Blind Identification of Stochastic Block Models from Dynamical Observations[*]

Michael T. Schaub[†], Santiago Segarra[‡], and John N. Tsitsiklis[§]

Abstract. We consider a blind identification problem in which we aim to recover a statistical model of a network without knowledge of the network's edges but based solely on nodal observations of a certain process. More concretely, we focus on observations that consist of single snapshots taken from multiple trajectories of a diffusive process that evolves over the unknown network. We model the network as generated from an independent draw from a latent stochastic block model (SBM), and our goal is to infer both the partition of the nodes into blocks and the parameters of this SBM. We discuss some nonidentifiability issues related to this problem and present simple spectral algorithms that provably solve the partition recovery and parameter estimation problems with high accuracy. Our analysis relies on recent results in random matrix theory and covariance estimation and on associated concentration inequalities. We illustrate our results with several numerical experiments.

1. Introduction. We consider the problem of inferring a statistical model of a network with unknown adjacency matrix $\boldsymbol{A} \in \{0,1\}^{n \times n}$ based on snapshots of multiple trajectories of a dynamical process that evolves over the network. Specifically, we are interested in the following setup (a more detailed description will be provided in section 3): We observe s independent samples $\boldsymbol{x}_T^{(1)}, \ldots, \boldsymbol{x}_T^{(s)}$ of a dynamical process at a fixed time T. Each sample is generated by the discrete-time system

$$(1.1) \qquad \boldsymbol{x}_{t+1}^{(i)} = f(\boldsymbol{A})\boldsymbol{x}_t^{(i)}, \qquad i = 1, \ldots, s,$$

initialized at some unknown $\boldsymbol{x}_0^{(i)}$, independently for each i. Each $\boldsymbol{x}_0^{(i)}$ is modeled as a random vector with unit covariance matrix ($\mathbb{E}[\boldsymbol{x}_0^{(i)}[\boldsymbol{x}_0^{(i)}]^\top] = \boldsymbol{I}$) drawn from some sub-Gaussian distribution (e.g., a standard multivariate Gaussian). Furthermore, we assume that the function f describing the dynamical network process is known and corresponds to a class of diffusive processes. Finally, $\boldsymbol{A}$ is assumed to be drawn independently according to a stochastic block

[†]Institute for Data, Systems and Society, Massachusetts Institute of Technology, Cambridge, MA 02139, and Department of Engineering Science, University of Oxford, Oxford, UK (michael.schaub@eng.ox.ac.uk).

[‡]Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 (segarra@rice.edu).

[§]Laboratory of Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 (jnt@mit.edu).

model (SBM), so that $\boldsymbol{A}$ is the same, albeit unknown, for each sample $\boldsymbol{x}_T^{(i)}$. Based on such data, our task is to identify the generative model underpinning the network: We wish to infer the parameters and the latent node partition of the SBM, without observing the network edges, by relying solely on the observations $\{\boldsymbol{x}_T^{(i)}\}_{i=1}^s$ at the nodes.

1.1. Motivation. In many applications we are confronted with the following scenario: We observe snapshots of the state of a system at particular times, and based on these observations we wish to infer the nature of the (dynamical) interactions between the entities that we observe. Examples abound and can be found across different fields [26, 27, 54].

Consider, e.g., measuring some node activities, such as the expression of opinions at s different, sufficiently separated instances of time in a social network. While the underlying social connections are typically not observable, the latent (unobserved) network will be essentially static and may be assumed to have emerged according to some homophilic interactions describable by an SBM. The measured opinion profiles will be shaped by both (random) initial personal opinions of the individuals, and certain network effects. Similar examples may arise when measuring (structurally stable) biological systems responding to different stimuli. For instance, we might have access to a set of measurements from certain regions of the brain and would like to infer the coupling between these regions. As another example, we might be interested in inferring a network of genetic interactions from multiple snapshots of gene expression, e.g., as obtained from single cell RNA-sequencing [43]. As taking these expression measurements will destroy the system under consideration, we are left with a few snapshots of data, from which we want to reconstruct (a model of) the underlying interactions. Finally, we may consider the monitoring of ecological populations from abundance data [13], where we often can only access snapshot information at specific instances of time, but we are interested in knowing the network of interactions between the species.

The above are canonical examples of problems where the system of interest is composed of a large number of different entities and can be conveniently conceptualized as a network. In this abstraction, nodes represent the individual entities, and the couplings between these nodes take place along the edges of the network. In many settings, we may have reason to believe that an observed dynamical process emerges from the interaction between nodes over such a network. While we may have some knowledge about the dynamics of each individual node, the edges of the network—which enable the interactions between nodes—are often unknown or only partially known. In such cases it is often desirable to learn the structure of the network based on observational data. This is a general class of network inference problems, which is a fundamental constituent of network analysis but has received limited attention.

On the other hand, there are many situations where inferring exactly the couplings in the network, e.g., identifying all edges, is out of reach for a number of reasons. First, we may not have access to sufficiently many samples. Second, the data and the network itself may be subject to some fluctuations with associated uncertainties. Finally, we may only be able to partially observe the dynamical system, so that the full model is nonidentifiable.

In such cases, it is then reasonable to aim for an *approximate* description of the system in terms of a statistical model that captures the coarse-grained features of the underlying network. This is precisely the type of problem addressed in this paper for the special case where the dynamics are of a linear, diffusive type, and the network model to be inferred is an SBM.

1.2. Contributions. We present a rigorous study of blind model identification for random networks. Our main contributions are as follows:

First, we present general algorithmic ideas on how to solve a blind identification problem using spectral methods, and we introduce simple algorithms to solve the concrete problems considered in this paper (see sections 4 and 5 and Algorithms 4.1 and 5.1).

Second, we establish a concentration inequality for the sample covariance matrix that is calculated on the basis of the observed samples (see Theorems 4.1 and 4.2).

Finally, we discuss how this concentration result translates into explicit bounds for the recovery error of the latent partitions (see Theorem 4.3), and we provide bounds on the estimation errors for the parameters of the SBM for the specific example of a planted partition model (see Theorem 5.3).

1.3. Related work. Many different network inference notions have been featured in the literature. For instance, the estimation of "functional" couplings based on statistical association measures, such as correlation or mutual information, is an analysis often encountered in the context of biological systems [14, 19, 36, 52]. Another type of network inference is causal inference, as discussed in, e.g., [5, 37, 53], in which one aims to infer directed graphs that encode direct causal influences between the entities. In contrast, the type of network inference we consider here is what may be called "topological" inference: Given a system of dynamical units, we want to infer their direct "physical" interactions. Some recent surveys include [12, 30, 54, 62]. More specifically, for an evolving dynamical process we might want to infer the underlying adjacency matrix. This may be viewed as a system identification problem, as considered in control theory [26, 27]. However, in the system identification literature, one can usually control the inputs to the system and observe the full system response. This is different from our setting, where there is no control, and we only observe a limited number of snapshots.

Problems of network identification that are somewhat similar to ours have received extensive attention in the literature recently. In many cases, formulations based on the optimization of a sparsity-promoting cost function [8, 16, 22], or formulations based on sparse dictionary learning [6, 65], are used. A number of methods also employ spectral characteristics of the observed outputs to identify (aspects of) the network topology [17, 31, 32]. In particular, the case of diffusive dynamics, similar to ours, is considered in [50, 51], which aim at inferring the network topology based on the (cross-)power spectral density of a consensus process driven by noise, and on a node knockout strategy. Methods combining spectral identification with optimization techniques are studied in [46, 48, 49, 59].

In all of the above works, however, the focus is on inferring the *exact* network topology, i.e., the complete specification of the network in terms of its adjacency or Laplacian matrices. In contrast, our goal is not to infer a specific network coupling but rather to identify a statistical model thereof. Stated differently, we wish to infer a distribution over networks that is compatible with the available observations. This scenario is particularly relevant if our ability to reconstruct the network is limited by the amount of samples that we can obtain. Closest to our approach is the conference paper [45], in which a related observational model is discussed, with a *new adjacency matrix* drawn for each sample in (1.1), and which provides results about the asymptotic consistency of partition recovery. In contrast, we consider the case of a constant but unknown adjacency matrix, provide nonasymptotic results in the number of samples, and, additionally, address the problem of parameter estimation.

Similar to this paper, the works [60, 61] study the "blind solution" of a ratio-cut problem (a common heuristic for partitioning graphs into dense, assortative clusters) based on observations of the output of a low-pass graph filter excited by a low-rank excitation signal. Those works provide results on how well the ratio-cut problem can be approximated on arbitrary, fixed graphs, under the assumption of a low-rank excitation structure. In contrast, here we aim to infer an explicit generative graph model and can thus provide probabilistic performance guarantees about the recovery of the true planted partition. The works by Hoffmann et al. [18] and Peixoto [39] discuss hierarchical Bayesian models and inference schemes for related problems but do not provide a theoretical analysis or statistical guarantees. Other related problems include the study of multireference alignment and synchronization problems (see, e.g., [4, 40]) and the study of state compression in Markov chains [67]. See also [33] for an example in which network reconstruction is attempted based on dynamics but without knowledge of temporal information.

Finally, the mathematical machinery that we use has connections to ideas from spectral methods for clustering on graphs and SBM recovery [1, 41, 42, 58] and is related to the statistics literature on so-called spiked random matrices, with a low-rank signal component. In particular, our development relies on recent results regarding the efficiency of estimating low-rank covariance matrices [7].

1.4. Outline. The rest of the paper is structured as follows. After briefly reviewing relevant preliminaries and notation in section 2, in section 3 we formally introduce the identification problem that we will study. We then present our main results and algorithmic ideas on recovering the latent *partitions* of the SBM in section 4. Building on those insights, in section 5 we discuss procedures for learning the *parameters* of the SBM. Section 6 contains the proofs of our main technical results, and in section 7 we present computational experiments to numerically assess the performance of our methods. We conclude with a brief discussion and outline possible avenues for future work.

2. Preliminaries.

2.1. Networks. An undirected network $\mathcal{G}$ consists of a node set $\mathcal{V}$ of known cardinality n, and an edge set $\mathcal{E}$ of unordered pairs of elements of $\mathcal{V}$. The edges can be conveniently represented as entries of a symmetric adjacency matrix $\boldsymbol{A} \in \{0,1\}^{n \times n}$, such that $A_{ij} = A_{ji} = 1$ if and only if $(i,j) \in \mathcal{E}$. We define the diagonal degree matrix $\boldsymbol{D} = \operatorname{diag}(\boldsymbol{A}\mathbf{1})$, where $\mathbf{1}$ denotes the vector of all ones, and $\operatorname{diag}(\boldsymbol{x})$ denotes the diagonal matrix whose entries are given by the components of the vector $\boldsymbol{x}$. We define the normalized adjacency matrix as $\boldsymbol{L} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$. Note that $\boldsymbol{L}$ is a shifted version of the normalized Laplacian matrix of the graph and is sometimes referred to in the literature simply as the "normalized Laplacian" [41, 42].

2.2. The stochastic block model. The SBM is a latent variable model that defines a probability measure over the set of undirected networks of fixed size n, represented by an adjacency matrix $\boldsymbol{A} \in \{0,1\}^{n \times n}$. In an SBM, the network is assumed to be partitioned into k nonempty groups of nodes, so that each node i is endowed with a latent group label $g_i \in \{1, \ldots, k\}$. Given these latent group labels, each link A_{ij} is an independent Bernoulli random variable that takes value 1 with probability Ω_{g_i, g_j} and value 0 otherwise; that is, the

probability of a link between two nodes depends only on their group labels:[1]

$$A_{ij} \sim \text{Ber}(\Omega_{g_i,g_j}).$$

To compactly describe the model, we collect link probabilities between the different groups in a symmetric affinity matrix $\boldsymbol{\Omega} = [\Omega_{ij}] \in [0,1]^{k \times k}$. Furthermore, we define a partition indicator matrix $\boldsymbol{G} \in \{0,1\}^{n \times k}$, with entries $G_{ij} = 1$ if node i is in group j, and $G_{ij} = 0$ otherwise. Based on these definitions, we can write the expected adjacency matrix under the SBM, given the partition and the group labels as[2]

$$(2.1) \qquad \boldsymbol{\mathcal{A}} := \mathbb{E}[\boldsymbol{A} \,|\, \boldsymbol{G}] = \boldsymbol{G}\boldsymbol{\Omega}\boldsymbol{G}^\top.$$

As $\boldsymbol{\mathcal{A}}$ encodes all of the parameters of an SBM, in what follows we will refer to such a stochastic block model as $\mathcal{M}(\boldsymbol{\mathcal{A}})$. Throughout the rest of the paper, and in order to avoid degenerate cases, we will always assume that $\boldsymbol{\mathcal{A}}$ has rank k. In particular, this excludes the case where two rows in $\boldsymbol{\mathcal{A}}$ are colinear, i.e., the case where the connection probabilities of two groups are identical or a linear scaling of each other.[3]

The SBM leads to a network whose expected adjacency matrix has a low-rank block structure (rank k). Due to the special structure induced by (2.1), there exists a simple relationship between the partition indicator matrix $\boldsymbol{G}$ and the eigenvectors of the normalized version of the expected adjacency matrix $\boldsymbol{\mathcal{L}}$, defined by

$$(2.2) \qquad \boldsymbol{\mathcal{L}} := \mathbb{E}[\boldsymbol{D}]^{-1/2}\mathbb{E}[\boldsymbol{A}]\mathbb{E}[\boldsymbol{D}]^{-1/2} = \text{diag}(\boldsymbol{\mathcal{A}}\mathbf{1})^{-1/2}\boldsymbol{\mathcal{A}}\,\text{diag}(\boldsymbol{\mathcal{A}}\mathbf{1})^{-1/2}.$$

The relationship between $\boldsymbol{G}$ and the eigenvectors of $\boldsymbol{\mathcal{L}}$ is given in Lemma 2.1 below and proved in the appendix. We denote the eigenvalues of $\boldsymbol{\mathcal{L}}$ by μ_i and assume that they are sorted in order of decreasing magnitude, so that $|\mu_1| \geq |\mu_2| \geq \cdots \geq |\mu_k| > |\mu_{k+1}| = \cdots = |\mu_n| = 0$.

Lemma 2.1. *Consider the matrix $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k]$ of eigenvectors of $\boldsymbol{\mathcal{L}}$ corresponding to the k nonzero eigenvalues. Then, there exists a $k \times k$ orthogonal matrix $\boldsymbol{\mathcal{U}}$ such that*

$$(2.3) \qquad \boldsymbol{V} = \boldsymbol{G}(\boldsymbol{G}^\top\boldsymbol{G})^{-1/2}\boldsymbol{\mathcal{U}}.$$

Moreover, if we define the row-normalized eigenvector matrix $\widetilde{\boldsymbol{V}} = \text{diag}(\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}\boldsymbol{V}$, then

$$(2.4) \qquad \widetilde{\boldsymbol{V}} = \boldsymbol{G}\boldsymbol{\mathcal{U}}.$$

Given an SBM with an expected adjacency matrix $\boldsymbol{\mathcal{A}}$ as described above, let us denote the expected degree of node i by $\delta_i := \mathbb{E}[\boldsymbol{D}_{ii}] = [\boldsymbol{\mathcal{A}}\mathbf{1}]_i$, and let $\delta_{\min}$ be the smallest expected degree. We will restrict attention to SBMs with a sufficiently large $\delta_{\min}$.

