

---

# Clustering Signed Networks with the Geometric Mean of Laplacians

---

Pedro Mercado<sup>1</sup>, Francesco Tudisco<sup>2</sup> and Matthias Hein<sup>1</sup>

<sup>1</sup>Saarland University, Saarbrücken, Germany

<sup>2</sup>University of Padua, Padua, Italy

## Abstract

Signed networks allow to model positive and negative relationships. We analyze existing extensions of spectral clustering to signed networks. It turns out that existing approaches do not recover the ground truth clustering in several situations where either the positive or the negative network structures contain no noise. Our analysis shows that these problems arise as existing approaches take some form of arithmetic mean of the Laplacians of the positive and negative part. As a solution we propose to use the geometric mean of the Laplacians of positive and negative part and show that it outperforms the existing approaches. While the geometric mean of matrices is computationally expensive, we show that eigenvectors of the geometric mean can be computed efficiently, leading to a numerical scheme for sparse matrices which is of independent interest.

## 1 Introduction

A signed graph is a graph with positive and negative edge weights. Typically positive edges model attractive relationships between objects such as similarity or friendship and negative edges model repelling relationships such as dissimilarity or enmity. The concept of balanced signed networks can be traced back to [10, 3]. Later, in [5], a signed graph is defined as  $k$ -balanced if there exists a partition into  $k$  groups where only positive edges are within the groups and negative edges are between the groups. Several approaches to find communities in signed graphs have been proposed (see [23] for an overview). In this paper we focus on extensions of spectral clustering to signed graphs. Spectral clustering is a well established method for unsigned graphs which, based on the first eigenvectors of the graph Laplacian, embeds nodes of the graphs in  $\mathbb{R}^k$  and then uses  $k$ -means to find the partition. In [16] the idea is transferred to signed graphs. They define the signed ratio and normalized cut functions and show that the spectrum of suitable signed graph Laplacians yield a relaxation of those objectives. In [4] other objective functions for signed graphs are introduced. They show that a relaxation of their objectives is equivalent to weighted kernel  $k$ -means by choosing an appropriate kernel. While they have a scalable method for clustering, they report that they can not find any cluster structure in real world signed networks.

We show that the existing extensions of the graph Laplacian to signed graphs used for spectral clustering have severe deficiencies. Our analysis of the stochastic block model for signed graphs shows that, even for the perfectly balanced case, recovery of the ground-truth clusters is not guaranteed. The reason is that the eigenvectors encoding the cluster structure do not necessarily correspond to the smallest eigenvalues, thus leading to a noisy embedding of the data points and in turn failure of  $k$ -means to recover the cluster structure. The implicit mathematical reason is that all existing extensions of the graph Laplacian are based on some form of arithmetic mean of operators of the positive and negative graphs. In this paper we suggest as a solution to use the geometric mean of the Laplacians of positive and negative part. In particular, we show that in the stochastic block model the geometric mean Laplacian allows in expectation to recover the ground-truth clusters in

any reasonable clustering setting. A main challenge for our approach is that the geometric mean Laplacian is computationally expensive and does not scale to large sparse networks. Thus a main contribution of this paper is showing that the first few eigenvectors of the geometric mean can still be computed efficiently. Our algorithm is based on the inverse power method and the extended Krylov subspace technique introduced by [8] and allows to compute eigenvectors of the geometric mean  $A\#B$  of two matrices  $A, B$  without ever computing  $A\#B$  itself.

In Section 2 we discuss existing work on Laplacians on signed graphs. In Section 3 we discuss the geometric mean of two matrices and introduce the geometric mean Laplacian which is the basis of our spectral clustering method for signed graphs. In Section 4 we analyze our and existing approaches for the stochastic block model. In Section 5 we introduce our efficient algorithm to compute eigenvectors of the geometric mean of two matrices, and finally in Section 6 we discuss performance of our approach on real world graphs. .

## 2 Signed graph clustering

Networks encoding positive and negative relations among the nodes can be represented by weighted signed graphs. Consider two symmetric non-negative weight matrices  $W^+$  and  $W^-$ , a vertex set  $V = \{v_1, \dots, v_n\}$ , and let  $G^+ = (V, W^+)$  and  $G^- = (V, W^-)$  be the induced graphs. A signed graph is the pair  $G^\pm = (G^+, G^-)$  where  $G^+$  and  $G^-$  encode positive and the negative relations, respectively.

The concept of community in signed networks is typically related to the theory of social balance. This theory, as presented in [10, 3], is based on the analysis of affective ties, where positive ties are a source of balance whereas negative ties are considered as a source of imbalance in social groups.

**Definition 1** ([5],  $k$ -balance). *A signed graph is  $k$ -balanced if the set of vertices can be partitioned into  $k$  sets such that within the subsets there are only positive edges, and between them only negative.*

The presence of  $k$ -balance in  $G^\pm$  implies the presence of  $k$  groups of nodes being both assortative in  $G^+$  and disassortative in  $G^-$ . However this situation is fairly rare in real world networks and expecting communities in signed networks to be a perfectly balanced set of nodes is unrealistic.

In the next section we will show that Laplacians inspired by Definition 1 are based on some form of arithmetic mean of Laplacians. As an alternative we propose the geometric mean of Laplacians and show that it is able to recover communities when either  $G^+$  is assortative, or  $G^-$  is disassortative, or both. Results of this paper will make clear that the use of the geometric mean of Laplacians allows to recognize communities where previous approaches fail.

### 2.1 Laplacians on Unsigned Graphs

Spectral clustering of undirected, unsigned graphs using the Laplacian matrix is a well established technique (see [19] for an overview). Given an unsigned graph  $G = (V, W)$ , the Laplacian and its normalized version are defined as

$$L = D - W \quad L_{\text{sym}} = D^{-1/2} L D^{-1/2} \quad (1)$$

where  $D_{ii} = \sum_{j=1}^n w_{ij}$  is the diagonal matrix of the degrees of  $G$ . Both Laplacians are positive semidefinite, and the multiplicity  $k$  of the eigenvalue 0 is equal to the number of connected components in the graph. Further, the Laplacian is suitable in assortative cases [19], *i.e.* for the identification of clusters under the assumption that the amount of edges inside clusters has to be larger than the amount of edges between them.

For disassortative cases, *i.e.* for the identification of clusters where the amount of edges has to be larger between clusters than inside clusters, the signless Laplacian is a better choice [18]. Given the unsigned graph  $G = (V, W)$ , the signless Laplacian and its normalized version are defined as

$$Q = D + W, \quad Q_{\text{sym}} = D^{-1/2} Q D^{-1/2} \quad (2)$$

Both Laplacians are positive semi-definite, and the smallest eigenvalue is zero if and only if the graph has a bipartite component [6].

## 2.2 Laplacians on Signed Graphs

Recently a number of Laplacian operators for signed networks have been introduced. Consider the signed graph  $G^\pm = (G^+, G^-)$ . Let  $D_{ii}^+ = \sum_{j=1}^n w_{ij}^+$  be the diagonal matrix of the degrees of  $G^+$  and  $\bar{D}_{ii} = \sum_{j=1}^n w_{ij}^+ + w_{ij}^-$  the one of the overall degrees in  $G^\pm$ .

The following Laplacians for signed networks have been considered so far

$$\begin{aligned} L_{BR} &= D^+ - W^+ + W^-, \quad L_{BN} = \bar{D}^{-1} L_{BR}, & (\text{balance ratio/normalized Laplacian}) \\ L_{SR} &= \bar{D} - W^+ + W^-, \quad L_{SN} = \bar{D}^{-1/2} L_{SR} \bar{D}^{-1/2}, & (\text{signed ratio/normalized Laplacian}) \end{aligned} \quad (3)$$

and spectral clustering algorithms have been proposed for  $G^\pm$ , based on these Laplacians [16, 4]. Let  $L^+$  and  $Q^-$  be the Laplacian and the signless Laplacian matrices of the graphs  $G^+$  and  $G^-$ , respectively. We note that the matrix  $L_{SR}$  blends the informations from  $G^+$  and  $G^-$  into (twice) the arithmetic mean of  $L^+$  and  $Q^-$ , namely the following identity holds

$$L_{SR} = L^+ + Q^-. \quad (4)$$

Thus, as an alternative to the normalization defining  $L_{SN}$  from  $L_{SR}$ , it is natural to consider the arithmetic mean of the normalized Laplacians  $L_{AM} = L_{\text{sym}}^+ + Q_{\text{sym}}^-$ . In the next section we introduce the geometric mean of  $L_{\text{sym}}^+$  and  $Q_{\text{sym}}^-$  and propose a new clustering algorithm for signed graphs based on that matrix. The analysis and experiments of next sections will show that blending the information from the positive and negative graphs through the geometric mean overcomes the deficiencies showed by the arithmetic mean based operators.

