

---

# A View of Exact Inference in Graphs from the Degree-4 Sum-of-Squares Hierarchy

---

**Kevin Bello**

Department of Computer Science  
Purdue University  
kbellome@purdue.edu

**Chuyang Ke**

Department of Computer Science  
Purdue University  
cke@purdue.edu

**Jean Honorio**

Department of Computer Science  
Purdue University  
jhonorio@purdue.edu

## Abstract

Performing inference in graphs is a common task within several machine learning problems, e.g., image segmentation, community detection, among others. For a given undirected connected graph, we tackle the statistical problem of *exactly* recovering an unknown ground-truth binary labeling of the nodes from a *single* corrupted observation of each edge. Such problem can be formulated as a quadratic combinatorial optimization problem over the boolean hypercube, where it has been shown before that one can (with high probability and in polynomial time) exactly recover the ground-truth labeling of graphs that have an isoperimetric number that grows with respect to the number of nodes (e.g., complete graphs, regular expanders). In this work, we apply a powerful hierarchy of relaxations, known as the sum-of-squares (SoS) hierarchy, to the combinatorial problem. Motivated by empirical evidence on the improvement in exact recoverability, we center our attention on the degree-4 SoS relaxation and set out to understand the origin of such improvement from a graph theoretical perspective. We show that the solution of the dual of the relaxed problem is related to finding edge weights of the Johnson and Kneser graphs, where the weights fulfill the SoS constraints and intuitively allow the input graph to increase its algebraic connectivity. Finally, as byproduct of our analysis, we derive a novel Cheeger-type lower bound for the algebraic connectivity of graphs with *signed* edge weights.

---

Proceedings of the 25<sup>th</sup> International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151. Copyright 2022 by the author(s).

## 1 INTRODUCTION

Inference in graphs spans several domains such as social networks, natural language processing, computational biology, computer vision, among others. For example, let  $\mathbf{X} \in \{-1, 0, +1\}^{n \times n}$  be some *noisy* observation, e.g., a social network (represented by a graph), where the output is a labeling  $\mathbf{y} \in \{-1, +1\}^n$  of the nodes, e.g., an assignment of each individual to a cluster. In the example, for the entries of  $\mathbf{X}$ , a value of 0 means no interaction (no edge) between two individuals (nodes), a value of +1 can represent an agreement of two individuals, while a value of -1 can represent disagreement. One can then predict a labeling  $\mathbf{y}$  by solving the following quadratic form over the hypercube  $\{\pm 1\}^n$ ,

$$\max_{\mathbf{y} \in \{\pm 1\}^n} \mathbf{y}^\top \mathbf{X} \mathbf{y}. \quad (1)$$

The above formulation is a well-studied problem that arises in multiple contexts, including Ising models from statistical physics (Barahona, 1982), finding the maximum cut of a graph (Goemans and Williamson, 1995), the Grothendieck problem (Grothendieck, 1956; Khot and Naor, 2011), stochastic block models (Abbe et al., 2016), and structured prediction problems such as image segmentation (Globerson et al., 2015), to name a few. However, the optimization problem above is NP-hard in general and only some cases are known to be exactly solvable in polynomial time. For instance, Chandrasekaran et al. (2008) showed that it can be solved exactly in polynomial time for a graph with low treewidth via the junction tree algorithm; Schraudolph and Kamenetsky (2009) showed that the inference problem can also be solved exactly in polynomial time for planar graphs via perfect matchings; while Boykov and Veksler (2006) showed that (1) can be solved exactly in polynomial time via graph cuts for binary labels and sub-modular pairwise potentials.

In this work, we consider a generative model proposed by Globerson et al. (2015) in the context of structured

prediction (specifically, image segmentation), and study which conditions on the graph allow for exact recovery (inference). In this case, following up with the example on social networks above, each individual can have an opinion labeled by  $-1$  or  $+1$ , and for each pair of individuals that are *connected*, we observe a *single* measurement of whether or not they have an agreement in opinion, but the value of each measurement is flipped with probability  $p$ . Since problem (1) is a hard computational problem, it is common to relax the problem to a convex one. In particular, (Abbe et al., 2016; Amini and Levina, 2018; Bello and Honorio, 2019, 2020) studied the sufficient conditions for a semidefinite programming relaxation (SDP) of problem (1) to achieve exact recovery. In contrast to those works, we will focus on the sum-of-squares (SoS) hierarchy of relaxations (Parrilo, 2000; Lasserre, 2001; Barak and Steurer, 2014), which is a sequential tightening of convex relaxations based on SDP. We study the SoS hierarchy because it is tighter than other known hierarchies such as the Sherali-Adams and Lovász-Schrijver hierarchies (Laurent, 2003). In addition, our motivation to study the level-2 or degree-4 SoS relaxation stems from three reasons. First, higher-levels of the hierarchy, while polynomial time solvable, are already computationally very costly. This is one of the reasons the SoS hierarchy have been mostly used as a proof system for finding lower bounds in hard problems (e.g., for the planted clique problem, see (Meka et al., 2015)). Second, little is still known about the level-2 SoS relaxation, where (Bandeira and Kunisky, 2018) and (Cifuentes et al., 2020) are attempts to understand its geometry. Third, there is empirical evidence on the *non-asymptotic* improvement in exact recoverability with respect to SDP, an example of which is depicted in Figure 1.

**Contributions.** While it is known that the level-2 SoS relaxation has a tighter search space than that of SDP, it is not obvious why it can perform better than SDP for *exact recovery*. In this work, we aim to understand the origin of such improvement from a graph theoretical perspective. We show that the solution of the dual of the relaxed problem is related to finding edge weights of the Johnson and Kneser graphs, where the weights fulfill the SoS constraints and intuitively allow the input graph to increase its algebraic connectivity (Theorem 2). Finally, as byproduct of our analysis, we derive a novel Cheeger-type lower bound for the algebraic connectivity of graphs with *signed* edge weights (Theorem 1), which can be of independent interest.

We emphasize that the objective of this work is on the *understanding* of exact recoverability by using the degree-4 SoS. Scalability of the SoS hierarchy is an

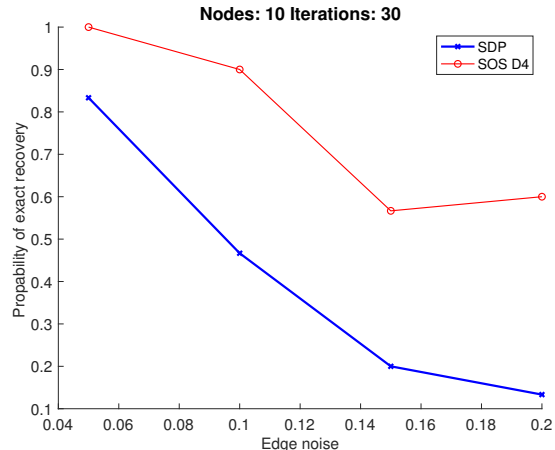


Figure 1: A comparison between the degree-4 SoS and SDP relaxations in the context of structured prediction. We observe that SoS attains a higher probability of exact recovery, for different levels of edge noise  $p$ . (See Section 2 for a formal problem definition.)

important open problem that is actively under study (Weisser et al., 2016; Erdogdu et al., 2017) and is beyond the scope of our work.

## 2 PRELIMINARIES

This section introduces the notation used throughout the paper and formally defines the problem under analysis.

Vectors and matrices are denoted by lowercase and uppercase bold faced letters respectively (e.g.,  $\mathbf{a}$ ,  $\mathbf{A}$ ), while scalars are in normal font weight (e.g.,  $a$ ). For a vector  $\mathbf{a}$ , and a matrix  $\mathbf{A}$ , their entries are denoted by  $a_i$  and  $A_{i,j}$  respectively. Indexing starts at 1, with  $\mathbf{A}_{i,:}$  and  $\mathbf{A}_{:,i}$  indicating the  $i$ -th row and  $i$ -th column of  $\mathbf{A}$  respectively. The eigenvalues of a  $n \times n$  matrix  $\mathbf{A}$  are denoted as  $\lambda_i(\mathbf{A})$ , where  $\lambda_1$  and  $\lambda_n$  correspond to the minimum and maximum eigenvalue respectively. Finally, the set of integers  $\{1, \dots, n\}$  is represented as  $[n]$ .

**Problem definition.** We aim to predict a vector of  $n$  node labels  $\mathbf{y} = (y_1, \dots, y_n)^\top$ , where  $y_i \in \{+1, -1\}$ , from a set of observations  $\mathbf{X}$ , where  $\mathbf{X}$  corresponds to noisy measurements of edges. These observations are assumed to be generated from a ground truth labeling  $\bar{\mathbf{y}}$  by a generative process defined via an undirected connected graph  $G = (V, E)$ , where  $V = [n]$ , and an edge noise  $p \in (0, 0.5)$ . For each edge  $(u, v) \in E$ , we have a *single* independent edge observation  $X_{u,v} = \bar{y}_u \bar{y}_v$  with probability  $1 - p$ , and  $X_{u,v} = -\bar{y}_u \bar{y}_v$  with

probability  $p$ . While for each edge  $(u, v) \notin E$ , the observation  $X_{u,v}$  is always 0. Thus, we have a *known* undirected connected graph  $G$ , an *unknown* ground truth label vector  $\bar{\mathbf{y}} \in \{+1, -1\}^n$ , noisy observations  $\mathbf{X} \in \{-1, 0, +1\}^{n \times n}$ . Given that we consider only edge observations, our goal is to understand when one can predict, in polynomial time and with high probability, a vector label  $\mathbf{y} \in \{-1, +1\}^n$  such that  $\mathbf{y} \in \{\bar{\mathbf{y}}, -\bar{\mathbf{y}}\}$ .

Given the aforementioned generative process, our focus will be to solve the following optimization problem, which stems from using maximum likelihood estimation (Globerson et al., 2015):

$$\max_{\mathbf{y}} \mathbf{y}^\top \mathbf{X} \mathbf{y}, \text{ subject to } y_i = \pm 1, \forall i \in [n]. \quad (2)$$

In general, the above combinatorial problem is NP-hard to compute, e.g., see results on grids by Barahona (1982). Let  $\mathbf{y}^{\text{dis}}$  denote the optimizer of eq.(2). It is clear that for any label vector  $\mathbf{y}$ , the negative label vector  $-\mathbf{y}$  attains the same objective value in eq.(2). Thus, we say that one can achieve exact recovery by solving eq.(2) if  $\mathbf{y}^{\text{dis}} \in \{\bar{\mathbf{y}}, -\bar{\mathbf{y}}\}$ . Given the computational hardness of solving eq.(2), in the next subsections we will revise approaches that relax problem (2) to one that can be solved in polynomial time. Then, our focus will be to understand the effects of the structural properties of the graph  $G$  in achieving, with high probability, exact recovery in the continuous problem.

