Technical Report

Department of Computer Science
and Engineering
University of Minnesota
4-192 Keller Hall
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 13-029

Modeling Cohesiveness and Enmity for Signed Social Communities

Pengkui Luo and Zhi-Li Zhang

September 20, 2013

Modeling Cohesiveness and Enmity for Signed Social Communities

Pengkui Luo and Zhi-Li Zhang
{pluo,zhzhang}@cs.umn.edu

Abstract

Social interactions can exhibit either positive or negative linkages, resulting in the abstraction of "signed graphs" or "signed social networks." The social balance theory implies that a signed social network tends to converge to a balanced state where all individuals are divided into two cohesive groups that are antagonistic to each other. In this paper, we take this implication as a premise, and model the cohesiveness within each community and the enmity between two communities. We evaluate our model using synthetic signed networks, online signed social networks, and the U.S. Senate Voting network, and demonstrate its effectiveness in modeling the cohesiveness and enmity.

1 Introduction

Social interactions – no matter online or in real-life – can be either *positive* or *negative*: an individual may establish a positive "link" with another to denote friendship, trust, support, proximity or similarity; on the other hand, one may also link negatively to another to express enmity, distrust, disapproval or dissimilarity. For example, on *Epinions* one user can express support or disagreement of another; users on *Slashdot* can label others as "friends" or "foes"; on *Wikipedia* participants can vote for or against those nominated administrators during the election. With more and more social web sites enabling negative linkage, the studies on both positives and negatives would be beneficial. Networks with such a mixture of positive edges and negative edges are called *signed graphs* [4] in graph theory. Quite a number of interesting questions have been raised on signed graphs, in the context of online social networks. For instance, Leskovec et al. [7] tackled on the link prediction problem, i.e., given a social network with signs on most edges, predict the signs on missing edges. Bansal et al. [1] initiated the study of clustering on signed graphs. Most of these studies were rooted in the social balance theory developed by Harary et al. [3], in order to account for the relationship between microscopic link signs and macroscopic community structures.

The social balance theory implies that a signed social network has a tendency in converging to a balanced state in which all individuals can be separated into two communities, where positive edges only exist within the same community to create intra-group cohesiveness, and negative edges only lies between members across the two communities to generate inter-group enmity. In this paper, we take the above implications from social balance theory as premises, and focus on studying the following problem: *given such a balanced (or nearly balanced) signed social network, how should we quantify the intra-group cohesiveness and the inter-group enmity?* To the best of our knowledge, this problem was not studied yet. Indeed, there exist many cohesiveness measures associating with different cohesive subgroup structures [9], but they all set at positively weighted graphs, not signed graphs. As for the enmity measures, we could not find one in the literature. Therefore, we believe that our work can make nontrivial contributions to the research community.

In the remainder of the paper, we mathematically model the community cohesiveness and enmity in Section 2, and evaluate our model using synthetic signed networks (Section 3), online social networks (Section 4) and the U.S. Senate voting network (Section 5), and conclude in Section 6.

2 Mathematical Modeling

A *signed graph* $\Gamma = (V, E)$ is an undirected, simple, connected, finite graph, with a symmetric adjacency matrix $A = [a_{i,j}]_{n\times n}$ where $a_{i,j} \in \{\pm 1, 0\}$. Its *signed Laplacian* is defined as $L := \text{diag}(d_1, \ldots, d_n) - A$ where $d_i := \sum_{j=1}^{n} |a_{i,j}|$. Let $L^+ = [l^+_{i,j}]_{n\times n}$ be the *pseudoinverse* of L. Both L and L^+ are symmetric and positive semi-definite (or positive definite if the graph is unbalanced [5]). Therefore, L^+ has the following eigen-decomposition, where $\Lambda^+ := \text{diag}(\lambda^+_1, \ldots, \lambda^+_n)$ comprises of decreasing non-negative eigenvalues $\lambda^+_1 \geq \cdots \geq \lambda^+_n \geq 0$ down the diagonal; and $Q = [q_{i,j}]_{n\times n}$ contains orthonormal columns $\{\vec{q}_j\}_{j=1}^{n}$ as corresponding eigenvectors.

$$L^+ = Q\Lambda^+ Q' = \sum\nolimits_{k=1}^{n} \lambda^+_k \vec{q}_k \vec{q}_k{}' \tag{1}$$

Define $X = (\vec{x}_1|\ldots|\vec{x}_n) := (\Lambda^+)^{1/2} Q'$, then we have $L^+ = X'X$. In particular, $l^+_{j,j} = \vec{x}_j{}'\vec{x}_j = \sum_k \lambda^+_k q^2_{j,k} > 0$. Therefore, each node j in the graph is embedded on $\mathbb{R}^n$, with its coordinates given by vector $\vec{x}_j$ w.r.t. the standard basis $(\vec{e}_1, \ldots, \vec{e}_n)$ where $\vec{e}_k \in \mathbb{R}^n$ for any k. We can thereby explore graph properties by studying the column vectors of X.