## 3 Geometric mean of Laplacians

We define here the geometric mean of matrices and introduce the geometric mean of normalized Laplacians for clustering signed networks. Let  $A^{1/2}$  be the unique positive definite solution of the matrix equation  $X^2 = A$ , where  $A$  is positive definite.

**Definition 2.** Let  $A, B$  be positive definite matrices. The geometric mean of  $A$  and  $B$  is the positive definite matrix  $A \# B$  defined by  $A \# B = A^{1/2} (A^{-1/2} B A^{-1/2})^{1/2} A^{1/2}$ .

One can prove that  $A \# B = B \# A$  (see [1] for details). Further, there are several useful ways to represent the geometric mean of positive definite matrices (see f.i. [1, 12])

$$A \# B = A(A^{-1}B)^{1/2} = (BA^{-1})^{1/2}A = B(B^{-1}A)^{1/2} = (AB^{-1})^{1/2}B \quad (5)$$

The next result reveals further consistency with the scalar case, in fact we observe that if  $A$  and  $B$  have some eigenvectors in common, then  $A + B$  and  $A \# B$  have those eigenvectors, with eigenvalues given by the arithmetic and geometric mean of the corresponding eigenvalues of  $A$  and  $B$ , respectively.

**Theorem 1.** Let  $\mathbf{u}$  be an eigenvector of  $A$  and  $B$  with eigenvalues  $\lambda$  and  $\mu$ , respectively. Then,  $\mathbf{u}$  is an eigenvector of  $A + B$  and  $A \# B$  with eigenvalue  $\lambda + \mu$  and  $\sqrt{\lambda\mu}$ , respectively.

*Proof.* Using the identities  $A\mathbf{u} = \lambda\mathbf{u}$  and  $B\mathbf{u} = \mu\mathbf{u}$  we have  $(A + B)\mathbf{u} = (\lambda + \mu)\mathbf{u}$ . For the geometric mean, observe that for any positive definite matrix  $M$ , if  $M\mathbf{x} = \lambda(M)\mathbf{x}$ , then  $M^{1/2}\mathbf{x} = \lambda(M)^{1/2}\mathbf{x}$ . In particular we have

$$A^{-1/2} B A^{-1/2} \mathbf{u} = \lambda^{-1/2} A^{-1/2} B \mathbf{u} = \lambda^{-1/2} \mu A^{-1/2} \mathbf{u} = (\mu/\lambda) \mathbf{u}$$

thus  $(A^{-1/2} B A^{-1/2})^{1/2} \mathbf{u} = \sqrt{\mu/\lambda} \mathbf{u}$ . As a consequence

$$(A \# B) \mathbf{u} = A^{1/2} (A^{-1/2} B A^{-1/2})^{1/2} A^{1/2} \mathbf{u} = \lambda^{1/2} A^{1/2} (A^{-1/2} B A^{-1/2})^{1/2} \mathbf{u} = (\sqrt{\lambda\mu}) \mathbf{u}$$

which concludes the proof.  $\square$

### 3.1 Geometric mean for signed networks clustering

Consider the signed network  $G^\pm = (G^+, G^-)$ . We define the normalized geometric mean Laplacian of  $G^\pm$  as

$$L_{GM} = L_{\text{sym}}^+ \# Q_{\text{sym}}^- \quad (6)$$

We propose Algorithm 1 for clustering signed networks, based on the spectrum of  $L_{GM}$ . By definition 2, the matrix geometric mean  $A\#B$  requires  $A$  and  $B$  to be positive definite. As both the Laplacian and the signless Laplacian are positive semi-definite, in what follows we shall assume that the matrices  $L_{\text{sym}}^+$  and  $Q_{\text{sym}}^-$  in (6) are modified by a small diagonal shift, ensuring positive definiteness. That is, in practice, we consider  $L_{\text{sym}}^+ + \varepsilon_1 I$  and  $Q_{\text{sym}}^- + \varepsilon_2 I$  being  $\varepsilon_1$  and  $\varepsilon_2$  small positive numbers. For the sake of brevity, we do not explicitly write the shifting matrices.

---

**Input:** Symmetric weight matrices  $W^+, W^- \in \mathbb{R}^{n \times n}$ , number  $k$  of clusters to construct.

**Output:** Clusters  $C_1, \dots, C_k$ .

- 1 Compute the  $k$  eigenvectors  $\mathbf{u}_1, \dots, \mathbf{u}_k$  corresponding to the  $k$  smallest eigenvalues of  $L_{GM}$ .
- 2 Let  $U = (\mathbf{u}_1, \dots, \mathbf{u}_k)$ .
- 3 Cluster the rows of  $U$  with  $k$ -means into clusters  $C_1, \dots, C_k$ .

**Algorithm 1:** Spectral clustering with  $L_{GM}$  on signed networks

---

The main bottleneck of Algorithm 1 is the computation of the eigenvectors in step 1. In Section 5 we propose a scalable Krylov-based method to handle this problem.

Let us briefly discuss the motivating intuition behind the proposed clustering strategy. Algorithm 1, as well as state-of-the-art clustering algorithms based on the matrices in (3), rely on the  $k$  smallest eigenvalues of the considered operator and their corresponding eigenvectors. Thus the relative ordering of the eigenvalues plays a crucial role. Assume the eigenvalues to be enumerated in ascending order. Theorem 1 states that the functions  $(A, B) \mapsto A + B$  and  $(A, B) \mapsto A\#B$  map eigenvalues of  $A$  and  $B$  having the same corresponding eigenvectors, into the arithmetic mean  $\lambda_i(A) + \lambda_j(B)$  and geometric mean  $\sqrt{\lambda_i(A)\lambda_j(B)}$ , respectively, where  $\lambda_i(\cdot)$  is the  $i^{\text{th}}$  smallest eigenvalue of the corresponding matrix. Note that the indices  $i$  and  $j$  are not the same in general, as the eigenvectors shared by  $A$  and  $B$  may be associated to eigenvalues having different positions in the relative ordering of  $A$  and  $B$ . This intuitively suggests that small eigenvalues of  $A + B$  are related to small eigenvalues of both  $A$  and  $B$ , whereas those of  $A\#B$  are associated with small eigenvalues of either  $A$  or  $B$ , or both. Therefore the relative ordering of the small eigenvalues of  $L_{GM}$  is influenced by the presence of assortative clusters in  $G^+$  (related to small eigenvalues of  $L_{\text{sym}}^+$ ) or by disassortative clusters in  $G^-$  (related to small eigenvalues in  $Q_{\text{sym}}^-$ ), whereas the ordering of the small eigenvalues of the arithmetic mean takes into account only the presence of both those situations.

In the next section, for networks following the stochastic block model, we analyze in expectation the spectrum of the normalized geometric mean Laplacian as well as the one of the normalized Laplacians previously introduced. In this case the expected spectrum can be computed explicitly and we observe that in expectation the ordering induced by blending the informations of  $G^+$  and  $G^-$  through the geometric mean allows to recover the ground truth clusters perfectly, whereas the use of the arithmetic mean introduces a bias which reverberates into a significantly higher clustering error.

## 4 Stochastic block model on signed graphs

In this section we present an analysis of different signed graph Laplacians based on the Stochastic Block Model (SBM). The SBM is a widespread benchmark generative model for networks showing a clustering, community, or group behaviour [22]. Given a prescribed set of groups of nodes, the SBM defines the presence of an edge as a random variable with probability being dependent on which groups it joins. To our knowledge this is the first analysis of spectral clustering on signed graphs with the stochastic block model. Let  $C_1, \dots, C_k$  be ground truth clusters, all having the same size  $|C|$ . We let  $p_{\text{in}}^+$  ( $p_{\text{in}}^-$ ) be the probability that there exists a positive (negative) edge between nodes in the same cluster, and let  $p_{\text{out}}^+$  ( $p_{\text{out}}^-$ ) denote the probability of a positive (negative) edge between nodes in different clusters.

Calligraphic letters denote matrices in expectation. In particular  $\mathcal{W}^+$  and  $\mathcal{W}^-$  denote the weight matrices in expectation. We have  $\mathcal{W}_{i,j}^+ = p_{\text{in}}^+$  and  $\mathcal{W}_{i,j}^- = p_{\text{in}}^-$  if  $v_i, v_j$  belong to the same cluster, whereas  $\mathcal{W}_{i,j}^+ = p_{\text{out}}^+$  and  $\mathcal{W}_{i,j}^- = p_{\text{out}}^-$  if  $v_i, v_j$  belong to different clusters. Sorting nodes according to the ground truth clustering shows that  $\mathcal{W}^+$  and  $\mathcal{W}^-$  have rank  $k$ .