---

[1]For simplicity, our model allows self-loops.

[2]Note the slight abuse of notation here with respect to the conditional expectation. We treat the partition indicator matrix $\boldsymbol{G}$ as a degenerate random variable with a deterministic but unknown value.

[3]Our main results rely on the analysis of a k-dimensional dominant subspace in the observed data, which is induced by the structure of $\boldsymbol{\mathcal{A}}$. If $\boldsymbol{\mathcal{A}}$ had rank $p < k$, then our arguments would need to be adjusted accordingly.

Definition 2.2 (SBMs with concentration property). *Given some $\epsilon > 0$, we define $\mathcal{M}_n(\epsilon)$ as the class of SBMs $\mathcal{M}(\boldsymbol{A})$ with n nodes, for which*

$$\delta_{min} > 27 \ln\left(\frac{4n}{\epsilon}\right).$$

The condition on the minimal degree essentially restricts $\mathcal{M}_n(\epsilon)$ to correspond to models such that all nodes that belong to the same group belong to the same connected component, with high probability.[4] This is what will allow us to obtain reliable group membership information from the link patterns. Indeed, without such an assumption, and even if we knew the adjacency matrix $\boldsymbol{A}$, the exact recovery (with high probability) of the partition would be impossible [1].

For the above class of models the following concentration result can be established.

Lemma 2.3 (concentration for normalized adjacency matrix). *Fix some $\epsilon > 0$. Consider a normalized adjacency matrix $\boldsymbol{L} := \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$, computed from an adjacency matrix $\boldsymbol{A}$ generated according to an SBM model $\mathcal{M}(\boldsymbol{A})$. If $\mathcal{M}(\boldsymbol{A}) \in \mathcal{M}_n(\epsilon)$, then the following bound holds with probability at least $1 - \epsilon$:*

$$\|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\| \le 3\sqrt{\frac{3\ln(4n/\epsilon)}{\delta_{min}}},$$

where we recall that $\boldsymbol{\mathcal{L}} := \mathbb{E}[\boldsymbol{D}]^{-1/2}\mathbb{E}[\boldsymbol{A}]\mathbb{E}[\boldsymbol{D}]^{-1/2}$.

Lemma 2.3 follows from a result by Chung and Radcliffe [9]. The proof, which parallels [9], is given in the appendix for completeness. In the following, we will be primarily interested in the case where $\delta_{\min} = \omega(\ln(n/\epsilon))$. In this case, when n is large, the condition in Definition 2.2 will be satisfied.[5] Lemma 2.3 then guarantees a tight spectral concentration for large networks, i.e., that $\|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\|$ will be small, with high probability.

2.3. Sub-Gaussian random vectors. For convenience, we recall here the definition of sub-Gaussian random variables and vectors.

Definition 2.4 (sub-Gaussian random variables and vectors). *A zero-mean random variable $x \in \mathbb{R}$ is sub-Gaussian if there exists a constant $\sigma > 0$ such that $\mathbb{E}[\exp(tx)] \le \exp(t^2\sigma^2/2)$ for all $t \in \mathbb{R}$. A zero-mean random vector $\boldsymbol{x} \in \mathbb{R}^n$ is sub-Gaussian if for any nonrandom $\boldsymbol{u} \in \mathbb{R}^n$, the random variable $z = \boldsymbol{u}^\top \boldsymbol{x}$ is sub-Gaussian.*

Note that any Gaussian random variable or vector is sub-Gaussian. We also define the sub-Gaussian norm as follows.

Definition 2.5 (sub-Gaussian norm [7]). *The sub-Gaussian norm of a sub-Gaussian random variable x is $\|x\|_{\Psi_2} := \sup_{k \ge 1} k^{-1/2}(\mathbb{E}|x|^k)^{1/k}$. For a sub-Gaussian random vector $\boldsymbol{x}$, the sub-Gaussian norm $\|\boldsymbol{x}\|_{\Psi_2}$ is defined as $\|\boldsymbol{x}\|_{\Psi_2} := \sup_{\boldsymbol{u} \in \mathbb{R}^n \setminus \{0\}} \|\boldsymbol{u}^\top \boldsymbol{x}\|_{\Psi_2}/\|\boldsymbol{u}\|$.*

---

[4]Note that for the special case of a single group, we are dealing with an Erdős–Rényi random graph, and it is well known that if $\delta_{\min} > c\ln(n)$, with $c > 1$, the random graph is connected with high probability, as $n \to \infty$.

[5]More precisely, the condition $\delta_{\min} = \omega(\ln(n/\epsilon))$ means that $\delta_{\min}/\ln(n/\epsilon) \to \infty$, as $n \to \infty$.

3. The blind identification problem. In this section, we introduce the specific inference problem that we will study. We first provide a brief discussion on the class of dynamical models we consider, and then describe the precise problem setup.

3.1. Dynamical model. We consider the dynamical model described in (1.1), where the mapping f is of the form

$$(3.1) \qquad f : \boldsymbol{A} \mapsto \sum_{\ell=1}^{R} \alpha_\ell \boldsymbol{L}^\ell,$$

for some known constants α_ℓ, and a known degree R, where $\boldsymbol{L} = \boldsymbol{D}^{-1/2} \boldsymbol{A} \boldsymbol{D}^{-1/2}$ is the (unknown) normalized adjacency matrix of the (unobserved) underlying graph. We assume that the system matrix satisfies $\|f(\boldsymbol{A})\| \leq 1$ and corresponds to a (marginally) stable dynamical system.

The above mapping f can be considered to describe some local interactions between the nodes, in that it allows for interactions between nodes that are up to R steps away in the graph. Alternatively, the mapping (3.1) f may be considered from the perspective of graph signal processing [30, 35] as a graph filter of order R that acts (instantaneously) on the state variables of the dynamical system (1.1) at each time step.

To make our discussion more concrete, in the following we will concentrate on the dynamical system

$$(3.2) \qquad \boldsymbol{x}_{t+1} = \boldsymbol{L} \boldsymbol{x}_t$$

as our main example. This discrete-time system, and by extension the dynamics in (3.3), is closely related to diffusion processes on networks [29, 47]. Indeed, $\boldsymbol{L}$ is related via a similarity transformation to the (column-stochastic) transition matrix $\boldsymbol{A}\boldsymbol{D}^{-1}$ (or diffusion kernel) of an unbiased random walk on the network. Such random walk dynamics have been used to model a variety of processes, such as the spreading of rumors and opinions, and have found applications in other areas of network analysis, such as ranking nodes or detecting communities [29]. Another interpretation of the model is in terms of (discrete-time) consensus processes, which are distributed averaging processes that have been considered extensively in the control theory literature [21, 34, 57].

It is well known that $\|\boldsymbol{A}\boldsymbol{D}^{-1}\| = 1$ and, accordingly, $\|\boldsymbol{L}\| = 1$, since $\boldsymbol{L}$ and $\boldsymbol{A}\boldsymbol{D}^{-1}$ are related via a similarity transform. Hence, the assumption $\|f(\boldsymbol{A})\| \leq 1$ is fulfilled by our example dynamics. We will make use of this fact in our calculations later.

Remark 3.1. By allowing for a large enough R we can view the observation $\boldsymbol{x}_T$ at time T as a signal $\boldsymbol{x}_0$ filtered by a graph filter with mapping $f'(\boldsymbol{A})$, where $f'(\boldsymbol{A})$ is obtained by composing T copies of the original function f. However, we refrain from considering this "static" filter perspective here, as this would disguise a trade-off between the sampling time and the number of samples required for our identification problem, which will become apparent in the following sections.

3.2. Problem description. Let us provide a detailed breakdown of the observation model we consider.

1. We have a discrete-time dynamical system with state vector $\boldsymbol{x}_t \in \mathbb{R}^n$, which evolves according to

$$(3.3) \qquad\qquad \boldsymbol{x}_{t+1} = f(\boldsymbol{A})\boldsymbol{x}_t,$$

   where f is of the form (3.1).

2. We assume that the normalized adjacency matrix $\boldsymbol{L}$ that appears in $f(\boldsymbol{A})$ is generated by first drawing a graph from an SBM with k blocks, (unknown) partition indicator matrix $\boldsymbol{G}$, and some (unknown) affinity matrix $\boldsymbol{\Omega}$, and then normalizing its adjacency matrix $\boldsymbol{A}$ to obtain $\boldsymbol{L}$ and $f(\boldsymbol{A})$.

3. We obtain s samples $\boldsymbol{x}_T^{(i)}$, $i = 1, \ldots, s$, at a known time T, associated to s different trajectories.

4. The (unobserved) initial vectors $\boldsymbol{x}_0^{(i)}$ for the different samples are independent and identically distributed (i.i.d.), drawn from a zero-mean sub-Gaussian vector distribution, where the components of the vector are uncorrelated, i.e., $\mathbb{E}[\boldsymbol{x}_0^{(i)}[\boldsymbol{x}_0^{(i)}]^\top] = \boldsymbol{I}$.

Under this setup, we consider the following two problems, which we refer to as the *partition recovery* and *parameter estimation* problems.

Problem 3.2 (SBM partition recovery from dynamics). *Given s independent samples, each one drawn at a fixed time T according to the above described model, recover the latent partition of the nodes into blocks.*

Problem 3.3 (SBM parameter estimation from dynamics). *Given s independent samples, each one drawn at a fixed time T according to the above described model, recover the affinity matrix $\boldsymbol{\Omega}$ describing the link probabilities in the underlying SBM.*

Note that as the assignment of labels to groups is arbitrary, we can only hope to recover the block structure and the affinity matrix $\boldsymbol{\Omega}$ up to a permutation of the group labels. Similarly, as will become clear in the following sections, in general, the parameters of the SBM will not be identifiable unless some additional assumptions are made on the underlying model or some additional side information is available on the structure of the network.

Moreover, it should be apparent that the difficulty of the above problems will depend on the specific realization of the initial conditions $\boldsymbol{x}_0^{(i)}$ and of the adjacency matrix $\boldsymbol{A}$. Thus, we cannot expect accurate inference for every possible realization. Instead we look for algorithms and results that, under certain assumptions, guarantee accurate inference with high probability. We make the following technical assumption about the data generation process.

Assumption 3.4 (bounded moments assumption). *We assume that there exists an absolute constant $c_0 > 0$ such that $\mathbb{E}[(\boldsymbol{u}^\top \boldsymbol{x}_0^{(i)})^2] \geq c_0 \|\boldsymbol{u}^\top \boldsymbol{x}_0^{(i)}\|_{\Psi_2}^2$ for all $\boldsymbol{u}$.*

As discussed in [7], Assumption 3.4 effectively bounds all moments of $\boldsymbol{x}_0^{(i)}$ as a function of the second moment of $\boldsymbol{x}_0^{(i)}$. As a special case, this assumption is satisfied by a standard Gaussian vector, with $c_0 = \pi/2$ [7]. Assumption 3.4 will enable us to invoke (via Lemma 6.2) the results of [7] on the estimation of low-rank covariance matrices, thereby facilitating our analysis.

Remark 3.5. If the number of available samples becomes large ($s \to \infty$), we could, in principle, try to infer the exact adjacency matrix $\boldsymbol{A}$. However, we are interested in a situation

---

Algorithm 4.1. Partition recovery.

Input: Samples $\{\boldsymbol{x}_T^{(i)}\}_{i=1}^s$, number of groups k
Output: Partition $\mathcal{P}$
1: Compute sample covariance $\widehat{\boldsymbol{\Sigma}}(T) := \frac{1}{s}\sum_{i=1}^s [\boldsymbol{x}_T^{(i)} - \bar{\boldsymbol{x}}_T][\boldsymbol{x}_T^{(i)} - \bar{\boldsymbol{x}}_T]^\top$
2: Compute top k eigenpairs: $\{(\lambda_i, \boldsymbol{w}_i)\}_{i=1}^k \leftarrow \text{eig}(\widehat{\boldsymbol{\Sigma}}(T))$, where $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_k|$
3: Form the matrix $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k] \in \mathbb{R}^{n \times k}$ and normalize its rows to have unit norm: $\widetilde{\boldsymbol{W}} = \text{diag}(\boldsymbol{W}\boldsymbol{W}^\top)^{-1/2}\boldsymbol{W}$
4: Perform k-means clustering on the rows of $\widetilde{\boldsymbol{W}}$: $\mathcal{P} \leftarrow k\text{-means}(\widetilde{\boldsymbol{W}}^\top)$

---

where s is comparatively small ($s \ll n$), and our goal is to infer the SBM model that generated $\boldsymbol{A}$ (and thus $\boldsymbol{L}$), that is, the latent partition and the affinity matrix $\boldsymbol{\Omega}$.

4. Partition recovery. In this section we first present our main results and an algorithm for solving the partition recovery problem (Problem 3.2) and then provide a broader discussion of the underlying algorithmic ideas.

4.1. Main results. Our algorithm for partition recovery relies on spectral clustering and is displayed as Algorithm 4.1. It builds upon the idea that the *sample covariance matrix* $\widehat{\boldsymbol{\Sigma}}(T)$ at time T, defined by

$$(4.1) \qquad \widehat{\boldsymbol{\Sigma}}(T) := \frac{1}{s}\sum_{i=1}^s [\boldsymbol{x}_T^{(i)} - \bar{\boldsymbol{x}}_T][\boldsymbol{x}_T^{(i)} - \bar{\boldsymbol{x}}_T]^\top,$$

where $\bar{\boldsymbol{x}}_T = \frac{1}{s}\sum_{i=1}^s \boldsymbol{x}_T^{(i)}$ is the sample mean, can serve as a proxy for what we call the *ensemble covariance matrix*

$$(4.2) \qquad \boldsymbol{S}(T) := f(\boldsymbol{\mathcal{A}})^{2T}.$$

Because the initial vector has identity covariance, it is seen that $\boldsymbol{S}(T)$ would be the covariance matrix of the vector $\boldsymbol{x}_T^{(i)}$ if the dynamics were evolving according to the matrix $f(\boldsymbol{\mathcal{A}})$, i.e., if we were to replace the adjacency matrix of the sampled network with its expected value. Intuitively, the matrix $f(\boldsymbol{\mathcal{A}})$ encodes the partition structure of the SBM, and since $\boldsymbol{S}(T)$ is simply a matrix power of $f(\boldsymbol{\mathcal{A}})$, it has the same dominant eigenspace. Being able to compute a close proxy for $\boldsymbol{S}(T)$ enables us to gain information about the partition structure. Thus, as long as the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}(T)$ is close to the ensemble covariance matrix $\boldsymbol{S}(T)$, we can expect good recovery performance.

To gain some first insight, Theorem 4.1 below addresses the relation between these two covariance matrices for our example dynamics $f(\boldsymbol{A}) = \boldsymbol{L}$.