If this signed graph of size n is *balanced* [4], then its vertex set can be partitioned into two subgroups each with n_1 and n_2 members, such that all edges within each group carry positive signs, whereas all edges connecting the two groups are negative.[1] Without loss of generality, we assume all nodes indexed from 1 to n_1 are in the first group, and the remaining n_2 nodes in the second group. Then L^+ can be written into blocks $[L^+_1, L^+_{12}; L^{+\prime}_{12}, L^+_2]$, and accordingly $X = [X_1|X_2]$, in which $X_1 = (\vec{x}_1|\ldots|\vec{x}_{n_1})$ and $X_2 = (\vec{y}_1|\ldots|\vec{y}_{n_2})$ where $\vec{y}_j := \vec{x}_{j+n_1}$ for $1 \leq j \leq n_2$. We have $L^+_1 = X_1{}'X_1$, $L^+_2 = X_2{}'X_2$, and $L^+_{21}{}' = L^+_{12} = X_1{}'X_2$. In particular, $l^+_{i,j+n_1} = \vec{x}_i{}'\vec{y}_j$. Therefore, L^+_1 and L^+_2 characterize the geometric properties of the two subgroups; whereas L^+_{12} encodes the pair-wise relationship between two subgroups.

To capture the intra-group cohesiveness, say, that of the first subgroup, we can study how "compact" these vectors $X_1 = (\vec{x}_1, \ldots, \vec{x}_{n_1})$ are in the $\mathbb{R}^n$ space. Therefore, we define the cohesiveness within subgroup i ($i \in \{1, 2\}$) as follows:

$$\text{Coh}_1 = \frac{1}{n_1} \sum\nolimits_{j=1}^{n_1} \|\vec{x}_j - \vec{c}_1\| \tag{2}$$

where $\vec{c}_1 := \frac{1}{n_1}\sum_{j=1}^{n_1} \vec{x}_j$ is the center of the first group. Since it is a Euclidean space, $\|\cdot\|$ is computed as the square root of the inner product on the vector itself. Similarly, the cohesiveness of the second subgroup can be formulated as $\frac{1}{n_2}\sum_{j=1}^{n_2} \|\vec{y}_j - \vec{c}_2\|$ where $\vec{c}_2 := \frac{1}{n_2}\sum_{j=1}^{n_2} \vec{y}_j$. The smaller the value is, the more compact or cohesive the group is. Similarly, we can define the cohesiveness measure of the second group.

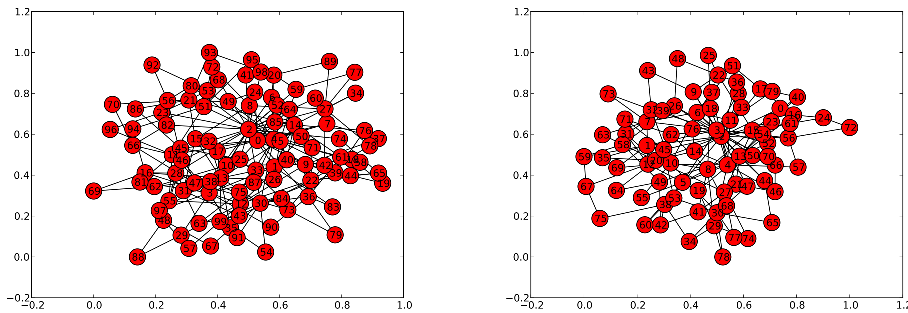
The enmity between the two groups can be modeled as the entry sum of the sub-matrix L^+_{12}. Since $l^+_{i,j+n_1} = \|\vec{x}_i\|\|\vec{y}_j\| \cos\beta_{ij}$, it sums over the *cosine distances* of all possible pairs of inter-group edges:

$$\text{Enm} = \sum L^+_{12} := \sum\nolimits_{i=1}^{n_1} \sum\nolimits_{j=1}^{n_2} l^+_{i,j+n_1} \tag{3}$$

