A general perturbation result

Let $\Gamma_u^i (i = 1,\dots,k)$ be the set of nodes in C_i that are newly connected to node u by perturbation E: $\Gamma_u^i = \{v : v \in C_i, e_{uv} = \pm 1\}$. Wu et al. (2011a) derived several theoretical results on general graph perturbation. We include the approximation of spectral coordinates below as a basis for our spectral analysis of signed graphs. Please refer to Wu et al. (2011a) for proof details.

Theorem 1: Let A be a block-wise diagonal matrix as shown in (2) and E be a symmetric perturbation matrix satisfying $\| E \|_2 \ll \lambda_k$. Let $\beta_{ij} = x_i^T E x_j$. For a graph with the adjacency matrix $B = A + E$, the spectral coordinate of an arbitrary node $u \in C_i$ can be approximated as

$$\tilde{\alpha}_u \approx x_{iu} r_i + \left(\sum_{v \in \Gamma_u^1} \frac{e_{uv} x_{1v}}{\lambda_1}, \dots, \sum_{v \in \Gamma_u^k} \frac{e_{uv} x_{kv}}{\lambda_k} \right) \tag{5}$$

where scalar x_{iu} is the only non-zero entry in its original spectral coordinate shown in (4), and r_i is the i^{th} row of matrix R in (6):

$$R = \begin{pmatrix} 1 & \frac{\beta_{12}}{\lambda_2 - \lambda_1} & \cdots & \frac{\beta_{1k}}{\lambda_k - \lambda_1} \\ \frac{\beta_{21}}{\lambda_1 - \lambda_2} & 1 & \cdots & \frac{\beta_{2k}}{\lambda_k - \lambda_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\beta_{k1}}{\lambda_1 - \lambda_k} & \frac{\beta_{k2}}{\lambda_2 - \lambda_k} & \cdots & 1 \end{pmatrix}. \tag{6}$$

3.3 Moderate inter-community edges

Proposition 1: Let $B = A + E$ where A has k disconnected communities and $\| E \|_2 \ll \lambda_k$ and E is non-positive. We have the following properties:

1 If node $u \in C_i$ is not connected to any C_j $(j \neq i)$, $\tilde{\alpha}_u$ lies on the half-line r_i that starts from the origin, where r_i is the i^{th} row of matrix R shown in (6). The k half-lines are approximately orthogonal to each other.

2 If node $u \in C_i$ is connected to node $v \in C_j$ $(j \neq i)$, $\tilde{\alpha}_u$ deviate from r_i. Moreover, the angle between $\tilde{\alpha}_u$ and r_j is an obtuse angle.

To illustrate Proposition 1, we now consider a two-balanced graph. Suppose that a graph has two communities and we add some sparse edges between two communities. For node $u \in C_1$ and $v \in C_2$, with (5), the spectral coordinates can be approximated as

$$\tilde{\alpha}_u \approx x_{1u} r_1 + \left(0, \frac{1}{\lambda_2} \sum_{v \in \Gamma_u^2} e_{uv} x_{2v} \right), \tag{7}$$

$$\tilde{\alpha}_v \approx x_{2v} r_2 + \left(\frac{1}{\lambda_1} \sum_{u \in \Gamma_v^1} e_{uv} x_{1u},\, 0 \right), \tag{8}$$

where $r_1 = \left(1, \frac{\beta_{12}}{\lambda_2 - \lambda_1}\right)$ and $r_2 = \left(\frac{\beta_{21}}{\lambda_1 - \lambda_2}, 1\right)$.

The item 1 of Proposition 1 is apparent from (7) and (8). For those nodes with no inter-community edges, the second parts of the right-hand side (RHS) of (7) and (8) are 0 since all e_{uv}'s are 0, and hence they lie on the two half-lines r_1 (nodes in C_1) and r_2 (nodes in C_2). Note that r_1 and r_2 are orthogonal since $r_1 r_2^T = 0$.

Next, we explain item 2 of Proposition 1. Consider the inner product

$$\langle \tilde{\alpha}_u, r_2 \rangle = \tilde{\alpha}_u r_2^T = \frac{1}{\lambda_2} \sum_{v \in \Gamma_u^2} e_{uv} x_{2v}.$$

If node $u \in C_1$ has some negative links to C_2 ($e_{uv} = -1$), $\langle \tilde{\alpha}_u, r_2 \rangle$ is thus negative. In other words, $\tilde{\alpha}_u$ lies outside the two half-lines r_1 and r_2.

Another interesting pattern is the direction of rotation of the two half lines. For the two-balanced graph, r_1 and r_2 rotate counter-clockwise from the axis ξ_1 and ξ_2. Notice that all the added edges are negative ($e_{uv} = -1$), and hence

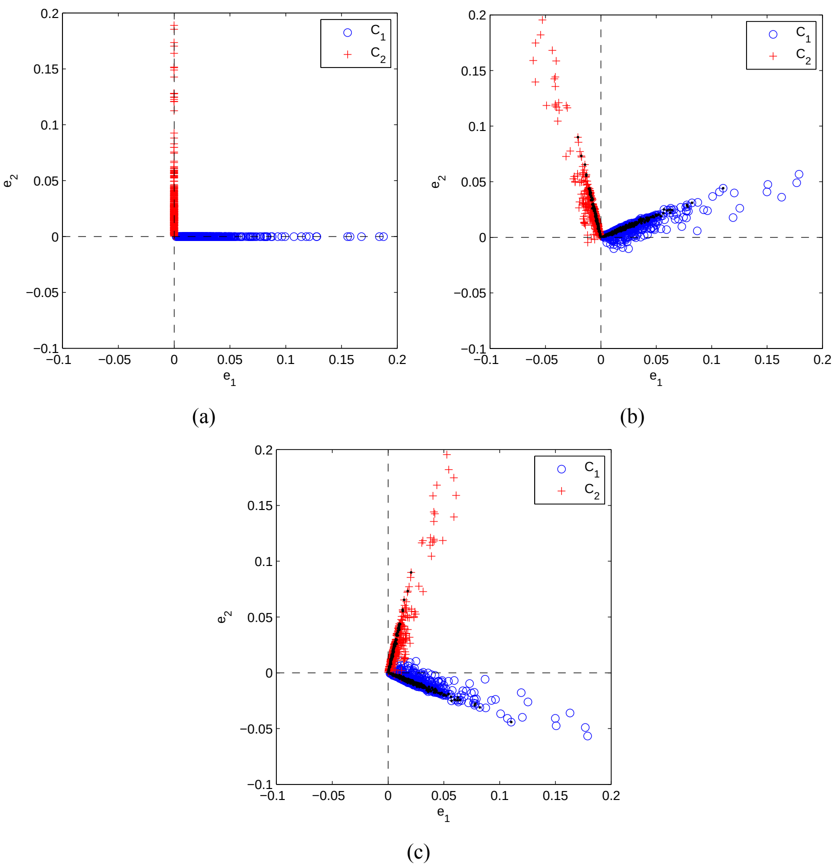
$$\beta_{12} = \beta_{21} = x_1^T E x_2 = \sum\nolimits_{u,v=1}^{n} e_{uv} x_{1u} x_{2v} < 0.$$