Theorem 4.1. *Let $\boldsymbol{L}$ be the normalized adjacency matrix constructed from a graph drawn from an SBM $\mathcal{M}(\boldsymbol{\mathcal{A}}) \in \mathcal{M}_n(\epsilon)$ with n nodes and minimal degree δ_{min}. Let $\boldsymbol{x}_0^{(i)}$ $i = 1, \ldots, s$, be i.i.d. zero-mean sub-Gaussian vectors that satisfy Assumption 3.4, with covariance matrix $\boldsymbol{I}$.*

Then, for the dynamics of the form (3.2), and with probability at least $1 - 5s^{-1} - 2\epsilon$, we have

$$(4.3) \qquad \left\|\widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{S}(T)\right\| \leq 2TM + B(T),$$

where $\widehat{\boldsymbol{\Sigma}}(T)$ is defined in (4.1), $\boldsymbol{S}(T)$ *is defined in* (4.2),

$$M := 3\sqrt{\frac{3\ln(4n/\epsilon)}{\delta_{min}}}, \qquad B(T) := C\left(1 + \sum_{i=2}^{k}\min\{1, |\mu_i| + M\}^{2T} + (n-k)M^{2T}\right)\sqrt{\frac{\ln s}{s}},$$

C is some absolute constant, and μ_i are the eigenvalues of $\boldsymbol{\mathcal{L}}$ ordered by decreasing magnitude.

The proof of Theorem 4.1 is given in subsection 6.1.[6]

For some interpretation, note that the term $2TM$ in (4.3) is independent of the number of samples. As we will see below, this term arises because the graph on which the dynamics takes place is only a proxy for the expected graph under the SBM and is drawn only once. We may think of this error term as a "baseline" insurmountable error that we will incur no matter how many samples we collect. In light of the condition on $\delta_{\min}$ in Definition 2.2, we have $M < 1$, but nevertheless the term $2TM$ in the error bound increases (linearly) with time. Hence, the later our observations (larger T), the worse the first term in our bound. From the perspective of our model (3.3), this is intuitive: Since $\lambda_i(\boldsymbol{L}) \leq 1$ for all i, the relevant signal that we observe will be attenuated over time, making it more difficult to extract information about the embedded partition from our observations when T is large.

In contrast to the first term, the second term $B(T)$ in (4.3) depends on both the time T and the number of samples s. The sample dependence is related to the convergence rate of the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}(T)$ to the true covariance. The time-dependence of $B(T)$ is again related to M but in a nonlinear fashion. Since $M < 1$, the error term $B(T)$ decreases with time.

The above discussion reveals a trade-off. If T is large, then the first term, $2TM$, is large and dominates. If T is small, then the second term, $B(T)$, can be large, especially if the number of samples, s, is moderate. In general, an intermediate value of T will yield the most favorable upper bound.

It is also insightful to consider the behavior of our bound both for the asymptotic regime in which we have a large number of samples $s \to \infty$ (even though this is not the setting of interest to us) and in the limit of large networks ($n \to \infty$). If we consider the case $s \to \infty$, and in the regime where $\delta_{\min} = \omega(\ln(4n/\epsilon))$, then the error term is optimal. This is exactly in line with the requirements of "exact recovery" of the partition for an SBM where the network (the matrix $\boldsymbol{A}$) is observed [1]. Similarly, assuming a sufficiently large $\delta_{\min}$ as $n \to \infty$, the dependency on the number of samples is aligned with the optimal rate for the estimation of low-rank covariance matrices [7].

For the general case, where f is given by (3.1), an analogous result can be established.

Theorem 4.2. *Let $\boldsymbol{L}$ be the normalized adjacency matrix constructed from a graph drawn from an SBM $\mathcal{M}(\boldsymbol{\mathcal{A}}) \in \mathcal{M}_n(\epsilon)$ with n nodes and minimal degree δ_{min}. Let $\boldsymbol{x}_0^{(i)}$, $i = 1, \ldots, s$, be i.i.d. zero-mean sub-Gaussian vectors that satisfy Assumption 3.4, with covariance matrix $\boldsymbol{I}$.*

---

[6]The probability $1 - 5s^{-1} - 2\epsilon$ in the statement of the theorem can be improved to $1 - 5s^{-1} - \epsilon$ by a more careful use of the union bound in the proof. The same comment applies to Theorems 4.2 and 4.3.

Then, for the dynamics of the form (3.3)*, and with probability at least* $1 - 5s^{-1} - 2\epsilon$*, we have*

$$(4.4) \qquad \left\| \widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{S}(T) \right\| \leq 2TM_f + B_f(T),$$

where $\widehat{\boldsymbol{\Sigma}}(T)$ *is defined in* (4.1)*,*

$$M_f := 3\sqrt{\frac{3\ln(4n/\epsilon)}{\delta_{min}}} \cdot \left(\sum_{\ell=1}^{R} \ell\alpha_\ell \right),$$

$$B_f(T) := C\left(1 + \sum_{i=2}^{k} \min\{1, |\zeta_i| + M_f\}^{2T} + (n-k)M_f^{2T} \right)\sqrt{\frac{\ln s}{s}},$$

C *is some absolute constant, and* ζ_i *are the eigenvalues of* $f(\boldsymbol{\mathcal{A}})$ *ordered by decreasing magnitude.*

The proof of this result is given in subsection 6.2.

Based on the concentration result in Theorem 4.2, we can establish the following guarantee for the partition recovery problem. This result is phrased in terms of the misclassification rate $q := |\mathbb{M}|/n$, where $\mathbb{M}$ is the set of "misclassified nodes" formally defined in subsection 6.3.[7]

Theorem 4.3. *Let* $\xi_1 \geq \cdots \geq \xi_k$ *be the top* k *eigenvalues of the ensemble covariance matrix* $\boldsymbol{S}(T) = f(\boldsymbol{\mathcal{A}})^{2T}$*, which we assume to be positive. Let* $\boldsymbol{W}$ *and* $\boldsymbol{V}$ *denote the matrices formed by the top* k *eigenvectors of* $\widehat{\boldsymbol{\Sigma}}(T)$ *and* $\boldsymbol{S}(T)$*, respectively, and denote their row-normalized versions by* $\widetilde{\boldsymbol{W}}$ *and* $\widetilde{\boldsymbol{V}}$*. Define* τ *to be the minimum* 2*-norm of any row of* $\boldsymbol{W}$ *and* $\boldsymbol{V}$*. Then, with probability* $1 - 5s^{-1} - 2\epsilon$ *the following bound for the misclassification rate* q *holds:*

$$(4.5) \qquad q \leq \frac{64k}{n\tau^2\xi_k^2}(2TM_f + B_f(T))^2,$$

where $B_f(T)$ *is as defined in Theorem* 4.2*.*

The proof is given in subsection 6.3.2. Theorems 4.1 to 4.3 provide performance guarantees that show the interplay between the number of groups k, the network size n, and the spectral properties of the network, as encapsulated in μ_k and $\|\widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{S}(T)\|$. Before continuing, let us add a few remarks on some aspects of these results.

1. The time T at which the samples are obtained is not a required input to the algorithm (cf. Algorithm 4.1). However, as is apparent from Theorems 4.1 to 4.3, the partition recovery performance is influenced by the sampling time. In particular, since the eigenvalues ξ_k of the ensemble covariance matrix $f(\boldsymbol{\mathcal{A}})^{2T}$ can only decrease with time, the bound we can obtain for the misclassification rate is weaker in comparison to the concentration result for the covariance matrix (Theorems 4.1 and 4.2).

2. In Algorithm 4.1, we require the number of blocks k to be an input to our algorithm. In principle, we can try to infer this parameter from the data, e.g., by looking at some appropriate statistics of the eigenvalue gaps; see, for example, [44, 58, 63], or the discussion in [7] on the detection of eigenvalue jumps in covariance matrix estimation. However, we will not carry out this line of analysis in this paper.

---

[7]Importantly, this definition of misclassified nodes assumes that we can solve the k-means procedure exactly. In practice, however, only a $(1 + \epsilon_0)$-approximation may be obtained efficiently [24].

3. The same approach can be used for other types of (linear) operators derived from a low-rank statistical network model such as the SBM. For instance, one may consider inference for other latent graph models, such as degree-corrected SBMs [10, 23].

Remark 4.4. Note that in Theorem 4.3, we assume that the first k eigenvalues of $\boldsymbol{S}(T)$ are positive. Under the premise that we are dealing with a k-group SBM $\boldsymbol{\mathcal{A}}$ with rank k, this will be generically true. However, if $\text{rank}(\boldsymbol{\mathcal{A}}) = p < k$, then the above theorem would need to be adjusted, as only the first p eigenvectors would carry information about the latent block structure, leading to possible nonidentifiability. The case $\text{rank}(\boldsymbol{\mathcal{A}}) < k$ could arise, for example, if we consider a block model with equal-sized groups, in which two rows of the affinity matrix $\boldsymbol{\Omega}$ are colinear. For a related discussion on some of these issues, see, e.g., [15]. Less trivial examples are also possible, e.g., if the rank of $\boldsymbol{\mathcal{A}}$ is k but $\text{rank}(f(\boldsymbol{\mathcal{A}})) < k$.

4.2. Algorithm and proof ideas. We now provide a more technical discussion of the proposed algorithm for partition recovery from a high-level perspective. The proofs in section 6 will make these ideas mathematically precise. To simplify our discussion below, we will describe the algorithmic ideas in terms of the simple dynamics $f(\boldsymbol{A}) = \boldsymbol{L}$. The general case is completely analogous.

The intuition underpinning our algorithm is that while we cannot observe the edges of the network, the network structure affects the second moments of the observed samples in ways that we can exploit.

Our analysis relies on a comparison of three different covariance matrices: (a) the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}(T)$, defined in (4.1); (b) the ensemble covariance matrix $\boldsymbol{S}(T)$, defined in (4.2); and (c) the covariance matrix $\boldsymbol{\Sigma}(T)$, defined by

$$(4.6) \qquad \boldsymbol{\Sigma}(T) := \mathbb{E}_{\boldsymbol{x}_0}[\boldsymbol{x}_T \boldsymbol{x}_T^\top] - \mathbb{E}_{\boldsymbol{x}_0}[\boldsymbol{x}_T]\mathbb{E}_{\boldsymbol{x}_0}[\boldsymbol{x}_T^\top] = [\boldsymbol{L}^T][\boldsymbol{L}^T]^\top = \boldsymbol{L}^{2T},$$

where the first equality uses the zero-mean property of $\boldsymbol{x}_T$, and the last equality uses the symmetry of $\boldsymbol{L}$.

As a first step, we establish that the sample covariance $\widehat{\boldsymbol{\Sigma}}(T)$ serves as a good estimator of $\boldsymbol{\Sigma}(T)$. This is done by leveraging the fact that the effective rank $r_e(\boldsymbol{\Sigma}(T)) = \text{trace}(\boldsymbol{\Sigma}(T))/\|\boldsymbol{\Sigma}(T)\|$ of $\boldsymbol{\Sigma}(T)$ is low, and then showing that $\widehat{\boldsymbol{\Sigma}}(T)$ indeed concentrates around $\boldsymbol{\Sigma}(T)$, even when the number of samples is relatively small (Lemma 6.1).

Second, we observe that the eigenvectors of $\boldsymbol{\Sigma}(T)$ are the same as the eigenvectors of $\boldsymbol{L}$, although they now correspond to the modified eigenvalues $\lambda_i(\boldsymbol{\Sigma}(T)) = \lambda_i^{2T}(\boldsymbol{L})$. This transformation does not change the ordering of the eigenvalues in terms of their absolute values, and thus the dominant eigenspaces of $\boldsymbol{\Sigma}(T)$ and $\boldsymbol{L}$ are the same. In particular this implies that the dominant eigenspace of $\boldsymbol{\Sigma}(T)$ allows us to approximate the dominant eigenspace of $\boldsymbol{L}$.

Finally, we show that $\boldsymbol{L}$ concentrates around $\boldsymbol{\mathcal{L}}$ (cf. Lemma 2.3), so that $\boldsymbol{\Sigma}(T)$ also concentrates around $\boldsymbol{S}(T)$. In particular, the dominant eigenspace of $\boldsymbol{\Sigma}(T)$ allows us to approximate the dominant eigenspace of $\boldsymbol{S}(T)$ and of $\boldsymbol{\mathcal{L}}$, which in turn allows us to estimate the partition structure.

From the above discussion we see that Algorithm 4.1 is subject to two sources of randomness: (a) The normalized adjacency matrix $\boldsymbol{L}$ is an approximation of its expected version $\boldsymbol{\mathcal{L}}$, and (b) the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}(T)$ is an approximation of the true covariance matrix

$\boldsymbol{\Sigma}(T)$. Theorems 4.1 and 4.2 allow us to bound these two effects, leading to the performance guarantees in Theorem 4.3.

Remark 4.5 (computational considerations). In Algorithm 4.1, we first form the sample covariance and then analyze its spectral properties via an eigenvalue decomposition. Equivalently, we can employ an SVD of the normalized and centered sample matrix $\boldsymbol{X} = \frac{1}{\sqrt{s}}[\widetilde{\boldsymbol{x}}_T^{(1)}, \dots, \widetilde{\boldsymbol{x}}_T^{(s)}]$, where $\widetilde{\boldsymbol{x}}_T^{(i)} = \boldsymbol{x}_T^{(i)} - \bar{\boldsymbol{x}}_T$. Using the SVD $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Theta}\boldsymbol{Q}^\top$, it is easy to see that $\widehat{\boldsymbol{\Sigma}}(T) = \boldsymbol{X}\boldsymbol{X}^\top = \boldsymbol{U}\boldsymbol{\Theta}^2\boldsymbol{U}^\top$, and hence all of the information required by the algorithm is contained in the SVD of $\boldsymbol{X}$. While these procedures are mathematically equivalent, working directly with the SVD of $\boldsymbol{X}$ avoids forming $\widehat{\boldsymbol{\Sigma}}(T)$ explicitly and thus reduces the memory requirements when $s \ll n$.

5. Parameter estimation. In this section we turn to the parameter estimation problem. We first provide an overview of the general algorithmic idea with the help of an example. We then discuss certain obstacles to its solution, including some insurmountable nonidentifiability issues. In subsection 5.2 we then provide a generic "meta"-algorithm for the solution of the parameter identification problem. Finally, in subsection 5.3 we provide some theoretical guarantees for its performance, for a planted partition model, which is a special case of the SBM.

5.1. Algorithm ideas. To estimate the *parameters* of the underlying SBM we can use a strategy similar to that for the recovery of the *partitions*. The ensemble covariance matrix $\boldsymbol{S}(T) = f(\boldsymbol{\mathcal{A}})^{2T}$ is a transformation of the normalized adjacency matrix. Hence, it contains relevant information not only about the partitions but also about the model parameters. Let us illustrate this fact with our example dynamics $f(\boldsymbol{A}) = \boldsymbol{L}$ on a graph drawn from a planted partition model, a simple form of an SBM.

