

MIT Open Access Articles

Blind Community Detection From Low-Rank Excitations of a Graph Filter

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Wai, Hoi-To et al. "Blind Community Detection From Low-Rank Excitations of a Graph Filter." IEEE Transactions on Signal Processing 68 (December 2019): 436 - 451 © 2020 IEEE

As Published: 10.1109/tsp.2019.2961296

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: <https://hdl.handle.net/1721.1/129825>

Version: Original manuscript: author's manuscript prior to formal peer review

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



Blind Community Detection from Low-rank Excitations of a Graph Filter

Hoi-To Wai, Santiago Segarra, Asuman E. Ozdaglar, Anna Scaglione, Ali Jadbabaie*

April 16, 2019

Abstract

This paper considers a new framework to detect communities in a graph from the observation of signals at its nodes. We model the observed signals as noisy outputs of an unknown network process, represented as a graph filter that is excited by a set of unknown low-rank inputs/excitations. Application scenarios of this model include diffusion dynamics, pricing experiments, and opinion dynamics. Rather than learning the precise parameters of the graph itself, we aim at retrieving the community structure directly. The paper shows that communities can be detected by applying a spectral method to the covariance matrix of graph signals. Our analysis indicates that the community detection performance depends on a ‘low-pass’ property of the graph filter. We also show that the performance can be improved via a low-rank matrix plus sparse decomposition method when the latent parameter vectors are known. Numerical experiments demonstrate that our approach is effective.

1 Introduction

The emerging field of *network science* and availability of *big data* have motivated researchers to extend signal processing techniques to the analysis of signals defined on graphs, propelling a new area of research referred to as *graph signal processing* (GSP) [1–3]. As opposed to signals on time defined on a regular topology, the properties of *graph signals* are intimately related to the generally irregular topology of the graph where they are defined. The goal of GSP is to develop mathematical tools to leverage this topological structure in order to enhance our understanding of graph signals. A suitable way to capture the graph’s structure is via the so-called *graph shift operator* (GSO), which is a matrix that reflects the local connectivity of the graph and is a generalization of the time shift or delay operator in classical discrete signal processing [1]. Admissible choices for the GSO include the graph’s adjacency matrix and the Laplacian matrix. When the GSO is known, the algebraic and spectral characteristics of a given graph signal can be analyzed in an analogous way as in time-series analysis [1]. Furthermore, signal processing tools such as sampling [4, 5], interpolation [6, 7] and filtering [8, 9] can be extended to the realm of graph signals.

*H.-T. Wai is with the Department of SEEM, The Chinese University of Hong Kong, Shatin, Hong Kong. E-mail: htwai@se.cuhk.edu.hk. S. Segarra is with Department of ECE, Rice University, TX, USA. E-mail: segarra@rice.edu. A. E. Ozdaglar is with LIDS, Massachusetts Institute of Technology, MA, USA. A. Jadbabaie is with IDSS, Massachusetts Institute of Technology, MA, USA. E-mails: {asuman,jadbabai}@mit.edu. A. Scaglione is with School of ECEE, Arizona State University, Tempe, AZ, USA. E-mail: Anna.Scaglione@asu.edu

This paper considers an *inverse* problem in GSP where our focus is to infer information about the GSO (or the graph) from the observed graph signals. Naturally, *graph or network inference* is relevant to network and data science, and has been studied extensively. Classical methods are based on partial correlations [10], Gaussian graphical models [11], and structural equation models [12], among others. Recently, GSP-based methods for graph inference have emerged, which tackle the problem as a system identification task. They postulate that the unknown graph is a structure encoded in the observed signals and the signals are obtained from observations of network dynamical processes defined on the graph [13, 14]. Different assumptions are put forth in the literature to aid the graph topology inference, such as smoothness of the observed signals [15–17], richness of the inputs to the network process [18–21], and partial knowledge of the network process [12, 22].

A drawback common to the prior GSP work on graph inference [15–21] is that they require the observed graph signals to be *full-rank*. Equivalently, the signals observed are results of a network dynamical process excited by a set of input signals that span a space with the same dimension as the number of nodes in the graph. Such assumption can be unnecessarily stringent for a number of applications, especially when the graph contains a large number of nodes. For example, whenever graph inference experiments can only be performed by exciting a few nodes on the graph (such as rumor spreading initiated by a small number of sources and the gene perturbation experiments in [23]); or the amount of data collected is limited due to cost and time constraints.

Oftentimes, inferring the entire graph structure is only the first step since the ultimate goal is to obtain *interpretable* information from the set of graph signals. To this end, a feature that is often sought in network science is the *community structure* [24] that offers a coarse description of graphs. For this task, applying conventional methods necessitates a *two-step* procedure which comprises of a graph learning and a community detection step. This paper departs from the conventional methods by developing a *direct analysis* framework to recover the communities based on the observation of graph signals. We consider a setting where the observations are graph signals modeled as the outputs of an unknown network process represented by a graph filter. Such signal model can be applicable to observations from, e.g., diffusion dynamics, pricing experiments in consumer networks [25, 26], and DeGroot dynamics [27] with stubborn agents. In addition, unlike the prior works on graph learning, we allow the excitations to the graph filter to be *low-rank*. This is a challenging yet practical scenario as we demonstrate later.

We propose and analyze two *blind community detection* (BlindCD) methods that do not require learning the graph topology nor knowing the dynamics governing the generation of graph signals explicitly. The first method applies spectral clustering on the sampled covariance matrix, which is akin to a common heuristics used in data clustering, e.g., [28]. Here our contribution lies in showing when sampled covariance carries information about the communities. Under a mild assumption that the underlying graph filter is *low-pass* with the GSO taken as graph Laplacian, we show that the covariance matrix of observed graph signals is a *sketch* of the Laplacian matrix that retains coarse topological features of the graph, like communities. We quantify the *suboptimality* of the BlindCD method compared to the minimizer of a convex relaxation of the RatioCut objective defined on the actual graph Laplacian. Our result helps in justifying the successful application of such heuristics on real data. Furthermore, the theoretical analysis of BlindCD identifies the key bottleneck in the spectral method applied to some GSP models. This leads to the development of our second method, called boosted BlindCD. The method works under an additional assumption that the latent parameter vectors are available and *boosts* the performance of the first method by leveraging a low-rank plus sparse structure in the linear transformation between excitations and observed graph

signals. Performance bound is also analyzed for this method.

The organization of this paper is as follows. In Section 2, we introduce notations by describing the graph model and a formal definition for communities on graph. Section 3 presents the GSP signal model with real world examples. In Section 4 and 5, we describe and analyze the proposed BlindCD method and its boosted version. In Section 6, we present numerical results on synthetic and real data to validate our findings.

Notation — We use boldface lower-case (*resp.* upper-case) letters to denote vectors (*resp.* matrices). For a vector \mathbf{x} , the notation x_i denotes its i th element and we use $\|\mathbf{x}\|_2$ to denote the standard Euclidean norm. For a matrix \mathbf{X} , the notation X_{ij} denotes its (i, j) th element whereas $[\mathbf{X}]_{i,:}$ denotes its i th row vector and $[\mathbf{X}]_{\mathcal{I},:}$ denotes the collection of its row vectors in \mathcal{I} . Also, $\mathcal{R}(\mathbf{X}) \subseteq \mathbb{R}^N$ denotes the range space of $\mathbf{X} \in \mathbb{R}^{N \times M}$. Moreover, $\|\mathbf{X}\|_F$ (*resp.* $\|\mathbf{X}\|_2$) denotes the Frobenius norm (*resp.* spectral norm). For a symmetric matrix \mathbf{E} , $\beta_i(\mathbf{E})$ denotes its i th largest eigenvalue. For a matrix $\mathbf{M} \in \mathbb{R}^{P \times N}$, $\sigma_i(\mathbf{M})$ denotes its i th largest singular value and $[\mathbf{M}]_K$ denotes its rank K approximation. Moreover, \mathbf{M} admits the partition $\mathbf{M} = [\mathbf{M}_K \ \mathbf{M}_{N-K}]$ where \mathbf{M}_K (*resp.* \mathbf{M}_{N-K}) denotes the matrix consisting of the *left-most* K (*resp.* *right-most* $N - K$) columns of \mathbf{M} . Similarly, $\mathbf{m} \in \mathbb{R}^N$ is partitioned into $\mathbf{m} = [\mathbf{m}_K; \mathbf{m}_{N-K}]$, where \mathbf{m}_K (*resp.* \mathbf{m}_{N-K}) consists of its *top* K (*resp.* *bottom* $N - K$) elements. For any integer K , we denote $[K] := \{1, \dots, K\}$.

2 Preliminaries

2.1 Graph Signal Processing

Consider an undirected graph $G = (V, E, \mathbf{A})$ with N nodes such that $V = [N] := \{1, \dots, N\}$ and $E \subseteq V \times V$ is the set of edges where $(i, i) \notin E$ for all i . The graph G is also associated with a symmetric and weighted adjacency matrix $\mathbf{A} \in \mathbb{R}_+^{N \times N}$ such that $A_{ij} = A_{ji} > 0$ if and only if $(i, j) \in E$. The graph Laplacian matrix for G is defined as $\mathbf{L} := \mathbf{D} - \mathbf{A}$, where $\mathbf{D} := \text{Diag}(\mathbf{A}\mathbf{1})$ is a diagonal matrix containing the weighted degrees of G . As \mathbf{L} is symmetric and positive semidefinite, it admits the following eigendecomposition

$$\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top, \quad (1)$$

where $\mathbf{\Lambda} = \text{Diag}([\lambda_1, \dots, \lambda_N])$ and λ_i is sorted in *ascending order* such that $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$.

A graph signal is defined as a function on the nodes of G , $f : V \rightarrow \mathbb{R}$, and can be equivalently represented as a vector $\mathbf{x} := [x_1, x_2, \dots, x_N] \in \mathbb{R}^N$, where x_i is the signal value at the i th node. The graph is endowed with a graph shift operator (GSO) that is set as the graph Laplacian \mathbf{L} . Note that it is also possible to define alternative GSOs such as the adjacency matrix \mathbf{A} and its normalized versions; see [1] for an overview on the subject, yet the analysis result in this paper may differ slightly for the latter cases. Having defined the GSO, the graph Fourier transform (GFT) [1] of \mathbf{x} is given by

$$\tilde{\mathbf{x}} := \mathbf{V}^\top \mathbf{x}. \quad (2)$$

The vector $\tilde{\mathbf{x}}$ is called the *frequency* domain representation of \mathbf{x} with respect to (w.r.t.) the GSO \mathbf{L} [1, 3].

The GSO can be used to define linear graph filters. These are linear graph signal operators that can be expressed as matrix polynomials on \mathbf{L} :

$$\mathcal{H}(\mathbf{L}) := \sum_{t=0}^{T_d-1} h_t \mathbf{L}^t = \mathbf{V} \left(\sum_{t=0}^{T_d-1} h_t \mathbf{\Lambda}^t \right) \mathbf{V}^\top, \quad (3)$$

where T_d is the *order* of the graph filter. Note that by the Cayley-Hamilton theorem, any matrix polynomial (even of infinite degree) can be represented using the form (3) with $T \leq N$. For a given excitation graph signal $\mathbf{x} \in \mathbb{R}^N$, the output of the filter is simply $\mathbf{y} = \mathcal{H}(\mathbf{L})\mathbf{x}$, and carries the classical interpretation of being a linear combination of shifted versions of the input. The graph filter $\mathcal{H}(\mathbf{L})$ may also be represented by its frequency response $\tilde{\mathbf{h}}$, defined as

$$\tilde{h}_i := h(\lambda_i) = \sum_{t=0}^{T_d-1} h_t \lambda_i^t. \quad (4)$$

We denominate the polynomial $h(\lambda) := \sum_{t=0}^{T_d-1} h_t \lambda^t$ as the *generating function* of the graph filter. From (3) it follows that the frequency representations of the input and the output of a filter are related by

$$\tilde{\mathbf{x}} = \tilde{\mathbf{h}} \odot \tilde{\mathbf{z}}, \quad (5)$$

where \odot denotes the element-wise product. This is analogous to the convolution theorem for time signals. In Section 3, we utilize GSP to model the relationship between the observed data and the unknown graph G .

2.2 Community Structure and its Detection

Intuitively, a community on the graph G is a subset of nodes, $\mathcal{C}_k^* \subseteq V$, that induces a densely connected subgraph while loosely connected with nodes not in \mathcal{C}_k^* . To formally describe a community structure, in this paper we refer to the common notion of *ratio-cut* [24] that measures the total *cut* weight across the boundary between a *disjoint* partition of $G = (V, E, \mathbf{A})$. In particular, for any disjoint K partition of V , *i.e.*, $V = \mathcal{C}_1 \cup \dots \cup \mathcal{C}_K$, define the function:

$$\text{RatioCut}(\mathcal{C}_1, \dots, \mathcal{C}_K) := \sum_{k=1}^K \frac{1}{|\mathcal{C}_k|} \sum_{i \in \mathcal{C}_k} \sum_{j \notin \mathcal{C}_k} A_{ij}. \quad (6)$$

Throughout this paper, we assume that there are K *non-overlapping communities* in G as given by $\mathcal{C}_1^*, \dots, \mathcal{C}_K^*$, where the latter is a minimizer to the ratio-cut function and it results in a small objective value. For instance,

$$\delta^* := \text{RatioCut}(\mathcal{C}_1^*, \dots, \mathcal{C}_K^*) \leq \text{RatioCut}(\mathcal{C}_1, \dots, \mathcal{C}_K) \quad (7)$$

where $\delta^* \ll 1$ indicates that the graph has K communities.

Having defined the above notion, the *community detection* problem is solved by *minimizing* (6) with the given number of communities K and graph adjacency matrix \mathbf{A} . However, the ratio-cut minimization problem is combinatorial and difficult to solve. As such, a popular remedy is to apply a convex relaxation – a method known as the *spectral clustering* [29, 30]. To describe the method, let us define the left- K eigenmatrix of the graph Laplacian \mathbf{L} as

$$\mathbf{V}_K := (\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_K) \in \mathbb{R}^{N \times K}, \quad (8)$$

where \mathbf{v}_i is the i th eigenvector of \mathbf{L} corresponding to the i th eigenvalue λ_i [cf. (1)]. The K -means method [31] is applied on the row vectors of \mathbf{V}_K , which seeks a partition $\mathcal{C}_1, \dots, \mathcal{C}_K$ that minimizes the distance of each row vector to their respective means. The spectral clustering minimizes

$$F(\mathcal{C}_1, \dots, \mathcal{C}_K) := \sum_{k=1}^K \sum_{i \in \mathcal{C}_k} \left\| \mathbf{v}_i^{\text{row}} - \frac{1}{|\mathcal{C}_k|} \sum_{j \in \mathcal{C}_k} \mathbf{v}_j^{\text{row}} \right\|_2^2, \quad (9)$$

where $\mathbf{v}_j^{\text{row}} := [\mathbf{V}_K]_{j,:}$ is the j th row vector of \mathbf{V}_K .

For general K , [32] proposed a polynomial-time algorithm that finds an $(1 + \epsilon)$ -optimal solution, $\tilde{\mathcal{C}}_1, \dots, \tilde{\mathcal{C}}_K$, to the K -means problem (9) satisfying

$$F(\tilde{\mathcal{C}}_1, \dots, \tilde{\mathcal{C}}_K) \leq (1 + \epsilon) \min_{\mathcal{C}_1, \dots, \mathcal{C}_K \subseteq V} F(\mathcal{C}_1, \dots, \mathcal{C}_K), \quad (10)$$

under some statistical assumptions on $\{\mathbf{v}_i^{\text{row}}\}_{i=1}^N$. The spectral clustering method is shown to be effective both in theory and in practice. In particular, when $K = 2$ and the graph of interest is drawn from a stochastic block model (SBM) satisfying certain spectral gap conditions, the spectral method exactly recovers the ground truth clusters in the SBM when $N \rightarrow \infty$ [which also gives a minimizer to (6)], see [29].