Therefore, $\frac{\beta_{12}}{\lambda_2 - \lambda_1} > 0, \frac{\beta_{21}}{\lambda_1 - \lambda_2} < 0$, which implies that r_1 and r_2 have a counter-clockwise rotation from the basis.

- *Comparison with adding positives edges:* when the added edges are all positive ($e_{uv} = 1$), we can derive the following two properties in a similar manner:
  1. nodes with no inter-community edges lie on the k half-lines (when $k = 2$, the two half-lines exhibit a clockwise rotation from the axes)
  2. for node $u \in C_i$ that connects to node $v \in C_j$, $\tilde{\alpha}_u$ and r_j form an acute angle.

Figure 1 shows the scatter plot of the spectral coordinates for a synthetic graph, *Synth*-2. *Synth*-2 is a two-balanced graph with 600 and 400 nodes in each community. We generate *Synth*-2 and modify its inter-community edges via the same method as synthetic dataset *Synth*-3 in Section 5.1. As we can see in Figure 1(a), when the two communities are disconnected, the nodes from C_1 and C_2 lie on the positive part of axis ξ_1 and ξ_2 respectively. We then add a small number of edges connecting the two communities ($p = 0.05$). When the added edges are all negative, as shown in Figure 1(b), the spectral coordinates of the nodes from the two communities form two half-lines respectively. The two quasi-orthogonal half-lines rotate counter-clockwise from the axes. Nodes with negative inter-community edges lie outside the two half-lines. On the contrary, if we add positive inter-community edges, as shown in Figure 1(c), the nodes from two communities display two half-lines with a clockwise rotation from the axes, and nodes with inter-community edges lie between the two half-lines.

Figure 1 *Synth*-2: rotation and deviation with inter-community edges ($p = 0.05$), (a) disconnected (b) add negative edges (c) add positive edges (see online version for colours)



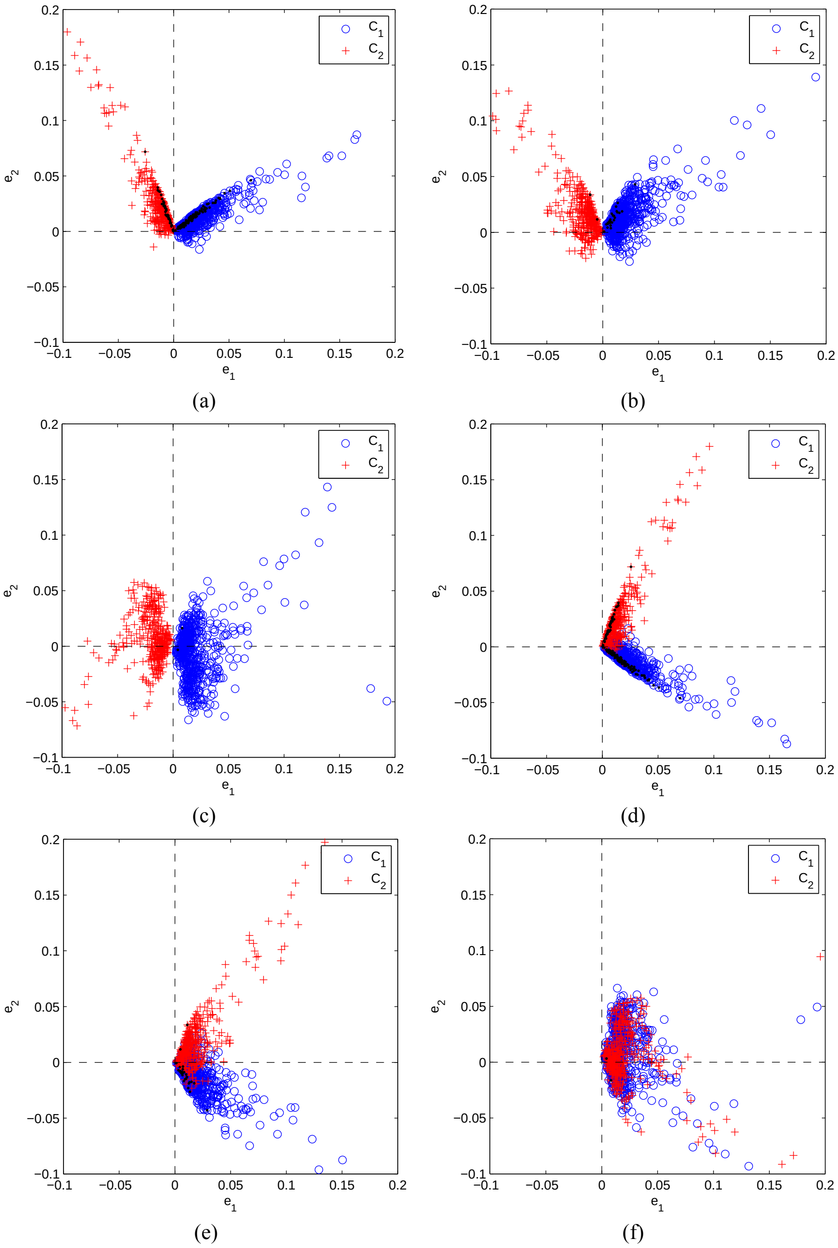
3.4 *Increase the number of inter-community edges*

Theorem 1 holds when the magnitude of perturbation is moderate. With perturbation of large magnitude, we can divide the perturbation matrix into several perturbation matrices of small magnitude and approximate the eigenvectors step by step. More general, the perturbed spectral coordinate of a node u can be approximated as

$$\tilde{\alpha}_u \approx \alpha_u R + \sum_{v=1}^{n} e_{uv} \alpha_v \Lambda^{-1}, \tag{9}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$.