$(E_+)$	$p_{\text{out}}^+ < p_{\text{in}}^+$	$(E_{\text{vol}})$	$p_{\text{in}}^- + (k-1)p_{\text{out}}^- < p_{\text{in}}^+ + (k-1)p_{\text{out}}^+$
$(E_-)$	$p_{\text{in}}^- < p_{\text{out}}^-$	$(E_{\text{conf}})$	$\left(\frac{kp_{\text{out}}^+}{p_{\text{in}}^+ + (k-1)p_{\text{out}}^+}\right) \left(\frac{kp_{\text{in}}^-}{p_{\text{in}}^- + (k-1)p_{\text{out}}^-}\right) < 1$
$(E_{\text{bal}})$	$p_{\text{in}}^- + p_{\text{out}}^+ < p_{\text{in}}^+ + p_{\text{out}}^-$	$(E_G)$	$\left(\frac{kp_{\text{out}}^+}{p_{\text{in}}^+ + (k-1)p_{\text{out}}^+}\right) \left(1 + \frac{p_{\text{in}}^- - p_{\text{out}}^-}{p_{\text{in}}^- + (k-1)p_{\text{out}}^-}\right) < 1$

Table 1: Conditions for the Stochastic Block Model analysis of Section 4

Consider the relations in Table 1. Conditions  $E_+$  and  $E_-$  describe the presence of assortative or disassortative clusters in expectation. Note that, by Definition 1, a graph is balanced if and only if  $p_{\text{out}}^+ = p_{\text{in}}^- = 0$ . We can see that if  $E_+ \cap E_-$  then  $G^-$  and  $G^+$  give information about the cluster structure. Further, if  $E_+ \cap E_-$  holds then  $E_{\text{bal}}$  holds. Similarly  $E_{\text{conf}}$  characterizes a graph where the relative amount of conflicts - *i.e.* positive edges between the clusters and negative edges inside the clusters - is small. Condition  $E_G$  is strictly related to such setting. In fact when  $E_- \cap E_G$  holds then  $E_{\text{conf}}$  holds. Finally condition  $E_{\text{vol}}$  implies that the expected volume in the negative graph is smaller than the expected volume in the positive one. This condition is therefore not related to any signed clustering structure.

Let

$$\chi_1 = \mathbf{1}, \quad \chi_i = (k-1)\mathbf{1}_{\mathcal{C}_i} - \mathbf{1}_{\overline{\mathcal{C}_i}}.$$

The use of  $k$ -means on  $\chi_i, i = 1, \dots, k$  identifies the ground truth communities  $\mathcal{C}_i$ . As spectral clustering relies on the eigenvectors corresponding to the  $k$  smallest eigenvalues (see Algorithm 1) we derive here necessary and sufficient conditions such that in expectation the eigenvectors  $\chi_i, i = 1, \dots, k$  correspond to the  $k$  smallest eigenvalues of the normalized Laplacians introduced so far. In particular, we observe that condition  $E_G$  affects the ordering of the eigenvalues of the normalized geometric mean Laplacian. Instead, the ordering of the eigenvalues of the operators based on the arithmetic mean is related to  $E_{\text{bal}}$  and  $E_{\text{vol}}$ . The latter is not related to any clustering, thus introduces a bias in the eigenvalues ordering which reverberates into a noisy embedding of the data points and in turn into a significantly higher clustering error.

**Theorem 2.** *Let  $\mathcal{L}_{BN}$  and  $\mathcal{L}_{SN}$  be the normalized Laplacians defined in (3) of the expected graphs. The following statements are equivalent:*

1.  $\chi_1, \dots, \chi_k$  are the eigenvectors corresponding to the  $k$  smallest eigenvalues of  $\mathcal{L}_{BN}$ .
2.  $\chi_1, \dots, \chi_k$  are the eigenvectors corresponding to the  $k$  smallest eigenvalues of  $\mathcal{L}_{SN}$ .
3. The two conditions  $E_{\text{bal}}$  and  $E_{\text{vol}}$  hold simultaneously.

*Proof.* We first prove that  $\chi_1, \dots, \chi_k$  are the eigenvectors corresponding to the  $k$  smallest eigenvalues of  $\mathcal{L}_{BN}$  if and only if the two conditions  $E_{\text{bal}}$  and  $E_{\text{vol}}$  hold simultaneously. It is simple to verify that  $\chi_i$  are eigenvectors of  $\mathcal{W}^+$  and  $\mathcal{W}^-$ , with eigenvalues denoted by  $\lambda_i^+$  and  $\lambda_i^-$ , respectively. Thus  $\mathcal{W}^+$  and  $\mathcal{W}^-$  are simultaneously diagonalizable, that is there exists a non-singular matrix  $\Sigma$  such that  $\Sigma^{-1}\mathcal{W}^\pm\Sigma = \Lambda^\pm$ , where  $\Lambda^+$  and  $\Lambda^-$  are diagonal matrices  $\Lambda^\pm = \text{diag}(\lambda_1^\pm, \dots, \lambda_k^\pm, 0, \dots, 0)$ . Observe that the eigenvalues  $\lambda_i^+$  and  $\lambda_i^-$  admits the following explicit representations

$$\begin{aligned} \lambda_1^+ &= |\mathcal{C}|(p_{\text{in}}^+ + (k-1)p_{\text{out}}^+), & \lambda_1^- &= |\mathcal{C}|(p_{\text{in}}^- + (k-1)p_{\text{out}}^-) \\ \lambda_i^+ &= |\mathcal{C}|(p_{\text{in}}^+ - p_{\text{out}}^+), & \lambda_i^- &= |\mathcal{C}|(p_{\text{in}}^- - p_{\text{out}}^-), \end{aligned} \tag{7}$$

for  $i = 2, \dots, k$ . As we assume clusters of the same size, the nodes have the same degree in expectation, inducing a regular graph. Hence the expected degrees of the graph are  $d^+ = \mathcal{W}^+\mathbf{1} = |\mathcal{C}|(p_{\text{in}}^+ + (k-1)p_{\text{out}}^+)\mathbf{1}$ ,  $d^- = \mathcal{W}^-\mathbf{1} = |\mathcal{C}|(p_{\text{in}}^- + (k-1)p_{\text{out}}^-)\mathbf{1}$  and  $\bar{d} = d^+ + d^-$ . With corresponding degree matrices  $\mathcal{D}^+ = d^+I$  and  $\bar{\mathcal{D}} = \bar{d}I$ . The expected balanced-ratio cut Laplacian operator is thus given by  $\mathcal{L}_{BR} = \Sigma(d^+I - \Lambda^+ + \Lambda^-)\Sigma^{-1}$ . It follows that the eigenvalues of  $\mathcal{L}_{BR}$  correspond to eigenvectors in the following way

$$\begin{cases} d^+ - \lambda_i^+ + \lambda_i^- & \text{with eigenvector } \chi_i, i = 1, \dots, k \\ d^+ & \text{corresponding to the remaining eigenvectors} \end{cases}$$

Thus, eigenvectors  $\chi_i, i = 1, \dots, k$  correspond to the smallest eigenvalues if and only if

$$d^+ - \lambda_i^+ + \lambda_i^- < d^+ \iff \lambda_i^- < \lambda_i^+$$

By Eqs. (7) we see that for the constant eigenvector we have

$$\lambda_1^- < \lambda_1^+ \iff d^- < d^+ \iff p_{\text{in}}^- + (k-1)p_{\text{out}}^- < p_{\text{in}}^+ + (k-1)p_{\text{out}}^+,$$

whereas for the eigenvectors  $\chi_i, i = 2, \dots, k$  the corresponding condition is

$$\lambda_i^- < \lambda_i^+ \iff p_{\text{in}}^- + p_{\text{out}}^+ < p_{\text{in}}^+ + p_{\text{out}}^-.$$

We deduce that the eigenvectors  $\chi_i, i = 1, \dots, k$  correspond to the smallest eigenvalues of  $\mathcal{L}_{BR}$  if and only if  $p_{\text{in}}^- + (k-1)p_{\text{out}}^- < p_{\text{in}}^+ + (k-1)p_{\text{out}}^+$  and  $p_{\text{in}}^- + p_{\text{out}}^+ < p_{\text{in}}^+ + p_{\text{out}}^-$ .

As  $\mathcal{L}_{BN}$  differs from  $\mathcal{L}_{BR}$  by a constant factor, the conditions hold for  $\mathcal{L}_{BN}$ . Conditions for  $\mathcal{L}_{SN}$  can be proved in the same way, as the only difference in the eigenvalues is a shift given by the degree vector  $\vec{d}$ .  $\square$

**Theorem 3.** Let  $\mathcal{L}_{GM} = \mathcal{L}_{\text{sym}}^+ \# \mathcal{Q}_{\text{sym}}^-$  be the geometric mean of the Laplacians of the expected graphs. Then  $\chi_1, \dots, \chi_k$  are the eigenvectors corresponding to the  $k$  smallest eigenvalues of  $\mathcal{L}_{GM}$  if and only if condition  $E_G$  holds.