3 Evaluation on Synthetic Networks

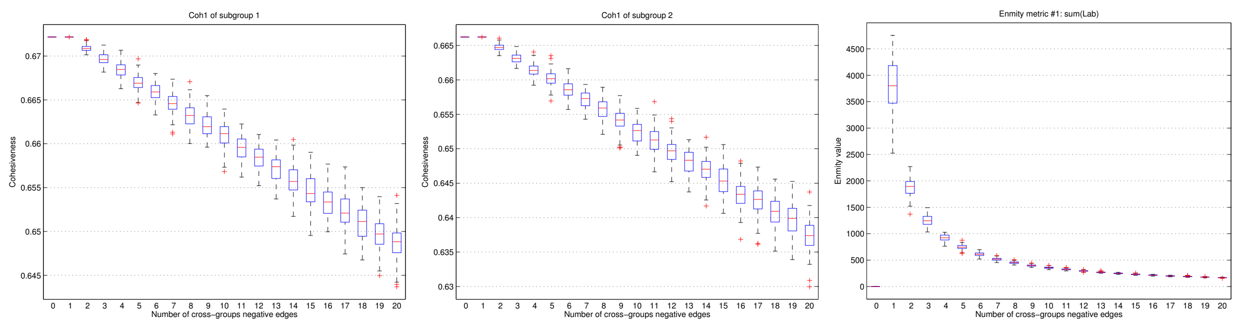
In this section, we apply our model to some synthetic networks, in order to verify if our proposed cohesiveness and enmity metrics indeed capture what real social communities should behave. As shown in Figure 1, we create two Barabási-Albert graphs [2] which better emulate the structure of our currently

[1] In the Appendix, we numerically visualize the distribution of λ_1 through λ_n, and show that q_1, the eigenvector corresponding to the leading eigenvalue λ_1, is a good membership indicator of the two subgroups.



(a) Community 1: 100-node BA graph (b) Community 2: 80-node BA graph

Figure 1: Two synthetic Barabási-Albert graphs with 2-edge preferential attaching scheme, both containing positive edges only. Extra positive/negative edges connect them into a signed graph.



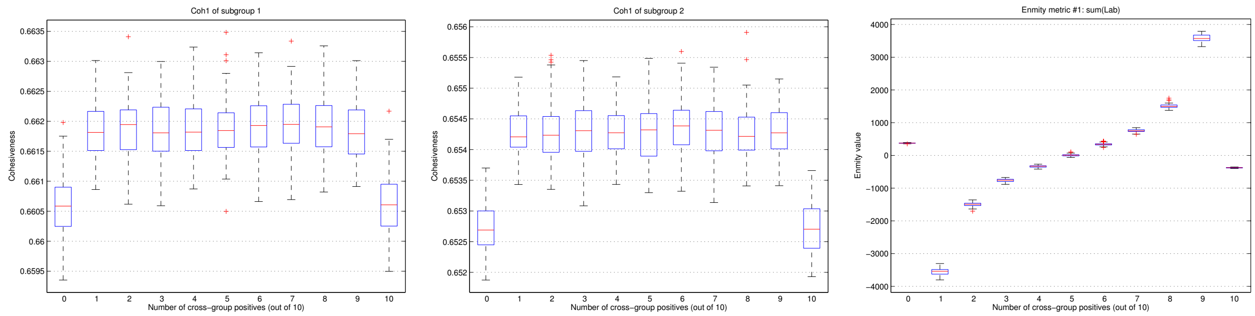
(a) Cohesiveness of Community 1 (b) Cohesiveness of Community 2 (c) Enmity between Two

Figure 2: Box-chart results of the 1st experiment: on balanced graphs, how the number of inter-community negatives affects our cohesiveness and enmity measures. Each box contains 100 random samples, the central mark being the median, and box edges being the 25th and 75th percentiles.

prevailing online social networks. They have all positive edges, and form the intended bipartition of our synthetic network. We then link the vertexes across the two subgroups in the following two different ways in order to form two different synthetic networks, and see how our model applies to each of them.

1. In the first experiment, we gradually introduce only negative edges across the two subgraphs in a random fashion, and observe how our proposed cohesiveness and enmity measures change as the number of negatives increases. Note that no matter how many negatives are added, the resulting signed graph is always balanced. The results in Figure 2 shows that both cohesiveness and enmity should be strengthened as the number of inter-community negatives increases (note: for enmity, smaller values indicate stronger effect). Therefore, our proposed measures indeed capture what cohesiveness and enmity should behave in balanced social communities.