Figure 2 *Synth*-2 with different types and sizes of inter-community edges, (a) negative edges ($p = 0.1$) (b) negative edges ($p = 0.3$) (c) negative edges ($p = 1$) (d) positive edges ($p = 0.1$) (e) positive edges ($p = 0.3$) (f) positive edges ($p = 1$) (see online version for colours)



One property implied by (9) is that, after adding negative inter-community edges, nodes from different communities are still separable in the spectral space. Note that R is close to an orthogonal matrix, and hence the first part of RHS of (9) specifies an orthogonal transformation. The second part of RHS of (9) specifies a deviation away from the position after the transformation. Note that when the inter-community edges are all negative ($e_{uv} = -1$), the deviation of α_u is just towards the negative direction of α_v (each dimension is weighted with λ_i^{-1}). Therefore, after perturbation, the nodes u and v are further separable from each other in the spectral space. The consequence of this repellency caused by adding negative edges is that nodes from different communities stay away from each other in the spectral space. As the magnitude of the noise increases, more nodes deviate from the half-lines r_i, and the line pattern eventually disappears.

- *Positive large perturbation:* When the added edges are positive, we can similarly observe the opposite phenomenon: more nodes from the two communities are 'pulled' closer to each other by the positive inter-community edges and are finally mixed together, indicating that the well-separable communities merge into one community.

Figure 2 shows the spectral coordinate of *Synth*-2 when we increase the number of inter-community edges (p = 0.1, 0.3 and 1). For Figures 2(a) to 2(c), we add negative inter-community edges in *Synth*-2, and for Figures 2(d) to 2(f), we add positive inter-community edges. As we add more and more inter-community edges, no matter positive or negative, more and more nodes deviate from the two half-lines, and finally the line pattern diminishes in Figures 2(c) or 2(f). When adding positive inter-community edges, the nodes deviate from the lines and hence finally mix together as show in Figure 2(f), indicating that two communities merge into one community. Different from adding positive edges, which mixes the two communities in the spectral space, adding negative inter-community edges 'pushes' the two communities away from each other. This is because nodes with negative inter-community edges lie outside the two half-lines as shown in Figures 2(a) and 2(b). Even when p = 1, as shown in Figure 2(c), two communities are still clearly separable in the spectral space.

4 Unbalanced signed graph

Signed networks in general are unbalanced and their topologies can be considered as perturbations on balanced graphs with some negative connections within communities and some positive connections across communities. Therefore, we can divide an unbalanced signed graph into three parts

$$B = A + E_{\text{in}} + E_{\text{out}}, \tag{10}$$

where A is a non-negative block-wise diagonal matrix as shown in (2), E_{in} represents the negative edges within communities, and E_{out} represents the both negative and positive inter-community edges.

- *Add negative inner-community edges:* for the block-wise diagonal matrix A, we first discuss the case where a small number of negative edges are added within the communities. E_{in} is also a block-wise diagonal. Hence, $\beta_{ij} = x_i^T E_{\text{in}} x_j = 0$ for all $i \neq j$, and matrix R caused by E_{in} in (6) is reduced to the identity matrix I.

  Consider that we add one negative inner-community edge between the nodes $u, v \in C_i$. Since $R = I$, only λ_i and x_i are involved in approximating $\tilde{\alpha}_u$ and $\tilde{\alpha}_v$:

  $$\tilde{\alpha}_u = (0, \ldots, 0, y_{iu}, 0, \ldots, 0), \quad y_{iu} \approx x_{iu} - \frac{x_{iv}}{\lambda_i}$$

  $$\tilde{\alpha}_v = (0, \ldots, 0, y_{iv}, 0, \ldots, 0), \quad y_{iv} \approx x_{iv} - \frac{x_{iu}}{\lambda_i}.$$

  Without loss of generality, assume $x_{iu} > x_{iv}$, and we have the following properties when adding $e_{uv} = -1$:

  1 both the nodes u and v move towards the negative part of axis ξ_i after perturbation: $y_{iu} < x_{iu}$ and $y_{iv} < x_{iu}$.

  2 node v moves farther than u after perturbation: $| y_{iv} - x_{iv} | > | y_{iu} - x_{iu} |$.

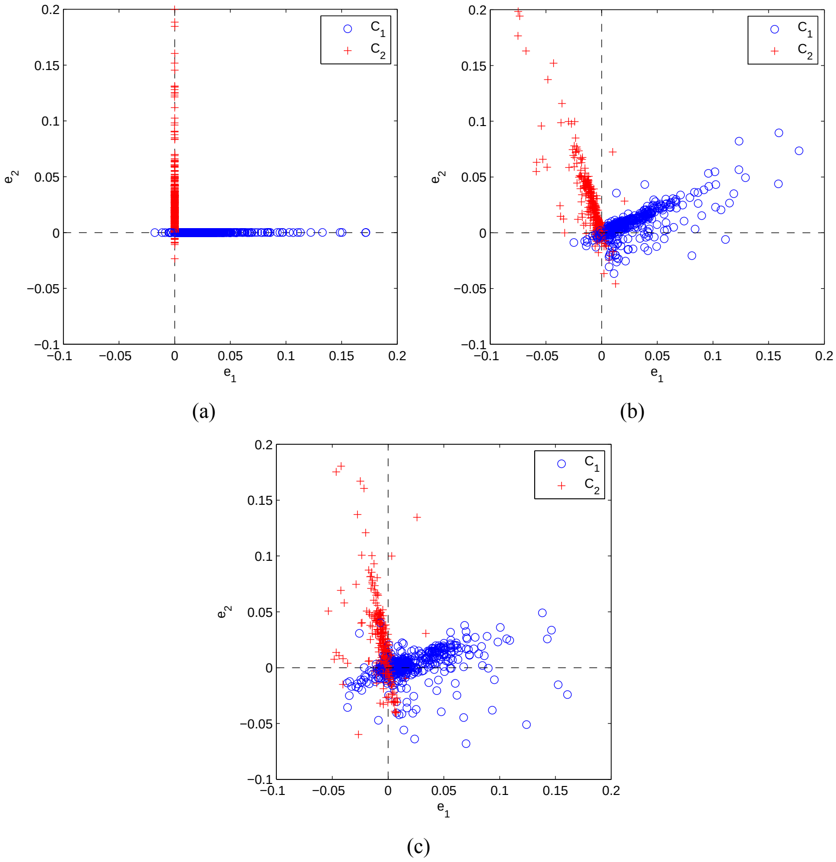
  The two preceding properties imply that, for those nodes close to the origin, adding negative edges would 'push' them towards the negative part of axis ξ_i, and a small number of nodes can thus lie on the negative part of axis ξ_i, i.e., $y_{iu} < 0$.