3 Graph Signal Model

Consider a graph signal $\mathbf{y}^\ell \in \mathbb{R}^N$ defined on the graph G described in Section 2.1. The graph signal is obtained by exciting the graph filter $\mathcal{H}(\mathbf{L})$ with an excitation $\mathbf{x}^\ell \in \mathbb{R}^N$,

$$\mathbf{y}^\ell = \mathcal{H}(\mathbf{L})\mathbf{x}^\ell + \mathbf{w}^\ell, \quad \ell = 1, \dots, L, \quad (11)$$

where $\mathbf{w}^\ell \in \mathbb{R}^N$ includes both the modeling and measurement error in data collection. We assume that \mathbf{w}^ℓ is zero mean and sub-Gaussian with $\mathbb{E}[\mathbf{w}^\ell(\mathbf{w}^\ell)^\top] = \sigma_w^2 \mathbf{I}$. Consider a low-rank excitation setting where $\{\mathbf{x}^\ell\}_{\ell=1}^L$ belong to an R -dimensional subspace of \mathbb{R}^N . Assume $K \leq R \ll N$, where K is the number of communities specified in Section 2.2. Let $\mathbf{B} \in \mathbb{R}^{N \times R}$ and

$$\mathbf{x}^\ell = \mathbf{B}\mathbf{z}^\ell, \quad (12)$$

where $\mathbf{z}^\ell \in \mathbb{R}^R$ is a latent parameter vector controlling the excitation signal. Under this model, the sampled covariance matrix of $\{\mathbf{y}^\ell\}_{\ell=1}^L$ is low rank with at most rank R . As mentioned, under such setting it is difficult to reconstruct \mathbf{L} from $\{\mathbf{y}^\ell\}_{\ell=1}^L$ using the existing methods [15–20].

Before discussing the proposed methods for inferring communities from $\{\mathbf{y}^\ell\}_{\ell=1}^L$ in Section 4 and 5, let us justify the model (11), (12) with three motivating examples.

3.1 Example 1: Diffusion Dynamics

The first example describes graph signals resulting from a diffusion process. For example, this model is commonly applied to temperatures within a geographical region [21]. Under this model, each node in the graph of Section 2.1 is a location and the weights $A_{ij} = A_{ji}$ represent the strengths of relative influence between i and j such that $\sum_{j=1}^N A_{ij} = 1$ for $i = 1, \dots, N$.

The ℓ th sample graph signal obtained is the result of a diffusion over T steps, described as

$$\begin{aligned} \mathbf{y}^\ell &= ((1 - \alpha)\mathbf{I} + \alpha\mathbf{A})^T \mathbf{x}^\ell + \mathbf{w}^\ell \\ &= (\mathbf{I} - \alpha\mathbf{L})^T \mathbf{x}^\ell + \mathbf{w}^\ell, \end{aligned} \quad (13)$$

where $\alpha \in (0, 1)$ is the speed of the diffusion process. As $(\mathbf{I} - \alpha\mathbf{L})^T$ is a polynomial of the graph's Laplacian, we observe that \mathbf{y}^ℓ is an output of a graph filter (11).

On the other hand, the excitation signal \mathbf{x}^ℓ may model the changes in temperature in the region due to a weather condition. The number of modes of temperature changes maybe limited, e.g., a typical hurricane in North America affects the east coast of the US. This effect can be captured by having a tall matrix \mathbf{B} , *i.e.*, the excitation lies in a low-dimensional space. The columns of \mathbf{B} represents the potential modes on which weather conditions may affect the region.

3.2 Example 2: Pricing Experiments in Consumers' Game

This example is concerned with graph signals obtained as the equilibrium consumption levels of a consumers' game subject to pricing experiments [25, 26]. Here, the graph described in Section 2.1 represents a network of N agents where $A_{ij} = A_{ji} \geq 0$ is the influence strength between agents i and j . We assume that $\mathbf{A}\mathbf{1} = c\mathbf{1}$ such that each agent experiences the same level of influence from the others.

It has been suggested in [26] that conducting a set of pricing experiments and observing the equilibrium behavior of agents can unveil the influence network between agents. Let ℓ be the index of a pricing experiment. Agent i chooses to consume y_i units of a product depending on (i) the price of the product p_i^ℓ and (ii) the consumption levels of other agents who are neighbors of him/her in the network, weighted by the influence strength A_{ij} . The consumption level y_i is determined by maximizing the utility

$$u_i(y_i, \mathbf{y}_{-i}, p_i^\ell) := ay_i - \frac{b}{2}y_i^2 + y_i \sum_{j=1}^N A_{ij}y_j - p_i^\ell y_i, \quad (14)$$

where $\mathbf{y}_{-i} := (y_j)_{j \neq i}$ and $a, b \geq 0$ are model parameters. As the utility function above depends on \mathbf{y}_{-i} , the equilibrium consumption level for the i th agent can be solved by the following network game:

$$y_i^\ell = \arg \max_{y_i \in \mathbb{R}_+} u_i(y_i, \mathbf{y}_{-i}^\ell, p_i^\ell), \quad \forall i. \quad (15)$$

Under the conditions that $b > \sum_{j=1}^N A_{ij}$ and $a > p_i^\ell$, the equilibrium to the above game is unique [25] and it satisfies

$$\mathbf{y}^\ell = (b\mathbf{I} - \mathbf{A})^{-1}(a\mathbf{1} - \mathbf{p}^\ell). \quad (16)$$

Removing the mean from \mathbf{y}^ℓ gives the graph signal:

$$\tilde{\mathbf{y}}^\ell := \frac{1}{L} \sum_{\tau=1}^L \mathbf{y}^\tau - \mathbf{y}^\ell = (b\mathbf{I} - \mathbf{A})^{-1} \tilde{\mathbf{p}}^\ell, \quad (17)$$

where $\tilde{\mathbf{p}}^\ell := \mathbf{p}^\ell - (1/L) \sum_{l=1}^L \mathbf{p}^l$ can be interpreted as a vector of discounts to agents during the ℓ th pricing experiment.

In fact, (17) can be interpreted as a filtered graph signal as in (11) by recognizing $\tilde{\mathbf{p}}^\ell$ as the excitation signal and $\tilde{\mathbf{y}}^\ell$ as the observed graph signal. Since $b > c$ and $\mathbf{A}\mathbf{1} = c\mathbf{1}$,

$$(b\mathbf{I} - \mathbf{A})^{-1} = \frac{1}{b-c} \sum_{t=0}^{\infty} \left(\frac{1}{b-c} \mathbf{L}\right)^t, \quad (18)$$

which is a matrix polynomial in \mathbf{L} . This shows that the linear operator $(b\mathbf{I} - \mathbf{A})^{-1}$ is indeed a graph filter.

Next, we study the types of discounts offered in the pricing experiment. A practical case is that due to the limitation of market, the pricing experiments only control the prices on R agents, while the prices of the rest are unchanged across experiments. This gives rise to a low-rank structure for the excitation signal. Note that $\tilde{\mathbf{p}}^\ell = \mathbf{B}\mathbf{z}^\ell$ holds with

$$[\mathbf{B}]_{\mathcal{I},:} = \mathbf{I}, [\mathbf{B}]_{[N] \setminus \mathcal{I},:} = \mathbf{0}, \quad (19)$$

where $\mathcal{I} \subset [N]$ is the index set of R agents whom prices are controlled, and $\mathbf{z}^\ell \in \mathbb{R}^R$ is simply a vector of the price variations from the mean. The latter can be assumed as known in a controlled experiment setting. The discount offered in the ℓ th experiment is a special case of low-rank excitation.

3.3 Example 3: DeGroot Dynamics with Stubborn Agents

The last example is related to a social network with N agents where the graph signals are opinions sampled from the agents on different topics, e.g., votes casted by Senators on different topics [33]. The network is represented by a *directed* graph $G = (V, E, \mathbf{A})$ such that $A_{ij} \geq 0$ captures the amount of ‘trust’ that agent i has on agent j . The agents are influenced by R *stubborn* agents in the sense that their opinions are not influenced by the others [34–36].

Consider the discussions on the ℓ th topic, the agents exchange opinions according to the DeGroot opinion dynamics [27] — let $y_i^\ell(\tau)$ (*resp.* z_j^ℓ) be the opinion of the i th agent (*resp.* j th stubborn agent) at time τ , e.g., $y_i^\ell(\tau) \in [0, 1]$ represents the probability for agent i to agree, we have

$$\mathbf{y}^\ell(\tau + 1) = \mathbf{A}\mathbf{y}^\ell(\tau) + \mathbf{B}\mathbf{z}^\ell, \quad \tau = 1, 2, \dots, \quad (20)$$

where $\mathbf{B} \in \mathbb{R}^{N \times R}$ is a weight matrix describing the bipartite graph that connects the stubborn agents to the agents in G . We assume that the concatenated matrix is stochastic such that $[\mathbf{A}, \mathbf{B}]\mathbf{1} = \mathbf{1}$ and therefore the updated opinions are convex combinations of the opinions of neighboring agents; see [22] for detailed description on the model. Note that it is possible to estimate the latent parameter \mathbf{z}^ℓ as well since the latter represents the opinions of stubborn agents.

Let us focus on the *steady-state* opinions, *i.e.*, the opinions when $\tau \rightarrow \infty$. Under mild assumptions, it holds [22, 37]

$$\begin{aligned} \mathbf{y}^\ell &:= \lim_{\tau \rightarrow \infty} \mathbf{y}^\ell(\tau) = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{B}\mathbf{z}^\ell \\ &= (\text{Diag}(\mathbf{1} - \mathbf{A}\mathbf{1}) + \mathbf{L})^{-1} \mathbf{B}\mathbf{z}^\ell \\ &\approx c^{-1} (\mathbf{I} + c^{-1} \mathbf{L})^{-1} \mathbf{B}\mathbf{z}^\ell, \end{aligned} \quad (21)$$

where the last approximation holds when there exists $c > 0$ such that $c\mathbf{1} \approx \mathbf{1} - \mathbf{A}\mathbf{1} = \mathbf{B}\mathbf{1}$, e.g., when the out-degrees of the stubborn agents are almost the same. From (21) it follows that the steady state opinions is a special case of (11), (12).

4 Blind Community Detection

We study the *blind community detection* problem, whose goal is to infer a disjoint partition of the nodes V that corresponds to the communities, $\mathcal{C}_1^*, \dots, \mathcal{C}_K^*$, in the graph $G = (V, E, \mathbf{A})$ as defined in Section 2.2, when the only given inputs are the observed graph signals $\{\mathbf{y}^\ell\}_{\ell=1}^L$ [cf. (11), (12)] and the desired number of communities K . Only in this section, we assume that the latent parameter vector \mathbf{z}^ℓ is a random, zero-mean, sub-Gaussian vector with $\mathbb{E}[\mathbf{z}^\ell(\mathbf{z}^\ell)^\top] = \mathbf{I}$. The covariance matrix of \mathbf{y}^ℓ is given by

$$\mathbf{C}_y := \mathbb{E}[\mathbf{y}^\ell(\mathbf{y}^\ell)^\top] = \mathcal{H}(\mathbf{L})\mathbf{B}\mathbf{B}^\top\mathcal{H}^\top(\mathbf{L}) + \sigma_w^2\mathbf{I}. \quad (22)$$

We also denote by $\overline{\mathbf{C}}_y := \mathcal{H}(\mathbf{L})\mathbf{B}\mathbf{B}^\top\mathcal{H}^\top(\mathbf{L})$ the covariance of \mathbf{y}^ℓ in the absence of measurement error. Observe that

$$\mathcal{H}(\mathbf{L})\mathbf{B} = \mathbf{V}\text{Diag}(\tilde{\mathbf{h}})\mathbf{V}^\top\mathbf{B}, \quad (23)$$

which is due to (3), (4). We can interpret $\mathcal{H}(\mathbf{L})\mathbf{B}$ as a *sketch* of the graph filter $\mathcal{H}(\mathbf{L})$, where \mathbf{B} is a sketch matrix that compresses the right dimension from N to R .

Algorithm 1 Blind Community Detection (BlindCD).

- 1: **Input:** Graph signals $\{\mathbf{y}^\ell\}_{\ell=1}^L$; desired number of communities K .
- 2: Compute the sample covariance $\hat{\mathbf{C}}_y$ as

$$\hat{\mathbf{C}}_y = (1/L) \sum_{\ell=1}^L \mathbf{y}^\ell (\mathbf{y}^\ell)^\top. \quad (24)$$

- 3: Find the top- K eigenvectors of $\hat{\mathbf{C}}_y$ (with the eigenvalues sorted in *descending* order). Denote the set of eigenvectors as $\hat{\mathbf{V}}_K \in \mathbb{R}^{N \times K}$.
- 4: Apply the K -means method, which seeks to optimize

$$\min_{\mathcal{C}_1, \dots, \mathcal{C}_K \subseteq V} \sum_{k=1}^K \sum_{i \in \mathcal{C}_k} \left\| \hat{\mathbf{v}}_i^{\text{row}} - \frac{1}{|\mathcal{C}_k|} \sum_{j \in \mathcal{C}_k} \hat{\mathbf{v}}_j^{\text{row}} \right\|_2^2, \quad (25)$$

where $\hat{\mathbf{v}}_i^{\text{row}} := [\hat{\mathbf{V}}_K]_{i,:} \in \mathbb{R}^K$.

- 5: **Output:** K communities $\hat{\mathcal{C}}_1, \dots, \hat{\mathcal{C}}_K$.
-

To perform blind community detection based on $\{\mathbf{y}^\ell\}_{\ell=1}^L$, let us gain intuition by considering the scenario when the noise is small ($\sigma_w^2 \approx 0$), the first K elements in $\tilde{\mathbf{h}}$ are non-zero which have larger magnitudes than the rest of elements, and the columns of \mathbf{B} span the same space as $\text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_K\}$. In this scenario, from (22) and (23), we observe that \mathbf{V}_K can be estimated (up to a rotation) by simply obtaining the top- K eigenvectors of \mathbf{C}_y . This intuition suggests that we can detect communities by applying spectral clustering on \mathbf{C}_y , similar to the one applied to the Laplacian \mathbf{L} in Section 2.2. The proposed BlindCD method is summarized in Algorithm 1.

The computation complexity of BlindCD is dominated by covariance estimation and eigenvalue decomposition in Line 2-3, which costs $\mathcal{O}(N^2(L + K))$ FLOPS for large N . This is significantly less complex than a two-step procedure using a sophisticated graph learning step, e.g., [18]. In addition to estimating the covariance, the latter requires a linear program with $\mathcal{O}(N^2)$ variables and constraints. This learning step entails a total complexity of $\mathcal{O}(N^2L + N^7 \log \epsilon_{\text{acc}}^{-1})$ FLOPS with the interior point method in [38]¹, where $\epsilon_{\text{acc}} > 0$ is the accuracy.

Similar method to the BlindCD method have been proposed in the data clustering literature [28], offering a simple interpretation of \mathbf{C}_y as the similarity graph between nodes. We provide a different interpretation here. Precisely, we view \mathbf{C}_y as a spectral sketch of the Laplacian \mathbf{L} and analyze the performance of BlindCD as an indirect algorithm to *approximately* find the ground truth communities in \mathbf{L} .

4.1 Low-pass Graph Filters

Following (23) and the ensuing discussion, the performance of BlindCD depends on $\tilde{\mathbf{h}}$, the frequency response of the graph filter. In particular, a desirable situation would be one where $\tilde{\mathbf{h}}$ contains only significant entries over the first K elements; in this way, the graph filter $\mathcal{H}(\mathbf{L})$ is approximately rank K and retains all the eigenvectors required for spectral clustering. To quantify the above conditions, we formally introduce the notion of a *low-pass graph filter* (LPGF) as follows.

¹In practice, the said linear program can be solved efficiently with a tailor-made solver such as [39].

Definition 1 A graph filter $\mathcal{H}(\mathbf{L})$ is a (K, η) -LPGF if

$$\eta := \frac{\max\{|\tilde{h}_{K+1}|, \dots, |\tilde{h}_N|\}}{\min\{|\tilde{h}_1|, \dots, |\tilde{h}_K|\}} < 1, \quad (26)$$

where \tilde{h}_i is defined in (4). The LPGF is ideal if $\eta = 0$.