Example 5.1. *Consider a special case of the SBM, known as the planted partition model. Conditional on the group labels, the planted partition model with n nodes is described by probabilities a/n and b/n (with $a \neq b$) of linking to a node in the same group or a different group, respectively.*[8] *The affinity matrix for such a model, with k blocks, is of the form $\boldsymbol{\Omega} = \frac{(a-b)}{n}\boldsymbol{I}_k + \frac{b}{n}\boldsymbol{1}_k\boldsymbol{1}_k^\top$, where $\boldsymbol{I}_k$ and $\boldsymbol{1}_k$ denote the k-dimensional identity matrix and the vector of all ones, respectively. Let us consider such a planted partition matrix with equally sized groups. This leads to an expected matrix $\boldsymbol{\mathcal{L}}$ with entries*

$$[\boldsymbol{\mathcal{L}}]_{ij} = \begin{cases} \frac{a}{n} \cdot \frac{k}{a+(k-1)b} & \text{if } g_i = g_j, \\ \frac{b}{n} \cdot \frac{k}{a+(k-1)b} & \text{if } g_i \neq g_j, \end{cases}$$

where, we recall, g_i represents the latent group label of node i.

Based on these calculations we can find expressions for $\boldsymbol{S}(T) = \boldsymbol{\mathcal{L}}^{2T}$:

$$(5.1) \qquad [\boldsymbol{\mathcal{L}}^{2T}]_{ij} = \begin{cases} \frac{k-1}{n}\left(\frac{a-b}{a+(k-1)b}\right)^{2T} + \frac{1}{n} & \text{if } g_i = g_j, \\ -\frac{1}{n}\left(\frac{a-b}{a+(k-1)b}\right)^{2T} + \frac{1}{n} & \text{if } g_i \neq g_j. \end{cases}$$

---

[8]In Theorem 5.3, we will assume that a, b, and n are such that $\mathcal{M}(\boldsymbol{\mathcal{A}}) \in \mathcal{M}_n(\epsilon)$. In the context of this example, this means a and b are assumed to be of order at most $\ln(n)$.

This example suggests the following generic strategy for the parameter recovery problem. First, we can derive the functional dependencies of the entries of $\boldsymbol{S}(T)$ on the latent parameters. Let us denote these relations by $F(\boldsymbol{\Omega}, \boldsymbol{G}) = \boldsymbol{S}(T)$. In the above example, $F(\boldsymbol{\Omega}, \boldsymbol{G})$ is given by (5.1). Note that it is possible to derive such relationships $F(\boldsymbol{\Omega}, \boldsymbol{G}) = \boldsymbol{S}(T)$ for any SBM with generic parameters $\boldsymbol{\Omega}$ and $\boldsymbol{G}$, even though the functional relationships will be more complicated. Second, in principle, we can then use the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}(T)$ as an approximation of $\boldsymbol{S}(T)$ and estimate the parameters by leveraging the functional relationships $F(\boldsymbol{\Omega}, \boldsymbol{G}) \approx \widehat{\boldsymbol{\Sigma}}(T)$. However, there are some difficulties to overcome to render this general idea operational.

First, the derived equations might be insufficient to identify the model parameters. Indeed, notice that the equations in (5.1) do not enable us to identify both model parameters a and b of the planted partition model. This is an artifact of the symmetry properties of the normalized adjacency matrix $\boldsymbol{L}$. In particular, this degeneracy arises from an invariance in the dynamical model (3.3). Indeed, if in Example 5.1 we replace (a, b) by $(\gamma a, \gamma b)$, where γ is a positive constant, the expected matrix $\boldsymbol{\mathcal{L}}$ remains the same. Thus the model parameters are not recoverable, even if we had access to the exact matrix $\boldsymbol{\mathcal{L}}$. In the context of our specific model (3.3), we thus have to make some additional assumptions to obtain a viable parameter estimation approach.[9] We remark that for other dynamical models such additional assumptions may not be necessary.

A second issue is that the obtained equations will be typically inconsistent if we just plug in the sample covariance matrix. Thus, instead of simply inverting the functional relations $F(\boldsymbol{\Omega}, \boldsymbol{G}) \approx \widehat{\boldsymbol{\Sigma}}(T)$, we will have to find approximate solutions. There are many ways of obtaining an approximate solution in such cases, e.g., by formulating an optimization problem $\min_{\boldsymbol{\Omega}} \|F(\boldsymbol{\Omega}, \boldsymbol{G}) - \widehat{\boldsymbol{\Sigma}}(T)\|$. Which approach would have better statistical properties in the face of the noise present in $\widehat{\boldsymbol{\Sigma}}(T)$ is an interesting question. We do not go into such detail but provide instead a generic "meta"-algorithm in the next section.

A third possible issue with the above approach is that it requires the knowledge of the group labels. In the context of Example 5.1, based on (5.1) we can estimate the parameters a and b directly from the entries of $\boldsymbol{S}(T)$—or rather its estimated version $\widehat{\boldsymbol{\Sigma}}(T)$—but only if the group labels g_i of the nodes are known. This implies that we would have to first recover the latent partition before we can estimate the parameters based on these relations. Although this is a reasonable procedure in practice, the recovery performance will depend strongly on the quality of the partition recovery in the first step.

However, in certain cases we can avoid the latter problem as follows. In particular, we can express the eigenvalues of $\boldsymbol{S}(T)$ as a function of the model parameters and obtain a set of equations $\lambda(\boldsymbol{S}(T)) = F_\lambda(\boldsymbol{\Omega})$ for the parameters $\boldsymbol{\Omega}$. Notice that these equations are independent of the group labels of the nodes. The reason is that a permutation of the nodes defines a similarity transformation of $\boldsymbol{\mathcal{L}}$, and hence the eigenvalues of $\boldsymbol{\mathcal{L}}$—and accordingly those of $\boldsymbol{S}(T)$—are not affected by the ordering of the rows and columns of the matrix.

To make these ideas concrete let us reconsider our earlier example.

---

[9] For a planted partition model with two equally sized blocks, there is additional symmetry in the problem in that we can exchange (a, b) with (b, a) and obtain exactly the same set of equations. In this case an additional assumption, e.g., of the form $a > b$, is thus necessary for identifiability.

Example 5.2 (continued). *In terms of spectral analysis, the nonzero eigenvalues of $\boldsymbol{S}(T)$ are*

$$(5.2) \qquad \lambda_i(\boldsymbol{S}(T)) = \begin{cases} 1 & \text{if } i = 1, \\ \left(\frac{a-b}{a+(k-1)b}\right)^{2T} & \text{if } i = 2, \ldots, k. \end{cases}$$

As expected, these equations are independent of the group labels. However, as outlined in our discussion above, as the equations are redundant, on their own they are insufficient to identify all of the parameters, but some additional information is needed. Nevertheless, the fact that these equations are independent of the group labels makes this "eigenvalue-based" parameter inference strategy more amenable to theoretical analysis, as we will see in subsection 5.3.

5.2. A meta-algorithm for parameter estimation. Despite the difficulties identified in the previous section, we can still obtain a viable parameter estimation approach. To avoid issues of nonidentifiability, we will have to assume some additional, partial prior knowledge on the model parameters. More concretely, suppose that we know of some additional constraints on $\boldsymbol{\Omega}$ (or alternatively $\boldsymbol{\mathcal{A}}$), of the form $h(\boldsymbol{\Omega}) = 0$, for some known function h. Such constraints could be, e.g., on the expected density ρ of the network, in which case $h(\boldsymbol{\Omega})$ would be of the form

$$h(\boldsymbol{\Omega}) = \rho - \frac{\boldsymbol{1}^\top \boldsymbol{\mathcal{A}} \boldsymbol{1}}{n^2} = 0.$$

As discussed earlier, based on our dynamical model we can obtain relations of the form $\boldsymbol{S}(T) = F(\boldsymbol{\Omega}, \boldsymbol{G})$ for the entries of the ensemble covariance matrix and the form $\lambda(\boldsymbol{S}(T)) = F_\lambda(\boldsymbol{\Omega})$ for the eigenvalues. Taking these equations and the constraints $h(\boldsymbol{\Omega})$ together, we obtain a system of (generally redundant) equations of the form

$$(5.3\text{a}) \qquad\qquad\qquad \boldsymbol{S}(T) = F(\boldsymbol{\Omega}, \boldsymbol{G}),$$
$$(5.3\text{b}) \qquad\qquad\qquad \lambda(\boldsymbol{S}(T)) = F_\lambda(\boldsymbol{\Omega}),$$
$$(5.3\text{c}) \qquad\qquad\qquad h(\boldsymbol{\Omega}) = 0.$$

Our approach is to "solve" this system of equations while replacing $\boldsymbol{S}(T)$ by its estimate $\widehat{\boldsymbol{\Sigma}}(T)$, as outlined in Algorithm 5.1. Observe that the sampling time T enters explicitly in the functional relations that we exploit, and therefore, unlike the partition recovery problem (cf. Algorithm 4.1), T must be known.

As discussed in the previous section, the word "solve" in the description of Algorithm 5.1 is to be understood as some not fully specified approach that yields an approximate solution of the system (5.3).

5.3. Theoretical guarantees for a planted partition model. In this subsection, we provide some theoretical results for a parameter estimation strategy in which we will ignore (5.3a) and concentrate on recovering the parameters solely based on the eigenvalues and the constraints. To this effect, we will leverage our concentration result in Theorem 4.1 to provide a bound between the estimated eigenvalues $\lambda_i(\widehat{\boldsymbol{\Sigma}}(T))$ and their ensemble counterparts $\lambda_i(\boldsymbol{S}(T))$. These bounds then translate immediately into (T-dependent) error bounds on the parameter recovery.

Algorithm 5.1. Model parameter estimation.

Input: Samples $\{\boldsymbol{x}_T^{(i)}\}_{i=1}^s$, number of groups k, time T
Output: Model parameters $\boldsymbol{\Omega}$
1: Compute the sample covariance $\widehat{\boldsymbol{\Sigma}}(T) := \frac{1}{s}\sum_{i=1}^s [\boldsymbol{x}_T^{(i)} - \bar{\boldsymbol{x}}_T][\boldsymbol{x}_T^{(i)} - \bar{\boldsymbol{x}}_T]^\top$
2: "Solve" the set of equations

$$\widehat{\boldsymbol{\Sigma}}(T) = F(\boldsymbol{\Omega}, \boldsymbol{G}),$$
$$\lambda(\widehat{\boldsymbol{\Sigma}}(T)) = F_\lambda(\boldsymbol{\Omega}),$$
$$h(\boldsymbol{\Omega}) = 0$$

to estimate model parameters $\boldsymbol{\Omega}$

For simplicity, we illustrate our approach by considering our example dynamics $f(\boldsymbol{\mathcal{A}}) = \boldsymbol{L}$ on a graph drawn from a two-group planted partition model (see Examples 5.1 and 5.2) subject to a density constraint on the network and the assumption that $a > b$. Similar results could be derived for other dynamics as well, using analogous arguments, but they would depend on the (possibly more complicated) parametrization of the model.

Recall from Example 5.1 that the second eigenvalue of $\boldsymbol{\mathcal{L}}^{2T}$ satisfies

$$(\lambda_2(\boldsymbol{\mathcal{L}}^{2T}))^{1/2T} = \frac{a-b}{a+(k-1)b}.$$

Given also prior knowledge of the density $\rho = \frac{a+(k-1)b}{nk}$, we can recover the parameter a for the case $k = 2$ as follows:

$$(5.4) \qquad a = (\lambda_2(\boldsymbol{\mathcal{L}}^{2T}))^{1/2T}\rho n + \rho n.$$

Then, to estimate a from samples, we can replace $\boldsymbol{\mathcal{L}}^{2T}$ in (5.4) by the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}(T)$. For this estimation strategy we can derive the following error bound.

Theorem 5.3. *Consider an SBM $\mathcal{M}(\boldsymbol{\mathcal{A}}) \in \mathcal{M}_n(\epsilon)$ with two equally sized blocks, and a parameter matrix of the form $\boldsymbol{\Omega} = \frac{(a-b)}{n}\boldsymbol{I}_2 + \frac{b}{n}\boldsymbol{1}_2\boldsymbol{1}_2^\top$, with $a > b$. Assume that $\rho = (a+b)/(2n)$ is given. Consider the estimator of a given by*

$$(5.5) \qquad \hat{a} = (\lambda_2(\widehat{\boldsymbol{\Sigma}}(T)))^{1/2T}\rho n + \rho n.$$

Then, using the notation of Theorem 4.1, the estimation error $\eta := |\hat{a} - a|/a$ satisfies the upper bound

$$(5.6) \qquad \eta \le \frac{(2TM + B(T))^{1/2T}}{\lambda_2(\boldsymbol{\mathcal{L}}) + 1},$$

with probability at least $1 - 5s^{-1} - 2\epsilon$.

The proof of Theorem 5.3 is given in subsection 6.4.

Note that the numerator in (5.6) indicates that the sampling time T plays a fundamental role in the estimation accuracy of $\hat{a}$. Moreover, this dependency is similar to the one encountered in Theorem 4.3 for partition recovery. In particular, there is a term that increases linearly with T, i.e., $2TM$, and there is another term that may decrease with T, i.e., $B(T)$ (cf. the discussion after Theorem 4.1). The denominator in (5.6) can also be explained intuitively. A larger $\lambda_2(\mathcal{L})$ is associated with a marked difference between a and b (cf. (5.2)), which translates into better delineated blocks within our SBM. In this scenario, estimating a should be easier, and this is captured in (5.6).

Remark 5.4. The above outlined eigenvalue-based parameter inference strategy is independent of the partition recovery problem and thus enables a clean theoretical analysis of the estimation error. Had we considered instead the general approach based on (5.3a), we would be solving simultaneously the partition recovery and the parameter estimation problems, and the estimation error would be affected by partition recovery errors. While a theoretical analysis of such an interplay may be feasible in principle, we do not pursue it further here but instead illustrate it with the help of numerical experiments in section 7.

6. Theoretical analysis. In the following, we provide the theoretical analysis of the spectral inference algorithms (Algorithms 4.1 and 5.1) discussed above. We first prove the concentration result in Theorem 4.1, which underlies the success of our spectral algorithms. The proof of this result follows from a series of lemmas that we establish in subsections 6.1.1 and 6.1.2. Afterwards, we show in subsection 6.3 how Theorem 4.1 implies the partition recovery performance guarantees stated in Theorem 4.3. We finally prove Theorem 5.3 in subsection 6.4.

6.1. Proof of Theorem 4.1: Concentration of the sample covariance matrix. In the next two subsections we prove Theorem 4.1 through a series of lemmas. The proof uses the triangle inequality

$$(6.1) \qquad \|\widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{S}(T)\| \le \|\widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{\Sigma}(T)\| + \|\boldsymbol{\Sigma}(T) - \boldsymbol{S}(T)\|$$

and separate bounds for the two terms on the right-hand side.