- *Add inter-community edges:* the spectral perturbation caused by adding E_{out} to matrix $A + E_{\text{in}}$ can be complicated. Notice that $(A + E_{\text{in}})$ is still a block-wise matrix, and we can still apply Theorem 1 and conclude that, when E_{out} is moderate, the major nodes from k communities form k lines in the spectral space and nodes with inter-community edges deviate from the lines.

  It is difficult to give the explicit form of the lines and the deviations, because x_{iu} and the inter-community edges can be either positive or negative. However, we expect that the effect of adding negative edges on positive nodes is still dominant in determining the spectral pattern, because most nodes lie along the positive part of the axes and the majority of inter-community edges are negative. Communities are still distinguishable in the spectral space. The majority of nodes in one community lie on the positive part of the line, while a small number of nodes may lie on the negative part due to negative connections within the community.

We make graph *Synth*-2 unbalanced by flipping the signs of a small proportion q of the edges. When the two communities are disconnected, as shown in Figure 3(a), after flipping $q = 0.1$ inner-community edges, a small number of nodes lie on the negative parts of the two axes. Figure 3(b) shows the spectral coordinates of the unbalanced graph generated from balanced graph *Synth*-2 ($p = 0.1$, $q = 0.1$). Since the size of the inter-community edges is small, we can still observe two orthogonal lines in the scatter plots. The negative edges within the communities cause a small number of nodes lie on the negative parts of the two lines. Nodes with inter-community edges deviate from the two lines. For Figure 3(c), we flip more edge signs ($p = 0.1$, $q = 0.2$). We can observe that more nodes lie on the negative parts of the lines, since more inner-community edges are changed to negative. The rotation angles of the two lines are smaller than that in Figure 3(b). This is because the positive inter-community edges 'pull' the rotation clockwise a little, and the rotation we observe depends on the effects from both positive and negative inter-community edges.

Figure 3 Spectral coordinates of unbalanced graphs generated from *Synth*-2, (a) two disconnected, $q = 0.1$ (b) $p = 0.1$, $q = 0.1$ (c) $p = 0.1$, $q = 0.2$ (see online version for colours)



Calculation of the eigenvectors of an $n \times n$ matrix takes in general a number of operations $O(n^3)$, which is almost inapplicable for large networks. However, in our framework, we only need to calculate the first k eigen-pairs. Furthermore, adjacency matrices in our context are usually sparse. The Arnoldi/Lanczos algorithm (Golub and Van Loan, 1996) generally needs $O(n)$ rather than $O(n^2)$ floating point operations at each iteration.

5 Evaluation

5.1 Synthetic balanced graph

Dataset *Synth*-3 is a synthetic three-balanced graph generated from the power law degree distribution with the scaling exponent 2.5. The three communities of *Synth*-3 contain 600, 500, 400 nodes, and 4,131, 3,179, 2,037 edges respectively. All the 13,027 inter-community edges are set to be negative. We delete the inter-community edges randomly until a proportion p of them remain in the graph. The parameter p is the ratio of the size of inter-community edges to that of the inner-community edges. If $p = 0$ there are no inter-community edges. If $p = 1$, inner- and inter-community edges have the same size.

Figure 4 shows the change of spectral coordinates of *Synth*-3, as we increase the size of inter-community edges. When there are no negative links ($p = 0$), the scatter plot of the spectral coordinates is shown in Figure 4(a). The disconnected communities display three orthogonal half-lines. Figure 4(b) shows the spectral coordinates when the size of inter-community edges is moderate ($p = 0.1$). We can see the nodes form three half-lines that rotate a certain angle, and some of the nodes deviate from the lines. Figures 4(c) and 4(d) show the spectral coordinates when we increase the size of inter-community edges ($p = 0.3, 1$). We can observe that, as more inter-community edges are added, more and more nodes deviate from the lines. However, nodes from different communities are still separable from each other in the spectral space.

We also add positive inter-community edges on *Synth*-3 for comparison, and the spectral coordinates are shown in Figures 4(e) and 4(f). We can observe that, different from adding negative edges, as the size of inter-community edges (p) increases, nodes from the three communities get closer to each other, and completely mix in one community in Figure 4(f).