## 2.1 Semidefinite Programming Relaxation

A popular approach for approximating problem (2) is to consider a larger search space that is simpler to describe and is convex. In particular, let  $\mathbf{Y} = \mathbf{y}\mathbf{y}^\top$ , that is,  $Y_{i,j} = y_i y_j$  and noting that  $\mathbf{Y}$  is a rank-1 positive semidefinite matrix. We can rewrite the objective of problem (2) in matrix terms as follows,  $\mathbf{y}^\top \mathbf{X} \mathbf{y} = \text{Tr}(\mathbf{X} \mathbf{y}\mathbf{y}^\top) = \text{Tr}(\mathbf{X} \mathbf{Y}) = \langle \mathbf{X}, \mathbf{Y} \rangle$ . Thus, we have

$$\begin{aligned} \max_{\mathbf{Y}} \quad & \langle \mathbf{X}, \mathbf{Y} \rangle \\ \text{subject to} \quad & \mathbf{Y} \succeq 0, \quad Y_{i,i} = 1, \forall i \in [n]. \end{aligned} \quad (3)$$

Let  $\mathbf{Y}^*$  denote the optimizer of the problem above, then, in this case, we say that exact recovery is realized by solving eq.(3) if  $\mathbf{Y}^* = \bar{\mathbf{y}}\bar{\mathbf{y}}^\top$ . The only constraint dropped in problem (3) with respect to problem (2) is the rank-1 constraint, which makes problem (3) convex. The above relaxation is known as semidefinite programming (SDP) relaxation and is typically used as an approximation algorithm. That is, after obtaining a continuous solution  $\mathbf{Y}^* \in \mathbb{R}^{n \times n}$ , a rounding procedure is performed to recover an approximate solution in  $\{\pm 1\}^n$ , e.g., see (Goemans and Williamson, 1995; Nesterov, 1998). However, SDP relaxations have also

been analyzed for exact inference, for instance, (Abbe et al., 2016) and (Amini and Levina, 2018) studied exact recovery in the context of stochastic block models, while, (Bello and Honorio, 2019, 2020) studied exact recovery in the context of structured prediction.

In the next subsection, we will introduce tighter levels of relaxations known as the SoS hierarchy, and we will see that it turns out that SDP relaxations correspond to the first level of the SoS hierarchy.

## 2.2 Sum-of-Squares Hierarchy

We start this section by introducing additional notation for describing the SoS hierarchy. Let  $[n]^{\leq d} = \{\emptyset\} \cup [n]^1 \cup \dots \cup [n]^d$  denote the set of (possibly empty) tuples, of length up to  $d$ , composed of the integers from 1 to  $n$ , e.g.,  $[2]^{\leq 2} = \{\emptyset, (1), (2), (1, 1), (1, 2), (2, 1), (2, 2)\}$ . Also, let the summation between two tuples be the concatenation of all the elements in them, e.g., for  $\mathcal{C}_1 = (1, 1, 2), \mathcal{C}_2 = (3, 1)$  we have  $\mathcal{C}_1 + \mathcal{C}_2 = (1, 1, 2, 3, 1)$ . We use  $\psi(\mathcal{C})$  to denote the tuple with elements from  $\mathcal{C}$  sorted in ascending order, e.g., for  $\mathcal{C} = (2, 1, 1, 3)$  we have  $\psi(\mathcal{C}) = (1, 1, 2, 3)$ . We also use  $|\mathcal{C}|$  to denote the cardinality of  $\mathcal{C}$ . For two distinct tuples  $\mathcal{C}_1$  and  $\mathcal{C}_2$ , the expression  $\mathcal{C}_1 < \mathcal{C}_2$  means that either  $|\mathcal{C}_1| < |\mathcal{C}_2|$ , or  $|\mathcal{C}_1| = |\mathcal{C}_2|$  and  $\exists i$  representing the first entry where they differ such that the  $i$ -th entry of  $\mathcal{C}_2$  is greater than the  $i$ -th entry of  $\mathcal{C}_1$ . Then, for a set of tuples  $\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_k\}$ , we say that  $\mathcal{C}$  is in lexicographical order if  $\mathcal{C}_i < \mathcal{C}_j$  for all  $i < j$ . Finally, for a matrix  $\mathbf{Y} \in \mathbb{R}^{[n]^{\leq \ell} \times [n]^{\leq \ell}}$ , we index its rows and columns by using tuples in  $[n]^{\leq \ell}$  ordered lexicographically, e.g., for  $\mathbf{Y} \in \mathbb{R}^{[5]^{\leq 3} \times [5]^{\leq 3}}$  we have that  $\mathbf{Y}_{(1,1,2),(5)}$  corresponds to the entry at row  $(1, 1, 2)$  and column  $(5)$ .

It is convenient to rewrite the objective of problem (2) as a polynomial optimization problem, i.e.,  $\sum_i \sum_j X_{i,j} y_i y_j$ , so that the standard machinery of SoS optimization (Lasserre, 2001; Parrilo, 2000; Laurent, 2009) can be applied to formulate the degree- $d$  relaxation. Then, for an even number  $d$ , the degree- $d$  (or level  $d/2$ ) SoS relaxation of problem (2) takes the form

$$\max_{\mathbf{Y} \in \mathbb{R}^{[n]^{\leq \frac{d}{2}} \times [n]^{\leq \frac{d}{2}}}} \sum_{i=1}^n \sum_{j=1}^n X_{i,j} Y_{(i),(j)}, \quad (4)$$

subject to  $\mathbf{Y} \succeq 0$ ,

$$\mathbf{Y}_{(\emptyset)(\emptyset)} = 1,$$

$$\mathbf{Y}_{(i)+\mathcal{C}_1, (i)+\mathcal{C}_2} = \mathbf{Y}_{\mathcal{C}_1, \mathcal{C}_2}, \quad \forall i \in [n], |\mathcal{C}_1|, |\mathcal{C}_2| \leq d/2 - 1,$$

$$\mathbf{Y}_{\mathcal{C}_1, \mathcal{C}_2} = \mathbf{Y}_{\mathcal{C}'_1, \mathcal{C}'_2},$$

$$\forall \psi(\mathcal{C}_1 + \mathcal{C}_2) = \psi(\mathcal{C}'_1 + \mathcal{C}'_2), |\mathcal{C}_1|, |\mathcal{C}_2|, |\mathcal{C}'_1|, |\mathcal{C}'_2| \leq d/2.$$

In the problem above, each entry of the matrix  $\mathbf{Y}$  corresponds to a reparametrization that takes the form

$\mathbf{Y}_{\mathcal{C}_1, \mathcal{C}_2} = \prod_{i \in \mathcal{C}_1} y_i \prod_{j \in \mathcal{C}_2} y_j = \prod_{i \in \mathcal{C}_1 + \mathcal{C}_2} y_i$ , which is also known as a pseudomoment matrix (Lasserre, 2001; Laurent, 2009). In problem (4), the second constraint can be thought as a normalization constraint. The third list of constraints corresponds to  $\prod_{j \in \mathcal{C}} y_j \cdot y_i^2 = \prod_{j \in \mathcal{C}} y_j, \forall |\mathcal{C}| \leq d - 2$ , which is equivalent to  $y_i = \pm 1$  in problem (2). Finally, the last list of constraints corresponds to  $\prod_{j \in \mathcal{C}_1 + \mathcal{C}_2} y_j = \prod_{j \in \mathcal{C}'_1 + \mathcal{C}'_2} y_j, \forall \psi(\mathcal{C}_1 + \mathcal{C}_2) = \psi(\mathcal{C}'_1 + \mathcal{C}'_2)$ , and  $|\mathcal{C}_1|, |\mathcal{C}_2|, |\mathcal{C}'_1|, |\mathcal{C}'_2| \leq d/2$ , which states that  $\mathbf{Y}_{\mathcal{C}_1, \mathcal{C}_2}$  should be invariant to all permutations of the tuple  $\mathcal{C}_1 + \mathcal{C}_2$ . One can note that, for  $d = 2$ , the degree-2 (or level 1) SoS relaxation is equivalent to the SDP relaxation in eq.(3). It is clear that for a larger  $d$ , the degree- $d$  SoS relaxation gives a tighter convex relaxation of problem (2). While one can solve problem (4) to a fixed accuracy using general-purpose SDP algorithms in polynomial time in  $n$ , the computational complexity will be of order  $n^{\mathcal{O}(d)}$ . Thus, it is important that  $d$  be of low order.

In the next section, we center our attention to the degree-4 SoS relaxation and in understanding how it can help improving the exact recovery rate with respect to the SDP (or degree-2 SoS) relaxation.

### 3 ON EXACT RECOVERY FROM THE DEGREE-4 SOS HIERARCHY

As the focus of this section will be on the degree-4 SoS relaxation, we start by formulating the corresponding optimization problem. In problem (4), for  $d = 4$ , the matrix  $\mathbf{Y}$  is in  $\mathbb{R}^{[n]^{\leq 2} \times [n]^{\leq 2}}$ , that is,  $\mathbf{Y}$  is a matrix of dimension  $(1 + n + n^2) \times (1 + n + n^2)$ . Bandeira and Kunisky (2018, Appendix A) showed that one can write an equivalent formulation by using only the principal submatrix of  $\mathbf{Y}$  indexed by  $[n]^2 \times [n]^2$  (i.e., a matrix of dimension  $n^2 \times n^2$ ). The reduced formulation takes the form:

$$\begin{aligned} \max_{\mathbf{Y} \in \mathbb{R}^{[n]^2 \times [n]^2}} & \sum_{i=1}^n \sum_{j=1}^n X_{i,j} Y_{(1,1), (i,j)}, \\ \text{subject to } & \mathbf{Y} \succeq 0, \\ & \mathbf{Y}_{(i,i)(j,j)} = 1, \quad \forall i, j \in [n], \\ & \mathbf{Y}_{(i,i)(j,k)} = \mathbf{Y}_{(i',i')(j,k)}, \quad \forall i, i', j, k \in [n], \\ & \mathbf{Y}_{(i,j)(k,\ell)} = \mathbf{Y}_{(\pi_1, \pi_2)(\pi_3, \pi_4)}, \\ & \quad \forall i, j, k, \ell \in [n], \pi \in \Pi(i, j, k, \ell), \end{aligned} \quad (5)$$

where  $\Pi(i, j, k, \ell)$  is the set of all permutations of  $(i, j, k, \ell)$ . We will go one step further in the reduction and show that one can indeed cast an equivalent formulation to problem (5) by using only the principal submatrix of  $\mathbf{Y} \in \mathbb{R}^{[n]^2 \times [n]^2}$  indexed by  $\binom{[n]}{2} \times \binom{[n]}{2}$ , i.e., a matrix of dimension  $\frac{n(n-1)}{2} \times \frac{n(n-1)}{2}$ . Here, it will be more

convenient to use sets instead of tuples for indexing the rows and columns of  $\mathbf{Y}$ , where  $\binom{[n]}{2}$  denotes the set of all unordered combinations of length 2 from the numbers in  $[n]$ , e.g.,  $\binom{[3]}{2} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ . For further distinction against the matrix  $\mathbf{Y} \in \mathbb{R}^{[n]^2 \times [n]^2}$ , we will use  $\tilde{\mathbf{Y}}$  to denote the matrix indexed by  $\binom{[n]}{2} \times \binom{[n]}{2}$ .

We will also make use of the next set of definitions, which are important for stating our results.

**Definition 1** (The level-2 vector). *For any vector  $\mathbf{v} \in \mathbb{R}^n$ , its level-2 vector, denoted by  $\mathbf{v}^{(2)} \in \mathbb{R}^{\binom{[n]}{2}}$  and indexed by  $\binom{[n]}{2}$ , is defined as  $v_{\{i,j\}}^{(2)} = v_i v_j$ .*

We also define the level-2 version of a graph as follows.