2. In the second experiment, we wire exactly 10 random pairs of nodes across the two subgroups. Within these 10 edges, x ($0 \leq x \leq 10$) edges are positive, and $10-x$ are negative. Figures 3 shows how the cohesiveness and enmity measures change as x increases from 0 to 10. In particular, $x = 0$ means all cross-group edges are negative, and the resulting graph is 2-balanced; and $x = 10$ results in all-positive graph, which is trivially 2-balanced. The remaining cases (i.e., $1 \leq x \leq 9$ are unbalanced cases). When the signed graph becomes unbalanced, as shown in the figure, although the enmity measure still behave as expected, the cohesiveness measures have two undesired behaviors: i) in Figure 3 (a)(b), when $1 \leq x \leq 10$, the cohesiveness measures for both communities are almost not affected by the proportion of positives; ii) when the graph



(a) Cohesiveness of Community 1 (b) Cohesiveness of Community 2 (c) Enmity between Two

Figure 3: Box-chart results of the 2nd experiment: when two communities are wired with a mixture of positives and negatives, how the proportion of negatives affects our proposed cohesiveness and enmity measures. x-axis: the number of *positive* inter-community edges (out of a total of 10), e.g., x=3 means that there are 3 positives and 7 negatives across two communities.

Table 1: Statistics of real-world social websites (data source: [7, 8, 6]) and the results of our model.

Dataset	Nodes	Edges	BalanceIndex	Cohesiveness1	Cohesiveness2	Enmity
Epinions	119,070	701,569	0.0043	0.350	0.324	0.205
Slashdot1(081106)	77,258	466,661	0.0058	0.516	0.449	0.220
Slashdot2(090216)	81,776	495,661	0.0061	0.563	0.632	0.269
Slashdot3(090221)	82,052	498,527	0.0065	0.544	0.545	0.310
Slashdot4	79,031	465,975	0.0062	0.698	0.611	0.365

transitions from balanced to unbalanced via the sign flip of only one edge (e.g., x changes from 0 to 1, or from 10 to 9), the cohesiveness has an abrupt change, indicating that our cohesiveness measure may be sensitive to the change of graph balance.

In sum, our proposed model can handle balanced graphs well, but may not behave totally as expected on unbalanced graphs. This also implies that we need to obtained balanced graph by removing some "noisy" edges before applying our model.

4 Evaluation on Social Websites

In this section, we apply our model to five real-world signed online social web sites: the first four datasets (Epinions and Slashdot 1 3) come from Leskovec et al. [7, 8], and the Slashdot4 comes from Kunegis et al. [6]. The statistics of each dataset and the results of our model are listed in Table 1. Note that some necessary preprocessing is performed on the original datasets. [2]

As we can see from the table, all four Slashdot datasets have roughly the same balance indices (around 0.006), which are larger than that of the Epinions dataset (0.0043). Since a smaller balance index indicates a more balanced graph (i.e., stronger bipartition effect), we see that Epinions has stronger bipartition tendency than Slashdot. This interpretation indeed matches our expectation,

[2] First, we convert all directed graphs into undirected graphs by removing all bidirectional edges with opposite signs, and then extract the largest connected component of the graph, so the nodes count and the edge count are for this processed graph, instead of the original one. Second, because our model only applies to balanced graphs, as illustrated in the previous section, before computing the balanced index, cohesiveness and enmity, we need to convert the graph obtained in the first step to balanced graphs. To do this, we use the signs of $\vec{q}_1$ as the indicator to bipartition the network into two parts, and then remove "noisy" edges (i.e., inter-community positives and intra-community negatives).

considering the natures of these two social sites. We further postulate that a social network with richer user interactions has a stronger tendency towards community bipartition (i.e., more balanced).

In addition, the intra-community cohesiveness and inter-community enmity computed in the last three columns are consistent with the balance index of each dataset (note that smaller cohesiveness or enmity measures denotes stronger effects). We see that Epinions has stronger cohesiveness within each community and stronger enmity between the two community, compared with all Slashdot datasets. Again, this is as expected. Therefore, our cohesiveness and enmity measures indeed capture the community behavior on real-world social web sites.

5 Evaluation on the Senate Voting Network

In this section, we would like to test our model on an *evolving* network. Online social networks may be a good choice, but they have two major limitations: i) when crawling the social sites, we only get a sampled subgraph, and this sample is most likely bias; ii) it is challenging to study the evolution of social sites through samples, as tracking a fixed user base would be generally infeasible.