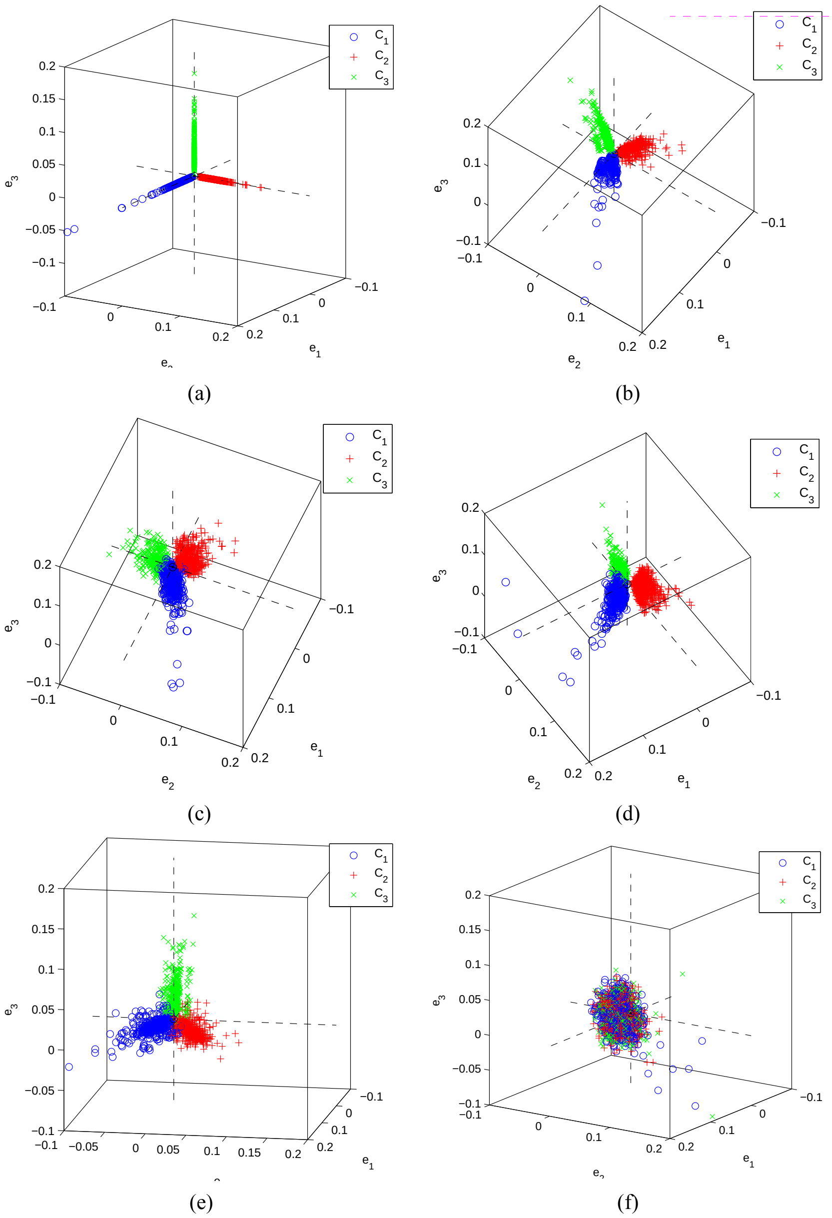
5.2 Synthetic unbalanced graph

To generate an unbalanced graph, we randomly flip the signs of a small proportion q of the inner- and inter-community edges of a balanced graph, i.e., the parameter q is the proportion of unbalanced edges given the partition. We first flip edge signs of a small size of inter-community edges. Figures 5(a) and 5(b) show the spectral coordinates after we flip $q = 10\%$ and $q = 20\%$ edge signs on *Synth*-3 with $p = 0.1$. We can observe that, even the graph is unbalanced, nodes from the three communities exhibit three lines starting from the origin, and some nodes deviate from the lines due to the inter-community edges.

We then flip edge signs of a large size of inter-community edges. Figure 5(c) shows the spectral coordinates after we flip $q = 20\%$ edge signs on *Synth*-3 with $p = 1$. We can observe that the line pattern diminishes because of the large number of inter-community edges. However, the nodes from three communities are separable in the spectral space, indicating that the unbalanced edges do not greatly change the patterns in the spectral space.

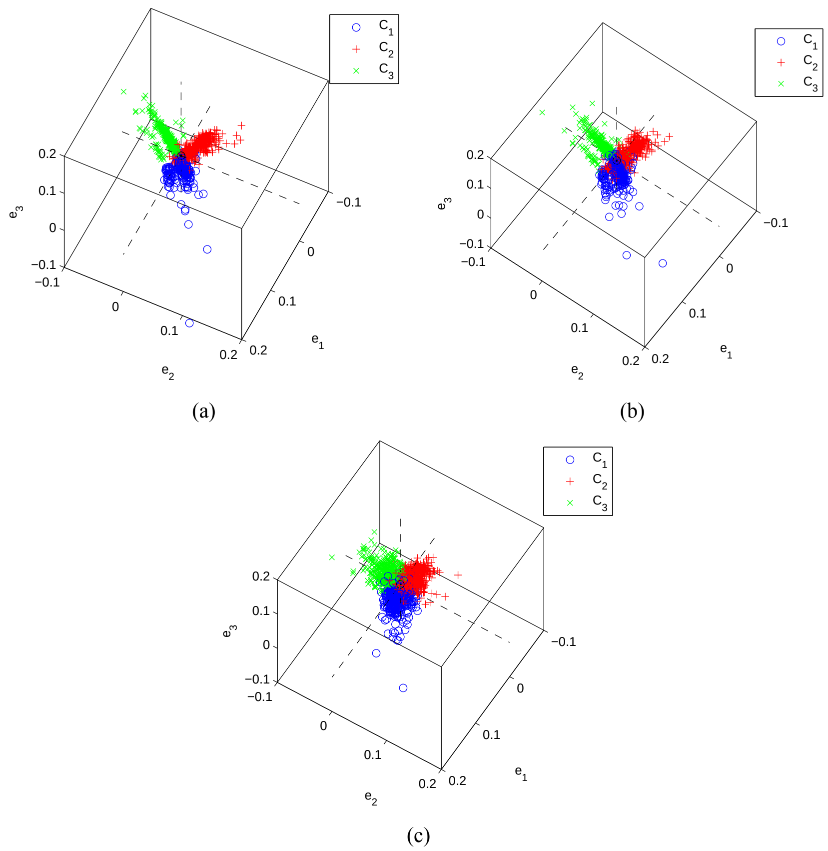
Figure 4 The spectral coordinates of the three-balanced graph *Synth*-3, (a) three disconnected communities (b) negative $p = 0.1$ (c) negative $p = 0.3$ (d) negative $p = 1$ (e) positive $p = 0.1$ (f) positive $p = 1$ (see online version for colours)

Comment [t1]: Author: Please note that e_2 in Figures 4(a), 4(e) and 4(f) was cut, please provide another version of theses figures.



Notes: Figures 4(b) to 4(d): add negative inter-community edges;
Figures 4(e) to 4(f): add positive inter-community edges.

Figure 5 The spectral coordinates of a unbalanced synthetic graph generated via flipping signs of inner- and inter-community edges of *Synth*-3 with $p = 0.1$ or 1, (a) $p = 0.1$, $q = 0.1$ (b) $p = 0.1$, $q = 0.2$ (c) $p = 1$, $q = 0.2$ (see online version for colours)