**Definition 2** (The level-2 graph). *Let  $G = (V, E)$ , where  $V = [n]$ , be any undirected graph of  $n$  nodes with adjacency matrix  $\mathbf{A} \in \{0, 1\}^{n \times n}$ . The level-2 graph of  $G$ , denoted by  $G^{(2)} = (\binom{[n]}{2}, E^{(2)})$  and with adjacency matrix  $\mathbf{A}^{(2)} \in \{0, 1\}^{\binom{[n]}{2} \times \binom{[n]}{2}}$ , has its adjacency matrix defined as  $A_{\{i,k\}, \{j,\ell\}}^{(2)} = 1$  if  $(i, j) \in E$  for all  $i < j < k \in [n]$ , and  $A_{\{i,j\}, \{k,\ell\}}^{(2)} = 0$  for all  $i < j < k < \ell \in [n]$ .*

The next type of graphs have been studied for several years within the graph theory community and we will later show how they relate to the solution of the level-2 SoS relaxation.

**Definition 3** (Johnson graph (Holton and Sheehan, 1993)). *For a set  $[n]$ , the Johnson graph  $\mathcal{J}(n, k)$  has all the  $k$ -element subsets of  $[n]$  as vertices, and two vertices are adjacent if and only if the intersection of the two vertices (subsets) contains  $(k - 1)$ -elements.*

**Definition 4** (Kneser graph (Lovász, 1978)). *For a set  $[n]$ , the Kneser graph  $\mathcal{K}(n, k)$  has all the  $k$ -element subsets of  $[n]$  as vertices, and two vertices are adjacent if and only if the two vertices (subsets) are disjoint.*

From Definitions 3 and 4, we are interested in  $\mathcal{J}(n, 2)$  and  $\mathcal{K}(n, 2)$ , where we first note that  $\mathcal{K}(n, 2)$  is the complement of  $\mathcal{J}(n, 2)$ . We also note that for a graph  $G$  of  $n$  nodes, by construction, the level-2 graph of  $G$  is always a subgraph of the Johnson graph  $\mathcal{J}(n, 2)$ , and is equal to  $\mathcal{J}(n, 2)$  if and only if  $G$  is the complete graph of  $n$  nodes. Finally, since our observation matrix  $\mathbf{X}$  depends on a graph  $G$ , one can also extend  $\mathbf{X}$  to a matrix in  $\{-1, 0, +1\}^{\binom{[n]}{2} \times \binom{[n]}{2}}$ . We will use  $\mathbf{X}^{(2)}$  to denote the level-2 version of  $\mathbf{X}$ . Specifically,  $X_{\{i,k\}, \{j,\ell\}}^{(2)} = X_{i,j}$  for all  $i < j < k \in [n]$ , and  $X_{\{i,j\}, \{k,\ell\}}^{(2)} = 0$  for all  $i < j < k < \ell \in [n]$ . For further clarity, we illustrate the level-2 construction of  $\mathbf{X}$  in Figure 2, where the input graph is a 2 by 2 grid.

Next, we present an optimization problem that is equivalent to problem (5) but in terms of the level-2 constructions defined above. For notational convenience, we

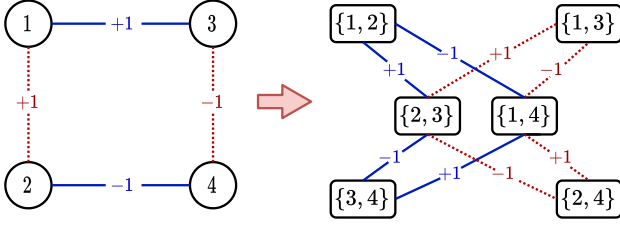


Figure 2: Illustration of the level-2 construction of  $\mathbf{X}$ . The edge values in the grid graph correspond to the observation  $\mathbf{X}$ , while the edge values on the right graph correspond to level-2 matrix  $\mathbf{X}^{(2)}$ . The solid blue and dotted red lines indicate that the observation is correct and corrupted, respectively.

will use  $\mathbf{S}_{-(stuv)}^{+(ijk\ell)}$  to denote a sparse *symmetric* matrix such that the only non-zero entries are  $\mathbf{S}_{\{i,j\},\{k,\ell\}} = 1$ , and  $\mathbf{S}_{\{s,t\},\{u,v\}} = -1$ .

$$\tilde{\mathbf{Y}} \in \mathbb{R}^{\binom{[n]}{2} \times \binom{[n]}{2}} \quad \frac{1}{n-2} \langle \mathbf{X}^{(2)}, \tilde{\mathbf{Y}} \rangle, \quad (6)$$

subject to  $\tilde{\mathbf{Y}} \succeq 0$ ,

$$\tilde{\mathbf{Y}}_{\mathcal{C},\mathcal{C}} = 1, \forall \mathcal{C} \in \binom{[n]}{2},$$

$$\langle \mathbf{S}_{-(ik'k'j)}^{+(ikkj)}, \tilde{\mathbf{Y}} \rangle = 0, \forall i < j < k < k',$$

$$\langle \mathbf{S}_{-\pi(ijkl)}^{+(ijk\ell)}, \tilde{\mathbf{Y}} \rangle = 0,$$

$$\forall i < j < k < l \in [n], \pi \in \Pi(i, j, k, l).$$

**Proposition 1.** *Problem (6) is equivalent to problem (5).*

All proofs are detailed in Appendix A.

**Remark 1.** *Let  $\bar{\mathbf{y}}^{(2)}$  be the level-2 vector of the ground-truth labeling  $\bar{\mathbf{y}}$ , and let  $\tilde{\mathbf{Y}}^*$  be the optimizer of problem (6). Then, we say that exact recovery is realized if  $\tilde{\mathbf{Y}}^* = \bar{\mathbf{y}}^{(2)}\bar{\mathbf{y}}^{(2)\top}$ .*

### 3.1 The Dual Problem

A key ingredient for our analysis is the dual formulation of problem (6), which takes the following form

$$\min_{\tilde{\mathbf{V}}, \boldsymbol{\mu}} \text{Tr}(\tilde{\mathbf{V}}), \quad (7)$$

subject to  $\tilde{\mathbf{V}}$  is diagonal,

$$\tilde{\boldsymbol{\Lambda}} = \tilde{\mathbf{V}} - \frac{\mathbf{X}^{(2)}}{n-2} + \left( \sum_{i < j < k < k'} \mu_{ik'k'j}^{ikkj} \mathbf{S}_{-(ik'k'j)}^{+(ikkj)} + \sum_{\substack{i < j < k < \ell \\ \pi \in \Pi(i, j, k, \ell)}} \mu_{\pi(ijkl)}^{ijk\ell} \mathbf{S}_{-\pi(ijkl)}^{+(ijk\ell)} \right) \succeq 0,$$

where  $\tilde{\mathbf{V}} \in \mathbb{R}^{\binom{[n]}{2} \times \binom{[n]}{2}}$ ,  $\mu_{ik'k'j}^{ikkj} \in \mathbb{R}$  and  $\mu_{\pi(ijkl)}^{ijk\ell} \in \mathbb{R}$  are the dual variables of the second constraint, and the third and fourth list of constraints from the primal formulation (6), respectively. The dual variable  $\boldsymbol{\mu}$  denotes all the scalars  $\mu_{ik'k'j}^{ikkj}$  and  $\mu_{\pi(ijkl)}^{ijk\ell}$ .

We have that if there exists  $\tilde{\mathbf{Y}}, \tilde{\mathbf{V}}, \boldsymbol{\mu}$  that satisfy the Karush-Kuhn-Tucker (KKT) conditions (Boyd and Vandenberghe, 2004), then  $\tilde{\mathbf{Y}}$  and  $\tilde{\mathbf{V}}, \boldsymbol{\mu}$  are primal and dual optimal, and strong duality holds in this case. Let  $\bar{\mathbf{y}}^{(2)}$  be the level-2 vector of the *ground-truth* labeling  $\bar{\mathbf{y}}$ . Since we are interested in exact recovery, we will consider the solution  $\tilde{\mathbf{Y}} = \bar{\mathbf{y}}^{(2)}\bar{\mathbf{y}}^{(2)\top}$  for the rest of our analysis, where it is clear that such setting satisfies the primal constraints. Let

$$\tilde{\mathbf{V}} = \frac{\text{diag}(\mathbf{X}^{(2)}\tilde{\mathbf{Y}})}{n-2} - \left( \text{diag} \left( \sum_{i < j < k < k'} \mu_{ik'k'j}^{ikkj} \mathbf{S}_{-(ik'k'j)}^{+(ikkj)} \tilde{\mathbf{Y}} \right) - \text{diag} \left( \sum_{\substack{i < j < k < \ell \\ \pi \in \Pi(ijkl)}} \mu_{\pi(ijkl)}^{ijk\ell} \mathbf{S}_{-\pi(ijkl)}^{+(ijk\ell)} \tilde{\mathbf{Y}} \right) \right), \quad (8)$$

where, for a matrix  $\mathbf{M}$ ,  $\text{diag}(\mathbf{M})$  denotes the diagonal matrix formed from the diagonal entries of  $\mathbf{M}$ . Complementary slackness and stationarity require the trace of  $\tilde{\mathbf{V}}$  to be equal to the trace of the r.h.s. of eq.(8), which is clearly satisfied by construction. Thus, if we find an assignment of  $\boldsymbol{\mu}$  such that  $\tilde{\boldsymbol{\Lambda}} \succeq 0$ , we would have an optimal solution since all KKT conditions are fulfilled. Nevertheless, we are also interested in  $\tilde{\mathbf{Y}} = \bar{\mathbf{y}}^{(2)}\bar{\mathbf{y}}^{(2)\top}$  being the *unique* optimal solution, where we note that having  $\lambda_2(\tilde{\boldsymbol{\Lambda}}) > 0$  suffices to guarantee a unique solution. The argument follows from the fact that, by the setting of eq.(8), we have  $\tilde{\boldsymbol{\Lambda}}\bar{\mathbf{y}}^{(2)} = 0$ . Thus, if  $\lambda_2(\tilde{\boldsymbol{\Lambda}}) > 0$  then  $\bar{\mathbf{y}}^{(2)}$  spans all of the null-space of  $\tilde{\boldsymbol{\Lambda}}$ . Combined with the KKT conditions, we have that  $\tilde{\mathbf{Y}}$  should be a multiple of  $\bar{\mathbf{y}}^{(2)}\bar{\mathbf{y}}^{(2)\top}$ . Since  $\tilde{\mathbf{Y}}$  has diagonal entries equal to 1, we must have that  $\tilde{\mathbf{Y}} = \bar{\mathbf{y}}^{(2)}\bar{\mathbf{y}}^{(2)\top}$ .