6.1.1. Bounding the first term in (6.1). We now establish a bound for the distance $\|\widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{\Sigma}(T)\|$ between the true and the sample covariances.

Lemma 6.1. *Under the conditions in Theorem 4.1, we have*

$$(6.2) \qquad \left\|\widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{\Sigma}(T)\right\| \le B(T),$$

with probability at least $1 - 5s^{-1} - \epsilon$, where $B(T)$ is as defined in Theorem 4.1.

For the proof of Lemma 6.1 we make use of the following result established in [7].

Lemma 6.2 (see [7, Theorem 2.1]). *Let $\boldsymbol{x} \in \mathbb{R}^n$ be a zero-mean sub-Gaussian vector satisfying Assumption 3.4, with covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(i)}$ be i.i.d. samples drawn from $\boldsymbol{x}$. Consider the sample covariance matrix $\widehat{\boldsymbol{\Sigma}} := \frac{1}{s} \sum_{i=1}^{s} [\boldsymbol{x}^{(i)} - \bar{\boldsymbol{x}}][\boldsymbol{x}^{(i)} - \bar{\boldsymbol{x}}]^{\top}$, where $\bar{\boldsymbol{x}} := \frac{1}{s} \sum_{i=1}^{s} \boldsymbol{x}^{(i)}$ is the sample mean. Then, with probability at least $1 - 5s^{-1}$ we have*

$$(6.3) \qquad \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_F \le c \cdot \|\boldsymbol{\Sigma}\| \cdot r_e(\boldsymbol{\Sigma}) \cdot \sqrt{\frac{\ln s}{s}},$$

where $r_e(\boldsymbol{\Sigma}) = \sum_i \lambda_i(\boldsymbol{\Sigma})/\|\Sigma\|$ is the effective rank of $\boldsymbol{\Sigma}$, c is a constant that depends on c_0 in Assumption 3.4, and $\|\cdot\|_F$ denotes the Frobenius norm.

Notice that we always have $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\| \le \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_F$ and, for the scenario considered here, $\|\boldsymbol{\Sigma}(T)\| = 1$, as by construction the dominant eigenvalue of $\boldsymbol{L}$ equals 1. Thus, Lemma 6.2 provides a useful tool to prove Lemma 6.1, provided that we can find an upper bound on the effective rank $r_e(\boldsymbol{\Sigma}(T))$.

Lemma 6.3 (bound on the effective rank). *Under the conditions stated in Theorem 4.1, the effective rank satisfies the bound*

$$(6.4) \qquad r_e(\boldsymbol{\Sigma}(T)) \le 1 + \sum_{i=2}^{k} \min\{1, |\mu_i| + M\}^{2T} + (n-k)M^{2T},$$

with probability at least $1 - \epsilon$.

Proof. Recall that we denote the eigenvalues of $\boldsymbol{L}$ by λ_i and the eigenvalues of $\boldsymbol{\mathcal{L}}$ by μ_i. Recall our convention that the eigenvalues have been sorted in descending order according to their magnitude, i.e., $|\lambda_1| \ge |\lambda_2| \ge \cdots \ge |\lambda_n|$ and $|\mu_1| \ge |\mu_2| \ge \cdots \ge |\mu_n|$, respectively. From the definition of the effective rank and the fact that $\|\boldsymbol{\Sigma}(T)\| = 1$, it follows that

$$r_e(\boldsymbol{\Sigma}(T)) = \sum_{i=1}^{n} \lambda_i(\boldsymbol{\Sigma}(T)) = \sum_{i=1}^{n} \lambda_i^{2T} = \sum_{i=1}^{n} |\lambda_i|^{2T}.$$

We can decompose this sum into three parts as follows:

$$r_e(\boldsymbol{\Sigma}(T)) = 1 + \sum_{i=2}^{k} |\lambda_i|^{2T} + \sum_{j=k+1}^{n} |\lambda_j|^{2T}.$$

In order to relate the (random) eigenvalues λ_i to quantities that can be expressed in terms of the model parameters, we make use of the triangle inequality $|\lambda_i| \le |\mu_i| + |\lambda_i - \mu_i|$. Weyl's Theorem (see, e.g., [20]) states that $|\lambda_i - \mu_i| \le \|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\|$ for all i. Furthermore, Lemma 2.3 implies that $\|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\| \le M$, with probability at least $1 - \epsilon$. Finally, we also know that $|\lambda_i| \le 1$ for all i, since $\|\boldsymbol{L}\| = 1$ (more generally, we assumed $\|f(\boldsymbol{A})\| \le 1$). We can thus deduce that $|\lambda_i| \le \min\{1, |\mu_i| + M\}$ for all i, with probability at least $1 - \epsilon$. Using, in addition, the facts that $\mu_i = 0$ for $i > k$ and $M < 1$ if $\mathcal{M}(\boldsymbol{\mathcal{A}}) \in \mathcal{M}_n(\epsilon)$, we conclude that

$$r_e(\boldsymbol{\Sigma}(T)) \le 1 + \sum_{i=1}^{k} \min\{1, |\mu_i| + M\}^{2T} + \sum_{j=k+1}^{n} \min\{1, |\mu_i| + M\}^{2T}$$

$$\le 1 + \sum_{i=2}^{k} \min\{1, |\mu_i| + M\}^{2T} + (n-k)M^{2T},$$

with probability at least $1 - \epsilon$. ∎

Proof of Lemma 6.1. By combining Lemmas 6.2 and 6.3, Lemma 6.1 follows. ∎

6.1.2. Bounding the second term in (6.1). We derive a bound for the difference $\|\mathbf{\Sigma}(T) - \mathbf{S}(T)\|$ between the true and ensemble covariances. We start with the following lemma.

Lemma 6.4. *For any integer $t > 0$, we have*

$$\|\boldsymbol{L}^t - \boldsymbol{\mathcal{L}}^t\| \leq t\|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\|. \tag{6.5}$$

Proof. We will show that for any integer $t > 0$,

$$\|\boldsymbol{L}^t - \boldsymbol{\mathcal{L}}^t\| \leq \|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\| + \|\boldsymbol{L}^{t-1} - \boldsymbol{\mathcal{L}}^{t-1}\|. \tag{6.6}$$

Equation (6.5) then follows directly from repeated application of (6.6). To show (6.6), notice that

$$
\begin{aligned}
\|\boldsymbol{L}^t - \boldsymbol{\mathcal{L}}^t\| &= \|(\boldsymbol{L} - \boldsymbol{\mathcal{L}})\,\boldsymbol{L}^{t-1} + \boldsymbol{\mathcal{L}}\left(\boldsymbol{L}^{t-1} - \boldsymbol{\mathcal{L}}^{t-1}\right)\| \\
&\leq \|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\|\,\|\boldsymbol{L}^{t-1}\| + \|\boldsymbol{\mathcal{L}}\|\,\|\boldsymbol{L}^{t-1} - \boldsymbol{\mathcal{L}}^{t-1}\| = \|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\| + \|\boldsymbol{L}^{t-1} - \boldsymbol{\mathcal{L}}^{t-1}\|,
\end{aligned}
$$

where the last equality follows from the fact that $\|\boldsymbol{\mathcal{L}}\| = \|\boldsymbol{L}\| = 1$. ∎

Concluding the proof of Theorem 4.1. Recall that $\boldsymbol{S}(T) = \boldsymbol{\mathcal{L}}^{2T}$ and $\boldsymbol{\Sigma}(T) = \boldsymbol{L}^{2T}$, and that $\|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\| \leq M$, with probability at least $1 - \epsilon$ (Lemma 2.3). Lemma 6.4 then implies that

$$\|\boldsymbol{\Sigma}(T) - \boldsymbol{S}(T)\| \leq 2T\|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\| \leq 2TM, \tag{6.7}$$

with probability at least $1 - \epsilon$. We bound the two terms on the right-hand side of (6.1) using (6.2) and (6.7), respectively, and hence establish the validity of Theorem 4.1. ∎

6.2. Proof of Theorem 4.2.

Proof. The proof is analogous to the proof of Theorem 4.1. Instead of using a bound for $\|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\|$, we develop a bound for $\|f(\boldsymbol{A}) - f(\boldsymbol{\mathcal{A}})\|$ and obtain that, with probability at least $1 - 5s^{-1} - \epsilon$,

$$\|f(\boldsymbol{A}) - f(\boldsymbol{\mathcal{A}})\| = \left\|\sum_{\ell=1}^{R} \alpha_\ell(\boldsymbol{L}^\ell - \boldsymbol{\mathcal{L}}^\ell)\right\| \leq \sum_{\ell=1}^{R} \alpha_\ell \left\|\boldsymbol{L}^\ell - \boldsymbol{\mathcal{L}}^\ell\right\| \leq \sum_{\ell=1}^{R} \ell\alpha_\ell \left\|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\right\| = M_f, \tag{6.8}$$

where the last inequality follows from Lemma 6.4. The rest of the argument remains the same but with M_f replacing M. ∎

6.3. Implications for partition recovery. In the following, we explain how the above concentration results translate into guarantees for the partition recovery problem. Subsection 6.3.1 provides intermediate results on the spectral geometry of the problem and discusses the precise definition of misclassified nodes. The proof of Theorem 4.3 is then given in subsection 6.3.2.

6.3.1. Partition recovery problem. In order to characterize recovery performance, we first clarify what we mean by a node being clustered correctly. To do this we follow the literature on SBMs [41, 42]. Recall that within our partition recovery algorithm, we perform a k-means clustering procedure on the rows of the matrix $\widetilde{\boldsymbol{W}}$ of (row-normalized) dominant eigenvectors

associated with the estimated covariance $\widehat{\boldsymbol{\Sigma}}(T)$. This procedure assigns each row i to a centroid vector $\boldsymbol{k}_i$ corresponding to a cluster.

Let us denote the "true" centroid vector that a node i is assigned to according to the ensemble covariance $\boldsymbol{S}(T)$ by $\boldsymbol{\kappa}_i$, where $i = 1, \ldots, n$. From the definition of $\boldsymbol{S}(T)$ it follows that each $\boldsymbol{\kappa}_i^\top$ corresponds to a row of the matrix $\widetilde{\boldsymbol{V}}$ of the (row-normalized) top k eigenvectors of $\mathcal{L}$. Moreover, we know from Lemma 2.1 that there are exactly k different population centroids and that the association of rows to centroids corresponds to the true group structure of the underlying SBM. However, if some of the eigenvalues of $\boldsymbol{S}(T)$ are repeated, the eigenvector matrix $\widetilde{\boldsymbol{V}}$ will not be uniquely defined, even though the subspace spanned by these vectors will be unique. As we are interested in the distance between the vectors $\widetilde{\boldsymbol{V}}$ and their approximation via $\widetilde{\boldsymbol{W}}$, our definition of misclassification has to include an orthogonal matrix $\boldsymbol{Q} \in \mathbb{R}^{k \times k}$ that minimizes the orthogonal Procrustes distance $\|\widetilde{\boldsymbol{W}} - \widetilde{\boldsymbol{V}}\boldsymbol{Q}\|_F$ (see also [41, 42]). In particular, this minimality property of $\boldsymbol{Q}$ will be essential for Lemma 6.8.

Definition 6.5 (set of misclassified nodes [41]). *The set of misclassified nodes* $\mathbb{M}$ *is defined as*

$$(6.9) \qquad \mathbb{M} := \{i : \|\boldsymbol{k}_i^\top - \boldsymbol{\kappa}_i^\top \boldsymbol{Q}\| > \|\boldsymbol{k}_i^\top - \boldsymbol{\kappa}_j^\top \boldsymbol{Q}\| \text{ for some } j \neq i\}.$$

In words, this means that we call a node i misclassified if the centroid $\boldsymbol{k}_i$ to which it has been assigned is closer to some other population centroid $\boldsymbol{\kappa}_j$ than it is to its true population centroid $\boldsymbol{\kappa}_i$. Note that this definition of misclassified nodes adopted from the literature on SBMs [41, 42] implicitly assumes that (i) we can solve the k-means clustering problem exactly, and (ii) the empirical centroids based on $\widetilde{\boldsymbol{W}}$ will be centered around the (appropriately rotated) population centroids. We refer the reader to the literature for additional discussion on the consistency of spectral clustering under the SBM [25] and limit theorems for the eigenvectors of random graphs (see, e.g., [2] for a survey on random dot product graphs).

By accounting for the special geometry of the eigenvectors $\widetilde{\boldsymbol{V}}$, we can establish the following necessary condition for a node to be misclassified.

Lemma 6.6 (see [41]). *If a node i is misclassified, then* $\|\boldsymbol{k}_i^\top - \boldsymbol{\kappa}_i^\top \boldsymbol{Q}\| \geq 1/\sqrt{2}$.

Proof. First, we know from Lemma 2.1 that $\widetilde{\boldsymbol{V}} = \boldsymbol{G}\boldsymbol{U}$ for some orthogonal matrix $\boldsymbol{U}$. Hence, all population centroids $\boldsymbol{\kappa}_i$ are of unit length and orthogonal to each other. In particular, for all j and i belonging to different blocks/groups, we have

$$(6.10) \qquad \|\boldsymbol{\kappa}_i^\top - \boldsymbol{\kappa}_j^\top\| = \|(\boldsymbol{g}_i^\top - \boldsymbol{g}_j^\top)\boldsymbol{U}\| = \sqrt{2},$$

where $\boldsymbol{g}_i$ is the true population cluster indicator vector corresponding to node i (see Lemma 2.1).

Second, by the triangle inequality and the orthogonality of $\boldsymbol{Q}$, we know that for all j and i in different blocks, we have

$$(6.11) \qquad \|\boldsymbol{\kappa}_i^\top - \boldsymbol{\kappa}_j^\top\| \leq \|\boldsymbol{\kappa}_i^\top \boldsymbol{Q} - \boldsymbol{k}_i^\top\| + \|\boldsymbol{k}_i^\top - \boldsymbol{\kappa}_j^\top \boldsymbol{Q}\|.$$

Plugging in the result from (6.10) into (6.11), we obtain that $\|\boldsymbol{k}_i^\top - \boldsymbol{\kappa}_j^\top \boldsymbol{Q}\| \geq \sqrt{2} - \|\boldsymbol{k}_i^\top - \boldsymbol{\kappa}_i^\top \boldsymbol{Q}\|$. From Definition 6.5, for a misclassified node i there exists some j such that

$$\|\boldsymbol{k}_i^\top - \boldsymbol{\kappa}_i^\top \boldsymbol{Q}\| > \|\boldsymbol{k}_i^\top - \boldsymbol{\kappa}_j^\top \boldsymbol{Q}\| \geq \sqrt{2} - \|\boldsymbol{k}_i^\top - \boldsymbol{\kappa}_i \boldsymbol{Q}\|,$$

where the last inequality is true by our argument above. The result follows after some simple algebraic manipulations. ∎

To arrive at our result for the misclassification rate in Theorem 4.3, we further follow closely the literature on SBMs [1, 41, 42] and establish the following two lemmas (Lemmas 6.7 and 6.8). The first lemma provides us with a relationship between the difference of the cluster centroids obtained from the k-means clustering and the true population cluster on the one side, and the difference between the eigenvectors of $\widehat{\boldsymbol{\Sigma}}(T)$ and $\boldsymbol{S}(T)$ on the other side.