Note that a small η implies a ‘good’ LPGF, since $\eta \ll 1$ implies that most of the energy is concentrated in the first K frequency bins of the graph filter. In fact, as we show later in Section 4.2, the low-pass coefficient η plays an important role in the performance of BlindCD.

We now survey a few graph filter designs that are LPGF and comment on their low-pass coefficients η .

Example 1 Consider the filter order $T_d < \infty$ and

$$\mathcal{H}_1(\mathbf{L}) = (\mathbf{I} - \alpha\mathbf{L})^{T_d-1}, \quad \alpha \in (0, 1/\lambda_N). \quad (27)$$

This filter models a discrete time diffusion process after $(T_d - 1)$ time instances on the graph [40]. In particular,

$$\eta_1 = \left(\frac{1 - \alpha\lambda_{K+1}}{1 - \alpha\lambda_K} \right)^{T_d-1}. \quad (28)$$

Observe that the coefficient η_1 improves exponentially with T_d .

Example 2 Consider

$$\mathcal{H}_2(\mathbf{L}) = (\mathbf{I} + c^{-1}\mathbf{L})^{-1}, \quad (29)$$

for some $c > 0$. This filter is analogous to a single-pole infinite impulse response (IIR) filter in classical signal processing. Its low-pass coefficient can be bounded as

$$\eta_2 = \frac{1 + c^{-1}\lambda_K}{1 + c^{-1}\lambda_{K+1}} = 1 - c^{-1} \frac{\lambda_{K+1} - \lambda_K}{1 + c^{-1}\lambda_{K+1}}. \quad (30)$$

Observe that the coefficient $\eta_2 \approx 1$ for $\lambda_{K+1} \gg 1$ or $c \gg 1$.

Example 1 is related to the diffusion dynamics in Section 3.1, while Example 2 is related to the consumers’ game and opinion dynamics in Sections 3.2 and 3.3. For further reference, an overview of graph filters and their relevant network processes can be found in [1, 3].

We conclude this subsection by characterizing the low-pass coefficient η from the properties of the generating function $h(\lambda)$. To simplify the analysis, we consider the class of filters such that $h(\lambda)$ satisfies the following assumption.

Assumption 1 The generating function $h(\lambda)$ is non-negative and non-increasing for all $\lambda \geq 0$.

Note that Assumption 1 holds for the graph filters in Examples 1 and 2. The following observation gives a bound on η using the first and second order derivatives of $h(\lambda)$.

Observation 1 Suppose that Assumption 1 holds and that $h(\lambda)$ is L_h -smooth and μ_h -strongly convex for $\lambda \in [\lambda_K, \lambda_{K+1}]$, where $0 \leq \mu_h \leq L_h$. Then, the graph filter $\mathcal{H}(\mathbf{L})$ is a (K, η) -LPGF with

$$\begin{aligned}\eta &\leq 1 - \frac{1}{\tilde{h}_K} \left(\frac{\mu_h}{2} \Delta\lambda_K^2 - h'(\lambda_{K+1}) \Delta\lambda_K \right), \\ \eta &\geq 1 - \frac{1}{\tilde{h}_K} \left(\frac{L_h}{2} \Delta\lambda_K^2 - h'(\lambda_{K+1}) \Delta\lambda_K \right),\end{aligned}\tag{31}$$

where $\Delta\lambda_K := \lambda_{K+1} - \lambda_K$ is the spectral gap of \mathbf{L} .

The observation can be verified using the definitions of L_h -smooth and μ_h -strongly convex functions [41]. Note that Assumption 1 implies that $h(\lambda)$ is convex and the derivative $h'(\lambda_{K+1})$ is non-positive. Consequently, the upper bound on η depends on the spectral gap $\Delta\lambda_K$ and the magnitude of \tilde{h}_K . In particular, for a constant spectral gap, a small \tilde{h}_K leads to $\eta \approx 0$ and thus a good LPGF.

4.2 Performance Analysis

This subsection shows that under the GSP model (11), (12) and using Definition 1, we can bound the ‘suboptimality’ of the communities obtained by BlindCD compared to the ‘optimal’ ones found using spectral clustering on \mathbf{L} [cf. Section 2.2]. Together with recent advances in the theoretical analysis of spectral clustering [29], this result allows us to quantify the accuracy of BlindCD to perform *blind community detection* and provides new insights on how to improve its performance.

To proceed, first let us take the K -means objective function $F(\cdot)$ in (9) constructed from eigenvectors of \mathbf{L} as our performance metric. Let us denote

$$F^* := \min_{\mathcal{C}_1, \dots, \mathcal{C}_K \subseteq V} F(\mathcal{C}_1, \dots, \mathcal{C}_K)\tag{32}$$

as the optimal objective value. Furthermore, $\widehat{\mathbf{C}}_y$ is the sampled covariance of $\{\mathbf{y}^\ell\}_{\ell=1}^L$ and $\overline{\mathbf{C}}_y$ is the covariance of \mathbf{y}^ℓ in the absence of noise. The ensuing performance guarantee follows:

Theorem 1 Under the following conditions:

1. $\mathcal{H}(\mathbf{L})$ is a (K, η) -LPGF [cf. Definition 1],
2. $\text{rank}(\mathbf{V}_K \text{diag}(\tilde{\mathbf{h}}_K) \mathbf{V}_K^\top \mathbf{B} \mathbf{Q}_K) = K$, where \mathbf{Q}_K is the top- K right singular vector of $\mathcal{H}(\mathbf{L})\mathbf{B}$.
3. $\text{rank}(\mathcal{H}(\mathbf{L})\mathbf{B}) \geq K$,
4. There exists $\delta > 0$ such that

$$\delta := \beta_K(\overline{\mathbf{C}}_y) - \beta_{K+1}(\overline{\mathbf{C}}_y) - \|\widehat{\mathbf{C}}_y - \overline{\mathbf{C}}_y\|_2 > 0,\tag{33}$$

where $\beta_K(\overline{\mathbf{C}}_y)$ is the K th largest eigenvalue of $\overline{\mathbf{C}}_y$.

For any $\epsilon > 0$, if the partition $\hat{\mathcal{C}}_1, \dots, \hat{\mathcal{C}}_K$ found by BlindCD is a $(1 + \epsilon)$ -optimal solution² to problem (25), then,

$$\begin{aligned}&\sqrt{F(\hat{\mathcal{C}}_1, \dots, \hat{\mathcal{C}}_K)} - \sqrt{(1 + \epsilon)F^*} \\ &\leq (2 + \epsilon)\sqrt{2K} \left(\sqrt{\frac{\gamma^2}{1 + \gamma^2}} + \frac{\|\widehat{\mathbf{C}}_y - \overline{\mathbf{C}}_y\|_2}{\delta} \right),\end{aligned}\tag{34}$$

²This means that the objective value obtained is at most $(1 + \epsilon)$ times the optimal value. See [32] for a polynomial-time algorithm achieving this.

where γ is bounded by

$$\gamma \leq \eta \|\mathbf{V}_{N-K}^\top \mathbf{B} \mathbf{Q}_K\|_2 \|(\mathbf{V}_K^\top \mathbf{B} \mathbf{Q}_K)^{-1}\|_2. \quad (35)$$

The proof (inspired by [42], also see [43]) can be found in Appendix A. Condition 1) requires that the graph filter involved is an LPGF. This natural requisite imposes that the frequency response must be higher for those eigenvectors that capture the community structure in the graph. Conditions 2) and 3) are technical requirements implying that the rank R of the excitation matrix \mathbf{B} cannot be smaller than the number of clusters K that we are trying to recover. Lastly, condition 4) imposes a restriction on the distance between the true covariance $\overline{\mathbf{C}}_y$ and the observed one $\widehat{\mathbf{C}}_y$. This condition may be violated if the spectral gap $\beta_K(\overline{\mathbf{C}}_y) - \beta_{K+1}(\overline{\mathbf{C}}_y)$ is small or, relying on Lemma 1, if the noise power σ_w^2 is large.

Moreover, Eq. (34) in Theorem 1 bounds the optimality gap for the communities found applying BlindCD compared to F^* in (32). We first observe that the performance decreases when the number of communities K increases, which is natural. This bound consists of the sum of two contributions. The first term is a function of γ , which in turn depends on the low-pass coefficient η of the LPGF involved as well as the alignment between the matrices $\mathbf{B} \mathbf{Q}_K$ and \mathbf{V}_{N-K} . From (35), the recovered communities are more accurate when: 1) the LPGF is close to ideal ($\eta \approx 0$) and 2) the distortion induced by \mathbf{B} on the relevant eigenvectors \mathbf{V}_K is minimal. The second term in (34) depends on the distance between $\widehat{\mathbf{C}}_y$ and $\overline{\mathbf{C}}_y$, capturing the combined effect of noise in the observations (via σ_w^2) as well as the finite sample size. To further control this term, if we define $\mathbf{\Delta} := \widehat{\mathbf{C}}_y - \overline{\mathbf{C}}_y$, the next result follows.

Lemma 1 [44, Remark 5.6.3, Exercise 5.6.4] *Suppose that i) $\mathbf{y}^1, \dots, \mathbf{y}^L$ are independent, and ii) they are bounded almost surely with $\|\mathbf{y}^\ell\|_2 \leq Y$. Let the effective rank of \mathbf{C}_y be $r := \text{Tr}(\mathbf{C}_y) / \|\mathbf{C}_y\|_2$, then for every $c > 0$ with probability at least $1 - c$, one has that*

$$\|\mathbf{\Delta}\|_2 \leq \sigma_w + C \left(\sqrt{\frac{Y^2 r \log(N/c)}{L}} + \frac{Y^2 r \log(N/c)}{L} \right), \quad (36)$$

for some constant C that is independent of N, r, L, c , and σ_y .

Condition ii) in Lemma 1 is satisfied if \mathbf{y}^ℓ is sub-Gaussian and $N \gg 1$. From Lemma 1 it follows that the error converges to σ_w at the rate of $\mathcal{O}(\sqrt{rK^2 \log(N)/L})$. For our model, it can be verified that $r \approx R \ll N$, where R is the rank of \mathbf{B} and the sampling complexity is significantly reduced compared to a signal model with full-rank excitations.

In a nutshell, Theorem 1 illustrates the effects that the observation noise, the finite number of observations, and the low-pass structure of the filter have on the suboptimality of the communities obtained. As discussed above, the low-pass coefficient η plays an important role in the performance of BlindCD. While η is determined by the dynamics that generates the graph signals $\{\mathbf{y}^\ell\}_{\ell=1}^L$, it is possible to improve this coefficient, as described in the next section.

5 Boosted Blind Community Detection

The performance analysis in the previous section shows that the performance of BlindCD depends on the low-pass filter coefficient η . While it is impossible to change the graph filter that generates the data, this section presents a ‘boosting’ technique that extracts an improved low-pass filtered

component, *i.e.*, one with a smaller η , from the observed graph signals. For the application of the boosting technique, we shall work with low-pass graph filters satisfying Assumption 1 and consider a data model where, apart from the access to the graph signals $\mathbf{y}^\ell \in \mathbb{R}^N$ we also have access to the latent parameter vector $\mathbf{z}^\ell \in \mathbb{R}^R$ [cf. (11), (12)]. This scenario can be justified in the example of pricing experiments [cf. Section 3.2] when the price discounts are directly controlled by the seller attempting to estimate the network; or in the example of DeGroot dynamics [cf. Section 3.3] where the latent parameter vectors are the opinions of the stubborn agents.

First, the input-output pairs $\{\mathbf{z}^\ell, \mathbf{y}^\ell\}_{\ell=1}^L$ enable us to estimate the $N \times R$ matrix $\mathcal{H}(\mathbf{L})\mathbf{B}$ via the least square estimator

$$\mathcal{H}^* \in \arg \min_{\hat{\mathcal{H}} \in \mathbb{R}^{N \times R}} \frac{1}{L} \sum_{\ell=1}^L \left\| \mathbf{y}^\ell - \hat{\mathcal{H}} \mathbf{z}^\ell \right\|_2^2, \quad (37)$$

where the solution is unique when $L \geq R$ and $\{\mathbf{z}^\ell\}_{\ell=1}^L$ spans \mathbb{R}^R . Importantly, we note the decomposition:

$$\mathcal{H}(\mathbf{L})\mathbf{B} = \tilde{\mathcal{H}}(\mathbf{L})\mathbf{B} + \tilde{h}_N \mathbf{B}, \quad (38)$$

where

$$\tilde{\mathcal{H}}(\mathbf{L}) := \mathcal{H}(\mathbf{L}) - \tilde{h}_N \mathbf{I} \quad (39)$$

is a graph filter with the generating function $\tilde{h}(\lambda) = h(\lambda) - \tilde{h}_N$. The graph filter $\tilde{\mathcal{H}}(\mathbf{L})$ is called a *boosted LPGF* as it has a smaller low-pass coefficient, denoted by $\tilde{\eta}$, than the low-pass coefficient of the original $\mathcal{H}(\mathbf{L})$. This can be seen since (i) the magnitude of the boosted K th frequency response is reduced to $\tilde{h}_K - \tilde{h}_N$; (ii) the first and second order derivatives of $\tilde{h}(\lambda)$ are the same as $h(\lambda)$. Applying Observation 1 it follows that $\tilde{\mathcal{H}}(\mathbf{L})$ has a smaller low-pass coefficient $\tilde{\eta}$ by replacing \tilde{h}_K by $\tilde{h}_K - \tilde{h}_N$ in (31). Concretely, we observe the example.

Example 3 (*Boosted single-pole IIR filter*). Consider

$$\mathcal{H}_3(\mathbf{L}) := \mathcal{H}_2(\mathbf{L}) - (1 + c^{-1}\lambda_N)^{-1} \mathbf{I}, \quad (40)$$

where $\mathcal{H}_2(\mathbf{L})$ was defined in (29) and we note that $\tilde{h}_N = (1 + c^{-1}\lambda_N)^{-1}$. We have

$$\eta_3 = \frac{\lambda_N - \lambda_{K+1}}{\lambda_N - \lambda_K} \frac{1 + c^{-1}\lambda_K}{1 + c^{-1}\lambda_{K+1}} = \left(\frac{\lambda_N - \lambda_{K+1}}{\lambda_N - \lambda_K} \right) \eta_2.$$

It follows that $\eta_3 \ll \eta_2$ whenever $\lambda_{K+1} \gg \lambda_K$.

In general, the discussion above shows that it is possible to reduce the low-pass coefficient η significantly by adjusting the constant level of the frequency responses in graph filters. As a result, applying spectral clustering based on the top- K left singular vectors of $\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$ will return a more accurate community detection result.

In order to estimate $\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$ from \mathcal{H}^* as in (38), one needs, in principle, to have access to \mathbf{B} and the frequency response \tilde{h}_N . However, our goal is to obtain a boosting effect in the absence of knowledge about \mathbf{B} and \tilde{h}_N . A key towards achieving this goal is to notice that $\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$ is close to a rank- K matrix since $\tilde{\mathcal{H}}(\mathbf{L})$ has a small low-pass coefficient $\tilde{\eta}$. Hence, for $R > K$, it follows from (38) that \mathcal{H}^* can be decomposed into a low-rank matrix and a scaled version of the sketch matrix \mathbf{B} . This motivates us to consider the noisy matrix decomposition problem proposed in [45]:

$$\begin{aligned} \min_{\hat{\mathcal{S}}, \hat{\mathbf{B}} \in \mathbb{R}^{N \times R}} \quad & \frac{1}{2} \left\| \mathcal{H}^* - \hat{\mathcal{S}} - \hat{\mathbf{B}} \right\|_F^2 + \kappa \|\hat{\mathcal{S}}\|_{\sigma,1} + \rho g(\hat{\mathbf{B}}) \\ \text{s.t.} \quad & g^*(\hat{\mathcal{S}}) \leq \alpha, \end{aligned} \quad (41)$$

Algorithm 2 Boosted BlindCD method.