Putting all pieces together, we have that under eq.(8), if for some  $\boldsymbol{\mu}$  we have that  $\tilde{\boldsymbol{\Lambda}} \succeq 0$  and  $\lambda_2(\tilde{\boldsymbol{\Lambda}}) > 0$ , then the optimizer of problem (6) is  $\bar{\mathbf{y}}^{(2)}\bar{\mathbf{y}}^{(2)\top}$ , i.e., we obtain exact recovery. Since  $\bar{\mathbf{y}}^{(2)}$  is an eigenvector of  $\tilde{\boldsymbol{\Lambda}}$  with eigenvalue zero, we focus on controlling the quantity  $\lambda_2(\tilde{\boldsymbol{\Lambda}}) = \min_{\mathbf{v} \perp \bar{\mathbf{y}}^{(2)}} \frac{\mathbf{v}^\top \tilde{\boldsymbol{\Lambda}} \mathbf{v}}{\mathbf{v}^\top \mathbf{v}}$ .<sup>1</sup> Also, as  $\tilde{\boldsymbol{\Lambda}}$  depends on the noisy observation  $\mathbf{X}^{(2)}$ , we have that  $\tilde{\boldsymbol{\Lambda}}$  is a random quantity. Then, by using Weyl's theorem on eigenvalues, we have

$$\begin{aligned} \lambda_2(\tilde{\boldsymbol{\Lambda}}) &= \lambda_2(\tilde{\boldsymbol{\Lambda}} - \mathbb{E}[\tilde{\boldsymbol{\Lambda}}] + \mathbb{E}[\tilde{\boldsymbol{\Lambda}}]) \\ &\geq \lambda_2(\mathbb{E}[\tilde{\boldsymbol{\Lambda}}]) + \lambda_1(\tilde{\boldsymbol{\Lambda}} - \mathbb{E}[\tilde{\boldsymbol{\Lambda}}]). \end{aligned} \quad (9)$$

<sup>1</sup>This expression comes from the variational characterization of eigenvalues.

In eq.(9), let  $t$  be a lower bound to  $\lambda_2(\mathbb{E}[\tilde{\Lambda}])$ , i.e.,  $\lambda_2(\mathbb{E}[\tilde{\Lambda}]) \geq t$ . Then, the second summand can be lower bounded by using matrix concentration inequalities. Specifically, by using matrix Bernstein inequality (Tropp, 2012), one can obtain that  $\mathbb{P}[\lambda_1(\tilde{\Lambda} - \mathbb{E}[\tilde{\Lambda}]) \leq -t] \leq \mathcal{O}(n^2 e^{-t})$ . Thus, we can now focus on the first summand, which will be lower bounded by a novel Cheeger-type inequality. In the next subsections, we look at the expected value of  $\tilde{\Lambda}$  in more detail.

### 3.2 The Relation between $\mathbb{E}[\tilde{\Lambda}]$ and the Algebraic Connectivity of the Level-2 Graph

In this section, we will show how  $\mathbb{E}[\tilde{\Lambda}]$  is related to the Laplacian matrix of  $G^{(2)}$  (the level-2 version of  $G$ ). To do so, we will use the following definitions and notation.

For a *signed* weighted graph  $H = (U, F)$ , we use  $\mathbf{W}^H$  to denote its weight matrix, that is, the entry  $W_{i,j}^H \in \mathbb{R}$  is the weight of edge  $(i, j) \in F$  and is zero if  $(i, j) \notin F$ . For any set  $T \subset U$ , its boundary is defined as  $\partial T = \{(i, j) \mid i \in T \text{ and } j \notin T\}$ ; while its boundary weight is defined as  $\omega(\partial T) = \sum_{i \in T, j \notin T} W_{i,j}^H$ . The number of nodes in  $T$  is denoted by  $|T|$ . The degree of a node is defined as  $\deg(i) = \sum_{j \neq i} W_{i,j}^H$ .

**Definition 5.** Let  $H$  be a graph with degree matrix  $\mathbf{D}^H$  and weight matrix  $\mathbf{W}^H$ , where  $\mathbf{D}^H$  is a diagonal matrix such that  $D_{i,i} = \deg(i)$ . The Laplacian matrix of  $H$  is defined as  $\mathbf{L}^H = \mathbf{D}^H - \mathbf{W}^H$ .

**Definition 6** (Cheeger constant (Cheeger, 1969)). For a graph  $H = (U, F)$  of  $n$  nodes, its Cheeger constant is defined as  $\phi(H) = \min_{T \subset U, |T| \leq n/2} \frac{\omega(\partial T)}{|T|}$ .

**Remark 2.** For unweighted graphs, the definitions above match the standard definitions for node degree, boundary of a set, and Laplacian matrix.

Next, we analyze the scenario where all the scalar dual variables in  $\boldsymbol{\mu}$  are zero, we defer the case when they are not for the next subsection.

**The  $\boldsymbol{\mu} = \mathbf{0}$  scenario.** From eq.(8) we have that  $\tilde{\mathbf{V}} = \text{diag}(\mathbf{X}^{(2)} \tilde{\mathbf{Y}}) / (n - 2)$ . Hence, for all  $i < j \in [n]$ , we have  $\mathbb{E}[\tilde{V}_{\{i,j\},\{i,j\}}] = (1-2p)/(n-2) \cdot \deg(\{i, j\})$ . In addition, we have  $\mathbb{E}[X_{\{i,k\},\{k,j\}}^{(2)}] = (1 - 2p) \bar{y}_i \bar{y}_j \mathbb{1}[(i, j) \in E]$ , for all  $i < j < k \in [n]$ .<sup>2</sup> Finally, since  $\boldsymbol{\mu} = \mathbf{0}$ , we have  $\tilde{\Lambda} = \tilde{\mathbf{V}} - \frac{\mathbf{X}^{(2)}}{n-2}$ . Therefore,

$$\mathbb{E}[\tilde{\Lambda}] = \frac{1 - 2p}{n - 2} \tilde{\mathbf{Y}} \mathbf{L}^{G^{(2)}} \tilde{\mathbf{Y}}, \quad (10)$$

<sup>2</sup>Recall that if  $(i, j) \in E$ , then  $X_{i,j} = -\bar{y}_i \bar{y}_j$  with probability  $p$ , and  $X_{i,j} = \bar{y}_i \bar{y}_j$  otherwise. If  $(i, j) \notin E$  then  $X_{i,j} = 0$ .

where  $\tilde{\mathbf{Y}}$  is a diagonal matrix with entries equal to the entries in  $\bar{\mathbf{y}}^{(2)}$ . Recall that  $\bar{y}_{\{i,j\}}^{(2)} = \bar{y}_i \bar{y}_j$  and  $\bar{y}_i \in \{\pm 1\}$  for all  $i \in [n]$ . Then, we have that  $\tilde{\mathbf{Y}}^{-1} = \tilde{\mathbf{Y}}$  and, thus, the matrix  $\mathbb{E}[\tilde{\Lambda}]$  and  $\frac{1-2p}{n-2} \mathbf{L}^{G^{(2)}}$  are similar. The latter means that both matrices share the same spectrum, i.e.,

$$\lambda_2(\mathbb{E}[\tilde{\Lambda}]) = \frac{1 - 2p}{n - 2} \lambda_2(\mathbf{L}^{G^{(2)}}). \quad (11)$$

Notice that the level-2 graph  $G^{(2)}$  is unweighted since  $G$  is unweighted. That implies that one can lower bound  $\lambda_2(\mathbb{E}[\tilde{\Lambda}])$  by using existing lower bounds for the second eigenvalue<sup>3</sup> of the Laplacian matrix of  $G^{(2)}$ . In particular, one can have (Mohar, 1991)

$$\begin{aligned} \lambda_2(\mathbb{E}[\tilde{\Lambda}]) &= \frac{1 - 2p}{n - 2} \lambda_2(\mathbf{L}^{G^{(2)}}) \\ &\geq \frac{(1 - 2p) \phi(G^{(2)})^2}{2(n - 2) \deg_{\max}^{G^{(2)}}}. \end{aligned} \quad (12)$$

Finally, we note that considering  $\boldsymbol{\mu} = \mathbf{0}$  is equivalent to not having the third and fourth list of constraints in problem (6). At this point, the reader might wonder if, setting  $\boldsymbol{\mu} = \mathbf{0}$  and solving problem (6) yields in any better chances of exact recovery than solving problem (3). We answer the latter in the negative.

**Proposition 2.** Without the third and fourth list of constraints, problem (6) does not improve exact recoverability with respect to problem (3).

The purpose of Proposition 2 is to highlight the role that a  $\boldsymbol{\mu} \neq \mathbf{0}$  will play in showing the improvement in exact recoverability of the degree-4 SoS relaxation with respect to the SDP relaxation, which is discussed next.

### 3.3 Connections to Systems of Sets and a Novel Cheeger-Type Lower Bound

We start this section by showing how the third and fourth list of constraints of problem (6) relate to finding edge weights of the Johnson and Kneser graphs, respectively, so that the Laplacian matrix of a *new* graph is positive semidefinite (PSD).

Note that the third and fourth list of constraints in the SoS relaxation (6) do not depend on the input graph, nor on the edge observations or the ground-truth node labels. Instead, they are constraints coming from the SoS relaxation, as explained in the subsequent paragraphs to problem (4). That means that they depend only on the number of nodes,  $n$ , and on the degree of the relaxation,  $d = 4$ . We will illustrate in

<sup>3</sup>The second eigenvalue of the Laplacian matrix is also known as the algebraic connectivity.



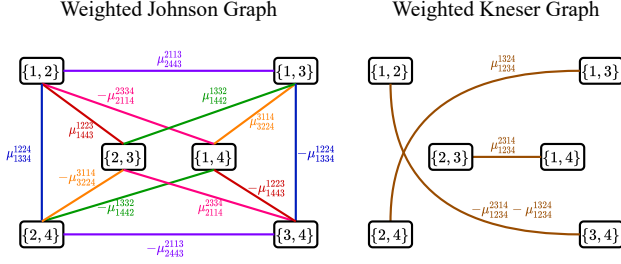


Figure 3: Johnson and Kneser graphs for  $n = 4$ , where each edge weight is related to some dual variables from the SoS constraints. Edge weights with the same color sum to zero, see eq.(13).

detail the case of  $n = 4$  as it is easier to generalize from there to any value of  $n$ .

Recall that  $\mathbf{S}_{-(stuv)}^{+(ijkl)}$  is a symmetric matrix that has non-zero entries  $\mathbf{S}_{\{i,j\},\{k,\ell\}} = 1$ , and  $\mathbf{S}_{\{s,t\},\{u,v\}} = -1$ . By taking advantage of the implicit symmetry constraint from  $\tilde{\mathbf{Y}} \succeq 0$ , for  $n = 4$ , one can realize that the third list of constraints in problem (6) has six different constraints in total (with their respective dual variables), which are:

$$\begin{aligned} \mu_{1224}^{1224} : \langle \mathbf{S}_{-(1224)}^{+(1224)}, \tilde{\mathbf{Y}} \rangle &= 0, & \mu_{2113}^{2113} : \langle \mathbf{S}_{-(2113)}^{+(2113)}, \tilde{\mathbf{Y}} \rangle &= 0, \\ \mu_{1332}^{1332} : \langle \mathbf{S}_{-(1332)}^{+(1332)}, \tilde{\mathbf{Y}} \rangle &= 0, & \mu_{3114}^{3114} : \langle \mathbf{S}_{-(3114)}^{+(3114)}, \tilde{\mathbf{Y}} \rangle &= 0, \\ \mu_{1443}^{1223} : \langle \mathbf{S}_{-(1443)}^{+(1223)}, \tilde{\mathbf{Y}} \rangle &= 0, & \mu_{2334}^{2334} : \langle \mathbf{S}_{-(2334)}^{+(2334)}, \tilde{\mathbf{Y}} \rangle &= 0. \end{aligned}$$

Similarly, from the fourth list of constraints we have:

$$\mu_{1234}^{1324} : \langle \mathbf{S}_{-(1234)}^{+(1324)}, \tilde{\mathbf{Y}} \rangle = 0, \quad \mu_{1234}^{2314} : \langle \mathbf{S}_{-(1234)}^{+(2314)}, \tilde{\mathbf{Y}} \rangle = 0.$$

In the dual formulation (7), for both lists above, the matrices  $\mathbf{S}$  are weighted by the dual variables  $\mu$ . Then, the two weighted summations can be thought of as weight matrices of some graphs. Interestingly, such graphs happen to be the Johnson and Kneser graphs<sup>4</sup> for the first and second list of constraints above, respectively. In Figure 3, we show an illustration of the Johnson and Kneser graphs with edge weights corresponding to the dual variables.