*Proof.* We use the same notation as in the proof of Theorem 2. Observing that  $\mathcal{L}_{\text{sym}}^+$  and  $\mathcal{Q}_{\text{sym}}^-$  have the same eigenvectors, it follows from Theorem 1 that

$$\mathcal{L}_{GM} = \Sigma \sqrt{(I - \hat{\Lambda}^+)(I + \hat{\Lambda}^-)} \Sigma^{-1} \quad (8)$$

where  $d^+ \hat{\Lambda}^+ = \Lambda^+$ , and  $d^- \hat{\Lambda}^- = \Lambda^-$ . We deduce that the eigenvalues of  $\mathcal{L}_{GM}$  correspond to eigenvectors in the following way

$$\begin{cases} \sqrt{\left(1 - \frac{\lambda_i^+}{d^+}\right)\left(1 + \frac{\lambda_i^-}{d^-}\right)} & \text{with eigenvector } \chi_i, i = 1, \dots, k \\ 1 & \text{corresponding to the remaining eigenvectors} \end{cases}$$

Thus, eigenvectors  $\chi_i, i = 1, \dots, k$  correspond to the smallest eigenvalues if and only if

$$\left(1 - \frac{\lambda_i^+}{d^+}\right)\left(1 + \frac{\lambda_i^-}{d^-}\right) < 1$$

By eqs. (7) we see that for the constant eigenvector  $\chi_1$  we have

$$\left(1 - \frac{\lambda_1^+}{d^+}\right)\left(1 + \frac{\lambda_1^-}{d^-}\right) = \left(1 - \frac{d^+}{d^+}\right)\left(1 + \frac{d^-}{d^-}\right) = 0 < 1.$$

For eigenvectors  $\chi_i, i = 2, \dots, k$  first observe that

$$1 - \frac{\lambda_i^+}{d^+} = (d^+ - \lambda_i^+)/d^+ = (d^+ - |\mathcal{C}|(p_{\text{in}}^+ - p_{\text{out}}^+))/d^+ = \frac{kp_{\text{out}}^+}{p_{\text{in}}^+ + (k-1)p_{\text{out}}^+}$$

In the same way we have

$$1 + \frac{\lambda_i^-}{d^-} = 1 + \frac{p_{\text{in}}^- - p_{\text{out}}^-}{p_{\text{in}}^- + (k-1)p_{\text{out}}^-}.$$

Thus, for the eigenvectors  $\chi_i, i = 2, \dots, k$  we have the following condition

$$\left(1 - \frac{\lambda_i^+}{d^+}\right)\left(1 + \frac{\lambda_i^-}{d^-}\right) < 1 \iff \left(\frac{kp_{\text{out}}^+}{p_{\text{in}}^+ + (k-1)p_{\text{out}}^+}\right)\left(1 + \frac{p_{\text{in}}^- - p_{\text{out}}^-}{p_{\text{in}}^- + (k-1)p_{\text{out}}^-}\right) < 1,$$

which implies in turn that the eigenvectors  $\chi_i, i = 1, \dots, k$  correspond to the smallest eigenvalues of  $\mathcal{L}_{GM}$  if and only if  $E_G$  holds.  $\square$

As mentioned above, in practical implementations one modifies the Laplacians defining  $\mathcal{L}_{GM}$  by adding a small diagonal shift. This is done to ensure positive definiteness of the matrices. The next theorem shows how to extend the previous result to the case of diagonally shifted Laplacians.

**Theorem 4.** Let  $\mathcal{L}_{GM} = (\mathcal{L}_{\text{sym}}^+ + \varepsilon_1 I) \# (\mathcal{Q}_{\text{sym}}^- + \varepsilon_2 I)$  be the geometric mean of the shifted Laplacians of the expected graphs. Then  $\chi_1, \dots, \chi_k$  are the eigenvectors corresponding to the  $k$  smallest eigenvalues of  $\mathcal{L}_{GM}$  if the following conditions hold.

1.  $\varepsilon_1 + \varepsilon_2 < 1$ .
2.  $\left( \frac{kp_{\text{out}}^+}{p_{\text{in}}^+ + (k-1)p_{\text{out}}^+} \right) \left( 1 + \frac{p_{\text{in}}^- - p_{\text{out}}^-}{p_{\text{in}}^- + (k-1)p_{\text{out}}^-} \right) + (\varepsilon_1 + \varepsilon_2) < 1$ .

*Proof.* We use the same notation as in the previous proof. Observing that  $\mathcal{L}_{\text{sym}}^+$  and  $\mathcal{Q}_{\text{sym}}^-$  have the same eigenvectors, it follows from Theorem 1 that

$$\mathcal{L}_{GM} = \Sigma \sqrt{(I - \hat{\Lambda}^+ + \varepsilon_1 I)(I + \hat{\Lambda}^- + \varepsilon_2 I)} \Sigma^{-1} \quad (9)$$

where  $d^+ \hat{\Lambda}^+ = \Lambda^+$ , and  $d^- \hat{\Lambda}^- = \Lambda^-$ . We deduce that the eigenvalues of  $\mathcal{L}_{GM}$  correspond to eigenvectors in the following way

$$\begin{cases} \sqrt{\left(1 - \frac{\lambda_i^+}{d^+} + \varepsilon_1\right) \left(1 + \frac{\lambda_i^-}{d^-} + \varepsilon_2\right)} & \text{with eigenvector } \chi_i, i = 1, \dots, k \\ (1 + \varepsilon_1)(1 + \varepsilon_2) & \text{corresponding to the remaining eigenvectors} \end{cases}$$

Thus, eigenvectors  $\chi_i, i = 1, \dots, k$  correspond to the smallest eigenvalues if and only if

$$\left(1 - \frac{\lambda_i^+}{d^+} + \varepsilon_1\right) \left(1 + \frac{\lambda_i^-}{d^-} + \varepsilon_2\right) < (1 + \varepsilon_1)(1 + \varepsilon_2) \quad (10)$$

Further, we can see that the previous equation holds if and only if

$$\left(1 - \frac{\lambda_i^+}{d^+}\right) \left(1 + \frac{\lambda_i^-}{d^-}\right) + \varepsilon_1 \left(1 + \frac{\lambda_i^-}{d^-}\right) + \varepsilon_2 \left(1 - \frac{\lambda_i^+}{d^+}\right) < 1 + \varepsilon_1 + \varepsilon_2$$

More over, as  $\left(1 - \frac{\lambda_i^+}{d^+}\right), \left(1 + \frac{\lambda_i^-}{d^-}\right) \in [0, 2]$ , we can see that eq.(10) holds if

$$\left(1 - \frac{\lambda_i^+}{d^+}\right) \left(1 + \frac{\lambda_i^-}{d^-}\right) + \varepsilon_1 + \varepsilon_2 < 1$$

By eqs. (7) we see that for the constant eigenvector  $\chi_1$  we have  $1 - \frac{\lambda_1^+}{d^+} = 1 - \frac{d^+}{d^+} = 0$ . Thus,

$$\left(1 - \frac{\lambda_1^+}{d^+}\right) \left(1 + \frac{\lambda_1^-}{d^-}\right) + \varepsilon_1 + \varepsilon_2 = \varepsilon_1 + \varepsilon_2 < 1$$

For eigenvectors  $\chi_i, i = 2, \dots, k$  first observe that

$$1 - \frac{\lambda_i^+}{d^+} = (d^+ - \lambda_i^+)/d^+ = (d^+ - |\mathcal{C}|(p_{\text{in}}^+ - p_{\text{out}}^+))/d^+ = \frac{kp_{\text{out}}^+}{p_{\text{in}}^+ + (k-1)p_{\text{out}}^+}$$

In the same way we have

$$1 + \frac{\lambda_i^-}{d^-} = 1 + \frac{p_{\text{in}}^- - p_{\text{out}}^-}{p_{\text{in}}^- + (k-1)p_{\text{out}}^-}.$$

Thus, for the eigenvectors  $\chi_i, i = 2, \dots, k$  we have the following condition

$$\begin{aligned} & \left(1 - \frac{\lambda_i^+}{d^+}\right) \left(1 + \frac{\lambda_i^-}{d^-}\right) + \varepsilon_1 + \varepsilon_2 < 1 \iff \\ & \left(\frac{kp_{\text{out}}^+}{p_{\text{in}}^+ + (k-1)p_{\text{out}}^+}\right) \left(1 + \frac{p_{\text{in}}^- - p_{\text{out}}^-}{p_{\text{in}}^- + (k-1)p_{\text{out}}^-}\right) + \varepsilon_1 + \varepsilon_2 < 1, \end{aligned}$$

This implies in turn that the eigenvectors  $\chi_i, i = 1, \dots, k$  correspond to the smallest eigenvalues of  $\mathcal{L}_{GM}$  if the following conditions hold

1.  $\varepsilon_1 + \varepsilon_2 < 1$ .
2.  $\left(\frac{kp_{\text{out}}^+}{p_{\text{in}}^+ + (k-1)p_{\text{out}}^+}\right) \left(1 + \frac{p_{\text{in}}^- - p_{\text{out}}^-}{p_{\text{in}}^- + (k-1)p_{\text{out}}^-}\right) + (\varepsilon_1 + \varepsilon_2) < 1$ .