Lemma 6.7. *Denote the matrix of cluster centroids obtained by performing k-means on the rows of $\widetilde{\boldsymbol{W}}$ as $\boldsymbol{K} := [\boldsymbol{k}_1, \ldots, \boldsymbol{k}_n]^\top \in \mathbb{R}^{n \times k}$. Then,*

$$(6.12) \qquad \|\boldsymbol{K} - \widetilde{\boldsymbol{V}}\boldsymbol{\mathcal{Q}}\|_F \leq 2\|\widetilde{\boldsymbol{W}} - \widetilde{\boldsymbol{V}}\boldsymbol{\mathcal{Q}}\|_F.$$

Proof. By the triangle inequality,

$$\|\boldsymbol{K} - \widetilde{\boldsymbol{V}}\boldsymbol{\mathcal{Q}}\|_F \leq \|\boldsymbol{K} - \widetilde{\boldsymbol{W}}\|_F + \|\widetilde{\boldsymbol{W}} - \widetilde{\boldsymbol{V}}\boldsymbol{\mathcal{Q}}\|_F.$$

From the definition of k-means, however, we know that $\|\widetilde{\boldsymbol{W}} - \boldsymbol{K}\|_F \leq \|\widetilde{\boldsymbol{W}} - \widetilde{\boldsymbol{V}}\boldsymbol{\mathcal{Q}}\|_F$, as $\boldsymbol{K}$ is chosen in order to minimize the distances of the centroids to the individual rows. Hence, the claim follows. ∎

The second lemma builds on the Davis–Kahan theorem [11, 64] and bounds the difference between the row-normalized eigenvectors of $\widehat{\boldsymbol{\Sigma}}(T)$ and $\boldsymbol{S}(T)$.

Lemma 6.8 (Davis–Kahan). *Following the definitions in Theorem 4.3, we have*

$$(6.13) \qquad \|\widetilde{\boldsymbol{W}} - \widetilde{\boldsymbol{V}}\boldsymbol{\mathcal{Q}}\|_F \leq \frac{\sqrt{8k}\|\widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{S}(T)\|}{\tau \xi_k},$$

where we recall (see Theorem 4.3) that ξ_k is the kth top eigenvalue of the ensemble covariance matrix, and τ is be the minimum 2-norm of any row of $\boldsymbol{W}$ and $\boldsymbol{V}$.

Proof. The fact that

$$\|\boldsymbol{W} - \boldsymbol{V}\boldsymbol{\mathcal{Q}}\|_F \leq \frac{\sqrt{8k}\|\widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{S}(T)\|}{\xi_k}$$

holds for the unnormalized versions of $\boldsymbol{W}$ and $\boldsymbol{V}$ is a restatement of the Davis–Kahan theorem as can be found in, e.g., [64, Theorem 2].

By definition, we have $\widetilde{\boldsymbol{V}} = \text{diag}(\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}\boldsymbol{V}$ and $\widetilde{\boldsymbol{W}} = \text{diag}(\boldsymbol{W}\boldsymbol{W}^\top)^{-1/2}\boldsymbol{W}$. Let us define $\tau_i := \min\{[\boldsymbol{V}\boldsymbol{V}^\top]_{ii}^{1/2}, [\boldsymbol{W}\boldsymbol{W}^\top]_{ii}^{1/2}\}$, which is the smaller of the 2-norms of the ith rows of $\boldsymbol{V}$ and $\boldsymbol{W}$, respectively. Based on this definition of τ_i we see that $\tau = \min_i \tau_i$. Now, as row-normalization corresponds to diagonal scaling, we obtain

$$\|\widetilde{\boldsymbol{W}} - \widetilde{\boldsymbol{V}}\boldsymbol{\mathcal{Q}}\|_F \leq \frac{1}{\tau}\|\boldsymbol{W} - \boldsymbol{V}\boldsymbol{\mathcal{Q}}\|_F \leq \frac{\sqrt{8k}\|\widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{S}(T)\|}{\tau \xi_k}. \qquad ∎$$

6.3.2. Proof of Theorem 4.3. Based on the results established thus far, we can provide a proof of the bound of the misclassification rate $q := |\mathbb{M}|/n$ as stated in Theorem 4.3. Our proof parallels results for the misclassification rate if the graph is observed [41].

Proof. We define the set $\mathbb{M}_2 := \{i : \|\boldsymbol{k}_i^\top - \boldsymbol{\kappa}_i^\top \boldsymbol{\mathcal{Q}}\| \geq 1/\sqrt{2}\} \supset \mathbb{M}$, which includes all misclassified nodes according to Lemma 6.6. From Lemmas 6.7 and 6.8 and Theorem 4.1, it follows that

$$q = \frac{|\mathbb{M}|}{n} \leq \frac{|\mathbb{M}_2|}{n} = \frac{1}{n} \sum_{i \in \mathbb{M}_2} 1 \leq \frac{2}{n} \sum_{i \in \mathbb{M}_2} \|\boldsymbol{k}_i^\top - \boldsymbol{\kappa}_i^\top \boldsymbol{\mathcal{Q}}\|^2 = \frac{2}{n} \|\boldsymbol{K} - \widetilde{\boldsymbol{V}} \boldsymbol{\mathcal{Q}}\|_F^2$$

$$\leq \frac{8}{n} \|\widetilde{\boldsymbol{W}} - \widetilde{\boldsymbol{V}} \boldsymbol{\mathcal{Q}}\|_F^2 \leq \frac{8}{n} \left(\frac{\sqrt{8k} \|\widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{S}(T)\|}{\tau \xi_k} \right)^2. \qquad \blacksquare$$

6.4. Implications for parameter estimation. Here we show how our concentration results allow us to prove Theorem 5.3.

Proof. Note that we can write the relative estimation error as

$$\frac{|\hat{a} - a|}{a} = \frac{|\lambda_2(\widehat{\boldsymbol{\Sigma}}(T))^{1/2T} \rho n - \lambda_2(\boldsymbol{S}(T))^{1/2T} \rho n|}{\lambda_2(\boldsymbol{S}(T))^{1/2T} \rho n + \rho n} = \frac{|\lambda_2(\widehat{\boldsymbol{\Sigma}}(T))^{1/2T} - \lambda_2(\boldsymbol{S}(T))^{1/2T}|}{\lambda_2(\boldsymbol{S}(T))^{1/2T} + 1}$$

$$\leq \frac{|\lambda_2(\widehat{\boldsymbol{\Sigma}}(T)) - \lambda_2(\boldsymbol{S}(T))|^{1/2T}}{\lambda_2(\boldsymbol{\mathcal{L}}) + 1} \leq \frac{\|\widehat{\boldsymbol{\Sigma}}(T) - \boldsymbol{S}(T)\|^{1/2T}}{\lambda_2(\boldsymbol{\mathcal{L}}) + 1}.$$

The first inequality follows from Jensen's inequality since the function $f(x) = x^{1/2T}$ is concave. The second inequality is a direct application of Weyl's inequality. Hence, using Theorem 4.1, it follows that with probability at least $1 - 5s^{-1} - 2\epsilon$,

$$\eta \leq \frac{(2TM + B(T))^{1/2T}}{\lambda_2(\boldsymbol{\mathcal{L}}) + 1}. \qquad \blacksquare$$

7. Numerical experiments. In this section we present some computational experiments with our algorithms for partition recovery and parameter estimation to demonstrate their performance in practice.

7.1. Partition recovery. To gain intuition about the qualitative features of the partition recovery problem, we first consider the following example SBM consisting of five groups with varying sizes $\boldsymbol{n}_g = [200, 300, 200, 300, 280]^\top$, and affinity matrix

$$\boldsymbol{\Omega} = \begin{bmatrix} 0.5 & 0.3 & 0.2 & 0.1 & 0.2 \\ 0.3 & 0.5 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.2 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.1 & 0.2 & 0.4 \end{bmatrix}.$$

The partition recovery performance is shown in Figure 1A for a varying number of samples s, as a function of the observation time T. Here, the recovery performance is measured via the
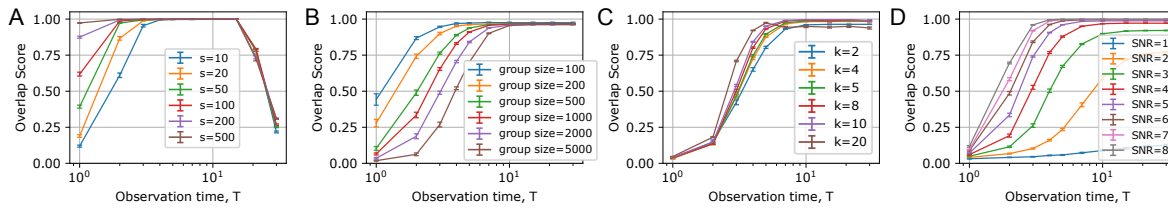
Figure 1. *Partition recovery performance for SBM inference.* **A–D.** *The recovery performance of* Algorithm 4.1 *is plotted against the observation time at which the samples are drawn. Each point in the plot corresponds to an average over* 30 *experiments, and the error bars denote the standard error of the mean. Within* Algorithm 4.1 *we run k-means* 10 *times and pick the best result in terms of the k-means objective function as our final partition.* **A.** *Recovery performance for different numbers of samples, for the SBM discussed in the text.* **B.** *Recovery performance in a two-group planted partition model (see* Example 5.1*) for different group sizes (SNR = 5, s = 50, $\rho = 30/n$).* **C.** *Recovery performance in a k-group planted partition model (see* Example 5.1*) for different numbers of groups (n = 5000, SNR = 5, s = 50, $\rho = 30/n$).* **D.** *Recovery performance in a k = 5 group planted partition model (see* Example 5.1*) for different SNRs (n = 2000, s = 50, $\rho = 30/n$).*

overlap score Z defined as

$$(7.1) \qquad\qquad Z = \frac{z_{\text{actual}} - z_{\text{chance}}}{1 - z_{\text{chance}}},$$

where z_{chance} is the probability of guessing the group assignments correctly and z_{actual} is the fraction of nodes that are correctly classified. The guessing probability is computed as $z_{\text{chance}} = \max(\boldsymbol{n}_g)/n$, where $\max(\boldsymbol{n}_g)$ is the size of the largest group. Note that z_{chance} is simply $\frac{1}{\#\text{groups}}$ for groups of equal size. The actual overlap z_{actual} is computed by solving a linear $k \times k$ optimal assignment problem between the inferred group labels and the planted partition.

Overall, we can infer the partitions remarkably well using our simple spectral inference strategy, even though we never observe the actual network. The results in Figure 1A display a couple of noteworthy features consistent with our technical results, which are worth examining. Specifically, we see that there is a pronounced effect of the observation time: Both for small and very large T, the performance is markedly worse than for intermediate times. Thus, if we have a fixed budget for the number of samples, choosing the "right" observation time can improve recovery performance. Obtaining more samples, in contrast, has a more moderate effect and is most beneficial for small times.

At an intuitive level, these results can be explained by viewing the network's effects as a linear filter. On the one hand, for short observation times, the original white noise input is only filtered weakly and thus still contains a large noise component, which is detrimental to recovery performance. On the other hand, as the output approaches its stationary state, the information that the signal carries about the network structure is completely washed out. For intermediate times, however, there is a trade-off between attenuating the noise within the observed signal and not losing too much information about the partition structure. Theorems 4.1 and 4.3 quantify precisely how this trade-off is governed by the spectral features of the network.

Using a planted partition model with an affinity matrix of the form $\boldsymbol{\Omega} = \frac{(a-b)}{n}\boldsymbol{I}_k + \frac{b}{n}\boldsymbol{1}_k\boldsymbol{1}_k^\top$ (see Example 5.1), we further investigate (see Figure 1B–D) numerically how the partition recovery performance is affected by the group sizes, the number of groups, and the signal-to-

noise ratio (SNR) defined by [1],

$$(7.2) \qquad\qquad \text{SNR} = \frac{(a-b)^2}{ka + k(k-1)b}.$$

We note that that $\text{SNR} = 1$ corresponds to the well-known detectability limit in the sparse regime of the SBM, for asymptotically large networks, with $n \to \infty$ (see [1] for a detailed review).

Figure 1B shows that for a fixed number of groups (here $k = 2$), the group sizes have a noticeable effect on recovery performance. On the other hand, if we fix the number of nodes, changing the number of groups does not have a noticeable effect. This hints at the fact that the number of nodes n (and thus the number of nodes per group) is one of the key drivers of the difficulty of partition recovery. Finally, in Figure 1D we plot the influence of the SNR on the detection performance. As expected, a smaller SNR results in a dramatically worse performance. Remarkably, for $\text{SNR} = 4$, using only $s = 50$ samples leads to an almost perfect partition recovery for a suitable observation time.

7.2. Model parameter estimation. We now demonstrate numerically that we can infer the model parameters of the SBM using only the observed outputs. For simplicity, we focus on the case of the planted partition model with $k = 2$ groups of equal size and assume that the average expected network density ρ or, equivalently, the average expected mean degree $d_{\text{av}} = \rho n$, is known a priori (see also the discussion in the context of Example 5.1 and subsection 5.3).

Since the estimation error is not directly related to the partition recovery performance, let us first focus on the eigenvalue-based parameter estimation strategy, using the estimator $\hat{a}$ for the parameter a given by (5.5) (see subsection 5.3).

Figure 2 shows the results of applying this parameter estimation scheme in numerical experiments. To quantify the parameter recovery performance we compute the relative recovery error $\eta = |a - \hat{a}|/a$. Similar to partition recovery, we see in Figure 2A that the observation time has a strong effect on the recovery performance, and we also see a marked decrease in the error from more than 10% initially to around 3% for longer observation times.

Figure 2B–C demonstrates the effect of the group size and the SNR on recovery performance. Similar to the partition recovery problem, we observe that smaller group sizes (a smaller network) and a larger SNR lead to better performance.

Let us now consider a different parameter estimation strategy based on first recovering the groups in the SBM and then trying to recover the parameters based on the entries of the sample covariance matrix (see discussion in previous sections). We concentrate again on the planted partition model with two equally sized groups and a given density $\rho = (a + b)/(2n)$ as above.