Let  $\oplus$  denote the symmetric difference of sets. Also, let  $\mathbf{W}^{\mathcal{J}}$  and  $\mathbf{W}^{\mathcal{K}}$  denote the weight matrices of the Johnson and Kneser graphs, respectively. Then, for any  $n$ , the third and fourth list of constraints of problem (6) translate to having the following constraints on  $\mathbf{W}^{\mathcal{J}}$  and  $\mathbf{W}^{\mathcal{K}}$ ,

$$\sum_{\substack{C_1, C_2 \\ C_1 \oplus C_2 = \{i, j\}}} \mathbf{W}_{C_1, C_2}^{\mathcal{J}} = 0, \quad \forall i < j \in [n],$$

<sup>4</sup>For any  $n$ , whenever we write the Johnson and Kneser graphs, we refer to  $\mathcal{J}(n, 2)$  and  $\mathcal{K}(n, 2)$ , respectively.

$$\sum_{\substack{C_1, C_2 \\ C_1 \oplus C_2 = \{i, j, k, \ell\}}} \mathbf{W}_{C_1, C_2}^{\mathcal{K}} = 0, \quad \forall i < j < k < \ell \in [n]. \quad (13)$$

Thus, by using the construction in eq.(8), we have that the PSD constraint of the dual formulation (7) can be rewritten in terms of  $\mathbf{W}^{\mathcal{J}}$  and  $\mathbf{W}^{\mathcal{K}}$  as follows,

$$\begin{aligned} \tilde{\mathbf{\Lambda}} &= \frac{\text{diag}(\mathbf{X}^{(2)} \tilde{\mathbf{Y}})}{n-2} - \frac{\mathbf{X}^{(2)}}{n-2} + \left( \left( \text{diag}(\mathbf{W}^{\mathcal{J}} \tilde{\mathbf{Y}}) - \mathbf{W}^{\mathcal{J}} \right) \right. \\ &\quad \left. + \left( \text{diag}(\mathbf{W}^{\mathcal{K}} \tilde{\mathbf{Y}}) - \mathbf{W}^{\mathcal{K}} \right) \right) \succeq 0. \end{aligned}$$

Let  $\tilde{\mathcal{G}} = G^{(2)} \cup \mathcal{J} \cup \mathcal{K}$  such that  $\mathbf{W}^{\tilde{\mathcal{G}}} = \frac{1-2p}{n-2} \mathbf{W}^{G^{(2)}} + \mathbf{W}^{\mathcal{J}} + \mathbf{W}^{\mathcal{K}}$ , and noting that w.l.o.g. one can multiply the weights in eq.(13) by  $\bar{y}_i \bar{y}_j$  and  $\bar{y}_i \bar{y}_j \bar{y}_k \bar{y}_\ell$ , respectively. We can use a similar argument to that of eq.(11) and obtain

$$\lambda_2(\mathbb{E}[\tilde{\mathbf{\Lambda}}]) = \lambda_2(\mathbf{L}^{\tilde{\mathcal{G}}}). \quad (14)$$

The subtlety for lower bounding eq.(14) is that, unless all edge weights are zero, the Johnson and Kneser graphs will both have at least one negative edge weight in order to fulfill eq.(13). In other words, the Laplacian matrix  $\mathbf{L}^{\tilde{\mathcal{G}}}$  is no longer guaranteed to be PSD. That fact alone rules out almost all existing results on lower bounding the algebraic connectivity as it is mostly assumed that all edge weights are positive. Among the few works that study the Laplacian matrix with negative weights, one can find (Zelazo and Bürger, 2014; Chen et al., 2016); however, their results focus on finding conditions for positive semidefiniteness of the Laplacian matrix in the context of electrical circuits and not in finding a lower bound. Our next result, generalizes the lower bound in (Mohar, 1991) by considering negative edge weights.

**Theorem 1.** *Let  $H = H^+ \cup H^-$  be a weighted graph such that  $H^+$  and  $H^-$  denote the disjoint subgraphs of  $H$  with positive and negative weights, respectively. Also, let  $\text{deg}_{\max}^{H^+}$  denote the maximum node degree of  $H^+$ . Then, we have that  $\lambda_2(\mathbf{L}^H) \geq \frac{\phi(H^+)^2}{2 \text{deg}_{\max}^{H^+}} + 2 \cdot \text{mincut}(H^-)$ .*

In Appendix B, we provide further discussion about Theorem 1. In the case when there are positive weights only, the theorem above yields the typical Cheeger bound (Mohar, 1991). When there is at least one negative weight, the bound shows an interesting trade-off between the Cheeger constant of the positive subgraph and the minimum cut of the negative subgraph. By applying Theorem 1 to eq.(14), we obtain

$$\lambda_2(\mathbb{E}[\tilde{\mathbf{\Lambda}}]) \geq \frac{\phi(\tilde{\mathcal{G}}^+)^2}{(2 \text{deg}_{\max}^{\tilde{\mathcal{G}}^+})} + 2 \cdot \text{mincut}(\tilde{\mathcal{G}}^-). \quad (15)$$

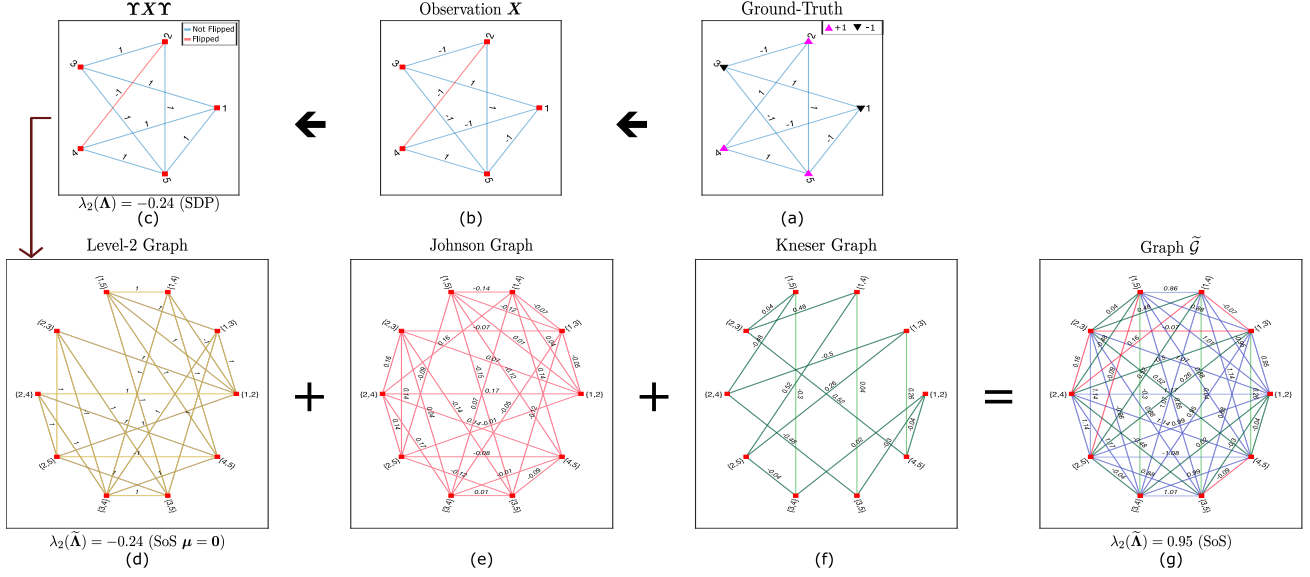


Figure 4: Detailed example of how the level-2 SoS relaxation improves the algebraic connectivity of the input graph through a combination of weights of its level-2 version, and the Johnson and Kneser graphs. In the final graph  $\tilde{\mathcal{G}}$ , green and red lines indicate that their weights remain unchanged w.r.t. the Kneser and Johnson edge weights, respectively; while blue lines indicate that their weights resulted from the summation of weights from the Level-2 and Johnson graphs.

Without the weights of the Johnson and Kneser graphs, the lower bound above is equal to that of eq.(12). Also, recall that, by construction, the edge set of the level-2 graph  $G^{(2)}$  is a subset of the edge set of the Johnson graph, and that the Kneser graph is the complement of the Johnson graph. That means that  $\tilde{\mathcal{G}}$  will be a complete graph of  $\binom{n}{2}$  vertices, where the edge weights of the Kneser graph are exclusively related to the dual variables  $\mu$ , while the edge weights of the Johnson graph might have an interaction between the noisy edge observations and the dual variables  $\mu$ . Finally, we provide the following result that compares the probabilities of exact recoverability of SDP (eq.(3)) vs SoS (eq.(6)).

**Theorem 2.** *Let  $G$  be any undirected connected graph with  $n$  nodes, whose level-2 lift is denoted by  $G^{(2)}$ . Let also  $\tilde{\mathcal{G}}$  be the union of  $G^{(2)}$  with the Johnson ( $\mathcal{J}$ ) and Kneser ( $\mathcal{K}$ ) graphs of the SoS dual variables  $\mu$ . Then, SDP (eq.(3)) and SoS (eq.(6)) recover exactly with probabilities of at least  $1 - \epsilon_{\text{sdp}}$  and  $1 - \epsilon_{\text{sos}}$ , respectively, where*

$$\epsilon_{\text{sdp}} = \mathcal{O}\left(n^2 \exp\left(-\frac{1-2p}{n-2} \frac{\phi(G^{(2)})^2}{2 \deg_{\max}^{G^{(2)}}}\right)\right),$$

$$\epsilon_{\text{sos}} = \mathcal{O}\left(n^2 \exp\left(-\frac{\phi(\tilde{\mathcal{G}}^+)^2}{2 \deg_{\max}^{\tilde{\mathcal{G}}^+}} + 2 \text{mincut}(\tilde{\mathcal{G}}^-)\right)\right),$$

$$\text{and } \mathbf{W}^{\tilde{\mathcal{G}}} = \frac{1-2p}{n-2} \mathbf{W}^{G^{(2)}} + \mathbf{W}^{\mathcal{J}} + \mathbf{W}^{\mathcal{K}}.$$

We note two things. First, from Proposition 2, we

know that when  $\mu = 0$ , SoS is equivalent to SDP, which Theorem 2 above reflects clearly. Moreover, that implies that SoS can never perform worse than SDP. Second, our bound shed lights on the trade-off between the Cheeger constant and the min-cut of the positive and negative subgraphs of  $\tilde{\mathcal{G}}$ , which is a *new insight* to the best of our knowledge. Next, for further clarity, we provide a detailed example of our analysis presented in this section.

## 4 EXAMPLE

The goal of this section is to provide a concrete example where the SoS relaxation (5) achieves exact recovery but the SDP relaxation (3) does not. Since for any input graph with  $n$  vertices, its level-2 version has  $\binom{n}{2}$  vertices, we select a value of  $n = 5$  so that the level-2 graph has 10 nodes and the plots can still be visually inspected in detail.