- 1: **Input:** Graph signals and excitation signals $\{\mathbf{y}^\ell, \mathbf{z}^\ell\}_{\ell=1}^L$; desired number of communities K .
 - 2: Solve the convex optimization problem (41) and denote its solution as $(\widehat{\mathcal{S}}^*, \widehat{\mathbf{B}}^*)$.
 - 3: Find the top- K left singular vectors of $\widehat{\mathcal{S}}^*$ and denote the set of singular vectors as $\widetilde{\mathcal{S}}_K \in \mathbb{R}^{N \times K}$.
 - 4: Apply the K -means method on the row vectors of $\widetilde{\mathcal{S}}_K$.
 - 5: **Output:** K communities $\widetilde{\mathcal{C}}_1, \dots, \widetilde{\mathcal{C}}_K$.
-

where $\|\widehat{\mathcal{S}}\|_{\sigma,1}$ is the trace norm of the matrix $\widehat{\mathcal{S}}$, \mathcal{H}^* is a solution to (37), $\alpha, \kappa, \rho > 0$ are predefined parameters, $g(\cdot)$ is a decomposable regularizer of $\widehat{\mathbf{B}}$, which is a norm chosen according to the prior knowledge on the unknown sketch matrix \mathbf{B} and $g^*(\cdot)$ is its dual norm. A few examples for choices of $g(\cdot)$ are listed below.

- *Localized excitation:* We set

$$g_1(\widehat{\mathbf{B}}) = \|\text{vec}(\widehat{\mathbf{B}})\|_1, \quad g_1^*(\widehat{\mathcal{S}}) = \|\text{vec}(\widehat{\mathcal{S}})\|_\infty. \quad (42)$$

This regularization forces the solution $\widehat{\mathbf{B}}^*$ to (41) to be an element-wise sparse matrix. This corresponds to the scenario where each element of the latent variables in \mathbf{z}^ℓ excites only a few of the nodes in our graph.

- *Small number of excited nodes:* Let $\widehat{\mathbf{b}}_i^{\text{row}}$ be the i th row vector of $\widehat{\mathbf{B}}$. We then set

$$g_2(\widehat{\mathbf{B}}) = \sum_{i=1}^N \|\widehat{\mathbf{b}}_i^{\text{row}}\|_2, \quad g_2^*(\widehat{\mathcal{S}}) = \max_{i=1, \dots, N} \|\widehat{\mathbf{s}}_i^{\text{row}}\|_2. \quad (43)$$

This regularization is motivated by the group-sparsity formulation in [46] which forces the solution $\widehat{\mathbf{B}}^*$ to (41) to be row-sparse. Notice that this is relevant when the graph filter is excited on a small number of nodes.

- *Small perturbation:* We set

$$g_3(\widehat{\mathbf{B}}) = \|\widehat{\mathbf{B}}\|_{\text{F}}, \quad g_3^*(\widehat{\mathcal{S}}) = \|\widehat{\mathcal{S}}\|_{\text{F}}. \quad (44)$$

This regularization models each entry of $\widetilde{h}_N \mathbf{B}$ as a Gaussian random variable of small, identical variance. This can be used when there is no prior knowledge on \mathbf{B} .

Notice that for every choice of the regularizer $g(\cdot)$ discussed, (41) is a convex problem that can be solved in polynomial time. Let the optimal solution to (41) be $\widehat{\mathcal{S}}^*, \widehat{\mathbf{B}}^*$. We apply spectral method on $\widehat{\mathcal{S}}^*$ based on its top- K left singular vectors. The boosted BlindCD method is overviewed in Algorithm 2.

5.1 Performance Analysis

This section analyzes the performance of the *boosted* BlindCD method, mimicking the ideas in Section 4.2. Due to the space limitation, we focus on the special case where \mathbf{B} is sparse and select $g_1(\widehat{\mathbf{B}})$ in (42) when solving (41).

Our first step towards deriving a theoretical bound for the performance of boosted BlindCD is to characterize the estimation error of $\mathcal{H}(\mathbf{L})\mathbf{B}$ when solving (37), defined as

$$\mathcal{E} := \mathcal{H}^* - \mathcal{H}(\mathbf{L})\mathbf{B}. \quad (45)$$

Lemma 2 *Suppose that $L \geq R$, $\{\mathbf{z}^\ell\}_{\ell=1}^L$ spans \mathbb{R}^R , and $\|\mathbf{w}^\ell(\mathbf{z}^\ell)^\top\| < \infty$ almost surely. For every $c > 0$ and with probability at least $1 - 2c$, it holds that*

$$\|\mathcal{E}\|_2 = \mathcal{O}\left(\frac{\sigma_w \log((N+R)/c)}{\sqrt{L}}\right). \quad (46)$$

The proof can be found in Appendix B. Lemma 2 captures the expected behavior of a vanishing estimation error when $L \rightarrow \infty$. Next, we show that $\widehat{\mathbf{L}}^*$ from (41) is close to $\widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$ by leveraging the fact that the latter is approximately rank- K .

Lemma 3 [45, Corollary 1] *Consider problem (41) with*

$$\begin{aligned} \kappa &\geq 4\|\mathcal{E}\|_2, \quad \rho \geq 4\left(\frac{\alpha}{\sqrt{NR}} + \|\text{vec}(\mathcal{E})\|_\infty\right), \\ \alpha &\geq \sqrt{NR} \|\text{vec}(\widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B})\|_\infty. \end{aligned} \quad (47)$$

Let $R \geq K$. There exists constants c_1, c_2 such that

$$\begin{aligned} \|\widehat{\mathbf{S}}^* - \widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B}\|_{\mathbb{F}}^2 + \|\widehat{\mathbf{B}}^* - \tilde{h}_N \mathbf{B}\|_{\mathbb{F}}^2 &\leq \\ c_1 \kappa^2 \left(K + \frac{1}{\kappa} \sum_{j=K+1}^R \sigma_j(\widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B})\right) + c_2 \rho^2 \|\text{vec}(\mathbf{B})\|_0. \end{aligned} \quad (48)$$

The term $\sum_{j=K+1}^R \sigma_j(\widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B})$ is negligible when $\widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$ is approximately rank- K . Therefore, the implication is that the distance between $\widehat{\mathbf{S}}^*$ and $\widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$ can be bounded by the sum of two terms — one that is dependent on \mathcal{E} , and one that is dependent on α/\sqrt{NR} . Overall, it shows that the error reduces when the excitation rank R and number of observations L increases. On the other hand, (47) suggests that one should set $\kappa = c_1/\sqrt{L}$, $\rho = c_2/\sqrt{RL}$ in (41) for some c_1, c_2 for the optimal performance.

Having established these results, the boosted BlindCD method is an approximation of BlindCD operating on the boosted LPGF $\widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$. Next, we define the SVD of $\widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$ as $\widetilde{\mathbf{V}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{Q}}^\top$ and analyze the performance of the boosted BlindCD through a minor modification of Theorem 1.

Corollary 1 *Suppose that Conditions 1 to 3 in Theorem 1 are met when replacing \mathbf{Q}_K by $\widetilde{\mathbf{Q}}_K$ and $\mathcal{H}(\mathbf{L})$ by $\widetilde{\mathcal{H}}(\mathbf{L})$. Let $\widetilde{\mathbf{\Delta}} := \widehat{\mathbf{S}}^* - \widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$ and assume that*

$$\widetilde{\delta} := \sigma_K(\widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B}) - \sigma_{K+1}(\widetilde{\mathcal{H}}(\mathbf{L})\mathbf{B}) - \|\widetilde{\mathbf{\Delta}}\|_2 > 0. \quad (49)$$

If Step 4 in the boosted BlindCD method finds an $(1 + \epsilon)$ optimal solution to the K -means problem, where $\epsilon > 0$, then,

$$\begin{aligned} \sqrt{F(\tilde{\mathcal{C}}_1, \dots, \tilde{\mathcal{C}}_K)} - \sqrt{(1 + \epsilon)F^*} &\leq \\ (2 + \epsilon)\sqrt{2K} \left(\sqrt{\frac{\tilde{\gamma}^2}{1 + \tilde{\gamma}^2}} + \frac{\|\widetilde{\mathbf{\Delta}}\|_2}{\widetilde{\delta}} \right), \end{aligned} \quad (50)$$

where $F(\cdot)$, F^* are defined in (9), (32), respectively, and

$$\tilde{\gamma} \leq \tilde{\eta} \|\mathbf{V}_{N-K}^\top \mathbf{B} \tilde{\mathbf{Q}}_K\|_2 \|(\mathbf{V}_K^\top \mathbf{B} \tilde{\mathbf{Q}}_K)^{-1}\|_2, \quad (51)$$

where $\tilde{\eta}$ is the low-pass coefficient of the boosted LPGF $\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$.

The proof of Corollary 1 can be found in Appendix C. We see that the performance of the boosted BlindCD method depends on $\tilde{\eta}$, the low-pass coefficient of the boosted LPGF. As $\tilde{\eta} \ll \eta$ due to our prior discussions, it is anticipated that the boosted method achieves a much better performance, especially when the original LPGF is not markedly low-pass.

While the bound in (50) is similar to that in Theorem 1, we observe that applying Lemma 2 and Lemma 3 yields

$$\|\tilde{\Delta}\|_2 \leq \|\tilde{\Delta}\|_{\text{F}} = \mathcal{O} \left(\sigma_{K+1}(\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}) + \frac{1}{\sqrt{L}} + \frac{1}{\sqrt{NR}} \right).$$

From the definition of $\tilde{\delta}$ we have

$$\frac{\|\tilde{\Delta}\|_2}{\tilde{\delta}} = \mathcal{O} \left(\frac{\sigma_{K+1}(\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}) + 1/\sqrt{L} + 1/\sqrt{NR}}{\sigma_K(\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}) - C\sigma_{K+1}(\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B})} \right), \quad (52)$$

for some constant C . Substituting (52) into (50) shows that the sub-optimality of boosted BlindCD can be minimized when 1) the spectral gap for the sketched matrix $\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$, 2) the number of samples L , and 3) the excitation rank R , are large. We remark that it is possible to undertake analogous performance analysis for the other proposed regularizers on \mathbf{B} [cf. (43) and (44)]. For example, this can be done using [47] and replacing Lemma 3 with the corresponding result. These extensions, however, are beyond the scope of the current paper.

6 Numerical Examples

To illustrate the efficacy of the BlindCD methods, we study three application examples that pertain to consensus dynamics, consumer networks, and social networks. Numerical examples will be given for these applications, which were introduced in Sections 3.1 through 3.3.

Unless otherwise specified, the graphs used in the simulations will be generated according to a *stochastic block model* (SBM) [48], denoted by $G \sim \text{SBM}(N, K, a, b)$, such that G has N nodes, K equal-sized non-overlapping communities and the intra (*resp.* inter) community connectivity probability is $a \in [0, 1]$ (*resp.* $b \in [0, 1]$). The weights on the graph, A_{ij} , are set to 1 if $(i, j) \in E$ and 0 otherwise. We use the ground truth community membership in generating the SBM graphs when evaluating the accuracies. The error rate is given by

$$P_e := \mathbb{E} \left[\frac{1}{N} \min_{\pi: [K] \rightarrow [K]} \sum_{i=1}^N \mathbb{1}_{\pi(c_i) \neq c_i^{\text{true}}} \right], \quad (53)$$

and the above is approximated via Monte-Carlo simulations, where $\mathbb{1}_{\mathcal{E}}$ is an indicator function for the event \mathcal{E} , $\pi: [K] \rightarrow [K]$ is a permutation function and $c_i \in [K]$ (*resp.* c_i^{true}) is the detected (*resp.* true) community membership of node i .

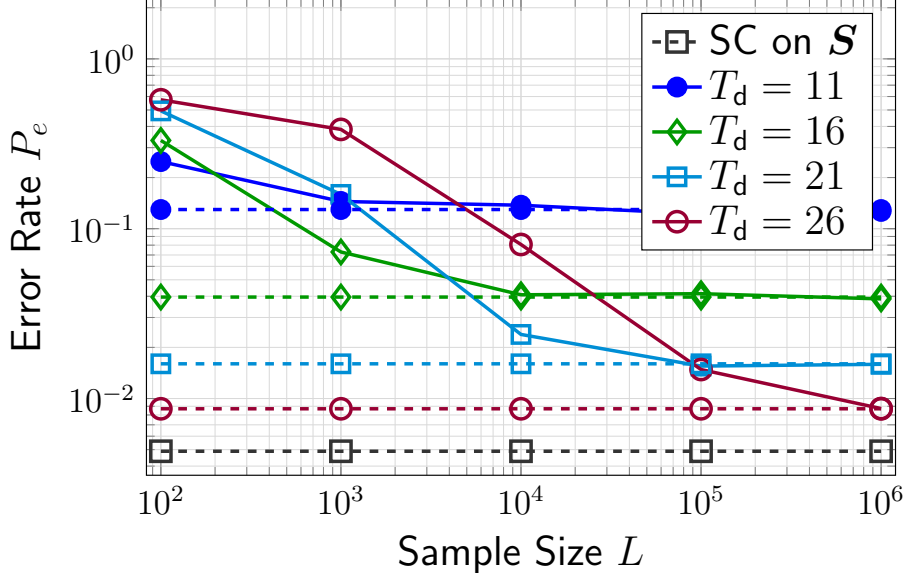


Figure 1: **Community detection performance versus sample size L** . We consider graphs generated as $G \sim \text{SBM}(N, K, 8 \log N/N, \log N/N)$ with $N = 150, K = 3$ and fix the excitation rank at $R = 15$. The solid (*resp.* dashed) lines show the performance of BlindCD on the sampled *output* covariance $\widehat{\mathbf{C}}_y$ (*resp.* true and noiseless covariance $\overline{\mathbf{C}}_y$).

6.1 Diffusion Dynamics

We first evaluate the performance of BlindCD using graph signals generated according to the observation model in (11). We focus on the diffusion dynamics in Section 3.1. We perform Monte-Carlo simulations to evaluate the community detection performance on random graphs. In this example, the SBM graphs generated are $G \sim \text{SBM}(N, K, 8 \log N/N, \log N/N)$ with $N = 150$ and $K = 3$. We simulate a scenario where the graph filter is excited on only R nodes. In this case, the sketch matrix \mathbf{B} is generated by first picking R rows uniformly from the N available rows, and the elements in each selected rows are set to one uniformly with probability $p_b = 0.5$. For the boosted BlindCD method, we test the formulation of (41) with regularizers $g_1(\widehat{\mathbf{B}})$ and $g_2(\widehat{\mathbf{B}})$ [cf. (42) and (43)] by setting $\kappa = 2/\sqrt{L}$ and $\rho = 0.5/\sqrt{RL}$. The variance of observation noise is $\sigma_w^2 = 10^{-2}$ and each element of \mathbf{z}^ℓ is generated independently as $[\mathbf{z}^\ell]_i \sim \mathcal{U}[-1, 1]$.

The first example examines the effect of the graph filter’s low-pass coefficient η and sample size L on the performance of BlindCD. In particular, Fig. 1 shows the performance of community detection for different filter orders T_d against the number of samples L accrued. Notice that the low-pass coefficient η decreases with the filter order T_d [cf. (28)]. As such, we observe that the performance improves with T_d . The error rate approaches that achieved by applying spectral clustering on the actual \mathbf{L} . An interesting observation is that for sample covariances, as T_d increases, the sample size L required to reach the performance of noiseless covariance also increases. This can be explained with the condition (33) in Theorem 1. In particular, as T_d increases, the absolute value of $\beta_K(\overline{\mathbf{C}}_y) - \beta_{K+1}(\overline{\mathbf{C}}_y)$ decreases, therefore restricting $\|\widehat{\mathbf{C}}_y - \overline{\mathbf{C}}_y\|_2$ to be smaller [cf. (33)]. The latter is satisfied when the number of samples accrued is sufficiently large.