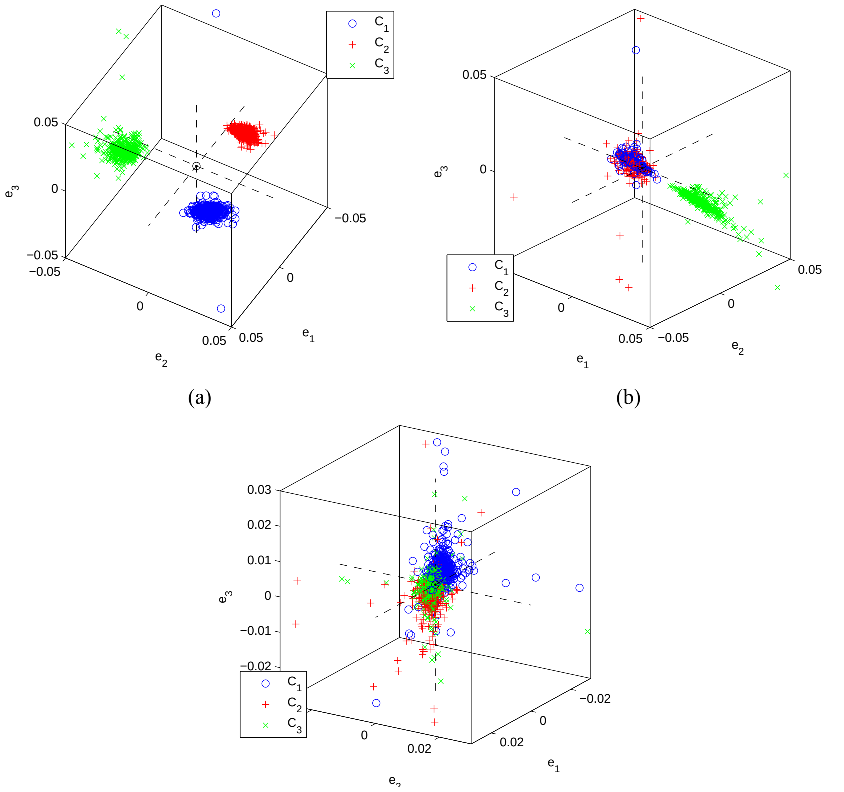
5.3 Comparison with Laplacian spectrum

The signed Laplacian matrix is defined as $L = \bar{D} - A$ where $\bar{D}_{n \times n}$ is a diagonal degree matrix with $\bar{D}_{ii} = \sum_{j=1}^{n} | A_{ij} |$ (Kunegis et al., 2010). Note that the unsigned Laplacian matrix is defined as $L = D - A$ where $D_{n \times n}$ is a diagonal degree matrix with $D_{ii} = \sum_{j=1}^{n} A_{ij}$. The eigenvectors corresponding to the k smallest eigenvalues of Laplacian matrix also reflect the community structure of a signed graph: the k communities form k clusters in the Laplacian spectral space. However, eigenvectors associated with the smallest eigenvalues are generally instable to noise according to the matrix perturbation theory (Stewart and Sun, 1990). Hence, when it comes to real-world networks, the communities may no longer form distinguishable clusters in the Laplacian spectral space.

Figure 6(a) shows the Laplacian spectrum of a balanced graph, *Synth*-3 with $p = 0.1$. We can see that the nodes from the three communities form three clusters in the spectral space. However, the Laplacian spectrum is less stable to the noise. Figures 6(b) and 6(c) plot the Laplacian spectra of the unbalanced graphs generated from *Synth*-3. We can observe that C_1 and C_2 are mixed together in Figure 6(b) and all the three communities are not separable from each other in Figure 6(c). For comparison, the adjacency spectra of the corresponding graphs were shown in Figure 5(b) and Figure 5(c) respectively where we can observe that the three communities are well-separable in the adjacency spectral space.

Figure 6 The Laplacian spectral space of signed graphs, (a) $p = 0.1$, $q = 0$ (balanced) (b) $p = 0.1$, $q = 0.2$ (c) $p = 1$, $q = 0.2$ (see online version for colours)

Comment [t2]: Author: Please note that e_2 in Figure 6(c) was cut please provide another version of this figure.



(c)

5.4 Case study on correlates of war

We conduct an empirical evaluation on a real-world signed graph taken from the Correlates of War dataset over the period 1993–2001 (Ghosn et al., 2004). The dataset contains international relations such as alliance and dispute among different countries and areas. The dataset Formal Alliances (v3.03) records formal alliance among different countries. There are three types of alliance: defence pact (Type I), neutrality and non-aggression pact (Type II), and ententes (Type III). The dataset Militarised Interstates Disputes (v3.1) records all instances of when one state threatened, displayed, or used force against another, e.g., border extension between Colombia and Venezuela and Turkish groups entering Iraqi territory. There are five levels of dispute: no militarised action (Level 1), threat to use force (Level 2), display of force (Level 3), use of force (Level 4), and war (Level 5). For those disputes that involve different levels of actions, we use the highest level to represent the level of dispute.

We construct a signed graph where military alliances are represented by positive edges and disputes by negative edges. We use the alliance of defence pact (Type I) to construct the cliques of positive edges and the disputes of use of force (Level 4) and war (Level 5) to construct the bipartite graph of negative edges. When the positive edge conflicts with the negative edge, we treat the negative edge with higher priority. This is because the use of force breaks the alliance. The adjacency matrix of the constructed signed graph A contains 159 nodes with 1,093 positive edges and 155 negative edges. USA has the maximum number of positive links and Yugoslavia has the maximum number of negative links.

Figure 7 shows the distribution of eigenvalues. We can see that the first five eigenvalues are significantly greater than the remaining ones. We project the spectral coordinate of each point onto the unit sphere and apply the classic k-means algorithm (setting $k = 5$) to partition this signed network. To measure the partition quality, we define $EDiff = p^{(i)} + n^{(o)} - p^{(o)} - n^{(i)}$, where $p^{(i)}$ and $p^{(o)}$ are the number of positive edges inside and outside each community whereas $n^{(i)}$ and $n^{(o)}$ are the number of negative edges inside and outside each community. Note that *EDiff* indicates the degree of balance of the signed graph. A higher *EDiff* value indicates a better community partition. Figure 8 shows the different *EDiff* values when we vary the number of partitions (k). We can see *EDiff* reaches its maximum when $k = 5$.

Figure 7 The eigenvalues of a from the Correlates of War (see online version for colours)