Figure (4a) shows the ground-truth labels of a graph with 5 nodes and 8 edges. Figure (4b) corresponds to the observation matrix  $\mathbf{X}$ . In this case, only one edge is corrupted (the red edge). Figure (4c) shows the graph where an edge label of  $-1$  or  $1$  indicates whether the observed edge value was corrupted or not, respectively. The latter graph is obtained by  $\Upsilon \mathbf{X} \Upsilon$ , where  $\Upsilon$  denotes a diagonal matrix with entries from  $\tilde{\mathbf{y}}$ , similar to the procedure in eq.(10). Let  $\mathbf{A}$  be the dual



variable of the PSD constraint in the SDP relaxation (3). Then, under a similar dual construction to the one in (Bello and Honorio, 2019; Abbe et al., 2016), we have that  $\lambda_2(\mathbf{\Lambda}) = \min_{\mathbf{v} \perp \bar{\mathbf{y}}} \frac{\mathbf{v}^\top \mathbf{\Lambda} \mathbf{v}}{\mathbf{v}^\top \mathbf{v}}$  is equal to the second eigenvalue of the Laplacian matrix of Figure (4c). Thus, we can observe that, for SDP, the ground-truth solution  $\bar{\mathbf{Y}}$  attains a value of  $\lambda_2(\mathbf{\Lambda}) = -0.24 < 0$ , hence, exact recovery fails.

In Figure (4d), we show the level-2 graph of  $\mathbf{Y} \mathbf{X} \mathbf{Y}$ , i.e.,  $\tilde{\mathbf{Y}} \mathbf{X}^{(2)} \tilde{\mathbf{Y}}$ . As argued by Proposition 2, by setting  $\boldsymbol{\mu} = \mathbf{0}$  the SoS does not do any better than SDP, which is verified by obtaining  $\lambda_2(\tilde{\mathbf{\Lambda}}) = -0.24 < 0$ , hence, exact recovery also fails in this case. However, by solving problem (6), we obtain  $\boldsymbol{\mu} \neq \mathbf{0}$  which, as discussed in Section 3.3, relates to edge weights in the Johnson and Kneser graphs. Those edge weights are depicted in Figures (4e) and (4f), respectively. Finally, after summing all the weights of the level-2 graph, Johnson and Kneser graphs, we obtain a complete graph depicted in Figure (4g). In the latter, we have that  $\lambda_2(\tilde{\mathbf{\Lambda}}) = 0.95 > 0$ , which guarantees that  $\tilde{\mathbf{Y}} = \bar{\mathbf{y}}^{(2)} \bar{\mathbf{y}}^{(2)\top}$ , i.e., exact recovery succeeds.

Motivated by eq.(13) and Theorem 1, in Appendix C, we show a non-trivial construction of the Kneser graph weights based on only the node degrees of the level-2 graph.

## 5 CONCLUDING REMARKS

We studied the statistical problem of exact recovery in graphs over the boolean hypercube. We considered a generative model similar to that of Globerson et al. (2015), which was motivated by an image segmentation problem. We analyzed the level-2 SoS relaxation (6) of problem (2), motivated by empirical evidence on the improvement in exact recoverability over the SDP relaxation (3). We showed how the dual formulation of the SoS relaxation relates to finding edge weights of Johnson and Kneser graphs so that the algebraic connectivity of the input graph increases. Finally, we characterized the improvement by deriving a novel lower bound to the algebraic connectivity of graphs with positive and negative weights, and provided a construction of the Kneser graph weights in Appendix C. It remains an interesting future work to study the recent mapping of degree-2 to degree-4 solutions in (Mohanty et al., 2020) for exact recovery.

### Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1716609-IIS.

### References

- Abbe, E., Bandeira, A. S. and Hall, G. (2016), ‘Exact recovery in the stochastic block model’, *IEEE Transactions on Information Theory*.
- Amini, A. and Levina, E. (2018), ‘On semidefinite relaxations for the block model’, *The Annals of Statistics*.
- Atay, F. M. and Liu, S. (2020), ‘Cheeger constants, structural balance, and spectral clustering analysis for signed graphs’, *Discrete Mathematics* **343**(1), 111616.
- Bandeira, A. S. and Kunisky, D. (2018), ‘A gramian description of the degree 4 generalized elliptope’, *ArXiv preprint ArXiv:1812.11583*.
- Barahona, F. (1982), ‘On the computational complexity of ising spin glass models’, *Journal of Physics A: Mathematical and General*.
- Barak, B. and Steurer, D. (2014), ‘Sum-of-squares proofs and the quest toward optimal algorithms’, *ArXiv 1404.5236*.
- Bello, K. and Honorio, J. (2019), ‘Exact inference in structured prediction’, *Advances in Neural Information Processing Systems*.
- Bello, K. and Honorio, J. (2020), ‘Fairness constraints can help exact inference in structured prediction’, *Advances in Neural Information Processing Systems*.
- Boyd, S. and Vandenberghe, L. (2004), *Convex optimization*, Cambridge university press.
- Boykov, Y. and Veksler, O. (2006), Graph cuts in vision and graphics: Theories and applications, in ‘Handbook of mathematical models in computer vision’, Springer, pp. 79–96.
- Chandrasekaran, V., Srebro, N. and Harsha, P. (2008), Complexity of inference in graphical models, in ‘Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence’.
- Cheeger, J. (1969), A lower bound for the smallest eigenvalue of the laplacian, in ‘Proceedings of the Princeton conference in honor of Professor S. Bochner’.
- Chen, Y., Khong, S. Z. and Georgiou, T. T. (2016), On the definiteness of graph laplacians with negative weights: Geometrical and passivity-based approaches, in ‘2016 American Control Conference (ACC)’, IEEE.
- Chiang, K.-Y., Whang, J. J. and Dhillon, I. S. (2012), Scalable clustering of signed networks using balance normalized cut, in ‘Proceedings of the 21st ACM international conference on Information and knowledge management’, pp. 615–624.

- Cifuentes, D., Harris, C. and Sturmfels, B. (2020), ‘The geometry of sdp-exactness in quadratic optimization’, *Mathematical Programming* **182**(1), 399–428.
- Cucuringu, M., Davies, P., Glielmo, A. and Tyagi, H. (2019), Sponge: A generalized eigenproblem for clustering signed networks, in ‘The 22nd International Conference on Artificial Intelligence and Statistics’, PMLR, pp. 1088–1098.
- Erdogdu, M. A., Deshpande, Y. and Montanari, A. (2017), ‘Inference in graphical models via semidefinite programming hierarchies’, *Advances in Neural Information Processing Systems* .
- Globerson, A., Roughgarden, T., Sontag, D. and Yildirim, C. (2015), How hard is inference for structured prediction?, in ‘International Conference on Machine Learning’.
- Goemans, M. and Williamson, D. P. (1995), ‘Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming’, *Journal of the ACM* .
- Grant, M. and Boyd, S. (2014), ‘CVX: Matlab software for disciplined convex programming, version 2.1’.
- Grothendieck, A. (1956), *Résumé de la théorie métrique des produits tensoriels topologiques*, Soc. de Matemática de São Paulo.
- Holton, D. and Sheehan, J. (1993), *The Petersen Graph*, Australian Mathematical Society Lecture Series, Cambridge University Press.
- Khot, S. and Naor, A. (2011), ‘Grothendieck-type inequalities in combinatorial optimization’, *arXiv preprint arXiv:1108.2464* .
- Knyazev, A. V. (2017), ‘Signed laplacian for spectral clustering revisited’, *arXiv preprint arXiv:1701.01394* **1**.
- Kunegis, J., Schmidt, S., Lommatzsch, A., Lerner, J., De Luca, E. W. and Albayrak, S. (2010), Spectral analysis of signed graphs for clustering, prediction and visualization, in ‘Proceedings of the 2010 SIAM International Conference on Data Mining’, SIAM, pp. 559–570.
- Lasserre, J. B. (2001), ‘Global optimization with polynomials and the problem of moments’, *SIAM Journal on optimization* .
- Laurent, M. (2003), ‘A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0–1 programming’, *Mathematics of Operations Research* .
- Laurent, M. (2009), Sums of squares, moment matrices and optimization over polynomials, in ‘Emerging applications of algebraic geometry’, Springer.
- Lovász, L. (1978), ‘Kneser’s conjecture, chromatic number, and homotopy’, *Journal of Combinatorial Theory* .
- Meka, R., Potechin, A. and Wigderson, A. (2015), Sum-of-squares lower bounds for planted clique, in ‘Proceedings of the forty-seventh annual ACM symposium on Theory of computing’, pp. 87–96.
- Mercado, P., Tudisco, F. and Hein, M. (2016), ‘Clustering signed networks with the geometric mean of laplacians’, *NIPS 2016-Neural Information Processing Systems* .
- Mohanty, S., Raghavendra, P. and Xu, J. (2020), Lifting sum-of-squares lower bounds: degree-2 to degree-4, in ‘Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing’, pp. 840–853.
- Mohar, B. (1991), ‘The laplacian spectrum of graphs’, *Graph theory, combinatorics, and applications* .
- Nesterov, Y. (1998), ‘Semidefinite relaxation and non-convex quadratic optimization’, *Optimization methods and software* .
- Parrilo, P. A. (2000), Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization, PhD thesis, California Institute of Technology.
- Schraudolph, N. N. and Kamenetsky, D. (2009), Efficient exact inference in planar ising models, in ‘Advances in Neural Information Processing Systems’, pp. 1417–1424.
- Tropp, J. A. (2012), ‘User-friendly tail bounds for sums of random matrices’, *Foundations of computational mathematics* .
- Weisser, T., Lasserre, J.-B. and Toh, K.-C. (2016), ‘A bounded degree sos hierarchy for large scale polynomial optimization with sparsity’.
- Zelazo, D. and Bürger, M. (2014), On the definiteness of the weighted laplacian and its connection to effective resistance, in ‘53rd IEEE Conference on Decision and Control’, IEEE.

---

# Supplementary Material: A View of Exact Inference in Graphs from the Degree-4 Sum-of-Squares Hierarchy

---

## A DETAILED PROOFS

In this section, we state the proofs of all propositions and theorem.

### A.1 Proof of Proposition 1

By construction of the level-2 matrix  $\mathbf{X}^{(2)}$ , we have that each entry  $X_{i,j}$  is repeated  $n - 2$  times. Thus, it follows that the objectives in problems (5) and (6) are equal.

Let  $\mathbf{Y}$  be a feasible solution to problem (5), then clearly the principal submatrix indexed by  $\binom{[n]}{2} \times \binom{[n]}{2}$  is a feasible solution to problem (6). It remains to verify that if  $\tilde{\mathbf{Y}}$  is a feasible solution to problem (6) then there exists a matrix  $\mathbf{Y}$  such that it is feasible to problem (5) and has  $\tilde{\mathbf{Y}}$  as a principal submatrix. We define the entries of  $\mathbf{Y}$  as follows,

$$\begin{aligned} Y_{(i,i)(j,j)} &= \tilde{Y}_{\{i,j\},\{i,j\}} \\ Y_{(i,i)(j,k)} &= \tilde{Y}_{\{i,j\},\{i,k\}} \\ Y_{(i,j)(k,\ell)} &= \tilde{Y}_{\{i,j\},\{k,\ell\}}. \end{aligned}$$

Clearly,  $\mathbf{Y}$  will fulfill the constraints of problem (5) if  $\tilde{\mathbf{Y}}$  is feasible to problem (6). In particular, one can verify that  $\mathbf{v}^\top \mathbf{Y} \mathbf{v} \geq 0$  for any  $\mathbf{v}$  if  $\tilde{\mathbf{Y}} \succeq 0$ , which concludes our proof.