The second example shows the effect of the excitation rank R . The results are shown in Fig. 2 where we have fixed $L = 10^3$ and $T_d = 16$. In this example, we have compared the performance of

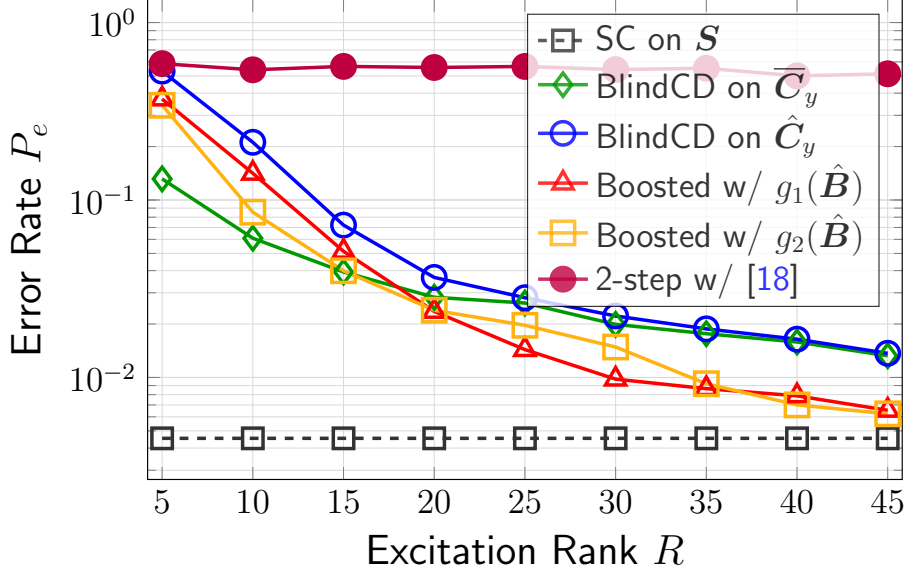


Figure 2: **Community detection performance versus excitation rank R .** We consider graphs generated as $G \sim \text{SBM}(N, K, 8 \log N/N, \log N/N)$ with $N = 150, K = 3$ and fixed $T_d = 16, L = 10^3$.

BlindCD to a 2-step procedure which uses [18] (with efficient implementation in [39]) to recover the GSO, then it applies spectral clustering on the recovered GSO to detect communities. For BlindCD, we observe that the performance improves with the rank R , while the 2-step procedure performs poorly³. As predicted by Corollary 1, the boosting technique enhances the performance of BlindCD.

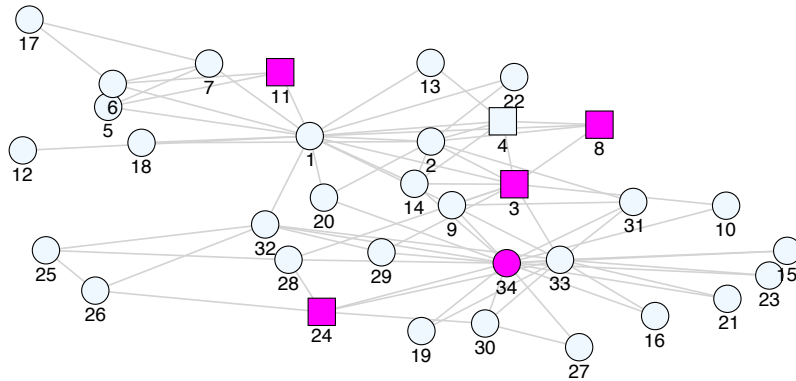
The example in Fig. 3 shows the performance of an instance of BlindCD on the Zachary’s Karate Club network when the graph signals are generated from the diffusion dynamics. To capitalize on the benefit of the *boosting* technique, we consider a scenario with a filter order of $T_d = 6$, observation rank of $R = 5$ (the graph is excited on just 5 nodes) and we observe $L = 10^3$ noisy samples of the graph signals. Observe that the low-pass coefficient for the filter may be close to 1 as T_d is small. This explains the poorer performance of BlindCD in Fig. 3.(b). The boosted BlindCD, instead, delivers good performance as it identifies the two communities in the network except for a miss-classification of agent 17. Through sorting the row sums of the estimated $\hat{\mathbf{B}}$, we also detected the sites of the excitations, as shown in the Fig. 3.a.

6.2 Network Dynamics Models

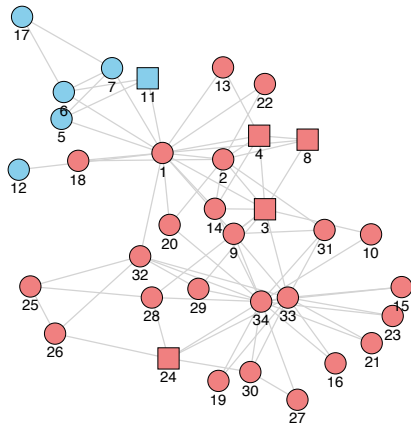
We describe applications of our BlindCD methods on detecting communities in consumer and social networks, where the models have been studied in Sections 3.2 and 3.3.

In the Monte-Carlo simulations below, we generate the graphs as $G \sim \text{SBM}(150, 3, 8 \log N/N, \log N/(2N))$, $N = 150$. For the *consumer* games, \mathbf{A} is taken as the binary adjacency matrix of G and \mathbf{B} is chosen as in (19) where the set of affected agents \mathcal{I} is selected uniformly. Furthermore, in the utility (14), we set $b = 2\|\mathbf{A}\mathbf{1}\|_\infty$ and $a = 2 \max_\ell \|\mathbf{p}^\ell\|_\infty$ such that the equilibrium always satisfies (16). For the *social* networks, we first generate the support of \mathbf{B} as a sparse bipartite graph with connectivity $2 \log N/N$, then the weights on \mathbf{A}, \mathbf{B} are assigned uniformly such that all the rows in the concatenated matrix $[\mathbf{A}, \mathbf{B}]$ sum up to one. This models a setting where the stubborn agents

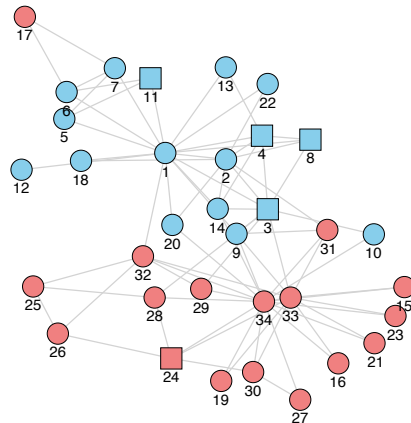
³The 2-step method with [18] provides accurate result only when $R \geq 100$.



(a) Zachary's karate club network. Highlighted nodes in magenta are the actual sites of excitations, while nodes marked as rectangles are the detected sites of excitations using boosted BlindCD. The only mismatches were node '4' and '34'.



(b) BlindCD result



(c) Boosted BlindCD result

Figure 3: **Experiments on Zachary's karate club network.** The network consists of $N = 34$ nodes and (approximately) $K = 2$ communities. The graph filter models a diffusion dynamics with an order of $T_d = 6$ and the graph signals observed have only rank $R = 5$ as only 5 nodes are injected with input signals. The bottom plots show the result of both BlindCD methods.

are connected sparsely to the others, *i.e.*, they are located at the *periphery of the communities*. Note the support of \mathbf{A} is symmetric with $A_{ij} \neq 0 \Leftrightarrow A_{ji} \neq 0$. Snapshots of the set-ups for both networks are found in Fig. 4.

Despite the similarity to the previous examples, it is important to note that for the social networks, the Laplacian matrix \mathbf{L} can be asymmetric. Nevertheless, we anticipate that BlindCD would work in this case provided that \mathbf{L} is approximately symmetric. This symmetry in \mathbf{L} is consistent with assuming that trust in social networks is of mutual nature⁴.

The consumption levels and steady-state opinions can both be generated from the graph filter

⁴Additional numerical experiments show that, on a directed graph where the trusts are not mutual, the BlindCD method recovers the same sets of nodes that are discovered by performing spectral clustering on the *eigenvectors* of \mathbf{L} . We omit these interesting results here since their interpretation requires a different notion of community for directed graphs, e.g., see [49]. Further investigations on this subject are left for future work.

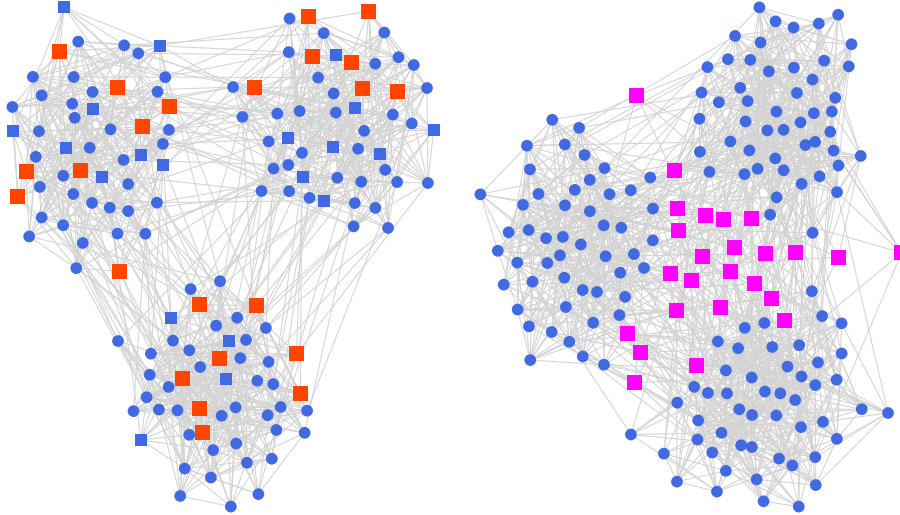


Figure 4: **Snapshots of set-ups for case studies on network dynamics.** (Left) A consumers network, where the highlighted nodes are the agents that the pricing experiments were performed on. (Right) A social network, where the highlighted nodes are the stubborn agents. Both networks are generated according to $\text{SBM}(150, 3, 8 \log N/N, \log N/(2N))$.

in Example 2. The difference between the two cases rests on the design of the sketch matrix \mathbf{B} . In the following, we fix the number of samples at $L = 10^4$ with a noise variance of $\sigma_w^2 = (10^{-1}/b^2)^2$ for consumer games and $\sigma_w^2 = 10^{-2}$ for social networks. For the boosted BlindCD method, we set $\kappa = 2/\sqrt{L}, \rho = 4/\sqrt{RL}$ for consumer games and $\kappa = 2/\sqrt{L}, \rho = 1/\sqrt{RL}$ for social networks; and we test the formulation of (41) with the regularizer $g_1(\hat{\mathbf{B}})$. For the social network, we included a comparison to a 2-step procedure which first recovers the graph topology using [22], and then applying spectral clustering on the inferred topology.

The results of our numerical experiments are shown in Fig. 5, where we compare the community detection performance as the excitation rank R increases in both systems. Similar to the previous experiment in Fig. 2, for both cases we observe that the performance improves with R and the boosted BlindCD method delivers the best performance consistently. Overall, the performance improvement with boosted BlindCD is greater than in the previous example [cf. Fig. 2]. The reason behind this is the fact that the IIR graph filter has a poor low-pass coefficient depending on the parameter $c \gg 1$ for the scenario we have considered. Another observation is that the community detection performance of the un-boosted BlindCD saturates at $R \approx 25$ for the opinion dynamics experiments while it continues to improve with R for pricing ones. This is due to the different model used for the sketch matrix \mathbf{B} . In particular, for the pricing experiments, \mathbf{B} is merely a sub-matrix of the identity matrix. Recall from Theorem 1 that the performance of BlindCD depend on the product $\|\mathbf{V}_{N-K}^\top \mathbf{B} \mathbf{Q}_K\|_2 \|(\mathbf{V}_K^\top \mathbf{B} \mathbf{Q}_K)^{-1}\|_2$, which is anticipated to decrease since \mathbf{B} approaches a permutation of \mathbf{I} as R approaches N , yielding a better performance. The same observation does not apply for opinion dynamics as the sketch matrix does not approximate the identity matrix as R grows.

We then illustrate an application on real network topologies for the two network dynamics. Fig. 6 shows an example of simulated pricing experiments on the network `highschool` from [50], which is a friendship network between $N = 70$ high school students with $|E| = 273$ undirected edges.

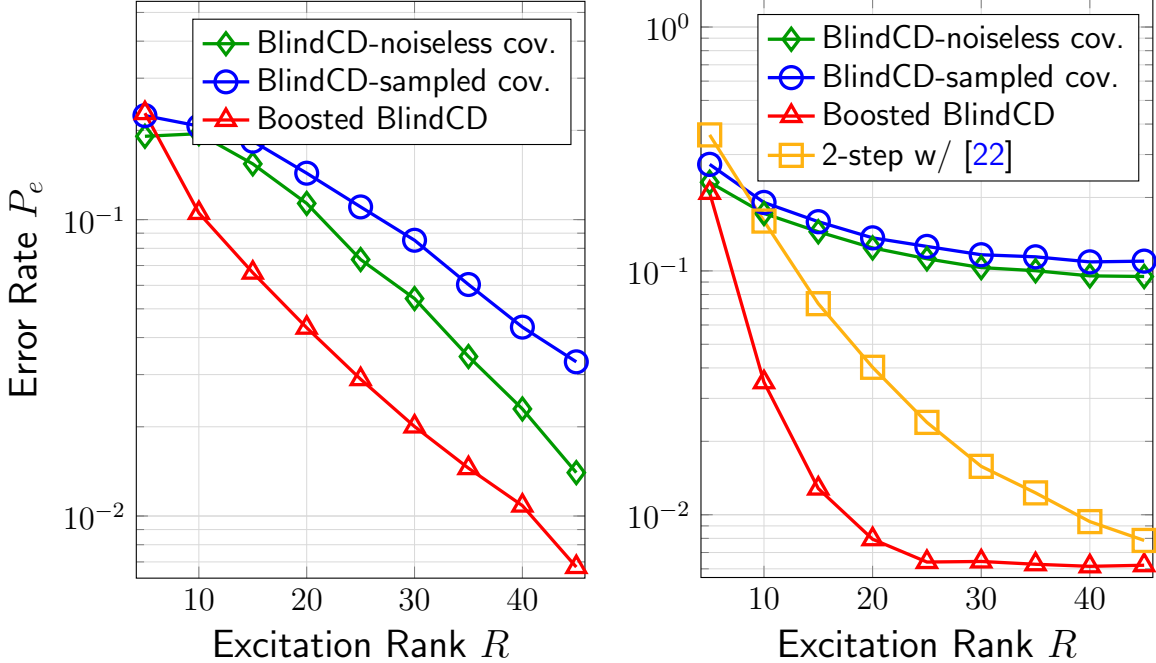


Figure 5: **Community detection performance on cases of network dynamics.** (Left) Pricing experiments on consumers network. (Right) Opinion dynamics with stubborn agents on social networks. In both cases, we consider networks generated as $G \sim \text{SBM}(N, 3, 8 \log N/N, \log N/(2N))$, $N = 150$.

On the other hand, Fig. 7 shows the case study for opinion dynamics on the Facebook network of ReedCollege [51], which is a friendship network with $N = 962$ college students with $|E| = 18,812$ undirected edges, and we influence the network using $R = 150$ stubborn agents. To handle the high dimensionality, we applied the fast algorithm from [52] to solve the robust PCA problem in (41). In both cases, we observe that the boosted BlindCD method recovers the communities in the networks, as evidenced from the illustrations and the ratio-cut scores.

6.3 Application to US Senate Rollcall Records

We consider applying the BlindCD methods to the US Senate rollcall record on <https://voteview.com> for the 110th congress. The dataset contains 657 rollcalls during the period from 2007 to 2009. To represent the opinions of the states during a rollcall, we consider the votes from the two Senators of a state by counting a ‘yay’ as 1, while a ‘nay’ or ‘absent’ is counted as 0. By treating each state as a node on a graph with 50 nodes, this results in $L = 657$ samples of graph signals with values $\{0, 1, 2\}$. As argued in [33], the rollcall data may be modeled as the equilibrium of an opinion dynamics process with stubborn agents. Therefore, we selected 4 states – Massachusetts (MA), New York (NY), Alabama (AL), Louisiana (LA), which are the most liberal/conservative states [53], as the ‘stubborn’ states modeled in Section 3.3. We then apply the BlindCD methods to detect communities for the remaining $N = 46$ states. For the boosted method, we use the sparse regularizer $g_1(\hat{\mathbf{B}})$ to promote sparsity in the $\hat{\mathbf{B}}^*$ component of the solution.