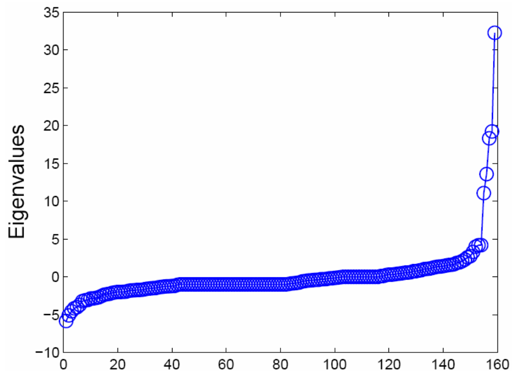
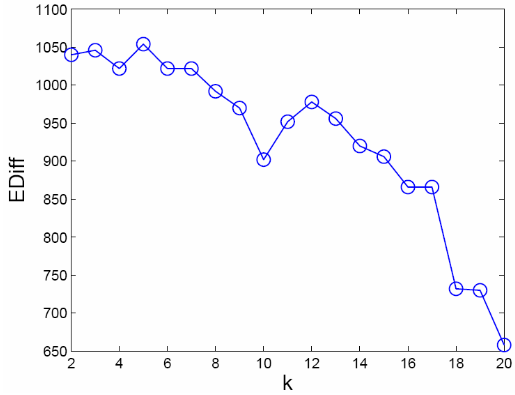
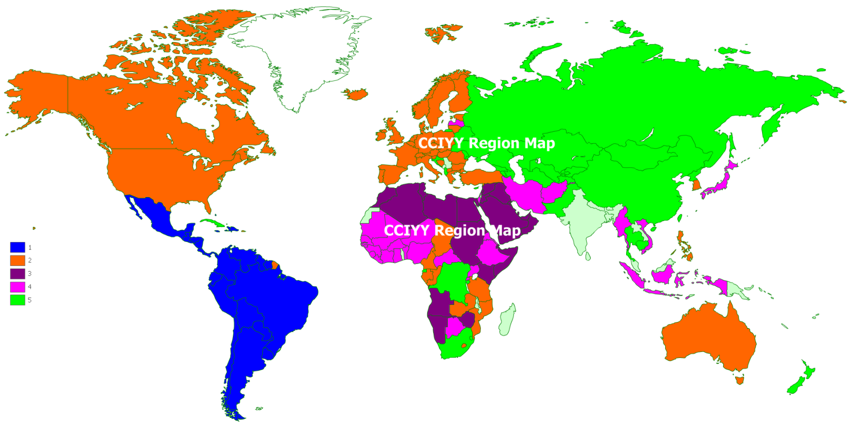


Figure 8 Quality of partition vs. *k* (see online version for colours)



We plot the world map based on our output with five communities in Figure 9. The power blocs can be identified as follows: Latin America, the West (including USA, Canada, Australia, and West European countries), Muslim World, West Africa, and China and the former Soviet Union. Our results match the configuration depicted in Huntington (1997) with a few notable exceptions. The West African power bloc is absent in the configuration of Huntington (1997). Some other noteworthy differences are South Korea is grouped with the West and South Africa is grouped with China and the Soviet Union.

Figure 9 Map of the communities in the conflict and alliance network found using $k = 5$ (see online version for colours)



6 Related work

Social networks have received much attention these days. To understand and utilise the information in a social network, researches have developed various methods to capture the structure and characteristics of the network from different perspectives (Kleinberg, 2007; Shiga et al., 2007; Tantipathananandh et al., 2007; Backstrom et al., 2006; Koren et al., 2006; Berger-Wolf and Saia, 2006; Kumar et al., 2006; Fast et al., 2005; Spertus et al., 2005; Baumes et al., 2004; Girvan and Newman, 2002; Kempe et al., 2003). Among them, spectral analysis of the adjacency matrix and its variants (e.g., Laplacian matrix and normal matrix) has shown intimate relationship between the combinatorial characteristics of a graph and the algebraic spectral properties of its matrix (Seary and Richards, 2003). For example, the eigenvalues of the adjacency matrix encode information about the cycles of a network as well as its diameter.

There is also a large literature on examining the eigenvectors of the graph Laplacian or normal matrix with various applications such as spectral clustering (Hagen and Kahng, 1992; Chan et al., 1993; Pothen et al., 1990; Shi and Malik, 2000; Ng et al., 2001; Ding et al., 2001; Huang et al., 2008) and graph visualisation (Belkin and Niyogi, 2002). In spectral analysis of the Laplacian matrix or the normal matrix, the coordinates are arranged to make the sum of all the distances between two nodes the 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, normalised cut, and min-max cut [refer to a recent tutorial (Von Luxburg, 2007)].

There are several studies on community partition in social networks with negative (or negatively weighted) edges (Yang et al., 2007; Bansal and Chawla, 2002; Demaine and Immorlica, 2003; Traag and Bruggeman, 2009). Bansal and Chawla (2002) introduced correlation clustering motivated by document clustering and agnostic learning and showed that it is an NP-hard problem to make a partition to a complete signed graph. Demaine and Immorlica (2003) gave an approximation algorithm based on a linear-programming rounding and the region-growing technique and showed that the problem is APX-hard, i.e., any approximation would require improving the best approximation algorithms known for minimum multi-cut. Traag and Bruggeman (2009) adapted the concept of modularity to detect communities in networks where both positive and negative links are present and also evaluated the social network of international disputes and alliances. However, the algorithm involves several parameters that are hard to configure in practice. Yang et al. (2007) investigated community partition from social networks with negative edges. Kunegis et al. (2009) presented a case study on the signed Slashdot Zoo corpus and analysed various measures (including signed clustering coefficient and signed centrality measures). Leskovec (2010) studied several signed online social networks (Epinions, Slashdot, and the voting network of Wikipedia) and developed a theory of status to explain the observed edge signs. Spectral graph analysis can help users capture topological structures in spectral spaces. Laplacian graph kernels that apply to signed graphs were described in Hou (2005) and Kunegis et al. (2010). However, the authors only focused on two-balanced signed graphs and many results (such as signed graphs' definiteness property) do not hold for general k-balanced graphs.

7 Conclusions

We conducted theoretical studies based on graph perturbation to examine spectral patterns of signed graphs, which extends our preliminary research reported in Wu et al. (2011b). Our results showed that communities in a *k*-balanced signed graph are distinguishable in the spectral space of its signed adjacency matrix even when connections between communities are dense. To our best knowledge, these are the first reported findings on showing separability of communities in the spectral space of the signed adjacency matrix. We applied our partition method to a real-world signed network and our findings match explanations of international relations. In our future work, we will conduct evaluations on more signed networks and compare with other clustering methods for signed networks such as Traag and Bruggeman (2009).

Acknowledgements

This research was supported in part by U.S. National Science Foundation (CCF-1047621, CNS-0831204) for L. Wu, X. Wu, and A. Lu and by the National Science Foundation of China (61073097, 61021062) for Z-H. Zhou.

References

Backstrom, L., Huttenlocher, D., Kleinberg, J. and Lan, X. (2006) 'Group formation in large social networks: membership, growth, and evolution', *KDD*, pp.44–54.

Bansal, N. and Chawla, S. (2002) 'Correlation clustering', *Machine Learning*, Vol. 56, pp.238–247.

Baumes, J., Goldberg, M., Magdon-Ismail, M. and Wallace, A. (2004) 'Discovering hidden groups in communication networks', *ISI*, pp.378–389.