Specifically, we estimate the parameter using the following approach. First, we estimate the group labels g_i of the nodes using Algorithm 4.1 and sort the indices of the sample covariance $\widehat{\boldsymbol{\Sigma}}(T)$ accordingly. Next, we take the tth root of $\widehat{\boldsymbol{\Sigma}}(T)$ (via an SVD), and then compute the average magnitude of all off-diagonal terms that relate different groups, to obtain the following statistic:

$$(7.3) \qquad\qquad x = \frac{1}{|\{(i,j) : g_i \neq g_j\}|} \sum_{(i,j):g_i \neq g_j} [\widehat{\boldsymbol{\Sigma}}(T)^{1/T}]_{ij}.$$
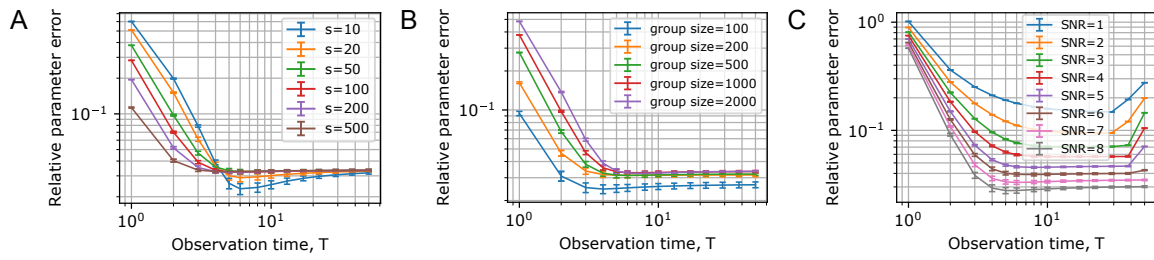
Figure 2. *Parameter recovery performance for SBM inference with an eigenvalue-based estimation strategy.* **A–C.** *The recovery performance of Algorithm 5.1 in terms of the relative error $\eta = |a - \hat{a}|/a$ is plotted against the observation time at which the samples are drawn. Each point corresponds to an average over 30 experiments, and the error bars denote the standard error of the mean.* **A.** *Parameter recovery performance for varying number of samples for the SBM with $k = 2$ groups as discussed in the text ($n = 2000, SNR = 7, \rho = 30/n$).* **B.** *Parameter recovery performance in a $k = 2$ group planted partition model (see Example 5.1) for varying group sizes ($SNR = 7, s = 50, \rho = 30/n$).* **C.** *Parameter recovery performance in a $k = 2$ group planted partition model (see Example 5.1) for different SNR ($n = 2000, s = 50, \rho = 30/n$).*

From (5.1) we know that

$$(7.4) \qquad [\widehat{\boldsymbol{\Sigma}}(T)^{1/T}]_{ij} \approx \begin{cases} \frac{1}{n}\left(\frac{a-b}{a+b}\right)^2 + \frac{1}{n} & \text{if } g_i = g_j, \\ -\frac{1}{n}\left(\frac{a-b}{a+b}\right)^2 + \frac{1}{n} & \text{if } g_i \neq g_j, \end{cases}$$

which implies that $x \approx -\frac{1}{n}\left(\frac{a-b}{a+b}\right)^2 + \frac{1}{n}$. We then use the following estimator for the parameter a:

$$(7.5) \qquad \hat{a}_{\mathrm{part}} = \rho n \sqrt{1 - nx} + \rho n.$$

Figure 3 displays the performance of this estimator in terms of the error $\eta_{\mathrm{part}} = |\hat{a}_{\mathrm{part}} - a|/a$, in comparison to the relative error $\eta_\lambda = |\hat{a} - a|/a$ of the eigenvalue-based estimation strategy.

7.3. Real-world networks. In this section, we examine the performance of our methods on real-world networks that are not known to be generated from an SBM. In this final set of numerical experiments, we test two real-world networks, on which we assume a diffusion dynamics of the form (3.2) takes place.

The first network we study is the well-known karate club network, which was initially charted by Zachary [66], describing the social interactions among 34 individuals of a karate club that eventually split into two factions. We use this split as our "ground-truth" partition to compare against. Although there is a priori no reason why the split should be correlated with the graph structure [38], this finding is now well supported in the literature.

The results of applying our procedure for a range of different sampling times and number of samples available are shown in Figure 4A. Interestingly, we find that already with three samples, we can achieve a recovery performance that is as good as applying a corresponding spectral clustering procedure to the full network—which in this case corresponds to making one classification error with respect to the reported fission of the club.
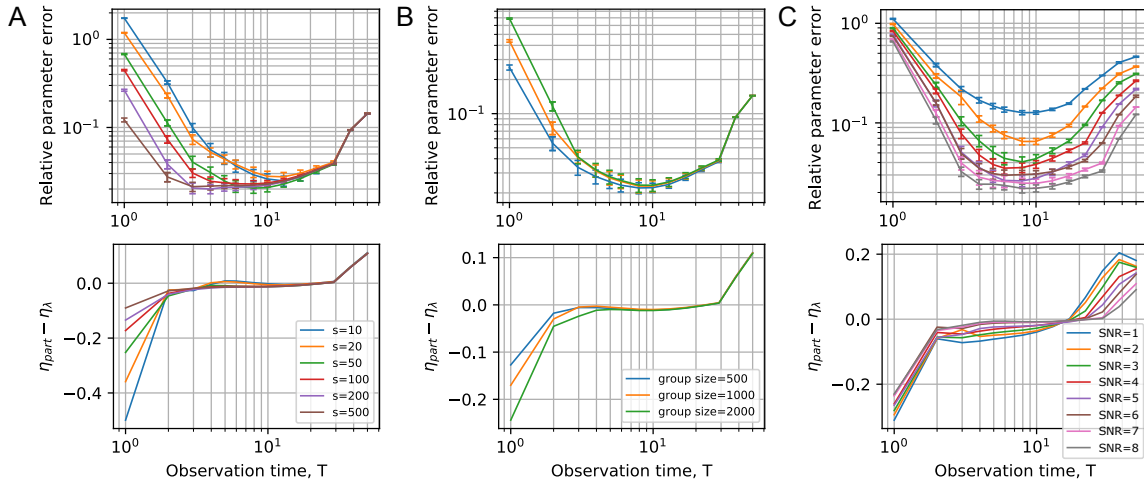
Figure 3. *Comparison of parameter recovery performance with different parameter estimation strategies.* **A**–**C.** *The recovery performance of* Algorithm 5.1 *in terms of the relative error* $\eta = |a - \hat{a}|/a$ *is plotted against the observation time at which the samples are drawn. Top row: Relative parameter estimation error* η_{part} *of the estimation procedure based on first inferring a partition of the underlying SBM. Each point corresponds to an average over* 30 *experiments, and the error bars denote the standard error of the mean. Bottom row: Difference between relative parameter estimation error* η_{part} *of the estimation procedure based on first inferring a partition from the data and the relative error* η_λ *of the eigenvalue-based parameter inference strategy (see text).* **A.** *Parameter recovery performance for varying number of samples for the SBM with* $k = 2$ *groups as discussed in the text* ($n = 4000, SNR = 7, \rho = 30/n$). **B.** *Parameter recovery performance in a* $k = 2$ *group planted partition model (see* Example 5.1*) for varying group sizes* ($SNR = 7, s = 50, \rho = 30/n$). **C.** *Parameter recovery performance in a* $k = 2$ *group planted partition model (see* Example 5.1*) for different SNR* ($n = 4000, s = 50, \rho = 30/n$).
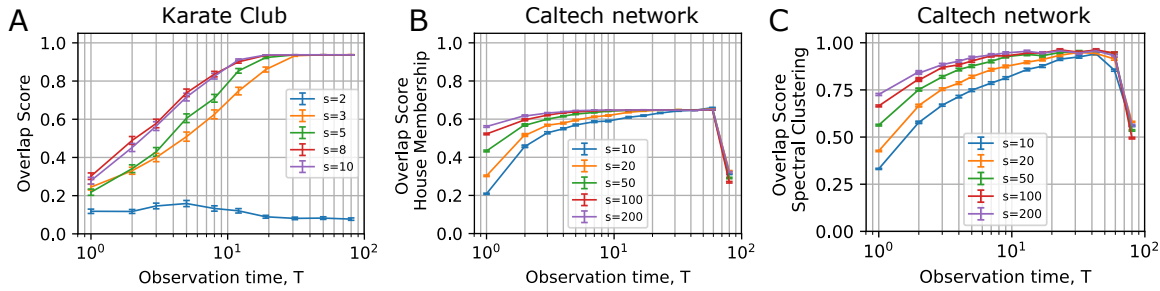
Figure 4. *Recovering partition structure from dynamics on real-world networks.* *Each point corresponds to an average over* 200 *experiments. The error bars denote the standard error of the mean.* **A.** *Partition recovery performance of our method on the karate club network with respect to the partition induced by the final split of the karate club, as reported by Zachary* [66] *(see text).* **B.** *Partition recovery performance of our method on the social network of Caltech students, with respect to the partition induced by the house membership of the students, as reported by Traud et al.* [55, 56]. **C.** *Partition recovery performance of our method on the social network of Caltech students, with respect to the partition obtained from spectral clustering on the fully observed network.*

As a second example, we study a social interaction network consisting of the friendship relationships between students at Caltech, a dataset that was originally studied by Traud et al. [55, 56] in a comparative study of a larger set of such networks. While there is again no clear "ground-truth" partition of the network, some additional node covariates, such as gender or the major of the students, are examined as part of their study [55, 56]. Traud et al. [55, 56] observed that in the context of the social network of Caltech students the node covariate "house membership" (dormitories) correlates strongly with a putative community structure in the network, even though the house membership cannot explain the structure of the network completely.

Motivated by these findings, we study the partition recovery performance with respect to the partition of the nodes induced by the house-membership data in Figure 4B–C. Here we limit ourselves to studying the largest connected component of the network, where we remove all of the nodes for which no house-membership label is available. This results in a network of 594 nodes split among eight houses (groups). As the structure of the network is not completely explained by the house-membership data, we recover the induced partition only with an overlap score of 0.65 (see Figure 4B). We also show our recovery performance relative to the partition we would obtain from the spectral clustering procedure that uses $\boldsymbol{L}$ (i.e., knowledge of the complete network) instead of $\widehat{\boldsymbol{\Sigma}}(T)$ for clustering. As can be seen in Figure 4C, our recovered partition is almost perfectly correlated with the result of the clustering-based procedure, highlighting that our method indeed performs as expected and that the low score with respect to the house-membership partition is largely due to a lack of correlation between the house-membership variable and the network block structure.

8. Discussion. Graph-based tools have become prevalent for the analysis of a range of different systems across the sciences. Network inference, the problem of determining the topology of a network based on a set of nodal observations, has accordingly become a paramount tool for enabling this type of analysis. In this work, we have advocated a fresh look into this problem for applications in which exact inference is impossible, e.g., due to sampling constraints. Instead of inferring the exact network in terms of its adjacency matrix, we infer a generative model of the network structure, thereby learning a probability distribution over networks. Importantly, for a low-dimensional statistical model such as the SBM discussed here, this procedure can be implemented using far fewer samples than would be required to infer the exact network structure.

The proposed method is especially relevant for scenarios in which we aim to analyze the network further from a statistical perspective, e.g., if our ultimate goal is to perform community detection on the network. An added benefit of having fitted a generative model is that we can employ the model for further tasks, e.g., to generate surrogate network data.

Most of our algorithmic ideas can be extended to more general functions f which inherit the low-rank structure of the (expected) SBM, provided an appropriate concentration result for the system matrix under consideration can be derived. For instance, this includes not only those functions describing a low-pass graph filter, as considered in [45, 61], but also functions of the type $f(\boldsymbol{A}) = \sum_\ell \alpha_\ell \boldsymbol{A}^\ell$ for some suitably chosen parameters α_ℓ.

To adjust the proposed algorithms to such setups, we would have to establish and prove appropriate relationships between the spectral properties of the system matrix on one side

and the partition structure and parameters of the model on the other. Specifically, step 3 of Algorithm 4.1 uses the particular structure of the eigenvectors of $\boldsymbol{L}$ (see Lemma 2.1). Of course, the set of equations relating the model parameters to the statistics of the observed outputs will in general be different and depend on the specific model chosen. In fact, some of the nonidentifiability problems we encountered here in terms of recovering the model parameters may not arise for other classes of system matrices $f(\boldsymbol{A})$.

Our work also opens up several additional avenues for future research. This includes the analysis of the inference of more realistic network models based on a setup as discussed here, ranging from degree-corrected SBMs [23] all the way to graphons [3, 28]. Likewise, it will be interesting to extend the above results to more general dynamical processes, including nonlinear dynamics. Another appealing line of future work would be to consider scenarios in which only a part of the network is observed [32] or additional (control) inputs are possible.

Appendix A. Omitted proofs.

A.1. Proof of Lemma 2.1. In order to prove Lemma 2.1, we provide a slightly more general discussion about the relationship between the eigenvectors of $\boldsymbol{\mathcal{L}}$ and the partition indicator matrix $\boldsymbol{G}$. Lemma 2.1 then follows directly from Corollary A.4 below.

Definition A.1 (normalized affinity matrix). *Consider an SBM with affinity matrix $\boldsymbol{\Omega}$ and partition indicator matrix $\boldsymbol{G}$. Let us denote the vector of group sizes by $\boldsymbol{n}_g = \boldsymbol{G}^\top \boldsymbol{1}$ and define $\boldsymbol{N}_g := \mathrm{diag}(\boldsymbol{n}_g) = \boldsymbol{G}^\top \boldsymbol{G}$. We then define the normalized affinity matrix as*

$$\boldsymbol{\Gamma} = \boldsymbol{N}_g^{1/2} \, \mathrm{diag}(\boldsymbol{\Omega} \boldsymbol{n}_g)^{-1/2} \boldsymbol{\Omega} \, \mathrm{diag}(\boldsymbol{\Omega} \boldsymbol{n}_g)^{-1/2} \boldsymbol{N}_g^{1/2}.$$

Note that $\boldsymbol{\mathcal{L}}$ can be written as

$$(\text{A.1}) \qquad \boldsymbol{\mathcal{L}} = \boldsymbol{G} \boldsymbol{N}_g^{-1/2} \boldsymbol{\Gamma} \boldsymbol{N}_g^{-1/2} \boldsymbol{G}^\top.$$

Lemma A.2. *Let $\boldsymbol{\mathcal{L}}$ be the expected normalized adjacency matrix of an SBM with a corresponding normalized affinity matrix $\boldsymbol{\Gamma}$ with eigenvalues λ_i and corresponding eigenvectors $\boldsymbol{u}_i$. Then, any eigenvector $\boldsymbol{u}_i$ of $\boldsymbol{\Gamma}$ gives rise to an eigenvector $\boldsymbol{v}_i = \boldsymbol{G} \boldsymbol{N}_g^{-1/2} \boldsymbol{u}_i$ of $\boldsymbol{\mathcal{L}}$ with the same eigenvalue.*

Proof. A simple computation shows that

$$\boldsymbol{\mathcal{L}} \boldsymbol{G} \boldsymbol{N}_g^{-1/2} \boldsymbol{u}_i = \boldsymbol{G} \boldsymbol{N}_g^{-1/2} \boldsymbol{\Gamma} \boldsymbol{N}_g^{-1/2} \boldsymbol{G}^\top \boldsymbol{G} \boldsymbol{N}_g^{-1/2} \boldsymbol{u}_i = \boldsymbol{G} \boldsymbol{N}_g^{-1/2} \boldsymbol{\Gamma} \boldsymbol{u}_i = \lambda_i \boldsymbol{G} \boldsymbol{N}_g^{-1/2} \boldsymbol{u}_i = \lambda_i \boldsymbol{v}_i. \quad \blacksquare$$