Instead, we use the U.S. Senate Roll Call Votes data (extracted from www.senate.gov) that spans from year 1989 through 2010. For each congress, a bipartite senator-bill graph can be extracted from the raw voting data, with positive edges being YEA votes and negative edges being NAYs. We then use some heuristics to convert this bipartite graph into a senator-senator weighted signed "social network", in which the signs indicate voting similarity or dissimilarity. We expect that such a senator-senator social network will exhibit a tendency of bipartition which more or less aligns with the actual party affiliation of senators. For example, Figure 4 shows the frequency distribution of the normalized co-voting similarity between three types of senator pairs (democrat vs democrat, republican vs republican, democrat vs republican) in congress year 2007-2008. We clearly see that statistically speaking, senators from the same party have high cohesiveness in terms of co-voting similarity, while senators from different parties tend to vote oppositely on most bills. This result is unsurprising, but it enables us to study the cohesiveness of both parties and the enmity between them through their voting behaviors.

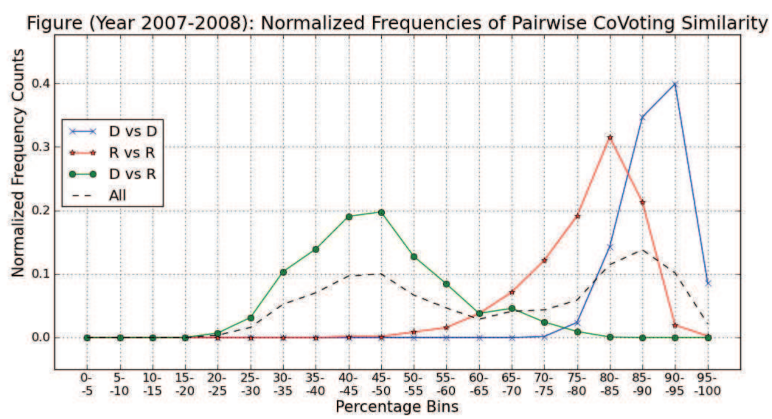


Figure 4: Frequencies count of the normalized co-voting similarity between three types of senator pairs (democrat vs democrat, republican vs republican, democrat vs republican) in congress year 2007-2008.

We apply our model to each of the 11 congresses in the data, and plot the cohesiveness and enmity measures as well as balance indices in Figure 5. There are two abrupt increase of the balance indices: congress 1993-1994, and congress 2009-2010. The former might reflect the economy crisis at the beginning of Clinton's adminstration; and the latter seems to align with the timing of the house crisis. We can also see from this figure that the Democrats have higher cohesiveness (smaller value) than Republicans, which is consistent with the similarity distribution in Figure 5.

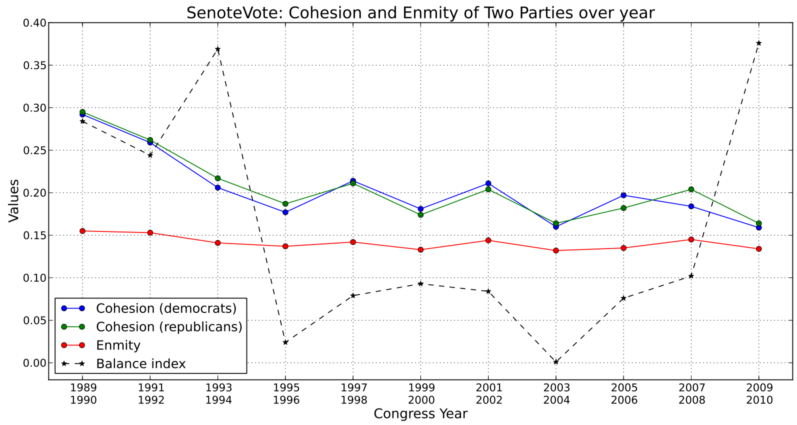


Figure 5: Evolution of the cohesiveness within each party, their enmity, and the balance index.

6 Conclusion and Future Work

In this paper, we take the implications of signed bipartition from social balance theory as premises, and propose a mathematical model to quantify the cohesiveness within each community and the enmity between two communities. To the best of our knowledge, this is the first work that tackles on this topic. We also validate the effectiveness of our model using synthetic signed networks, online signed social networks, and the U.S. Senate Voting network.

As acknowledged in section 3, our model works well on balanced networks, but may exhibit some undesired properties on unbalanced networks. Although we can remove noisy edges to obtain balanced graphs in the first place, as we did in section 4 and 5, we are not sure about the side effects of such removals. In our future work, we plan to revise our mathematical model such that it can naturally handle all types of signed graphs.