□

Intuition suggests that a good model should easily identify clusters when  $E_+ \cap E_-$ . However, unlike condition  $E_G$ , condition  $E_{\text{vol}} \cap E_{\text{bal}}$  is not directly satisfied under that regime. Specifically, we have

**Corollary 1.** *Assume that  $E_+ \cap E_-$  holds. Then  $\chi_1, \dots, \chi_k$  are eigenvectors corresponding to the  $k$  smallest eigenvalues of  $\mathcal{L}_{GM}$ . Let  $p(k)$  denote the proportion of cases where  $\chi_1, \dots, \chi_k$  are the eigenvectors of the  $k$  smallest eigenvalues of  $\mathcal{L}_{SN}$  or  $\mathcal{L}_{BN}$ , then  $p(k) \leq \frac{1}{6} + \frac{2}{3(k-1)} + \frac{1}{(k-1)^2}$ .*

*Proof.* The event  $E_{\text{vol}}$  is defined as

$$E_{\text{vol}} = \{(p_{\text{in}}^-, p_{\text{out}}^-, p_{\text{in}}^+, p_{\text{out}}^+) \in [0, 1]^4 \mid p_{\text{in}}^- + (k-1)p_{\text{out}}^- < p_{\text{in}}^+ + (k-1)p_{\text{out}}^+\}$$

We can rewrite the inequality as

$$p_{\text{out}}^- - p_{\text{out}}^+ < \frac{1}{k-1} (p_{\text{in}}^+ - p_{\text{in}}^-) < \frac{1}{k-1}.$$

Thus the event  $\tilde{E}_{\text{vol}}$  defined as

$$\tilde{E}_{\text{vol}} = \{(p_{\text{in}}^-, p_{\text{out}}^-, p_{\text{in}}^+, p_{\text{out}}^+) \in [0, 1]^4 \mid p_{\text{out}}^- - p_{\text{out}}^+ < \frac{1}{k-1}\},$$

satisfies  $E_{\text{vol}} \subset \tilde{E}_{\text{vol}}$ . Then with

$$E_3 = E_+ \cap E_- = \{(p_{\text{in}}^-, p_{\text{out}}^-) \in [0, 1]^2 \mid p_{\text{in}}^- < p_{\text{out}}^-\} \cap \{(p_{\text{in}}^+, p_{\text{out}}^+) \in [0, 1]^2 \mid p_{\text{out}}^+ < p_{\text{in}}^+\}$$

$$E_B = \{(p_{\text{in}}^-, p_{\text{out}}^-, p_{\text{in}}^+, p_{\text{out}}^+) \in [0, 1]^4 \mid p_{\text{in}}^- + p_{\text{out}}^+ < p_{\text{in}}^+ + p_{\text{out}}^-\}$$

we observe  $E_3 \subset E_B$ . Then

$$\begin{aligned} p(k) &= P(E_B \cap E_{\text{vol}} \mid E_3) = \frac{P(E_B \cap E_{\text{vol}} \cap E_3)}{P(E_3)} = \frac{P(E_{\text{vol}} \cap E_3)}{P(E_3)} \\ &\leq \frac{P(\tilde{E}_{\text{vol}} \cap E_3)}{P(E_3)} \end{aligned}$$

Then we get with  $(x_1, x_2, x_3, x_4)$  corresponding to  $(p_{\text{in}}^+, p_{\text{out}}^+, p_{\text{out}}^-, p_{\text{in}}^-)$

$$\begin{aligned} P(\tilde{E}_{\text{vol}} \cap E_3) &\leq \int_0^1 \left( \int_0^{x_1} \left( \int_0^{x_2 + \frac{1}{k-1}} \left( \int_0^{x_3} dx_4 \right) dx_3 \right) dx_2 \right) dx_1 \\ &= \int_0^1 \left( \int_0^{x_1} \left( \int_0^{x_2 + \frac{1}{k-1}} x_3 dx_3 \right) dx_2 \right) dx_1 \\ &= \int_0^1 \left( \int_0^{x_1} \left( \frac{1}{2} \left( x_2 + \frac{1}{k-1} \right)^2 \right) dx_2 \right) dx_1 \\ &= \int_0^1 \left[ \frac{1}{6} \left( x_1 + \frac{1}{k-1} \right)^3 \right]_0^{x_1} dx_1 \\ &= \int_0^1 \left[ \frac{x_1^3}{6} + \frac{x_1^2}{2(k-1)} + \frac{x_1}{2(k-1)^2} \right] dx_1 \\ &= \left[ \frac{x_1^4}{24} + \frac{x_1^3}{6(k-1)} + \frac{x_1^2}{4(k-1)^2} \right]_0^1 \\ &= \frac{1}{24} + \frac{1}{6(k-1)} + \frac{1}{4(k-1)^2} \end{aligned}$$

The first inequality comes from the fact that we do not ensure that the integration upper border for  $x_3$  is smaller or equal to one. Thus with  $P(E_3) = \frac{1}{4}$  we get

$$p(k) \leq \frac{1}{6} + \frac{2}{3(k-1)} + \frac{1}{(k-1)^2}.$$

□



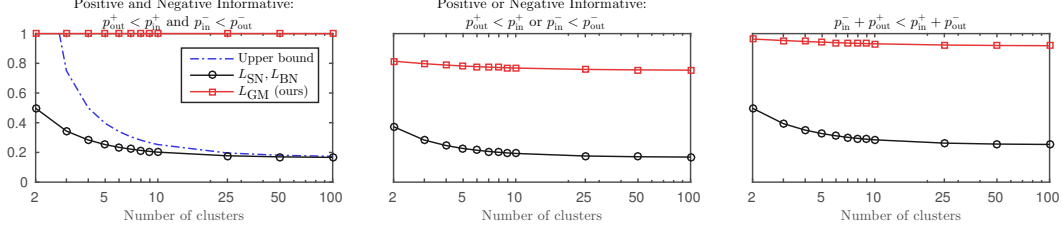


Figure 1: Fraction of cases where in expectation  $\chi_1, \dots, \chi_k$  correspond to the  $k$  smallest eigenvalues under the SBM.

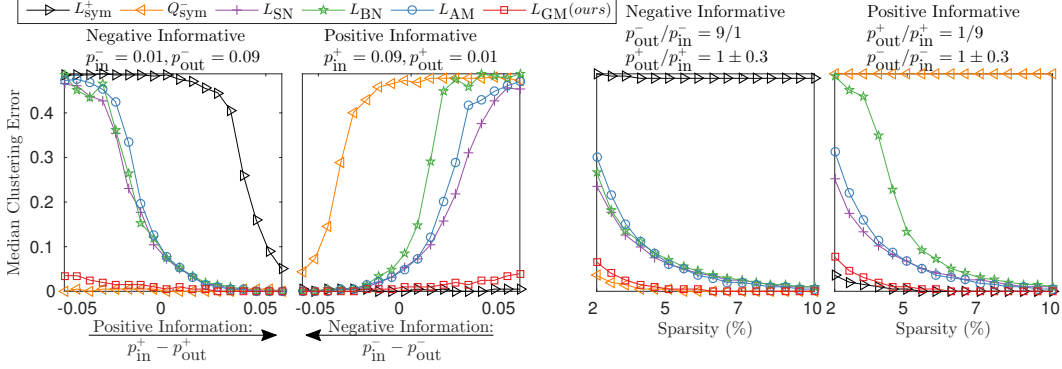


Figure 2: Median clustering error under the stochastic block model over 50 runs.

In order to grasp the difference in expectation between  $L_{BN}$ ,  $L_{SN}$  and  $L_{GM}$ , in Fig 1 we present the proportion of cases where Theorems 2 and 3 hold under different contexts. Experiments are done with all four parameters discretized in  $[0, 1]$  with 100 steps. The expected proportion of cases where  $E_G$  holds (Theorem 3) is far above the corresponding proportion for  $E_{vol} \cap E_{bal}$  (Theorem 2), showing that in expectation the geometric mean Laplacian is superior to the other signed Laplacians. In Fig. 2 we present experiments on sampled graphs with  $k$ -means on top of the  $k$  smallest eigenvectors. In all cases we consider clusters of size  $|\mathcal{C}| = 100$  and present the median of clustering error (i.e., error when clusters are labeled via majority vote) of 50 runs. The results show that the analysis made in expectation closely resembles the actual behavior. In fact, even if we expect only one noisy eigenvector for  $L_{BN}$  and  $L_{SN}$ , the use of the geometric mean Laplacian significantly outperforms any other previously proposed technique in terms of clustering error.  $L_{SN}$  and  $L_{BN}$  achieve good clustering only when the graph resembles a  $k$ -balanced structure, whereas they fail even in the ideal situation where either the positive or the negative graphs are informative about the cluster structure. As shown in Section 6, the advantages of  $L_{GM}$  over the other Laplacians discussed so far allow us to identify a clustering structure on the Wikipedia benchmark real world signed network, where other clustering approaches have failed.