Belkin, M. and Niyogi, P. (2002) 'Laplacian eigenmaps and spectral techniques for embedding and clustering', *NIPS*, Vol. 1, pp.585–592.

Berger-Wolf, T. and Saia, J. (2006) 'A framework for analysis of dynamic social networks', *KDD*, pp.523–528.

Chan, P., Schlag, M. and Zien, J. (1993) 'Spectral k-way ratio-cut partitioning and clustering', *Proceedings of the 30th international Design Automation Conference*, pp.749–754.

Davis, J. (1967) 'Clustering and structural balance in graphs', *Human Relations*, Vol. 20, No. 2, pp.181–187.

Demaine, E. and Immorlica, N. (2003) 'Correlation clustering with partial information', *The 6th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, pp.1–13.

Ding, C., He, X., Zha, H., Gu, M. and Simon, H. (2001) 'A min-max cut algorithm for graph partitioning and data clustering', *ICDM*, pp.107–114.

Fast, A., Jensen, D. and Levine, B. (2005) 'Creating social networks to improve peer-to-peer networking', *KDD*, pp.568–573.

Ghosn, F., Palmer, G. and Bremer, S. (2004) 'The MID3 data set, 19932001: procedures, coding rules, and description', *Conflict Management and Peace Science*, Vol. 21, No. 2, pp.133–154.

Girvan, M. and Newman, M. (2002) 'Community structure in social and biological networks', *Proc. Natl. Acad. Sci.*, Vol. 99, No. 12, pp.7821–7826, USA.

Golub, G. and Van Loan, C. (1996) *Matrix Computations*, The Johns Hopkins University Press, Baltimore, London.

Hage, P. and Harary, F. (1983) *Structural Models in Anthropology*, pp.56–60, Cambridge University Press, UK.

Hagen, L. and Kahng, B. (1992) 'New spectral methods for ratio cut partitioning and clustering', *IEEE Transactions on Computer-aided Design*, Vol. 11, No. 9, pp.1074–1085.

Hou, Y.P. (2005) 'Bounds for the least Laplacian eigenvalue of a signed graph', *Acta Mathematica Sinica*, Vol. 21, No. 4, pp.955–960.

Huang, L., Yan, D., Jordan, M. and Taft, N. (2008) 'Spectral clustering with perturbed data', *Advances in Neural Information Processing Systems (NIPS)*, pp.705–712.

Huntington, S. (1997) *The Clash of Civilizations and the Remaking of World Order*, Simon and Schuster, New York.

Inohara, T. (2002) 'Characterization of clusterability of signed graph in terms of newcombs balance of sentiments', *Applied Mathematics and Computation*, Vol. 133, No. 1, pp.93–104.

Kempe, D., Kleinberg, J. and Tardos, É. (2003) 'Maximizing the spread of influence through a social network', *KDD*, pp.137–146.

Kleinberg, J. (2007) 'Challenges in mining social network data: processes, privacy, and paradoxes', *KDD*, pp.4–5.

Leskovec, J., Huttenlocher, D. and Kleinberg, J. (2010) 'Signed networks in social media', *CHI*, pp.1361–1370.

Koren, Y., North, S. and Volinsky, C. (2006) 'Measuring and extracting proximity in networks', *KDD*, pp.245–255.

Kumar, R., Novak, J. and Tomkins, A. (2006) 'Structure and evolution of online social networks', *KDD*, pp.611–617.

Kunegis, J., Lommatzsch, A. and Bauckhage, C. (2009) 'The slashdot zoo: mining a social network with negative edges', *WWW*, pp.741–750.

Kunegis, J., Schmidt, S., Lommatzsch, A. and Lerner, J. (2010) 'Spectral analysis of signed graphs for clustering, prediction and visualization', *SDM*, edited by De Luca, E. and Albayrak, S., pp.559–570.

Ng, A., Jordan, M. and Weiss, Y. (2001) 'On spectral clustering: analysis and an algorithm', *NIPS*, pp.849–856.

Pothen, A., Simon, H. and Liou, K. (1990) 'Partitioning sparse matrices with eigenvectors of graphs', *SIAM Journal on Matrix Analysis and Applications*, Vol. 11, No. 3, pp.430–452.

Prakash, B., Sridharan, A., Seshadri, M., Machiraju, S. and Faloutsos, C. (2010) 'EigenSpokes: surprising patterns and scalable community chipping in large graphs', *PAKDD*, pp.435–448.

Seary, A. and Richards, W. (2003) 'Spectral methods for analyzing and visualizing networks: an introduction', *National Research Council, Dynamic Social Network Modelling and Analysis: Workshop Summary and Papers*, pp.209–228.

Shi, J. and Malik, J. (2000) 'Normalized cuts and image segmentation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp.888–905.

Shiga, M., Takigawa, I. and Mamitsuka, H. (2007) 'A spectral clustering approach to optimally combining numerical vectors with a modular network', *KDD*, pp.647–656.

Spertus, E., Sahami, M. and Buyukkokten, O. (2005) 'Evaluating similarity measures: a large-scale study in the orkut social network', *KDD*, pp.678–684.

Stewart, G. and Sun, J. (1990) *Matrix Perturbation Theory*, Academic Press, San Diego, CA.

Tantipathananandh, C., Berger-Wolf, T. and Kempe, D. (2007) 'A framework for community identification in dynamic social networks', *KDD*, pp.717–726.

Traag, V. and Bruggeman, J. (2009) 'Community detection in networks with positive and negative links', *Physics Review E*, Vol. 80, No. 3, p.036115.

Von Luxburg, U. (2007) 'A tutorial on spectral clustering', *Statistics and Computing*, Vol. 17, No. 4, pp.395–416.

Wu, L., Ying, X., Wu, X. and Zhou, Z-H. (2011a) 'Line orthogonality in adjacency eigenspace with application to community partitions', *IJCAI*, pp.2349–2354.

Wu, L., Ying, X., Wu, X., Lu, A. and Zhou, Z-H. (2011b) 'Spectral analysis of k-balanced signed graphs', *PAKDD*, pp.1–12.

Yang, B., Cheung, W. and Liu, J. (2007) 'Community mining from signed social networks', *IEEE Trans. on Knowl. and Data Eng.*, Vol. 19, No. 10, pp.1333–1348.

Ying, X. and Wu, X. (2009) 'On randomness measures for social networks', *SDM*, pp.709–720.