References

[1] N. Bansal, A. Blum, and S. Chawla. Correlation clustering. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, volume 25, pages 238–250, Vancouver, Canada, Nov. 2002.

[2] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

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

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

[5] Y. Hou. Bounds for the least Laplacian eigenvalue of a signed graph. *Acta Mathematica Sinica, English Series*, 21(4):955–960, Dec. 2004.

[6] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De, and L. S. Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proceedings of SIAM International Conference on Data Mining*, 2010.

[7] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. In *Proceedings of the 19th international conference on World Wide Web*, WWW '10, pages 641–650, New York, USA, 2010. ACM Press.

[8] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, New York, USA, 2010. ACM Press.

[9] S. Wasserman and K. Faust. *Social network analysis: methods and applications (Structural analysis in the social sciences)*. Cambridge University Press, 1994.

Appendix

To study the Laplacian eigen spectrum of 2-balanced signed graphs, and to seek for a means to separate the two subgroups, we conduct two sets of numerical experiments. **In the first experiment set**, we create two *complete* (sub-)graphs each containing 25 nodes and only positive edges, and connect the two subgraphs using negative edges into a 2-balanced signed graph. We vary the number of these inter-group negative edges, denoted as W, and investigate how W impacts the Laplacian spectrum. The results are shown in Figure 6. Each row in the figure represents the results for a specific W value. The 1st column is the Laplacian spectrum of the signed graph, namely, the eigenvalues of L^+ for this signed graph, sorted in an increasing order. The 2nd column is the zoom-in view of the leading eigenvalues from the 1st column. The 3rd column "Signed: (v1,v2)" is the 2-dimensional embedding of v_1 and v_2 for the signed graph, where v_1 is the eigenvector corresponding to the leading eigenvalue λ_1, and v_2 is for λ_2. The 4th column "Unsigned: (v1,v2)" is the same v_1-v_2 embedding, but for the corresponding unsigned graph (i.e., by converting all negative edges into positive). The 5th the 6th columns are the embedding results on v_3 and v_4.

In the second experiment set, as shown in Figure 7, we study the impact of the edge-completeness (in percentage) of each positive and *incomplete* (sub-)graph, while fixing the number of inter-group negative edges to be 50. We still use 25 nodes for each positive sub-graph, but vary the percentage of the edge-completeness for both sub-graphs – note that on each level of percentage, both sub-graphs have the same percentage, and the edge distribution is uniformed while keeping each sub-graph connected. The meanings of the columns in Figure 7 are the same as those in Figure 6.

The most important insight from these plots is that v_1**, the eigenvector corresponding to the largest Laplacian eigenvalue, always gives the membership indicator for signed 2-balanced graphs, no matter each of the two subgraphs is complete or not.** Apart from this, we also have the following general observations:

1. As the number of inter-group links increase, eigenvalue λ_2 of the signed Laplacian will increase. Actually all eigenvalues except for the first one will be boosted up, until reaching the maximum $n_1 + n_2$ where n_1 and n_2 are the number of nodes of the two complete graphs.

2. For two groups each with fixed number of nodes (n_1 for Group1, n_2 for Group2) and a fixed number of inter-group links, as the edge completeness (e.g., $m_1/(n_1(n_1 - 1)/2)$ where m_1 is the actual number of edges in Group1) of both groups increases, all Laplacian eigenvalues except for the first one will get boosted up.

3. A 2-balanced signed graph and its corresponding unsigned graph always have the same eigenvalues, and the eigenvector matrix of the unsigned graph is given by flipping the signs of the lower rows (i.e., those for Group2) of the eigenvector matrix of the signed graph.

Knowing that v_1 is a good membership indicator of a 2-balanced signed graph, we are interested in investigating whether there exists a good membership indicator for *the corresponding unsigned graph*. When we look into another vector v_2, however, the answer is negative. We elaborate our observations on v_2 as follows.