## 5 Krylov-based inverse power method for small eigenvalues of $L_{sym}^+ \# Q_{sym}^-$

The computation of the geometric mean  $A \# B$  of two positive definite matrices of moderate size has been discussed extensively by various authors [20, 11, 12, 13]. However, when  $A$  and  $B$  have large dimensions, the approaches proposed so far become unfeasible, in fact  $A \# B$  is in general a full matrix even if  $A$  and  $B$  are sparse. In this section we present a scalable algorithm for the computation of the smallest eigenvectors of  $L_{sym}^+ \# Q_{sym}^-$ . The method is discussed for a general pair of matrices  $A$  and  $B$ , to emphasize its general applicability which is therefore interesting in itself. We remark that the method takes advantage of the sparsity of  $A$  and  $B$  and does not require to explicitly compute the matrix  $A \# B$ . To our knowledge this is the first effective method explicitly built for the computation of the eigenvectors of the geometric mean of two large and sparse positive definite matrices.

Given a positive definite matrix  $M$  with eigenvalues  $\lambda_1 \leq \dots \leq \lambda_n$ , let  $\mathcal{H}$  be any eigenspace of  $M$  associated to  $\lambda_1, \dots, \lambda_t$ . The inverse power method (IPM) applied to  $M$  is a method that converges to an eigenvector  $\mathbf{x}$  associated to the smallest eigenvalue  $\lambda_{\mathcal{H}}$  of  $M$  such that  $\lambda_{\mathcal{H}} \neq \lambda_i, i = 1, \dots, t$ .

The pseudocode of IPM applied to  $A\#B = A(A^{-1}B)^{1/2}$  is shown in Algorithm 2. Given a vector  $\mathbf{v}$  and a matrix  $M$ , the notation  $\text{solve}\{M, \mathbf{v}\}$  is used to denote a procedure returning the solution  $\mathbf{x}$  of the linear system  $M\mathbf{x} = \mathbf{v}$ . At each step the algorithm requires the solution of two linear systems. The first one (line 2) is solved by the preconditioned conjugate gradient method, where the preconditioner is obtained by the incomplete Cholesky decomposition of  $A$ . Note that the conjugate gradient method is very fast, as  $A$  is assumed sparse and positive definite, and it is matrix-free, *i.e.* it requires to compute the action of  $A$  on a vector, whereas it does not require the knowledge of  $A$  (nor its inverse). The solution of the linear system occurring in line 3 is the major inner-problem of the proposed algorithm. Its efficient solution is performed by means of an extended Krylov subspace technique that we describe in the next section. The proposed implementation ensures the whole IPM is matrix-free and scalable.

### 5.1 Extended Krylov subspace method for the solution of the linear system $(A^{-1}B)^{1/2}\mathbf{x} = \mathbf{y}$

We discuss here how to apply the technique known as Extended Krylov Subspace Method (**EKSM**) for the solution of the linear system  $(A^{-1}B)^{1/2}\mathbf{x} = \mathbf{y}$ . Let  $M$  be a large and sparse matrix, and  $\mathbf{y}$  a given vector. When  $f$  is a function with a single pole, EKSM is a very effective method to approximate the vector  $f(M)\mathbf{y}$  without ever computing the matrix  $f(M)$  [8]. Note that, given two positive definite matrices  $A$  and  $B$  and a vector  $\mathbf{y}$ , the vector we want to compute is  $\mathbf{x} = (A^{-1}B)^{-1/2}\mathbf{y}$ , so that our problem boils down to the computation of the product  $f(M)\mathbf{y}$ , where  $M = A^{-1}B$  and  $f(X) = X^{-1/2}$ . The general idea of EKSM  $s$ -th iteration is to project  $M$  onto the subspace

$$\mathbb{K}^s(M, \mathbf{y}) = \text{span}\{\mathbf{y}, M\mathbf{y}, M^{-1}\mathbf{y}, \dots, M^{s-1}\mathbf{y}, M^{1-s}\mathbf{y}\},$$

and solve the problem there. The projection onto  $\mathbb{K}^s(M, \mathbf{y})$  is realized by means of the Lanczos process, which produces a sequence of matrices  $V_s$  with orthogonal columns, such that the first column of  $V_s$  is a multiple of  $\mathbf{y}$  and  $\text{range}(V_s) = \mathbb{K}^s(M, \mathbf{y})$ . Moreover at each step we have

$$MV_s = V_s H_s + [\mathbf{u}_{s+1}, \mathbf{v}_{s+1}][\mathbf{e}_{2s+1}, \mathbf{e}_{2s+2}]^T \quad (11)$$

where  $H_s$  is  $2s \times 2s$  symmetric tridiagonal,  $\mathbf{u}_{s+1}$  and  $\mathbf{v}_{s+1}$  are orthogonal to  $V_s$ , and  $\mathbf{e}_i$  is the  $i$ -th canonical vector. The solution  $\mathbf{x}$  is then approximated by  $\mathbf{x}_s = V_s f(H_s) \mathbf{e}_1 \|\mathbf{y}\| \approx f(M)\mathbf{y}$ . If  $n$  is the order of  $M$ , then the exact solution is obtained after at most  $n$  steps. However, in practice, significantly fewer iterations are enough to achieve a good approximation, as the error  $\|\mathbf{x}_s - \mathbf{x}\|$  decays exponentially with  $s$  (Thm 3.4 and Prop. 3.6 in [14]).

The pseudocode for the extended Krylov iteration is presented in Algorithm 3. We use the stopping criterion proposed in [14]. It is worth pointing out that at step 4 of the algorithm we can freely choose any scalar product  $\langle \cdot, \cdot \rangle$ , without affecting formula (11) nor the convergence properties of the method. As  $M = A^{-1}B$ , we use the scalar product  $\langle \mathbf{u}, \mathbf{v} \rangle_A = \mathbf{u}^T A \mathbf{v}$  induced by the positive definite matrix  $A$ , so that the computation of the tridiagonal matrix  $H_s$  in the algorithm simplifies to  $V_s^T B V_s$ . We refer to [9] for further details. As before, the solve procedure is implemented by means of the preconditioned conjugate gradient method, where the preconditioner is obtained by the incomplete Cholesky decomposition of the coefficient matrix. Figure 3 shows that we are able to compute the smallest eigenvector of  $L_{\text{sym}}^+ \# Q_{\text{sym}}^-$  being just a constant factor worse than the computation of the eigenvector of the arithmetic mean, whereas the direct computation of the geometric mean followed by the computation of the eigenvectors is unfeasible for large graphs.

<b>Input:</b> $\mathbf{x}_0$ , eigenspace $\mathcal{H}$ of $A\#B$ .	<b>Input:</b> $\mathbf{u}_0 = \mathbf{y}$ , $V_0 = [\cdot]$
<b>Output:</b> Eigenpair $(\lambda_{\mathcal{H}}, \mathbf{x})$ of $A\#B$	<b>Output:</b> $\mathbf{x} = (A^{-1}B)^{-1/2}\mathbf{y}$
1 <b>repeat</b>	1 $\mathbf{v}_0 \leftarrow \text{solve}\{B, A\mathbf{u}_0\}$
2 $\mathbf{u}_k \leftarrow \text{solve}\{A, \mathbf{x}_k\}$	2 <b>for</b> $s = 0, 1, 2, \dots, n$ <b>do</b>
3 $\mathbf{v}_k \leftarrow \text{solve}\{(A^{-1}B)^{1/2}, \mathbf{u}_k\}$	3 $\tilde{V}_{s+1} \leftarrow [V_s, \mathbf{u}_s, \mathbf{v}_s]$
4 $\mathbf{y}_k \leftarrow \text{project } \mathbf{u}_k \text{ over } \mathcal{H}^\perp$	4 $V_{s+1} \leftarrow \text{Orthogonalize columns of } \tilde{V}_{s+1} \text{ w.r.t. } \langle \cdot, \cdot \rangle_A$
5 $\mathbf{x}_{k+1} \leftarrow \mathbf{y}_k / \ \mathbf{y}_k\ _2$	5 $H_{s+1} \leftarrow V_{s+1}^T B V_{s+1}$
6 <b>until tolerance reached</b>	6 $\mathbf{x}_{s+1} \leftarrow H_{s+1}^{-1/2} \mathbf{e}_1$
7 $\lambda_{\mathcal{H}} \leftarrow \mathbf{x}_{k+1}^T \mathbf{x}_k$ , $\mathbf{x} \leftarrow \mathbf{x}_{k+1}$	7 <b>if tolerance reached then break</b>
<b>Algorithm 2:</b> IPM applied to $A\#B$ .	8 $\mathbf{u}_{s+1} \leftarrow \text{solve}\{A, B V_{s+1} \mathbf{e}_1\}$
	9 $\mathbf{v}_{s+1} \leftarrow \text{solve}\{B, A V_{s+1} \mathbf{e}_2\}$
	10 <b>end</b>
	11 $\mathbf{x} \leftarrow V_{s+1} \mathbf{x}_{s+1}$
	<b>Algorithm 3:</b> EKSM for the computation of $(A^{-1}B)^{-1/2}\mathbf{y}$