### A.2 Proof of Proposition 2

We will show the equivalence between problem (5), *without the third and fourth list of constraints*, and problem (3). Then, by Proposition 1, our claim follows.

The proof is similar to that of Section A.1, where it is clear that the objectives in problems (3) and (5) are equal. Let  $\mathbf{Y}^{\text{sdp}}$  be a feasible solution to problem (3), then we define  $\mathbf{Y}^{\text{sos}}$  as follows,

$$\begin{aligned} Y_{(i,i)(j,k)}^{\text{sos}} &= Y_{i,j}^{\text{sdp}} \\ Y_{(i,j)(k,\ell)}^{\text{sos}} &= 0. \end{aligned}$$

Since  $\mathbf{Y}^{\text{sdp}} \succeq 0$ , it follows that  $\mathbf{Y}^{\text{sos}} \succeq 0$  and, thus,  $\mathbf{Y}^{\text{sos}}$  is feasible to problem (5) without the third and fourth list of constraints. Similarly, in the other direction, let  $\mathbf{Y}^{\text{sos}}$  be a feasible solution to problem (5) without the third and fourth list of constraints, and define  $\mathbf{Y}^{\text{sdp}}$  to be the principal submatrix of  $\mathbf{Y}^{\text{sos}}$  with the first  $n$  rows and columns. Then, it follows that if  $\mathbf{Y}^{\text{sos}} \succeq 0$  then  $\mathbf{Y}^{\text{sdp}} \succeq 0$ , which is feasible to problem (3).

### A.3 Proof of Theorem 1

For simplicity, let  $\mathbf{W}$  and  $\mathbf{L}$  be the weight matrix and Laplacian matrix of an undirected connected graph  $H$  of  $m$  nodes. Also, let  $\mathbf{W}^+$  and  $\mathbf{W}^-$  be the weight matrices of  $H^+$  and  $H^-$ . For a matrix  $\mathbf{M}$  and vector  $\mathbf{v}$ , we use  $R_{\mathbf{M}}(\mathbf{v})$  to denote their Rayleigh quotient, i.e.,  $R_{\mathbf{M}}(\mathbf{v}) = \frac{\mathbf{v}^\top \mathbf{M} \mathbf{v}}{\mathbf{v}^\top \mathbf{v}}$ . It follows that  $R_{\mathbf{L}}(\mathbf{v}) := \frac{\mathbf{v}^\top \mathbf{L} \mathbf{v}}{\mathbf{v}^\top \mathbf{v}} = \frac{\sum_{i < j} W_{i,j} (v_i - v_j)^2}{\mathbf{v}^\top \mathbf{v}}$ , and  $\lambda_2(\mathbf{L}) = \min_{\mathbf{v} \perp \mathbf{1}} R_{\mathbf{L}}(\mathbf{v})$ . Similarly, we define  $R_{\mathbf{L}}^+(\mathbf{v}) := \frac{\sum_{i < j} W_{i,j}^+ (v_i - v_j)^2}{\mathbf{v}^\top \mathbf{v}}$ ,  $R_{\mathbf{L}}^-(\mathbf{v}) := \frac{\sum_{i < j} W_{i,j}^- (v_i - v_j)^2}{\mathbf{v}^\top \mathbf{v}}$ . Note that  $R_{\mathbf{L}}(\mathbf{v}) = R_{\mathbf{L}}^+(\mathbf{v}) + R_{\mathbf{L}}^-(\mathbf{v})$ . Next, we state a lemma that will be of use for the proof of Theorem 1.

**Lemma 1.** Let  $\mathbf{L}$  be a Laplacian matrix of dimension  $m \times m$ . Let also  $\mathbf{1}$  denote a vector of ones. Then, for any  $\delta \in \mathbb{R}, \mathbf{v} \in \mathbb{R}^m, \sum_i v_i \geq 0$ , it follows that

$$R_{\mathbf{L}}^+(\mathbf{v}) \geq R_{\mathbf{L}}^+(\mathbf{v} + \delta \mathbf{1}).$$

*Proof.* Starting from the right-hand side, we have

$$\begin{aligned} R_{\mathbf{L}}^+(\mathbf{v} + \delta \mathbf{1}) &= \frac{\sum_{i < j} W_{i,j}^+ ((v_i + \delta) - (v_j + \delta))^2}{\sum_i (v_i + \delta)^2} \\ &= \frac{\sum_{i < j} W_{i,j}^+ (v_i - v_j)^2}{\sum_i (v_i + \delta)^2} \\ &= \frac{\sum_{i < j} W_{i,j}^+ (v_i - v_j)^2}{\sum_i (v_i^2 + \delta^2 + 2\delta v_i)} \\ &= \frac{\sum_{i < j} W_{i,j}^+ (v_i - v_j)^2}{\sum_i v_i^2 + m\delta^2 + 2\delta \sum_i v_i} \\ &\stackrel{(a)}{\leq} \frac{\sum_{i < j} W_{i,j}^+ (v_i - v_j)^2}{\sum_i v_i^2 + m\delta^2} \\ &\leq \frac{\sum_{i < j} W_{i,j}^+ (v_i - v_j)^2}{\sum_i v_i^2} \\ &= R_{\mathbf{L}}^+(\mathbf{v}), \end{aligned}$$

where (a) holds by the fact that  $\sum_i v_i \geq 0$ . □

We now present the proof of Theorem 1.

*Proof.* Let  $\mathbf{v}$  be the eigenvector related to the eigenvalue  $\lambda_2(\mathbf{L})$ . Without loss of generality, we assume  $\|\mathbf{v}\| = 1$  and  $v_1 \leq v_2 \leq \dots \leq v_m$ . Recall that  $\mathbf{1}^\top \mathbf{v} = 0$ . Then, we have that

$$\lambda_2(\mathbf{L}) = R_{\mathbf{L}}(\mathbf{v}) = R_{\mathbf{L}}^+(\mathbf{v}) + R_{\mathbf{L}}^-(\mathbf{v}).$$

**Lower bounding  $R_{\mathbf{L}}^+(\mathbf{v})$ .** Set  $\delta = v_1$  and denote  $\mathbf{u} = \mathbf{v} - \delta \mathbf{1}$ . Then, we have that  $0 = u_1 \leq \dots \leq u_m$ . Also note that  $\delta^2 \leq 1$ . Then, by Lemma 1, it follows that  $R_{\mathbf{L}}^+(\mathbf{v}) \geq R_{\mathbf{L}}^+(\mathbf{u})$ .

We now define a random variable  $t$  on the support  $[0, u_m]$ , with probability density function  $f(t) = \frac{2}{u_m^2}t$ . One can verify that  $\int_{t=0}^{u_m} \frac{2}{u_m^2}t dt = 1$ , thus  $f(t)$  is a valid probability density function. Then, for any interval  $[a, b]$ , it follows that the probability of  $t$  falling in the interval is

$$\mathbb{P}[a \leq t \leq b] = \int_{t=a}^b \frac{2}{u_m^2}t dt = \frac{1}{u_m^2}(b^2 - a^2).$$

Next, for some  $t$ , construct a random set  $S_t = \{i \mid u_i \geq t\}$ . Let  $\omega^+(\partial S_t) = \sum_{i \in S_t, j \notin S_t} W_{i,j}^+$ .

It follows that

$$\begin{aligned} \mathbb{E}[\omega^+(\partial S_t)] &= \mathbb{E}\left[\sum_{i \in S_t, j \notin S_t} W_{i,j}^+\right] \\ &= \sum_{i < j} \mathbb{P}[u_j \leq t \leq u_i] W_{i,j}^+ \\ &= \frac{1}{u_m^2} \sum_{i < j} (u_i - u_j)(u_i + u_j) W_{i,j}^+ \\ &\leq \frac{1}{u_m^2} \sqrt{\sum_{i < j} (u_i - u_j)^2 W_{i,j}^+} \sqrt{\sum_{i < j} (u_i + u_j)^2 W_{i,j}^+} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{u_m^2} \sqrt{R_{\mathbf{L}}^+(\mathbf{u}) \sum_i u_i^2} \sqrt{\sum_{i<j} (u_i + u_j)^2 W_{i,j}^+} \\
 &\leq \frac{1}{u_m^2} \sqrt{R_{\mathbf{L}}^+(\mathbf{u}) \sum_i u_i^2} \sqrt{2 \sum_i u_i^2 \deg^{H^+}(i)} \\
 &\leq \frac{1}{u_m^2} \sqrt{R_{\mathbf{L}}^+(\mathbf{u}) \sum_i u_i^2} \sqrt{2 \deg_{\max}^{H^+} \sum_i u_i^2} \\
 &= \frac{\sum_i u_i^2}{u_m^2} \sqrt{2 \deg_{\max}^{H^+} R_{\mathbf{L}}^+(\mathbf{u})}.
 \end{aligned}$$

Also note that  $\mathbb{E}[|S_t|] = \sum_i \mathbb{P}[u_i \geq t] = \sum_i \frac{u_i^2}{u_m^2}$ . As a result we obtain

$$\mathbb{E}[\omega^+(\partial S_t)] \leq \mathbb{E}[|S_t|] \sqrt{2 \deg_{\max}^{H^+} R_{\mathbf{L}}^+(\mathbf{u})}.$$

Thus, we have  $\mathbb{E}\left[\omega^+(\partial S_t) - |S_t| \sqrt{2 \deg_{\max}^{H^+} R_{\mathbf{L}}^+(\mathbf{u})}\right] \leq 0$ . This implies that  $\exists S_t$  such that  $\omega^+(\partial S_t) - |S_t| \sqrt{2 \deg_{\max}^{H^+} R_{\mathbf{L}}^+(\mathbf{u})} \leq 0$ . Rearranging we have,

$$R_{\mathbf{L}}^+(\mathbf{v}) \geq R_{\mathbf{L}}^+(\mathbf{u}) \geq \frac{\omega^+(\partial S_t)^2}{2 \deg_{\max}^{H^+} |S_t|^2} \quad (16)$$

**Lower bounding  $R_{\mathbf{L}}^-(\mathbf{v})$ .** Set  $\alpha = \sqrt{\frac{1}{v_1^2 + v_m^2}}$  and denote  $\mathbf{u} = \alpha \mathbf{v}$ . Then, we have that  $u_1^2 + u_m^2 = 1$ . Note also that  $R_{\mathbf{L}}^-(\mathbf{v}) = R_{\mathbf{L}}^-(\mathbf{u})$ .