Fig. 8 shows the $K = 2$ communities detected using the proposed methods. We observe that the boosted BlindCD method successfully identifies Maine to be in the same community as Texas, where both states were controlled by Republicans in this congress. Fig. 9 shows the inferred $\hat{\mathbf{B}}^*$

matrix modeled in Section 3.3, where we labeled the rows as the stubborn states and the columns as the regular states. A large number in the table indicates strong influence from the stubborn to regular state. We observe consistent results, e.g., NY (*resp.* LA) positively influencing Illinois (*resp.* Arkansas) as both are Democrat (*resp.* Republican) states in this congress; NY is negatively influencing Idaho (Republican in this congress).

7 Conclusions

This paper proposes two *blind* community detection methods for inferring community structure from graph signals. We consider a challenging and realistic setting where the observed graph signals are outcomes of a graph filter with low-rank excitations. The BlindCD methods rely on an intrinsic *low-pass* property of the graph filters that models the network dynamics. This property holds for common network processes and the accuracy of BlindCD is analyzed by viewing the graph signals as sketches of the graph filters. We propose a boosting technique to improve the performance of BlindCD. The technique leverages the latent ‘low-rank plus sparse’ structure related to the graph signals. Extensive numerical experiments verify our findings.

A Proof of Theorem 1

To simplify the notations while proving the theorem, let us define the following indicator matrices for the communities. Firstly, the matrix $\widehat{\mathbf{X}} \in \mathbb{R}^{N \times K}$ is associated with the communities $\{\hat{\mathcal{C}}_1, \dots, \hat{\mathcal{C}}_K\}$ found with BlindCD and defined as

$$\hat{X}_{ij} := \begin{cases} 1/\sqrt{|\hat{\mathcal{C}}_j|}, & \text{if } i \in \hat{\mathcal{C}}_j, \\ 0, & \text{otherwise.} \end{cases} \quad (54)$$

We have

$$\|\widehat{\mathbf{V}}_K - \widehat{\mathbf{X}}\widehat{\mathbf{X}}^\top \widehat{\mathbf{V}}_K\|_{\mathbb{F}}^2 = \sum_{k=1}^K \sum_{i \in \hat{\mathcal{C}}_k} \left\| \hat{v}_i^{\text{row}} - \frac{1}{|\hat{\mathcal{C}}_k|} \sum_{j \in \hat{\mathcal{C}}_k} \hat{v}_j^{\text{row}} \right\|_2^2.$$

Define \mathcal{X} as the set of all possible indicator matrices of partitions. Using Condition 1 in Theorem 1, we have that

$$\begin{aligned} \|\widehat{\mathbf{V}}_K - \widehat{\mathbf{X}}\widehat{\mathbf{X}}^\top \widehat{\mathbf{V}}_K\|_{\mathbb{F}}^2 &\leq (1 + \epsilon) \min_{\mathbf{X} \in \mathcal{X}} \|\widehat{\mathbf{V}}_K - \mathbf{X}\mathbf{X}^\top \widehat{\mathbf{V}}_K\|_{\mathbb{F}}^2 \\ &\leq (1 + \epsilon) \|\widehat{\mathbf{V}}_K - \mathbf{X}^*(\mathbf{X}^*)^\top \widehat{\mathbf{V}}_K\|_{\mathbb{F}}^2, \end{aligned} \quad (55)$$

where we have defined $\mathbf{X}^* \in \mathbb{R}^{N \times K}$ by replacing $\hat{\mathcal{C}}_i$ in (54) with \mathcal{C}_i^* such that $\mathcal{C}_1^*, \dots, \mathcal{C}_K^*$ is an optimal set of communities found by minimizing $F(\mathcal{C}_1, \dots, \mathcal{C}_K)$ [cf. (9)]. On the other hand, by the definition,

$$\begin{aligned} \|\mathbf{V}_K - \mathbf{X}^*(\mathbf{X}^*)^\top \mathbf{V}_K\|_{\mathbb{F}}^2 &= \min_{\mathbf{X} \in \mathcal{X}} \|\mathbf{V}_K - \mathbf{X}\mathbf{X}^\top \mathbf{V}_K\|_{\mathbb{F}}^2 \\ &= \min_{\mathcal{C}_1, \dots, \mathcal{C}_K} F(\mathcal{C}_1, \dots, \mathcal{C}_K) = F^*, \end{aligned} \quad (56)$$

and furthermore $\|\mathbf{V}_K - \widehat{\mathbf{X}}\widehat{\mathbf{X}}^\top \mathbf{V}_K\|_{\mathbb{F}}^2 = F(\hat{\mathcal{C}}_1, \dots, \hat{\mathcal{C}}_K)$.

Define the error matrix as $\mathbf{E} := \mathbf{V}_K \mathbf{V}_K^\top - \widehat{\mathbf{V}}_K \widehat{\mathbf{V}}_K^\top$. We observe the following chain of inequalities:

$$\begin{aligned} \|\mathbf{V}_K - \widehat{\mathbf{X}} \widehat{\mathbf{X}}^\top \mathbf{V}_K\|_{\text{F}} &= \|(\mathbf{I} - \widehat{\mathbf{X}} \widehat{\mathbf{X}}^\top)(\widehat{\mathbf{V}}_K \widehat{\mathbf{V}}_K^\top + \mathbf{E})\|_{\text{F}} \\ &\leq \|(\mathbf{I} - \widehat{\mathbf{X}} \widehat{\mathbf{X}}^\top) \widehat{\mathbf{V}}_K \widehat{\mathbf{V}}_K^\top\|_{\text{F}} + \|(\mathbf{I} - \widehat{\mathbf{X}} \widehat{\mathbf{X}}^\top) \mathbf{E}\|_{\text{F}} \\ &\leq \|(\mathbf{I} - \widehat{\mathbf{X}} \widehat{\mathbf{X}}^\top) \widehat{\mathbf{V}}_K \widehat{\mathbf{V}}_K^\top\|_{\text{F}} + \|\mathbf{E}\|_{\text{F}}, \end{aligned} \quad (57)$$

where the first equality is due to $\mathbf{V}_K^\top \mathbf{V}_K = \mathbf{I}$ and the last inequality is due to $\mathbf{I} - \widehat{\mathbf{X}} \widehat{\mathbf{X}}^\top$ is a projection matrix. Using (55), we have that

$$\begin{aligned} &\|(\mathbf{I} - \widehat{\mathbf{X}} \widehat{\mathbf{X}}^\top) \widehat{\mathbf{V}}_K \widehat{\mathbf{V}}_K^\top\|_{\text{F}} + \|\mathbf{E}\|_{\text{F}} \\ &\leq \sqrt{1 + \epsilon} \|(\mathbf{I} - \mathbf{X}^* (\mathbf{X}^*)^\top) (\mathbf{V}_K \mathbf{V}_K^\top - \mathbf{E})\|_{\text{F}} + \|\mathbf{E}\|_{\text{F}} \\ &\leq \sqrt{1 + \epsilon} \|(\mathbf{I} - \mathbf{X}^* (\mathbf{X}^*)^\top) \mathbf{V}_K \mathbf{V}_K^\top\|_{\text{F}} \\ &\quad + \sqrt{1 + \epsilon} \|(\mathbf{I} - \mathbf{X}^* (\mathbf{X}^*)^\top) \mathbf{E}\|_{\text{F}} + \|\mathbf{E}\|_{\text{F}} \\ &\leq \sqrt{1 + \epsilon} \|(\mathbf{I} - \mathbf{X}^* (\mathbf{X}^*)^\top) \mathbf{V}_K \mathbf{V}_K^\top\|_{\text{F}} + (2 + \epsilon) \|\mathbf{E}\|_{\text{F}} \\ &= \sqrt{(1 + \epsilon) F^*} + (2 + \epsilon) \|\mathbf{E}\|_{\text{F}}, \end{aligned} \quad (58)$$

where we have used the fact $\mathbf{I} - \mathbf{X}^* (\mathbf{X}^*)^\top$ is a projection matrix and $\sqrt{1 + \epsilon} \leq 1 + \epsilon$ in the third inequality. The final step is to bound $\|\mathbf{E}\|_{\text{F}}$, where we rely on the following results.

Lemma 4 [42, Lemma 7] For any $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N \times K}$ with $N \geq K$ and $\mathbf{A}^\top \mathbf{A} = \mathbf{B}^\top \mathbf{B} = \mathbf{I}$, it holds that

$$\|\mathbf{A} \mathbf{A}^\top - \mathbf{B} \mathbf{B}^\top\|_{\text{F}}^2 \leq 2K \|\mathbf{A} \mathbf{A}^\top - \mathbf{B} \mathbf{B}^\top\|_2^2. \quad (59)$$

Proposition 1 Under Conditions 2 to 4 in Theorem 1, we have

$$\|\overline{\mathbf{V}}_K \overline{\mathbf{V}}_K^\top - \mathbf{V}_K \mathbf{V}_K^\top\|_2^2 = (1 + \gamma^2)^{-1} \gamma^2, \quad (60)$$

where the columns of $\overline{\mathbf{V}}_K$ are the top K eigenvectors of $\overline{\mathbf{C}}_y$ and γ is bounded as stated in (35).

Proposition 2 Under Condition 5 in Theorem 1, it holds that

$$\|\overline{\mathbf{V}}_K \overline{\mathbf{V}}_K^\top - \widehat{\mathbf{V}}_K \widehat{\mathbf{V}}_K^\top\|_2 \leq \|\widehat{\mathbf{C}}_y - \overline{\mathbf{C}}_y\|_2 / \delta. \quad (61)$$

The proofs of the propositions can be found in the subsections A and B of this appendix. Applying Lemma 4 we obtain that

$$\|\mathbf{E}\|_{\text{F}} \leq \sqrt{2K} \|\mathbf{V}_K \mathbf{V}_K^\top - \widehat{\mathbf{V}}_K \widehat{\mathbf{V}}_K^\top\|_2. \quad (62)$$

Combining (60), (61) and using the triangle inequality yields

$$\begin{aligned} \sqrt{F(\hat{\mathbf{C}}_1, \dots, \hat{\mathbf{C}}_K)} &= \|(\mathbf{I} - \widehat{\mathbf{X}} \widehat{\mathbf{X}}^\top) \mathbf{V}_K \mathbf{V}_K^\top\|_{\text{F}} \\ &\leq \sqrt{(1 + \epsilon) F^*} + (2 + \epsilon) \sqrt{2K} \left(\sqrt{\frac{\gamma^2}{1 + \gamma^2}} + \frac{\|\widehat{\mathbf{C}}_y - \overline{\mathbf{C}}_y\|_2}{\delta} \right), \end{aligned}$$

concluding the proof.

A.1 Proof of Proposition 1

We begin our proof by establishing the relationships between $\bar{\mathbf{V}}_K, \mathbf{V}_K$ and the left singular vectors of $\mathcal{H}(\mathbf{L})\mathbf{B}$. Denote the rank- K approximation to $\mathcal{H}(\mathbf{L})$ as $[\mathcal{H}(\mathbf{L})]_K := \mathbf{V}_K \text{diag}(\tilde{\mathbf{h}}_K) \mathbf{V}_K^\top$. This expression is valid due to the low pass property of $\mathcal{H}(\mathbf{L})$. Define $\tilde{\mathbf{B}} := \mathbf{B}\mathbf{Q}_K$, we observe that

$$\mathcal{R}([\mathcal{H}(\mathbf{L})]_K) = \mathcal{R}([\mathcal{H}(\mathbf{L})]_K \tilde{\mathbf{B}}), \quad (63)$$

which is due to Condition 3 in Theorem 1 such that the linear transformation $\tilde{\mathbf{B}}$ does not modify the range space of $[\mathcal{H}(\mathbf{L})]_K$. Similarly, $[\bar{\mathbf{C}}_y]_K := \bar{\mathbf{V}}_K \text{diag}(\boldsymbol{\sigma}_K)^2 \bar{\mathbf{V}}_K^\top$ is the rank K approximation to $\bar{\mathbf{C}}_y$. We observe the equivalences

$$\mathcal{R}([\bar{\mathbf{C}}_y]_K) = \mathcal{R}([\mathcal{H}(\mathbf{L})\mathbf{B}]_K) = \mathcal{R}(\mathcal{H}(\mathbf{L})\tilde{\mathbf{B}}) \quad (64)$$

where the last equality is due to $\mathcal{H}(\mathbf{L})\tilde{\mathbf{B}} = \mathcal{H}(\mathbf{L})\mathbf{B}\mathbf{Q}_K = \bar{\mathbf{V}}_K \text{diag}(\boldsymbol{\sigma}_K)$, as we recall that the columns of \mathbf{Q}_K are the top K right singular vectors of $\mathcal{H}(\mathbf{L})\mathbf{B}$. Furthermore,

$$\mathcal{R}([\mathcal{H}(\mathbf{L})]_K \tilde{\mathbf{B}}) \perp \mathcal{R}((\mathcal{H}(\mathbf{L}) - [\mathcal{H}(\mathbf{L})]_K) \tilde{\mathbf{B}}). \quad (65)$$

Let the columns of $\tilde{\tilde{\mathbf{V}}}_K$ and $\tilde{\bar{\mathbf{V}}}_K$ be respectively the top- K singular vectors of $[\mathcal{H}(\mathbf{L})]_K \tilde{\mathbf{B}}$ and $\mathcal{H}(\mathbf{L})\tilde{\mathbf{B}}$, therefore (63) and (64) imply that $\mathbf{V}_K \mathbf{V}_K^\top = \tilde{\tilde{\mathbf{V}}}_K \tilde{\tilde{\mathbf{V}}}_K^\top$ and $\bar{\mathbf{V}}_K \bar{\mathbf{V}}_K^\top = \tilde{\bar{\mathbf{V}}}_K \tilde{\bar{\mathbf{V}}}_K^\top$. Invoking (65) with [54, Lemma 8] through setting $\mathbf{D} = \mathcal{H}(\mathbf{L})\tilde{\mathbf{B}}$, $\mathbf{C} = [\mathcal{H}(\mathbf{L})]_K \tilde{\mathbf{B}}$ and $\mathbf{E} = (\mathcal{H}(\mathbf{L}) - [\mathcal{H}(\mathbf{L})]_K) \tilde{\mathbf{B}}$ therein, and applying [55, Theorem 2.6.1], we obtain that

$$\begin{aligned} \|\mathbf{V}_K \mathbf{V}_K^\top - \bar{\mathbf{V}}_K \bar{\mathbf{V}}_K^\top\|_2^2 &= \|\tilde{\tilde{\mathbf{V}}}_K \tilde{\tilde{\mathbf{V}}}_K^\top - \tilde{\bar{\mathbf{V}}}_K \tilde{\bar{\mathbf{V}}}_K^\top\|_2^2 \\ &= 1 - \beta_K \left([\mathcal{H}(\mathbf{L})]_K \tilde{\mathbf{B}} \boldsymbol{\Pi}^\dagger ([\mathcal{H}(\mathbf{L})]_K \tilde{\mathbf{B}})^\top \right), \end{aligned} \quad (66)$$

where we have defined $\boldsymbol{\Pi} := (\mathcal{H}(\mathbf{L})\tilde{\mathbf{B}})^\top \mathcal{H}(\mathbf{L})\tilde{\mathbf{B}}$ and $\beta_K(\cdot)$ denotes the K th largest eigenvalue. Under Condition 4 in Theorem 1, the $K \times K$ matrix $\boldsymbol{\Pi}$ is non-singular. We observe the following chain of equalities

$$\begin{aligned} &\beta_K \left([\mathcal{H}(\mathbf{L})]_K \tilde{\mathbf{B}} \boldsymbol{\Pi}^{-1} ([\mathcal{H}(\mathbf{L})]_K \tilde{\mathbf{B}})^\top \right) \\ &= \beta_K \left(\text{diag}(\tilde{\mathbf{h}}_K) \mathbf{V}_K^\top \tilde{\mathbf{B}} \boldsymbol{\Pi}^{-1} (\text{diag}(\tilde{\mathbf{h}}_K) \mathbf{V}_K^\top \tilde{\mathbf{B}})^\top \right) \\ &= \frac{1}{\beta_1 \left((\text{diag}(\tilde{\mathbf{h}}_K) \mathbf{V}_K^\top \tilde{\mathbf{B}})^{-\top} \boldsymbol{\Pi} (\text{diag}(\tilde{\mathbf{h}}_K) \mathbf{V}_K^\top \tilde{\mathbf{B}})^{-1} \right)}, \end{aligned} \quad (67)$$

where the first equality is due to $\beta_K(\mathbf{U}\mathbf{A}\mathbf{U}^\top) = \beta_K(\mathbf{A})$ for any symmetric \mathbf{A} and $\mathbf{U} \in \mathbb{R}^{N \times K}$ with orthogonal columns, and the second equality follows since the argument in $\beta_K(\cdot)$ is of rank K . Moreover, $\boldsymbol{\Pi}$ admits the decomposition

$$\begin{aligned} \boldsymbol{\Pi} &= (\mathcal{H}(\mathbf{L})\tilde{\mathbf{B}})^\top \mathcal{H}(\mathbf{L})\tilde{\mathbf{B}} = \tilde{\mathbf{B}}^\top \mathcal{H}(\mathbf{L})^\top \mathcal{H}(\mathbf{L})\tilde{\mathbf{B}} \\ &= \tilde{\mathbf{B}}^\top \mathbf{V}_K \text{diag}(\tilde{\mathbf{h}}_K)^2 \mathbf{V}_K^\top \tilde{\mathbf{B}} \\ &\quad + \tilde{\mathbf{B}}^\top \mathbf{V}_{N-K} \text{diag}(\tilde{\mathbf{h}}_{N-K})^2 \mathbf{V}_{N-K}^\top \tilde{\mathbf{B}}. \end{aligned} \quad (68)$$