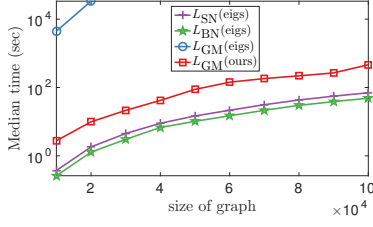


Figure 3: Median execution time of 10 runs for different Laplacians. Graphs have two perfect clusters and 2.5% of edges among nodes.  $L_{GM}(ours)$  uses Algs 2 and 3, whereas we used Matlab’s `eigs` for the other matrices. The use of `eigs` on  $L_{GM}$  is prohibitive as it needs the matrix  $L_{GM}$  to be built (we use the toolbox provided in [2]), destroying the sparsity of the original graphs. Experiments are performed using one thread.

## 5.2 On the computational cost of the method

Let  $c(n)$  denote the computational cost to compute the solution of a linear system with coefficient matrix either  $L_{sym}^+$  or  $Q_{sym}^-$ . Standard iterative techniques allows to compute the smallest eigenvector of  $L_{sym}^+$  or  $Q_{sym}^-$  at a cost of  $O(c(n))$  operations per step. We show that the use of Algorithms 2 and 3 allows to compute the eigenvectors of  $L_{sym}^+ \# Q_{sym}^-$  with the same order of operations.

First of all it is important to realize that the matrix  $H_s = V_s^T B V_s$ , defined in line 5, can be defined iteratively and does not require any additional matrix multiplication [14]. Thus the cost of each iteration of Algorithm 3 dominated by lines 8 and 9, and requires  $O(2c(n))$  ops. The algorithm converges exponentially, namely if  $[a, b]$  is any interval containing the eigenvalues of  $A^{-1}B$ , then  $\|x_s - x\| = O(\exp(-2s \sqrt[4]{a/b}))$ , where  $x = (A^{-1}B)^{-1/2}y$ . See f.i. [14] for details. Thus  $O(s_\varepsilon)$  iterations are enough to reach the prescribed tolerance  $\varepsilon > 0$ , where  $s_\varepsilon = \lceil \log \varepsilon / 2 \sqrt[4]{a/b} \rceil$ . However it is worth pointing out that in practice, at least for the matrices considered in this work, much less iterations than  $O(s_\varepsilon)$  are enough. Therefore the proposed IPM technique allows to compute the smallest eigenvector of  $L_{sym}^+ \# Q_{sym}^-$  at a cost of  $O(c(n)) + O(2s_\varepsilon c(n))$  operations per step. This shows that the method is scalable. A final important remark concerns step 6. The matrix  $H_s$  is tridiagonal of size  $2s \times 2s$ , thus the function  $H_s^{-1/2}$  can be implemented directly using a method for dense matrices, without any notable change to the overall algorithm cost.

Next Figure 3 shows that, despite the computationally ugly definition of  $L_{sym}^+ \# Q_{sym}^-$ , we are able to compute its smallest eigenvector with a constant factor overcome, whereas the naïve direct computation would be extremely prohibitive or unfeasible.

In Fig. 3 we show the median execution time for the computation of the smallest eigenvector of the signed ratio/normalized cut Laplacians, the balance ratio/normalized cut Laplacians and the geometric mean  $L_{sym}^+ \# Q_{sym}^-$ . We randomly generate graphs with a sparsity of 2.5% under the perfect stochastic case, i.e.  $p_{in}^+ = p_{out}^- = 1$  and  $p_{out}^+ = p_{in}^- = 0$ , where the size of graphs goes from 10,000 to 100,000 in steps of 10,000. For each setting we report the median execution time out of 10 runs. Experiments are performed using one thread.

For the computation of the smallest eigenvector of the signed ratio/normalized cut Laplacians and the balance ratio/normalized cut Laplacians we compute the Laplacian matrix (i.e.  $L_{SR}$ ,  $L_{SN}$ ,  $L_{BR}$  and  $L_{BN}$ ) and use the function `eigs` from Matlab. For the computation of the smallest eigenvector of the geometric mean we consider two approaches: one approach is based on the computation of the geometric mean  $L_{sym}^+ \# Q_{sym}^-$  using the Matlab toolbox provided by [12] and then the use of the function `eigs` from Matlab (in Fig. 3 denoted as  $L_{GM}(eigs)$ ). The second approach is based on the Inverse Power Method of Algorithm 2 together with the extended Krylov method of Algorithm 3 (in Fig. 3 denoted as  $L_{GM}(ours)$ ).

We can see that the execution time of for signed Laplacians is rather similar. One can observe that the execution time for the geometric mean with Matlab’s `eigs` is truncated for graphs that have more than 20,000 nodes. This happens as the computation of the geometric mean does not fit into memory. On the other side, the time execution for the geometric mean with the Inverse Power Method and extended Krylov methods (Algorithms 2 and 3) is comparable with the one of the signed Laplacians that use `eigs`. In particular it is noticeable that the time executions differs just by a constant factor.

## 6 Experiments

**Sociology Networks** We evaluate signed Laplacians  $L_{SN}$ ,  $L_{BN}$ ,  $L_{AM}$  and  $L_{GM}$  through three real-world and moderate size signed networks: Highland tribes (Gahuku-Gama) network [21], Slovene Parliamentary Parties Network [15] and US Supreme Court Justices Network [7]. For the sake of comparison we take as ground truth the clustering that is stated in the corresponding references. We observe that all signed Laplacians yield zero clustering error.

**Experiments on Wikipedia signed network.** We consider the Wikipedia adminship election dataset from [17], which describes relationships that are positive, negative or non existent. We use Algs. 1–3 and look for 30 clusters. Positive and negative adjacency matrices sorted according to our clustering are depicted in Figs. 4(a) and 4(b). We can observe the presence of a large relatively empty cluster. Zooming into the denser portion of the graph we can see a  $k$ -balanced behavior (see Figs. 4(c) and 4(d)), *i.e.* the positive adjacency matrix shows assortative groups - resembling a block diagonal structure - while the negative adjacency matrix shows a disassortative setting. Using  $L_{AM}$  and  $L_{BN}$  we were not able to find any clustering structure, which corroborates results reported in [4]. This further confirms that  $L_{GM}$  overcomes other clustering approaches. To the knowledge of the authors, this is the first time that clustering structure has been found in this dataset.

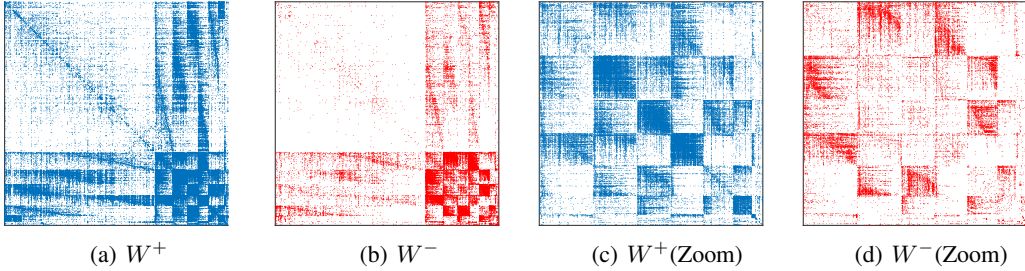


Figure 4: Wikipedia weight matrices sorted according to the clustering obtained with  $L_{GM}$  (Alg. 1).