We now define a random variable  $t$  on the support  $[u_1, u_m]$ , with probability density function  $f(t) = 2|t|$ . One can verify that  $\int_{t=u_1}^{u_m} 2|t| dt = 1$ , thus  $f(t)$  is a valid probability density function. Then, for any interval  $[a, b]$ , it follows that the probability of  $t$  falling in the interval is

$$\mathbb{P}[a \leq t \leq b] = \int_a^b 2|t| dt = b^2 \text{sign}(b) - a^2 \text{sign}(a).$$

Since  $[u_1, u_m] \subset [-1, 1]$ , one can verify that  $(a - b)^2/2 \leq \mathbb{P}[a \leq t \leq b]$ . Let  $\omega^-(\partial S_t) = \sum_{i \in S_t, j \notin S_t} W_{i,j}^-$ . For some  $t$ , construct a random set  $S_t = \{i \mid u_i \leq t\}$ . It follows that

$$\begin{aligned}
 \mathbb{E}[\omega^-(\partial S_t)] &= \mathbb{E}\left[\sum_{i \in S_t, j \notin S_t} W_{i,j}^-\right] \\
 &= \sum_{i<j} \mathbb{P}[u_i \leq t \leq u_j] W_{i,j}^- \\
 &\leq \frac{1}{2} \sum_{i<j} (u_i - u_j)^2 W_{i,j}^- \\
 &= \frac{1}{2} R_{\mathbf{L}}^-(\mathbf{u}) \sum_i u_i^2 \\
 &\leq \frac{1}{2} R_{\mathbf{L}}^-(\mathbf{u}),
 \end{aligned}$$

where the last inequality follows from having  $\sum_i u_i^2 \geq 1$  and  $R_{\mathbf{L}}^-(\mathbf{u}) \leq 0$ . Thus, we have  $\mathbb{E}[\omega^-(\partial S_t) - \frac{1}{2} R_{\mathbf{L}}^-(\mathbf{u})] \leq 0$ . This implies that  $\exists S_t$  such that  $\omega^-(\partial S_t) - \frac{1}{2} R_{\mathbf{L}}^-(\mathbf{u}) \leq 0$ . Rearranging we have,

$$R_{\mathbf{L}}^-(\mathbf{v}) = R_{\mathbf{L}}^-(\mathbf{u}) \geq 2\omega^-(\partial S_t). \quad (17)$$

By minimizing (16) and (17) independently, and combining them, we conclude our proof.  $\square$

#### A.4 Proof of Theorem 2

The proof follows by using matrix Bernstein inequality (Tropp, 2012). That is, in eq.(9), let  $t$  be a lower bound to  $\lambda_2(\mathbb{E}[\tilde{\mathbf{A}}])$ , i.e.,  $\lambda_2(\mathbb{E}[\tilde{\mathbf{A}}]) \geq t$ . Then, the second summand can be lower bounded by using matrix Bernstein inequality and thus we obtain  $\mathbb{P}[\lambda_1(\tilde{\mathbf{A}} - \mathbb{E}[\tilde{\mathbf{A}}]) \leq -t] \leq \mathcal{O}(n^2 e^{-t})$ . The theorem statement follows by plugging Theorem 1 into the inequality, where the SDP bound corresponds to  $\boldsymbol{\mu} = 0$ .

## B FURTHER DISCUSSION ON THEOREM 1

We remark that the reason we do not consider other versions of the Laplacian matrix (e.g., the normalized Laplacian matrix which is guaranteed to be PSD even in the presence of negative weights) is because how our primal/dual construction (see Section 3.1) leads to a valid solution of the constraints in eq.(7) and that also satisfies the KKT conditions. That is, using other notions of Laplacian matrix (see e.g., Kunegis et al. (2010); Mercado et al. (2016); Cucuringu et al. (2019); Chiang et al. (2012); Knyazev (2017); Atay and Liu (2020)) would **not** satisfy the optimality conditions needed for exact recovery (in particular, stationarity and complementary slackness). In fact, one of the challenges we face in our analysis is that by having the standard Laplacian matrix, its minimum eigenvalue can be negative, as shown in our example in Section 4 and also discussed in Knyazev (2017), which motivated the search of a more general lower bound for the algebraic connectivity of signed graphs (Theorem 1).

We also highlight that Theorem 1 sheds light on the subtle trade-off between the Cheeger constant of the positive subgraph and the minimum cut of the negative subgraph. That is, intuitively, the SoS solution will try to find negative weights for the Johnson and Kneser graphs of as low magnitude as possible, so that the minimum-cut of the negative subgraph does not make the algebraic connectivity negative.

## C A DEGREE-BASED CONSTRUCTION OF THE KNESER GRAPH

In Section 4, we used CVX (Grant and Boyd, 2014) to solve problem (6) and, thus, obtain the dual variables  $\boldsymbol{\mu}$  in problem (7) from which we construct the weights of the Johnson and Kneser graphs. Motivated by the trade-off between Cheeger constants of the positive and negative subgraphs, shown in Theorem 1, we show a simple non-trivial way (not necessarily optimal) to directly construct the weights of the Kneser graph. The reason why we focus in the Kneser graph weights is because the fourth list of constraints in problem (6) can be expressed by two constraints for any  $i < j < k < \ell$ , as noted in Section 3.3. The latter fact implies that, for any  $i < j < k < \ell$ , the edge weights  $W_{\{i,j\},\{k,\ell\}}^{\mathcal{K}}$ ,  $W_{\{i,k\},\{j,\ell\}}^{\mathcal{K}}$ , and  $W_{\{i,\ell\},\{j,k\}}^{\mathcal{K}}$  need to sum to zero in order to fulfill the SoS constraints. As also noted in Section 3.3, at least one of the previous weights need to be negative unless all three are zero. With these considerations, we present our construction in Algorithm 1, which relies only on the node degrees and a constant real value.

The intuition behind Algorithm 1 is that the negative weight will be assigned to the edge that connects the two nodes that have the highest combined node degree. In Lines 8-9, if all three edges have the same combined node degree then we set all three weights to zero. In Lines 10-11, the edge with lowest combined node degree is set to  $2c$ , while the other edges that attain the same combined node degree are set to  $-c$ . In Line 13, the edge with highest combined node degree is set to  $-2c$ , while the other edges are set to  $c$ . It is clear that the SoS constraints will be fulfilled for each quadruple  $i < j < k < \ell$ . Finally, we note that if  $c = 0$  then simply the same input,  $\mathbf{M}$ , is returned. The latter implies that, for the *optimal* value of  $c$ , Algorithm 1 cannot return a weight matrix with lower algebraic connectivity than that of  $\mathbf{M}$ .<sup>5</sup>

Recall that  $\lambda_2(\tilde{\mathbf{A}}) = \lambda_2(\mathbf{L}^{\tilde{\mathcal{G}}})$ . In Figure 5, we ran Algorithm 1 with input graph  $\tilde{\mathbf{Y}} \mathbf{X}^{(2)} \tilde{\mathbf{Y}}$  equal to the graph in Figure (4d), and  $c \in [0, 0.6]$ . For each  $c$ , we plotted the algebraic connectivity of our construction. We observe that when  $c = 0$ , in effect  $\lambda_2(\tilde{\mathbf{A}}) = -0.24$  as pointed in Figure (4d). In this example, the optimal value of  $c$  is 0.32 and attains a  $\lambda_2(\tilde{\mathbf{A}})$  of 0.9368, which is very close to the value  $\lambda_2(\tilde{\mathbf{A}}) = 0.95$  found by CVX (see Figure (4g)). Finally, we also plot the Kneser graph weights for  $c = 0.32$  following the construction in Algorithm 1.

---

<sup>5</sup>Recall that the SDP problem (3) would attain an algebraic connectivity equal to that of  $\mathbf{M}$  if the optimal solution is  $\overline{\mathbf{y}} \overline{\mathbf{y}}^{\top}$ .



**Algorithm 1** A construction of Kneser graph weights

**Input:** Level-2 weight matrix  $M = \tilde{\mathbf{Y}}\mathbf{X}^{(2)}\tilde{\mathbf{Y}}$ , constant  $c \in \mathbb{R}$ .  
 1:  $\deg(\mathcal{C}_1) \leftarrow \sum_{\mathcal{C}_2} M_{\mathcal{C}_1, \mathcal{C}_2}, \forall \mathcal{C}_1 \in \binom{[n]}{2}$   
 2: Initialize  $\mathbf{W}^{\mathcal{K}}$  as a zero matrix  
 3: **for all**  $i < j < k < \ell \in [n]$  **do**  
 4:   Assign the following such that  $\psi_1 \geq \psi_2 \geq \psi_3$   
 5:    $\psi_1 \leftarrow \deg(\{i, j\}) + \deg(\{k, \ell\})$   
 6:    $\psi_2 \leftarrow \deg(\{i, k\}) + \deg(\{j, \ell\})$   
 7:    $\psi_3 \leftarrow \deg(\{i, \ell\}) + \deg(\{j, k\})$   
 8:   **if**  $\psi_1 = \psi_2 = \psi_3$  **then**  
 9:      $W_{\{i,j\},\{k,\ell\}}^{\mathcal{K}} \leftarrow 0, W_{\{i,k\},\{j,\ell\}}^{\mathcal{K}} \leftarrow 0, W_{\{i,\ell\},\{j,k\}}^{\mathcal{K}} \leftarrow 0$   
 10:   **else if**  $\psi_1 = \psi_2$  **then**  
 11:      $W_{\{i,j\},\{k,\ell\}}^{\mathcal{K}} \leftarrow -c, W_{\{i,k\},\{j,\ell\}}^{\mathcal{K}} \leftarrow -c, W_{\{i,\ell\},\{j,k\}}^{\mathcal{K}} \leftarrow 2c$   
 12:   **else**  
 13:      $W_{\{i,j\},\{k,\ell\}}^{\mathcal{K}} \leftarrow -2c, W_{\{i,k\},\{j,\ell\}}^{\mathcal{K}} \leftarrow c, W_{\{i,\ell\},\{j,k\}}^{\mathcal{K}} \leftarrow c$   
 14:   **end if**  
 15: **end for**  
 16:  $\mathbf{W}^{\mathcal{K}} \leftarrow \mathbf{W}^{\mathcal{K}} + (\mathbf{W}^{\mathcal{K}})^{\top}$  {To symmetrize.}  
**Output:**  $\mathbf{W}^{\tilde{\mathcal{G}}} \leftarrow M + \mathbf{W}^{\mathcal{K}}$

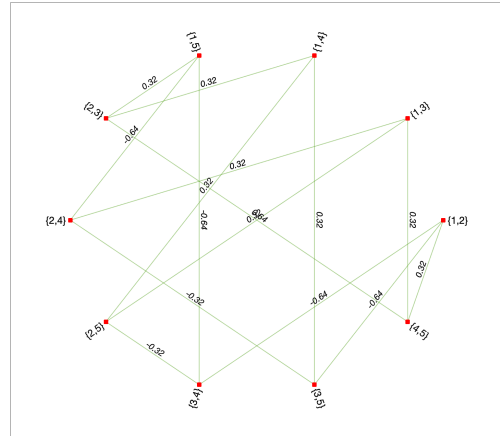
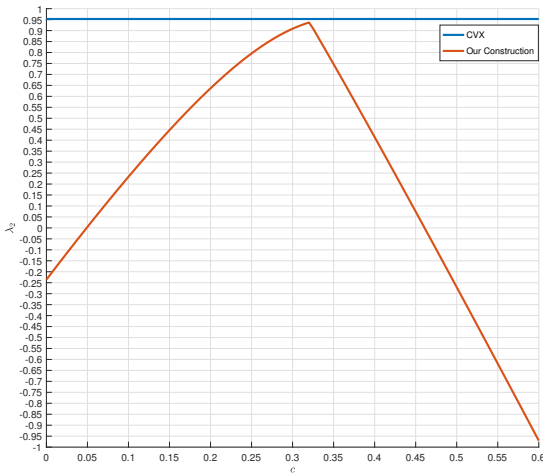


Figure 5: (Left) The blue line is the algebraic connectivity found by CVX, i.e., 0.95 as pointed in Figure (4g). The red line is the algebraic connectivity of our construction in Algorithm 1 for different values of  $c \in [0, 0.6]$ . (Right) The Kneser graph weights for the optimal  $c = 0.32$ , which in effect differs from the weights found by CVX in Figure (4f).