1. A *sufficient* but *not necessary* condition for v_2 to contain *no opposite signs* for a 2-balanced signed graph is that both subgraphs are complete. The "sufficient" aspect is evident by the "Signed:(v1,v2)" plots in Figure 6; whereas the "not necessary" aspect is evident by the "Signed:(v1,v2)" plots in Figure 7. In addition, we have to following three remarks. **Remark A:** "Laplacian eigenvector v_2 for the signed graph containing no opposite signs" is equivalent to "v_2 for the corresponding unsigned graph being the membership indicator." **Remark B:** "No opposite signs" is more accurate than "all positive" because some entries may be strictly zero. Those nodes whose v_2 values are zero can no longer be separated on the unsigned graph using the v_2 signs alone (although v_1 can still do the job on the original signed graph). So strictly speaking, v_2 is not a "good" membership indicator on the corresponding unsigned graph because it may give no membership indications for some nodes, when the number of inter-group links increases to some extent. **Remark C:** When two complete subgraphs (with n_1 and n_2 nodes respectively) are fully connected by $n_1 * n_2$ inter-group negative links, the signed Laplacian eigen spectrum contains eigenvalue 0 with multiplicity 1, and eigenvalue $n_1 + n_2$ with multiplicity $n_1 + n_2 - 1$. If one of these $n_1 * n_2$ inter-group links is deleted (i.e., $W = n_1 * n_2 - 1$), then except for the two nodes that would otherwise form this deleted link, all other $n_1 + n_2 - 2$ nodes all have zero v_2 and are therefore indistinguishable on the corresponding unsigned graph. If we continue to delete these inter-group links, nodes seem to depart from zero v_2 one by one, and eigenvalues seem to get "dragged" down from the value $n_1 + n_2$ one by one.

2. When at least one of the subgraphs in a 2-balanced signed graph is incomplete, e.g., the "Signed:(v1,v2), C=20%" plot in Figure 7, v_2 may contain opposite signs. We remark that when the number of inter-group links is *comparable* to the number of intra-group links, these inter-group links tend to "pull" those nodes with relatively weaker intra-group connectivity or stronger inter-group connectivity to the other group, and melt the two groups together such that they can no longer be separated on the unsigned graph using v_2. Combined with the remarks in the previous point, we conclude that being a complete graph is the strongest possible connection for a group of nodes, and it guarantees to prevent their members from being pulled away to the other group. We also conjecture that when at least one of the two subgraphs is incomplete, there always exists a set of inter-group links that makes the Laplacian eigenvector v_2 contain opposite signs. We leave the proof or disproof of this conjecture to our future work.