**Experiments on UCI datasets.** We evaluate our method  $L_{GM}$  (Algs. 1–3) against  $L_{SN}$ ,  $L_{BN}$ , and  $L_{AM}$  with datasets from the UCI repository (see Table. 2). We build  $W^+$  from a symmetric  $k^+$ -nearest neighbor graph, whereas  $W^-$  is obtained from the symmetric  $k^-$ -farthest neighbor graph. For each dataset we test all clustering methods over all possible choices of  $k^+, k^- \in \{3, 5, 7, 10, 15, 20, 40, 60\}$ . In Table 2 we report the fraction of cases where each method achieves the best and strictly best clustering error over all the 64 graphs, per each dataset. We can see that our method outperforms other methods across all datasets.

		iris	wine	ecoli	optdig	USPS	pendig	MNIST
# vertices		150	178	310	5620	9298	10992	70000
# classes		3	3	3	10	10	10	10
$L_{SN}$	Best (%)	23.4	40.6	18.8	28.1	10.9	10.9	12.5
	Str. best (%)	10.9	21.9	14.1	28.1	9.4	10.9	12.5
$L_{BN}$	Best (%)	17.2	21.9	7.8	0.0	1.6	3.1	0.0
	Str. best (%)	7.8	4.7	6.3	0.0	1.6	3.1	0.0
$L_{AM}$	Best (%)	12.5	28.1	14.1	0.0	0.0	1.6	0.0
	Str. best (%)	10.9	14.1	12.5	0.0	0.0	1.6	0.0
$L_{GM}$	Best (%)	<b>59.4</b>	<b>42.2</b>	<b>65.6</b>	<b>71.9</b>	<b>89.1</b>	<b>84.4</b>	<b>87.5</b>
	Str. best (%)	<b>57.8</b>	<b>35.9</b>	<b>60.9</b>	<b>71.9</b>	<b>87.5</b>	<b>84.4</b>	<b>87.5</b>

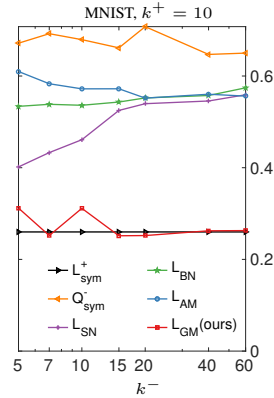


Table 2: Experiments on UCI datasets. Left: fraction of cases where methods achieve best and strictly best clustering error. Right: clustering error on MNIST dataset.

In the figure on the right of Table 2 we present the clustering error on MNIST dataset fixing  $k^+ = 10$ . With  $Q_{sym}^-$  one gets the highest clustering error, which shows that the  $k^-$ -farthest neighbor graph is a source of noise and is not informative. In fact, we observe that a small subset of nodes is the farthest

neighborhood of a large fraction of nodes. The noise from the  $k^-$ -farthest neighbor graph is strongly influencing the performances of  $L_{SN}$  and  $L_{BN}$ , leading to a noisy embedding of the datapoints and in turn to a high clustering error. On the other hand we can see that  $L_{GM}$  is robust, in the sense that its clustering performances are not affected negatively by the noise in the negative edges. Similar behaviors have been observed for the other datasets in Table 2.

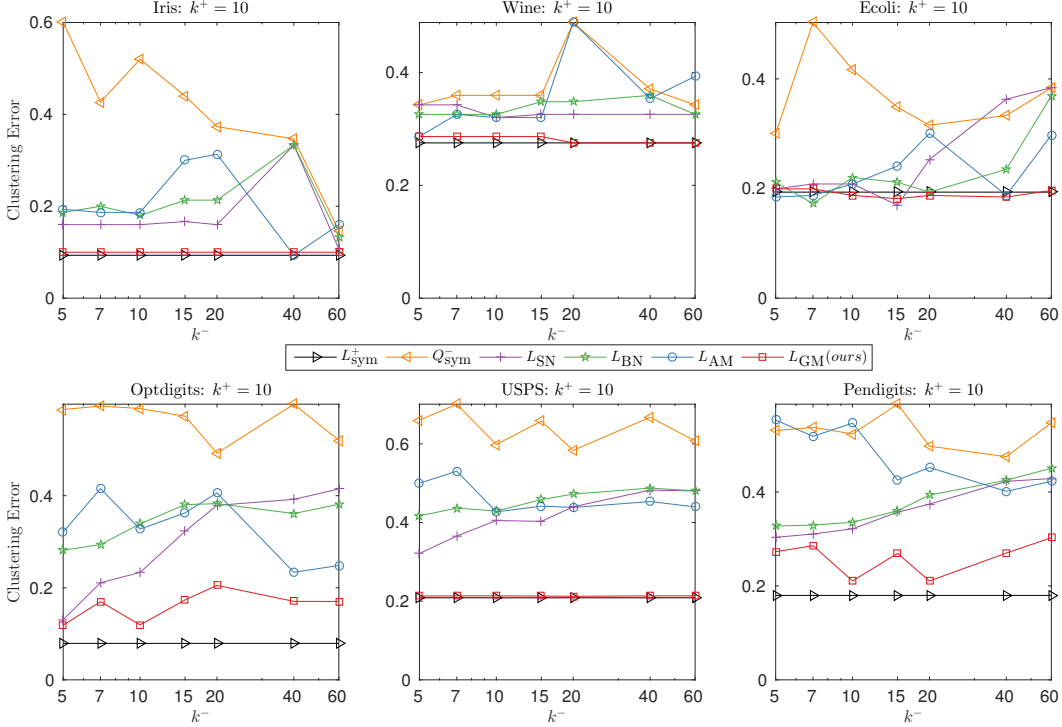


Figure 5: Clustering error on UCI datasets, for  $k^+ = 10$ .

**Acknowledgments.** The authors acknowledge support by the ERC starting grant NOLEPRO

## References

- [1] R. Bhatia. *Positive definite matrices*. Princeton University Press, 2009.
- [2] D. Bini and B. Iannazzo. The Matrix Means Toolbox. <http://bezout.dm.unipi.it/software/mmttoolbox/>, May 2015.
- [3] D. Cartwright and F. Harary. Structural balance: a generalization of Heider’s theory. *Psychological Review*, 63(5):277–293, 1956.
- [4] K. Chiang, J. Whang, and I. Dhillon. Scalable clustering of signed networks using balance normalized cut. *CIKM*, pages 615–624, 2012.
- [5] J. A. Davis. Clustering and structural balance in graphs. *Human Relations*, 20:181–187, 1967.
- [6] M. Desai and V. Rao. A characterization of the smallest eigenvalue of a graph. *Journal of Graph Theory*, 18(2):181–194, 1994.
- [7] P. Doreian and A. Mrvar. Partitioning signed social networks. *Social Networks*, 31(1):1–11, 2009.
- [8] V. Druskin and L. Knizhnerman. Extended Krylov subspaces: approximation of the matrix square root and related functions. *SIAM J. Matrix Anal. Appl.*, 19:755–771, 1998.
- [9] M. Fasi and B. Iannazzo. Computing the weighted geometric mean of two large-scale matrices and its inverse times a vector. *MIMS EPrint: 2016.29*.

- [10] F. Harary. On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 2:143–146, 1953.
- [11] N. J. Higham, D. S. Mackey, N. Mackey, and F. Tisseur. Functions preserving matrix groups and iterations for the matrix square root. *SIAM J. Matrix Anal. Appl.*, 26:849–877, 2005.
- [12] B. Iannazzo. The geometric mean of two matrices from a computational viewpoint. *Numer. Linear Algebra Appl.*, to appear, 2015.
- [13] B. Iannazzo and M. Porcelli. The Riemannian Barzilai-Borwein method with nonmonotone line-search and the Karcher mean computation. *Optimization online*, December 2015.
- [14] L. Knizhnerman and V. Simoncini. A new investigation of the extended Krylov subspace method for matrix function evaluations. *Numer. Linear Algebra Appl.*, 17:615–638, 2009.
- [15] S. Kropivnik and A. Mrvar. An Analysis of the Slovene Parliamentary Parties Networks. *Development in Statistics and Methodology*, pages 209–216, 1996.
- [16] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. Luca, and S. Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *ICDM*, pages 559–570, 2010.
- [17] J. Leskovec and A. Krevl. SNAP Datasets: Stanford Large Network Dataset Collection. <http://snap.stanford.edu/data>, June 2014.
- [18] S. Liu. Multi-way dual cheeger constants and spectral bounds of graphs. *Advances in Mathematics*, 268:306 – 338, 2015.
- [19] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, Dec. 2007.
- [20] M. Raïssouli and F. Leazizi. Continued fraction expansion of the geometric matrix mean and applications. *Linear Algebra Appl.*, 359:37–57, 2003.
- [21] K. E. Read. Cultures of the Central Highlands, New Guinea. *Southwestern Journal of Anthropology*, 10(1):pp. 1–43, 1954.
- [22] K. Rohe, S. Chatterjee, B. Yu, et al. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.
- [23] J. Tang, Y. Chang, C. Aggarwal, and H. Liu. A survey of signed network mining in social media. *arXiv preprint arXiv:1511.07569*, 2015.