Thus, yielding that

$$\begin{aligned}
& \beta_K \left([\mathcal{H}(\mathbf{L})]_K \tilde{\mathbf{B}} \boldsymbol{\Pi}^{-1} ([\mathcal{H}(\mathbf{L})]_K \tilde{\mathbf{B}})^\top \right) \\
&= \left(1 + \beta_1 \left((\text{diag}(\tilde{\mathbf{h}}_K) \mathbf{V}_K^\top \tilde{\mathbf{B}})^{-\top} \tilde{\mathbf{B}}^\top \mathbf{V}_{N-K} \right. \right. \\
&\quad \left. \left. \text{diag}(\tilde{\mathbf{h}}_{N-K})^2 \mathbf{V}_{N-K}^\top \tilde{\mathbf{B}} (\text{diag}(\tilde{\mathbf{h}}_K) \mathbf{V}_K^\top \tilde{\mathbf{B}})^{-1} \right) \right)^{-1} \\
&= \frac{1}{1 + \|\text{diag}(\tilde{\mathbf{h}}_{N-K}) \mathbf{V}_{N-K}^\top \tilde{\mathbf{B}} (\text{diag}(\tilde{\mathbf{h}}_K) \mathbf{V}_K^\top \tilde{\mathbf{B}})^{-1}\|_2^2} \\
&= \left(1 + \gamma^2 \right)^{-1},
\end{aligned}$$

where we have defined γ such that

$$\begin{aligned}
\gamma &:= \|\text{diag}(\tilde{\mathbf{h}}_{N-K}) \mathbf{V}_{N-K}^\top \tilde{\mathbf{B}} (\text{diag}(\tilde{\mathbf{h}}_K) \mathbf{V}_K^\top \tilde{\mathbf{B}})^{-1}\|_2 \\
&\leq \eta \|\mathbf{V}_{N-K}^\top \mathbf{B} \mathbf{Q}_K\|_2 \|(\mathbf{V}_K^\top \mathbf{B} \mathbf{Q}_K)^{-1}\|_2.
\end{aligned} \tag{69}$$

Substituting the above into (66) concludes the proof.

A.2 Proof of Proposition 2

Denote the SVD of the sampled covariance as $\hat{\mathbf{C}}_y = \hat{\mathbf{V}} \hat{\boldsymbol{\Sigma}} \hat{\mathbf{V}}^\top$. The left hand side of (61) can be written as

$$\|\bar{\mathbf{V}}_K \bar{\mathbf{V}}_K^\top - \hat{\mathbf{V}}_K \hat{\mathbf{V}}_K^\top\|_2 = \|\hat{\mathbf{V}}_{N-K}^\top \bar{\mathbf{V}}_K\|_2, \tag{70}$$

where the equality is due to [55, Theorem 2.6.1].

Define $\boldsymbol{\Delta} := \hat{\mathbf{C}}_y - \bar{\mathbf{C}}_y$. Condition 5 in Theorem 1 implies that the largest eigenvalue in $\hat{\boldsymbol{\Sigma}}_{N-K}$ will never exceed $\beta_K(\bar{\mathbf{C}}_y) - \delta$ since

$$\begin{aligned}
\beta_{\max}(\hat{\boldsymbol{\Sigma}}_{N-K}) &= \beta_{K+1}(\hat{\mathbf{C}}_y) \leq \beta_{K+1}(\bar{\mathbf{C}}_y) + \beta_1(\boldsymbol{\Delta}) \\
&\leq \beta_{K+1}(\bar{\mathbf{C}}_y) + \|\boldsymbol{\Delta}\|_2,
\end{aligned} \tag{71}$$

where the first inequality is due to Weyl's inequality [55]. The perturbed matrix $\hat{\mathbf{C}}_y$ thus satisfies the requirement of the Davis-Kahan's $\sin(\Theta)$ theorem [56]

$$\|\hat{\mathbf{V}}_{N-K}^\top \bar{\mathbf{V}}_K\|_2 \leq \delta^{-1} \|\hat{\mathbf{V}}_{N-K}^\top \boldsymbol{\Delta} \bar{\mathbf{V}}_K\|_2. \tag{72}$$

The inequality in (61) is obtained by observing that both $\bar{\mathbf{V}}_K$ and $\hat{\mathbf{V}}_{N-K}$ are orthogonal matrices.

B Proof of Lemma 2

Fix $1 \geq c > 0$. Under the conditions stated in the lemma, the least-squares optimization (37) admits a closed form solution

$$\mathcal{H}^* - \mathcal{H}(\mathbf{L})\mathbf{B} = \left(\sum_{\ell=1}^L \mathbf{w}^\ell (\mathbf{z}^\ell)^\top \right) \left(\sum_{\ell=1}^L \mathbf{z}^\ell (\mathbf{z}^\ell)^\top \right)^{-1}, \tag{73}$$

where \mathbf{w}^ℓ was introduced in (11). Denoting the right hand side in (73) by \mathcal{E} , we have that

$$\begin{aligned}\|\mathcal{E}\|_2 &= \left\| \left(\frac{1}{L} \sum_{\ell=1}^L \mathbf{w}^\ell (\mathbf{z}^\ell)^\top \right) \left(\frac{1}{L} \sum_{\ell=1}^L \mathbf{z}^\ell (\mathbf{z}^\ell)^\top \right)^{-1} \right\|_2 \\ &\leq \left\| \frac{1}{L} \sum_{\ell=1}^L \mathbf{w}^\ell (\mathbf{z}^\ell)^\top \right\|_2 \left\| \left(\frac{1}{L} \sum_{\ell=1}^L \mathbf{z}^\ell (\mathbf{z}^\ell)^\top \right)^{-1} \right\|_2.\end{aligned}\tag{74}$$

Observe that $\frac{1}{L} \sum_{\ell=1}^L \mathbf{z}^\ell (\mathbf{z}^\ell)^\top$ converges to \mathbf{I} such that with probability at least $1 - c$,

$$\left\| \frac{1}{L} \sum_{\ell=1}^L \mathbf{z}^\ell (\mathbf{z}^\ell)^\top - \mathbf{I} \right\|_2 \leq C_0 \sqrt{\frac{R \log(1/c)}{L}},\tag{75}$$

for some constant C_0 . Applying [57, Proposition 2.1] we get that

$$\begin{aligned}\left\| \left(\frac{1}{L} \sum_{\ell=1}^L \mathbf{z}^\ell (\mathbf{z}^\ell)^\top \right)^{-1} \right\|_2 &\leq \left(1 - \left\| \frac{1}{L} \sum_{\ell=1}^L \mathbf{z}^\ell (\mathbf{z}^\ell)^\top - \mathbf{I} \right\|_2 \right)^{-1} \\ &\leq \left(1 - C_0 \sqrt{\frac{R \log(1/c)}{L}} \right)^{-1}.\end{aligned}$$

On the other hand, observe that $\mathbb{E}[\mathbf{w}^\ell (\mathbf{z}^\ell)^\top] = 0$ and $\|\mathbf{w}^\ell (\mathbf{z}^\ell)^\top\| \leq C_w$ almost surely. Applying the matrix Bernstein's inequality [58, Theorem 1.6] shows that with probability at least $1 - c$ and for sufficiently large L ,

$$\left\| \frac{1}{L} \sum_{\ell=1}^L \mathbf{w}^\ell (\mathbf{z}^\ell)^\top \right\|_2 \leq C_1 \sqrt{\frac{\sigma_w^2 \log((N+R)/c)}{L}},\tag{76}$$

for some constant C_1 . Finally, with probability at least $1 - 2c$,

$$\|\mathcal{E}\|_2 \leq \frac{C_1 \sqrt{\sigma_w^2 \log((N+R)/c)}}{\sqrt{L} - C_0 \sqrt{R \log(1/c)}} = \mathcal{O}(\sigma_w / \sqrt{L}).\tag{77}$$

C Proof of Corollary 1

Let $\tilde{\mathbf{V}}_K$ and $\tilde{\mathbf{S}}_K$ be the top K left singular vectors of $\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$ and $\tilde{\mathbf{S}}^\star$, respectively. We can repeat the proof for Theorem 1 up to (58) by re-defining the error matrix \mathbf{E} therein as $\tilde{\mathbf{E}} = \mathbf{V}_K \mathbf{V}_K^\top - \tilde{\mathbf{S}}_K \tilde{\mathbf{S}}_K^\top$. This entails

$$\sqrt{F(\tilde{\mathcal{C}}_1, \dots, \tilde{\mathcal{C}}_K)} - \sqrt{(1+\epsilon)F^\star} \leq (2+\epsilon) \|\tilde{\mathbf{E}}\|_F.\tag{78}$$

Next, we bound $\|\tilde{\mathbf{E}}\|_F$. Applying Lemma 4 and using the triangle inequality we get that

$$\begin{aligned}\|\tilde{\mathbf{E}}\|_F &\leq \sqrt{2K} \|\mathbf{V}_K \mathbf{V}_K^\top - \tilde{\mathbf{S}}_K \tilde{\mathbf{S}}_K^\top\|_2 \\ &\leq \sqrt{2K} (\|\mathbf{V}_K \mathbf{V}_K^\top - \tilde{\mathbf{V}}_K \tilde{\mathbf{V}}_K^\top\|_2 + \|\tilde{\mathbf{V}}_K \tilde{\mathbf{V}}_K^\top - \tilde{\mathbf{S}}_K \tilde{\mathbf{S}}_K^\top\|_2),\end{aligned}$$

Proposition 1 implies that

$$\|\mathbf{V}_K \mathbf{V}_K^\top - \tilde{\mathbf{V}}_K \tilde{\mathbf{V}}_K^\top\|_2 \leq \sqrt{\tilde{\gamma}/(1+\tilde{\gamma})},\tag{79}$$

where $\tilde{\gamma}$ is bounded as in (51). Our remaining task is to bound $\|\tilde{\mathbf{V}}_K \tilde{\mathbf{V}}_K^\top - \tilde{\mathbf{S}}_K \tilde{\mathbf{S}}_K^\top\|_2$. Observe that

$$\|\tilde{\mathbf{V}}_K \tilde{\mathbf{V}}_K^\top - \tilde{\mathbf{S}}_K \tilde{\mathbf{S}}_K^\top\|_2 = \|\tilde{\mathbf{S}}_{R-K}^\top \tilde{\mathbf{V}}_K\|_2 \quad (80)$$

and

$$\sigma_K(\hat{\mathbf{S}}^*) \geq \sigma_K(\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}) - \|\tilde{\mathbf{\Delta}}\|_2, \quad (81)$$

where we recalled the definition $\tilde{\mathbf{\Delta}} = \hat{\mathbf{S}}^* - \tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}$ and applied the Weyl's inequality [55]. From (49), we have that

$$\sigma_K(\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}) - \|\tilde{\mathbf{\Delta}}\|_2 = \sigma_{K+1}(\tilde{\mathcal{H}}(\mathbf{L})\mathbf{B}) + \tilde{\delta}, \quad (82)$$

with $\tilde{\delta} > 0$. Finally, applying the Wedin theorem [59] yields

$$\|\tilde{\mathbf{S}}_{R-K}^\top \tilde{\mathbf{V}}_K\|_2 \leq (\tilde{\delta})^{-1} \|\tilde{\mathbf{\Delta}}\|_2. \quad (83)$$

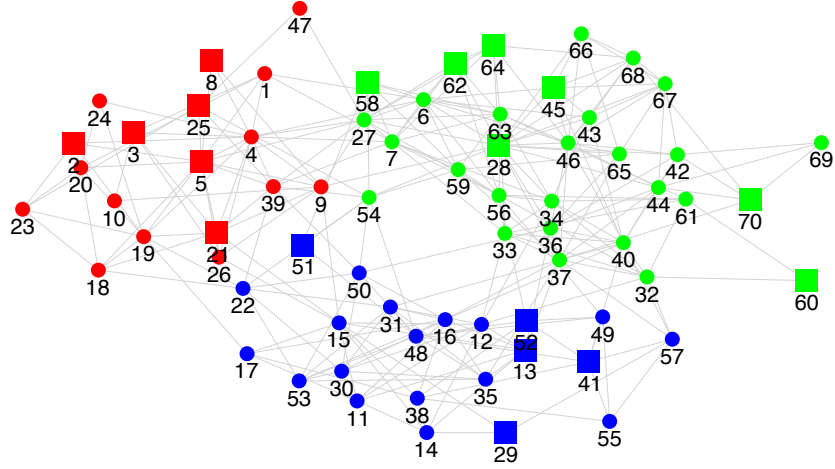
References

- [1] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, 2013.
- [2] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [3] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing," *arXiv preprint arXiv:1712.00468*, 2017.
- [4] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1832–1843, April 2016.
- [5] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec 2015.
- [6] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 764–778, Feb 2017.
- [7] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4363–4378, Aug 2016.
- [8] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4117–4131, Aug 2017.
- [9] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Autoregressive moving average graph filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 274–288, Jan 2017.
- [10] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

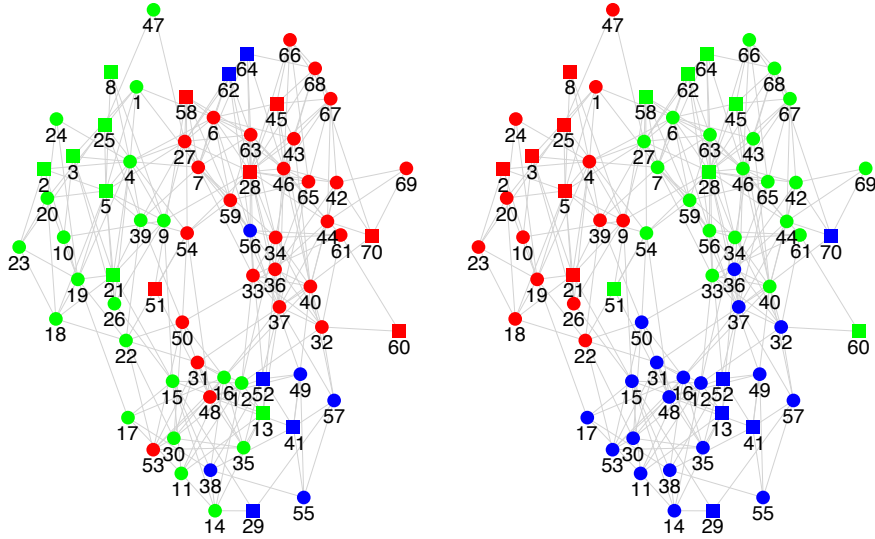
- [11] E. Pavez and A. Ortega, “Generalized Laplacian precision matrix estimation for graph signal processing,” in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Shanghai, China, Mar. 20-25, 2016.
- [12] Y. Shen, B. Baingana, and G. B. Giannakis, “Kernel-based structural equation models for topology identification of directed networks,” *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2503–2516, 2017.
- [13] X. Cai, J. A. Bazerque, and G. B. Giannakis, “Sparse structural equation modeling for inference of gene regulatory networks exploiting genetic perturbations,” *PLoS, Computational Biology*, Jun. 2013.
- [14] H.-T. Wai, A. Scaglione, U. Harush, B. Barzel, and A. Leshem, “Rids: Robust identification of sparse gene regulatory networks from perturbation experiments,” *arXiv preprint arXiv:1612.06565*, 2016.
- [15] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, “Learning Laplacian matrix in smooth graph signal representations,” *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, Dec 2016.
- [16] V. Kalofolias, “How to learn a graph from smooth signals,” in *Intl. Conf. Artif. Intel. Stat. (AISTATS)*. J Mach. Learn. Res., 2016, pp. 920–929.
- [17] S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero, “Learning sparse graphs under smoothness prior,” in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, March 2017, pp. 6508–6512.
- [18] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, “Network topology inference from spectral templates,” *IEEE Trans. Signal and Info. Process. over Networks*, vol. 3, no. 3, pp. 467–483, 2017.
- [19] S. Segarra, G. Mateos, A. G. Marques, and A. Ribeiro, “Blind identification of graph filters,” *IEEE Trans. Signal Process.*, vol. 65, no. 5, pp. 1146–1159, March 2017.
- [20] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, “Network topology inference from non-stationary graph signals,” in *Proc. ICASSP*, March 2017, pp. 5870–5874.
- [21] H. E. Egilmez, E. Pavez, and A. Ortega, “Graph learning from filtered signals: Graph system and diffusion kernel identification,” *IEEE Transactions on Signal and Information Processing over Networks*, 2018.
- [22] H.-T. Wai, A. Scaglione, and A. Leshem, “Active sensing of social networks,” *IEEE Trans. Signal and Info. Process. over Networks*, vol. 2, no. 3, pp. 406–419, 2016.
- [23] D. Marbach, J. C. Costello, R. Küffner, N. M. Vega, R. J. Prill, D. M. Camacho, K. R. Allison, A. Aderhold, R. Bonneau, Y. Chen *et al.*, “Wisdom of crowds for robust gene network inference,” *Nature methods*, vol. 9, no. 8, p. 796, 2012.
- [24] S. Fortunato, “Community detection in graphs,” *Physics reports*, vol. 486, no. 3, pp. 75–174, 2010.