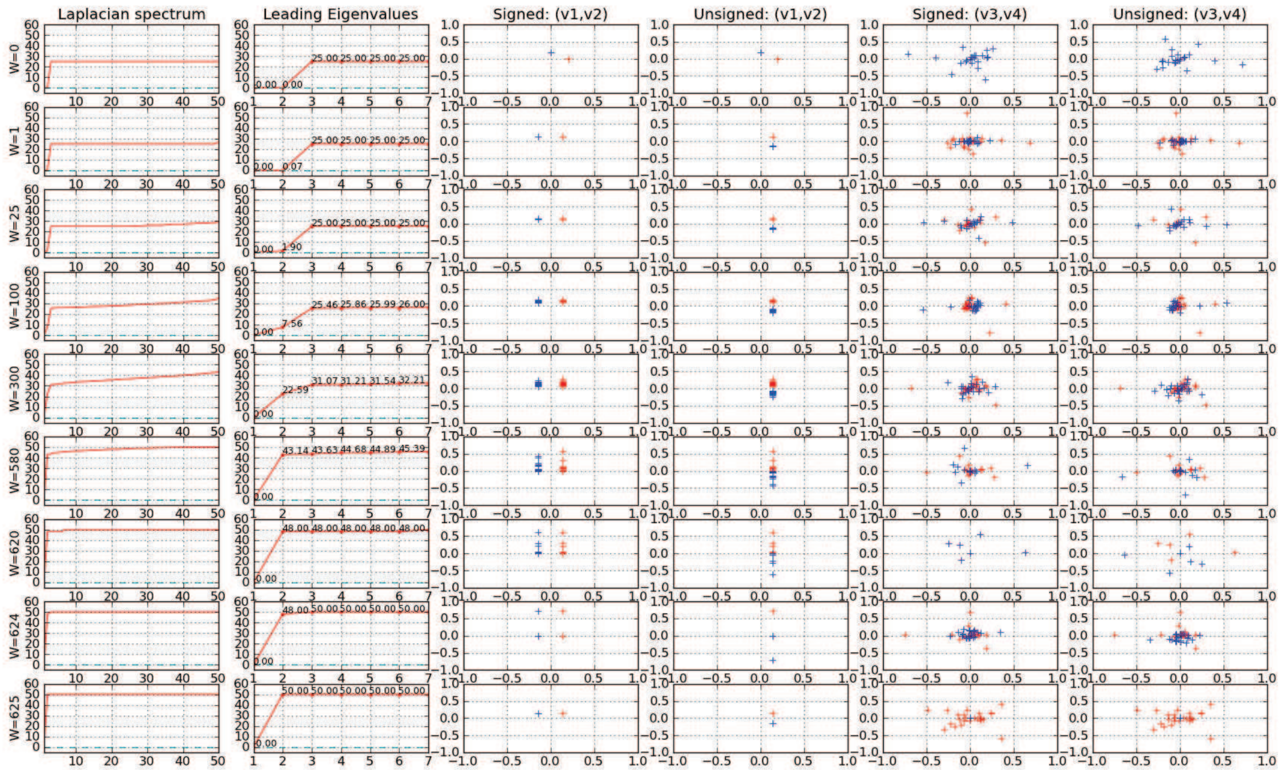


Figure 6: Laplacian spectrum and eigenvector embeddings for a 2-balanced graph that consists of two 25-node positive *complete* (sub-)graph connected by *a varying number of inter-group negative edges*.

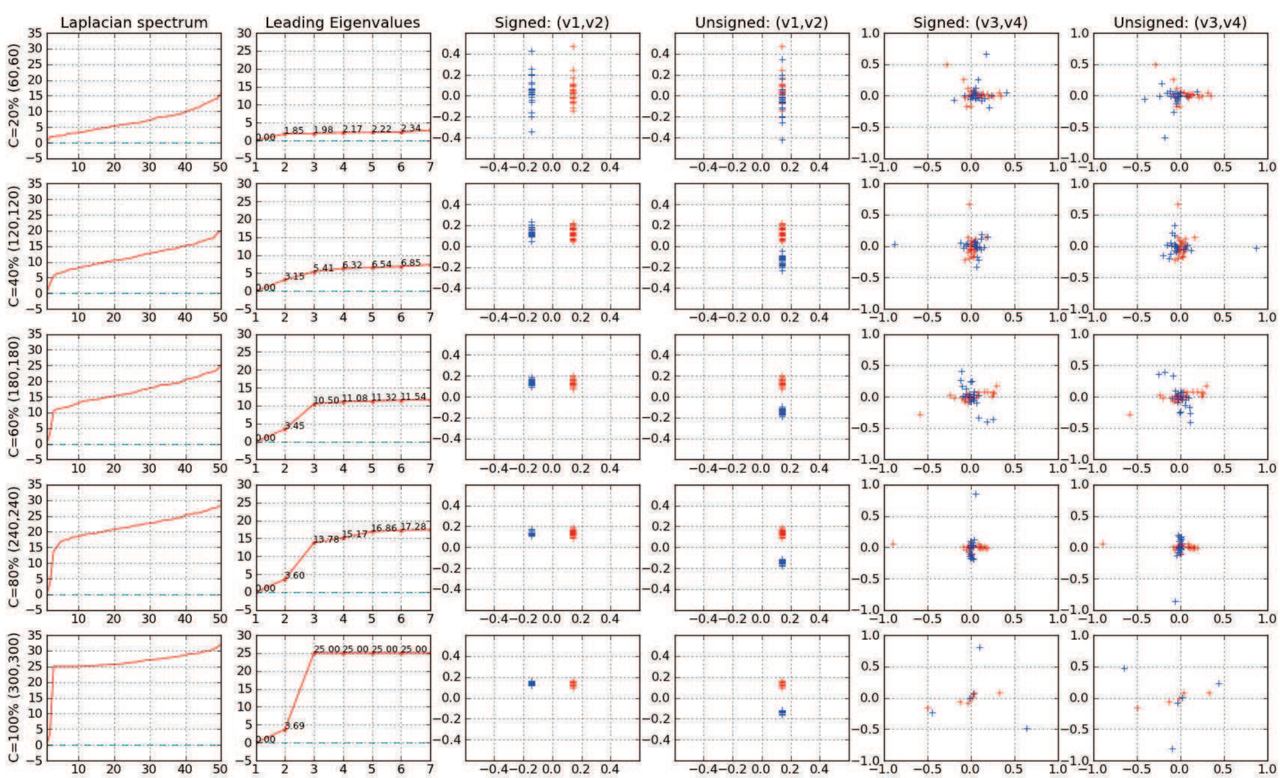


Figure 7: Laplacian spectrum and eigenvector embeddings for a 2-balanced graph that consists of two 25-node positive (sub-)graph of *varying edge-completeness* connected by 50 negative inter-group edges.