- [25] O. Candogan, K. Bimpikis, and A. Ozdaglar, “Optimal pricing in networks with externalities,” *Operations Research*, vol. 60, no. 4, pp. 883–905, 2012.
- [26] B. Ata, A. Belloni, and O. Candogan, “Latent agents in networks: Estimation and pricing,” *arXiv:1808.04878v1*, August 2018.
- [27] M. H. DeGroot, “Reaching a consensus,” *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.
- [28] H. Zha, X. He, C. Ding, M. Gu, and H. D. Simon, “Spectral relaxation for k-means clustering,” in *Advances in neural information processing systems*, 2002, pp. 1057–1064.
- [29] E. Abbe, “Community detection and stochastic block models: recent developments,” *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6446–6531, 2017.
- [30] U. von Luxburg, “A tutorial on spectral clustering,” *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, Dec 2007.
- [31] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A k-means clustering algorithm,” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.
- [32] A. Kumar, Y. Sabharwal, and S. Sen, “A simple linear time $(1+\epsilon)$ -approximation algorithm for k-means clustering in any dimensions,” in *FOCS*. IEEE, 2004, pp. 454–462.
- [33] S. X. Wu, H.-T. Wai, and A. Scaglione, “Estimating social opinion dynamics models from voting records,” *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4193–4206, 2018.
- [34] D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi, “Spread of (mis) information in social networks,” *Games and Economic Behavior*, vol. 70, no. 2, pp. 194–227, 2010.
- [35] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione, “Binary opinion dynamics with stubborn agents,” *ACM Transactions on Economics and Computation (TEAC)*, vol. 1, no. 4, p. 19, 2013.
- [36] P. Jia, A. MirTabatabaei, N. E. Friedkin, and F. Bullo, “Opinion dynamics and the evolution of social power in influence networks,” *SIAM review*, vol. 57, no. 3, pp. 367–397, 2015.
- [37] M. E. Yildiz and A. Scaglione, “Computing along routes via gossiping,” *IEEE Trans. on Signal Process.*, vol. 58, no. 6, pp. 3313–3327, 2010.
- [38] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Siam, 2001, vol. 2.
- [39] R. Shafipour, A. Hashemi, G. Mateos, and H. Vikalo, “Online topology inference from streaming stationary graph signals,” *submitted to DSW 2019*, 2019.
- [40] J. Tsitsiklis, “Problems in decentralized decision making and computation,” Ph.D. dissertation, Dept. of Electrical Engineering and Computer Science, M.I.T., Boston, MA, 1984.
- [41] D. P. Bertsekas, *Nonlinear programming*. Athena scientific Belmont, 1999.

- [42] C. Boutsidis, P. Kambadur, and A. Gittens, “Spectral clustering via the power method-provably,” in *International Conference on Machine Learning*, 2015, pp. 40–48.
- [43] N. Tremblay, G. Puy, R. Gribonval, and P. Vandergheynst, “Compressive spectral clustering,” in *International Conference on Machine Learning*, 2016, pp. 1002–1011.
- [44] R. Vershynin, *High-Dimensional Probability*. Cambridge University Press, 2017.
- [45] A. Agarwal, S. Negahban, and M. J. Wainwright, “Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions,” *The Annals of Statistics*, pp. 1171–1197, 2012.
- [46] J. Huang, T. Zhang, and D. Metaxas, “Learning with structured sparsity,” *JMLR*, vol. 12, no. Nov, pp. 3371–3412, 2011.
- [47] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, “A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers,” *Statistical Science*, pp. 538–557, 2012.
- [48] P. W. Holland, K. B. Laskey, and S. Leinhardt, “Stochastic blockmodels: First steps,” *Social networks*, vol. 5, no. 2, pp. 109–137, 1983.
- [49] F. D. Malliaros and M. Vazirgiannis, “Clustering and community detection in directed networks: A survey,” *Physics Reports*, vol. 533, no. 4, pp. 95–142, 2013.
- [50] J. S. Coleman *et al.*, “Introduction to mathematical sociology.” *Introduction to mathematical sociology.*, 1964.
- [51] A. L. Traud, P. J. Mucha, and M. A. Porter, “Social structure of Facebook networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 16, pp. 4165–4180, 2012.
- [52] A. Aravkin, S. Becker, V. Cevher, and P. Olsen, “A variational approach to stable principal component pursuit,” in *UAI*, July 2014.
- [53] P. R. Center, “Political ideology by state.” [Online]. Available: <http://www.pewforum.org/religious-landscape-study/compare/political-ideology/by/state/>
- [54] A. Gittens, P. Kambadur, and C. Boutsidis, “Approximate spectral clustering via randomized sketching,” *arXiv:1311.2854v1*, 2013.
- [55] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012, vol. 3.
- [56] C. Davis and W. M. Kahan, “The rotation of eigenvectors by a perturbation. iii,” *SIAM Journal on Numerical Analysis*, vol. 7, no. 1, pp. 1–46, 1970.
- [57] R. Vershynin, “How close is the sample covariance matrix to the actual covariance matrix?” *Journal of Theoretical Probability*, vol. 25, no. 3, pp. 655–686, 2012.
- [58] J. A. Tropp, “User-friendly tail bounds for sums of random matrices,” *Foundations of computational mathematics*, vol. 12, no. 4, pp. 389–434, 2012.
- [59] P.-A. Wedin, “Perturbation bounds in connection with singular value decomposition,” *BIT Numerical Mathematics*, vol. 12, no. 1, pp. 99–111, 1972.



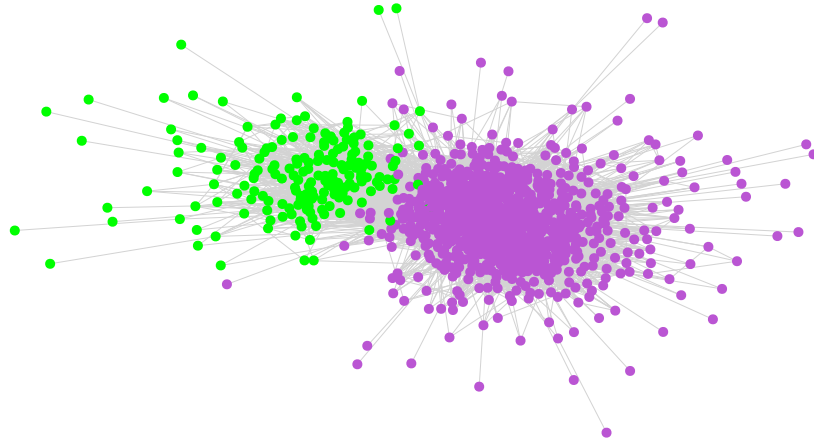
(a) Highschool network. The pricing experiments modify prices for the agents marked with *square*. The above clustering on the network is computed from the true Laplacian matrix L , which has a RatioCut of 3.618.



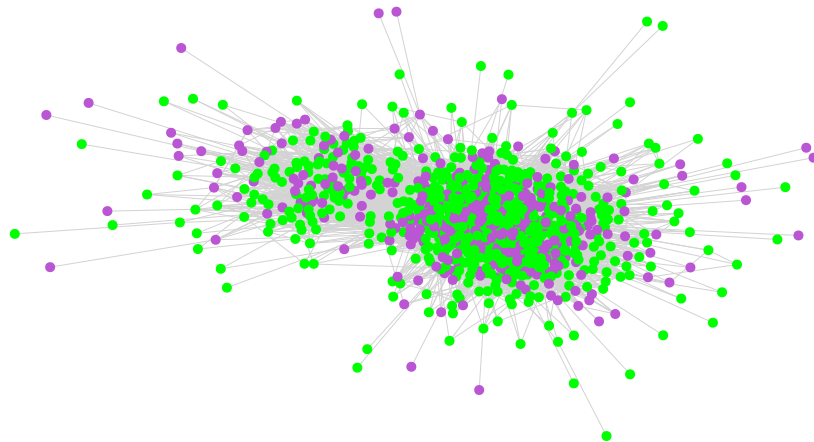
(b) BlindCD, RatioCut= 6.769.

(c) Boosted BlindCD, RatioCut= 4.701.

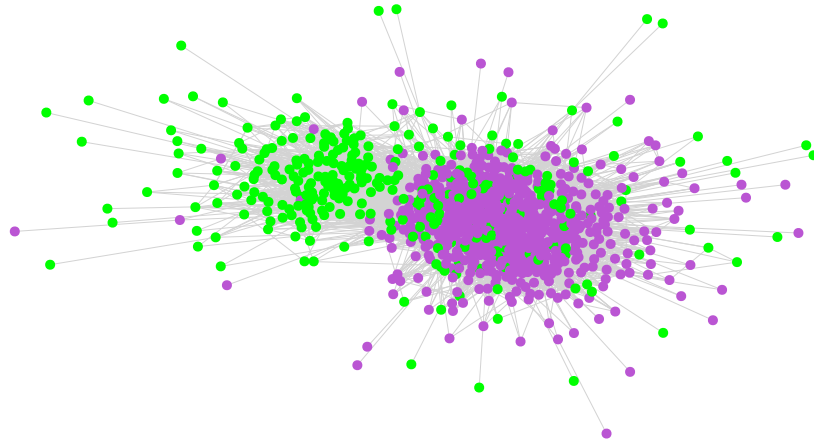
Figure 6: **Pricing experiments on the Highschool network.** The network comprises of $N = 70$ agents and (approximately) $K = 3$ communities. The utility parameters in (14) are $a = R$ and $b = 2\|\mathbf{A}\mathbf{1}\|_\infty$. The network's equilibrium consumption levels are collected for $L = 10^3$ samples, each observed with a noise variance of $\sigma_w^2 = 10^{-2}/b^4$. The pricing experiments are conducted through controlling the prices for $R = 18$ agents.



(a) ReedCollege network. The clustering above is found by applying spectral clustering on the true L . The obtained RatioCut is 0.1419 w.r.t. A .



(b) Applying BlindCD method. The obtained RatioCut is 0.8953 w.r.t. A .



(c) Applying Boosted BlindCD method. The obtained RatioCut is 0.5249 w.r.t. A .

Figure 7: **Opinion dynamics experiments on ReedCollege network.** The network comprises of $N = 962$ agents and (approximately) $K = 2$ communities. The network's steady-state opinions are collected with $L = 10^4$ samples, each observed with a noise variance of $\sigma_w^2 = 10^{-2}$. There are $R = 150$ stubborn agents in the experiments, which are connected to the social network through a random bi-partite graph with connectivity $\log N/N$.

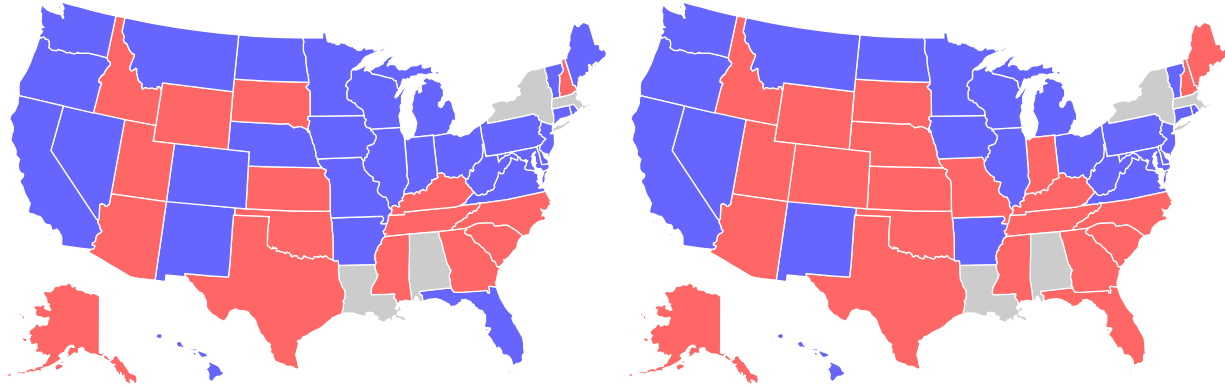


Figure 8: Applying BlindCD methods on the 110th US Senate Rollcall records. The states marked in red/blue are found to be in different communities; while the states marked in gray are marked as the ‘stubborn’ states as explained in the text. (Left) Results of BlindCD. (Right) Results of boosted BlindCD.

	CT	ME	NH	RI	VT	DE	NJ	PA	IL	IN	MI	OH	WI	IA	KS	MN
MA	0	0	0	9	4	0	0	8	0	0	0	5	0	0	0	0
NY	11	3	0	5	4	0	14	0	29	9	6	0	0	0	0	0
AL	0	0	31	0	0	0	0	0	0	0	0	0	0	5	5	0
LA	0	0	0	-2	0	0	0	0	0	2	13	0	0	0	0	0
	TN	WV	AZ	CO	ID	MT	NV	NM	UT	WY	CA	OR	WA	AK	HI	
MA	2	8	0	4	6	0	0	0	2	0	0	6	0	0	0	
NY	0	0	0	0	-9	0	0	0	0	0	10	0	0	0	0	
AL	19	0	7	0	19	0	18	0	28	34	0	0	0	9	0	
LA	0	5	0	0	0	28	0	0	0	0	0	0	0	0	0	
	OK	MO	NE	ND	SD	VA	AR	FL	GA	MS	NC	SC	TX	KY	MD	
MA	-7	6	0	0	1	0	1	0	0	-2	0	0	0	0	0	
NY	0	0	0	0	0	0	0	8	0	0	0	0	0	0	8	
AL	15	0	0	0	0	0	0	0	18	16	11	26	4	12	0	
LA	0	19	11	20	2	0	34	5	0	5	0	0	5	0	0	

Figure 9: Illustrating the $\hat{\mathbf{B}}^*$ matrix found with boosted BlindCD method [cf. (20),(41)]. Note that the values within the table have been rescaled.