Corollary A.3. *The partition indicator matrix $\boldsymbol{G}$ can be written as a linear combination of the eigenvectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ corresponding to the k nonzero eigenvectors of $\boldsymbol{\mathcal{L}}$.*

Proof. From the previous lemma we know that the eigenvectors of $\boldsymbol{\mathcal{L}}$ can be written as $\boldsymbol{v}_i = \boldsymbol{G} \boldsymbol{N}_g^{-1/2} \boldsymbol{u}_i$, where $\boldsymbol{u}_i$ are the eigenvectors of $\boldsymbol{\Gamma}$. Assembling these relationships into matrices, we obtain

$$(\text{A.2}) \qquad \boldsymbol{V} = \boldsymbol{G} \boldsymbol{N}_g^{-1/2} \boldsymbol{U}.$$

Since $\boldsymbol{U}$ is an orthogonal matrix and $\boldsymbol{N}_g^{1/2}$ is diagonal, both matrices are invertible. It then follows that

$$(\text{A.3}) \qquad \boldsymbol{G} = \boldsymbol{V} \boldsymbol{U}^\top \boldsymbol{N}_g^{1/2}. \qquad \blacksquare$$

Corollary A.4. *Consider the row-normalized eigenvector matrix $\widetilde{V} = \mathrm{diag}(VV^\top)^{-1/2}V$. There exists an orthogonal transformation $\mathcal{U}$ such that $\widetilde{V} = G\mathcal{U}$.*

Proof. Note that

$$\mathrm{diag}(VV^\top) = \mathrm{diag}(GN_g^{-1/2}UU^\top N_g^{-1/2}G^\top) = \mathrm{diag}(GN_g^{-1}G^\top).$$

Combining this equation with (A.2) and the fact that G is an indicator matrix, it follows that

$$\widetilde{V} = \mathrm{diag}(GN_g^{-1}G^\top)^{-1/2}V = \mathrm{diag}(GN_g^{-1}G^\top)^{-1/2}GN_g^{-1/2}U = GN_g^{1/2}N_g^{-1/2}U = GU.$$

Since U corresponds to the eigenvector basis of the symmetric normalized affinity matrix Γ, we know that U is orthogonal. ∎

A.2. Proof of Lemma 2.3.

Proof. We define the matrices $\mathcal{D} := \mathbb{E}[D]$ and $H := \mathcal{D}^{-1/2}A\mathcal{D}^{-1/2}$ for notational convenience. From the triangle inequality we have

$$(A.4) \qquad\qquad \|L - \mathcal{L}\| \leq \|L - H\| + \|H - \mathcal{L}\|.$$

To obtain our desired concentration inequality, we bound the two terms on the right-hand side of (A.4) separately.

For the first term we note that since $\|L\| \leq 1$,

$$\|L - H\| = \|D^{-1/2}AD^{-1/2} - \mathcal{D}^{-1/2}A\mathcal{D}^{-1/2}\| = \|L - \mathcal{D}^{-1/2}D^{1/2}LD^{1/2}\mathcal{D}^{-1/2}\|$$

$$= \left\|(I - \mathcal{D}^{-1/2}D^{1/2})LD^{1/2}\mathcal{D}^{-1/2} + L(I - D^{1/2}\mathcal{D}^{-1/2})\right\|$$

$$(A.5) \qquad \leq \|I - \mathcal{D}^{-1/2}D^{1/2}\|\|D^{1/2}\mathcal{D}^{-1/2}\| + \|I - \mathcal{D}^{-1/2}D^{1/2}\|.$$

We can now analyze the term $\|I - \mathcal{D}^{-1/2}D^{1/2}\|$ as follows:

$$\|I - \mathcal{D}^{-1/2}D^{1/2}\| = \max_i \left|1 - \sqrt{D_{ii}/\mathcal{D}_{ii}}\right| \leq \max_i |1 - D_{ii}/\mathcal{D}_{ii}|,$$

where we have used the fact that the spectral norm of a diagonal matrix is its maximal diagonal entry. To make use of the above result, we need to obtain a concentration result for the degrees. Recalling that $\delta_i = \mathbb{E}(d_i) = \mathcal{D}_{ii}$, we can use Chernoff's inequality to obtain a bound of the form

$$\mathbb{P}(|d_i - \delta_i| \geq t\delta_i) \leq \frac{\epsilon}{2n} \quad \text{for } t \geq \sqrt{\frac{\ln(4n/\epsilon)}{\delta_i}}.$$

If we choose $t_0 = \sqrt{\ln(4n/\epsilon)/\delta_{\min}}$, the above bound holds for all i, and by a simple transformation we obtain $\mathbb{P}(|d_i/\delta_i - 1| \geq t_0) \leq \epsilon/2n$ for all i. Using the union bound, we thus have with probability at least $1 - \epsilon/2$

$$\|I - \mathcal{D}^{-1/2}D^{1/2}\| = \max_i \left|1 - \sqrt{\frac{D_{ii}}{\mathcal{D}_{ii}}}\right| \leq \sqrt{\frac{\ln(4n/\epsilon)}{\delta_{\min}}}.$$

Therefore, plugging this result into (A.5), with probability $1 - \epsilon/2$ it holds that

$$(\text{A.6}) \qquad \|\boldsymbol{L} - \boldsymbol{H}\| \leq t_0(t_0 + 1) + t_0.$$

For the second term $\|\boldsymbol{H} - \boldsymbol{\mathcal{L}}\|$, we can use the following matrix concentration inequality.

Lemma A.5 (Bernstein matrix inequality [9]). *Let $\boldsymbol{X}_i$ be independent $n \times n$ dimensional random Hermitian matrices such that $\|\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i]\| \leq C$ for all i, and $\boldsymbol{X} = \sum_i \boldsymbol{X}_i$. Then for any $a > 0$,*

$$\mathbb{P}(\|\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}]\| > a) \leq 2n \exp\left(\frac{-a^2}{2v^2 + 2Ca/3}\right),$$

where $v^2 = \|\sum_i \mathrm{var}(\boldsymbol{X}_i)\| = \|\sum_i \mathbb{E}[\boldsymbol{X}_i^2] - \mathbb{E}[\boldsymbol{X}_i]^2\|$.

To apply this result, note that we can decompose $[\boldsymbol{H} - \boldsymbol{\mathcal{L}}] = \sum_{i \leq j} \boldsymbol{X}_{ij}$, where $\boldsymbol{X}_{ij}$ are the Hermitian matrices

$$\boldsymbol{X}_{ij} = \begin{cases} \frac{A_{ij} - p_{ij}}{\sqrt{\delta_i \delta_j}}(\boldsymbol{E}^{ij} + \boldsymbol{E}^{ij}) & \text{for } i \neq j, \\ \frac{A_{ii} - p_{ii}}{\delta_i}\boldsymbol{E}^{ii} & \text{for } i = j. \end{cases}$$

Here, $\boldsymbol{E}^{ij}$ is the canonical basis matrix with entries $[\boldsymbol{E}^{ij}]_{ij} = 1$ and 0 otherwise. As $\mathbb{E}[\boldsymbol{X}_{ij}] = 0$ we compute $\mathrm{var}(\boldsymbol{X}_{ij}) = \mathbb{E}[\boldsymbol{X}_{ij}^2]$ as follows:

$$\mathrm{var}(\boldsymbol{X}_{ij}) = \begin{cases} \frac{1}{\delta_i \delta_j}(p_{ij} - p_{ij}^2)(\boldsymbol{E}^{ii} + \boldsymbol{E}^{jj}) & \text{for } i \neq j, \\ \frac{1}{\delta_i^2}(p_{ii} - p_{ii}^2)\boldsymbol{E}^{ii} & \text{for } i = j. \end{cases}$$

From this we can bound v^2 as

$$v^2 = \left\|\sum_{i=1}^n \sum_{j=1}^n \frac{1}{\delta_i \delta_j}(p_{ij} - p_{ij}^2)\boldsymbol{E}^{ii}\right\| = \max_i \left(\sum_{j=1}^n \frac{1}{\delta_i \delta_j}p_{ij} - \sum_{j=1}^n \frac{1}{\delta_i \delta_j}p_{ij}^2\right)$$

$$\leq \max_i \left(\frac{1}{\delta_{\min}}\sum_{j=1}^n \frac{1}{\delta_i}p_{ij}\right) = \frac{1}{\delta_{\min}},$$

where we have used the bound $1/\delta_{\min} \geq 1/\delta_j$.

Based on the above calculations, we select $a = \sqrt{3\ln(4n/\epsilon)/\delta_{\min}} = \sqrt{3}t_0$, which leads to our desired bound for the second term

$$\mathbb{P}(\|\boldsymbol{H} - \boldsymbol{\mathcal{L}}\| > a) \leq 2n \exp\left(\frac{-a^2}{2/\delta_{\min} + 2a/(3\delta_{\min})}\right)$$

$$(\text{A.7}) \qquad \leq 2n \exp\left(-\frac{3\ln(4n/\epsilon)}{3}\right) \leq \epsilon/2,$$

where the first inequality follows from the fact that $C \leq 1/\delta_{\min}$ (cf. Lemma A.5), and for the second inequality we used the fact that $\mathcal{M}(\boldsymbol{\mathcal{A}}) \in \mathcal{M}_n(\epsilon)$ to determine a lower bound on $\delta_{\min}$. Finally, by combining the results in (A.6) and (A.7) we obtain that with probability $1 - \epsilon$,

$$\|\boldsymbol{L} - \boldsymbol{\mathcal{L}}\| \leq \|\boldsymbol{L} - \boldsymbol{H}\| + \|\boldsymbol{H} - \boldsymbol{\mathcal{L}}\| \leq a + t_0^2 + 2t_0 \leq 3a. \qquad \blacksquare$$

REFERENCES

[1] E. ABBE, *Community detection and stochastic block models: Recent developments*, J. Mach. Learn. Res., 18 (2018), pp. 1–86.

[2] A. ATHREYA, D. E. FISHKIND, M. TANG, C. E. PRIEBE, Y. PARK, J. T. VOGELSTEIN, K. LEVIN, V. LYZINSKI, AND Y. QIN, *Statistical inference on random dot product graphs: A survey*, J. Mach. Learn. Res., 18 (2017), pp. 8393–8484.

[3] M. AVELLA-MEDINA, F. PARISE, M. T. SCHAUB, AND S. SEGARRA, *Centrality measures for graphons: Accounting for uncertainty in networks*, IEEE Trans. Network Sci. Engrg., 7 (2018), pp. 520–537, https://doi.org/10.1109/TNSE.2018.2884235.

[4] A. S. BANDEIRA, B. BLUM-SMITH, J. KILEEL, A. PERRY, J. WEED, AND A. S. WEIN, *Estimation under Group Actions: Recovering Orbits from Invariants*, preprint, https://arxiv.org/abs/1712.10163, 2017.

[5] A. BOLSTAD, B. D. VAN VEEN, AND R. NOWAK, *Causal network inference via group sparse regularization*, IEEE Trans. Signal Process., 59 (2011), pp. 2628–2641.

[6] S. L. BRUNTON, J. L. PROCTOR, AND J. N. KUTZ, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proc. Natl. Acad. Sci. USA, 113 (2016), pp. 3932–3937.

[7] F. BUNEA AND L. XIAO, *On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA*, Bernoulli, 21 (2015), pp. 1200–1230.

[8] E. J. CANDÈS, M. B. WAKIN, AND S. P. BOYD, *Enhancing sparsity by reweighted ℓ_1 minimization*, J. Fourier Anal. Appl., 14 (2008), pp. 877–905.

[9] F. CHUNG AND M. RADCLIFFE, *On the spectra of general random graphs*, Electron. J. Combin., 18 (2011), P215.

[10] A. DASGUPTA, J. E. HOPCROFT, AND F. MCSHERRY, *Spectral analysis of random graphs with skewed degree distributions*, in Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science, Rome, 2004, pp. 602–610.

[11] C. DAVIS AND W. M. KAHAN, *The rotation of eigenvectors by a perturbation* III, SIAM J. Numer. Anal., 7 (1970), pp. 1–46, https://doi.org/10.1137/0707001.

[12] G. B. GIANNAKIS, Y. SHEN, AND G. V. KARANIKOLAS, *Topology identification and learning over graphs: Accounting for nonlinearities and dynamics*, Proc. IEEE, 106 (2018), pp. 787–807.

[13] C. GRAY, D. J. BAIRD, S. BAUMGARTNER, U. JACOB, G. B. JENKINS, E. J. O'GORMAN, X. LU, A. MA, M. J. POCOCK, N. SCHUWIRTH, M. THOMPSON, AND G. WOODWARD, *Ecological networks: The missing links in biomonitoring science*, J. Appl. Ecol., 51 (2014), pp. 1444–1449.

[14] M. D. GREICIUS, K. SUPEKAR, V. MENON, AND R. F. DOUGHERTY, *Resting-state functional connectivity reflects structural connectivity in the default mode network*, Cerebral Cortex, 19 (2009), pp. 72–78.

[15] L. GULIKERS, M. LELARGE, AND L. MASSOULIÉ, *A spectral method for community detection in moderately sparse degree-corrected stochastic block models*, Adv. Appl. Probab., 49 (2017), pp. 686–721.

[16] D. HAYDEN, Y. H. CHANG, J. GONCALVES, AND C. J. TOMLIN, *Sparse network identifiability via compressed sensing*, Automatica, 68 (2016), pp. 9–17, https://doi.org/10.1016/j.automatica.2016.01.008.

[17] D. HAYDEN, Y. YUAN, AND J. GONÇALVES, *Network identifiability from intrinsic noise*, IEEE Trans. Automat. Control, 62 (2017), pp. 3717–3728, https://doi.org/10.1109/TAC.2016.2640219.

[18] T. HOFFMANN, L. PEEL, R. LAMBIOTTE, AND N. S. JONES, *Community detection in networks without observing edges*, Sci. Adv., 6 (2020), eeav1478, https://doi.org/10.1126/sciadv.aav1478.

[19] C. HONEY, O. SPORNS, L. CAMMOUN, X. GIGANDET, J.-P. THIRAN, R. MEULI, AND P. HAGMANN, *Predicting human resting-state functional connectivity from structural connectivity*, Proc. Natl. Acad. Sci. USA, 106 (2009), pp. 2035–2040.

[20] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, 1990.

[21] A. JADBABAIE, J. LIN, AND A. S. MORSE, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Trans. Automat. Control, 48 (2003), pp. 988–1001.

[22] A. JULIUS, M. ZAVLANOS, S. BOYD, AND G. J. PAPPAS, *Genetic network identification using convex programming*, IET Syst. Biol., 3 (2009), pp. 155–166.

[23] B. KARRER AND M. E. NEWMAN, *Stochastic blockmodels and community structure in networks*, Phys. Rev. E, 83 (2011), 016107.

[24] A. KUMAR, Y. SABHARWAL, AND S. SEN, *A simple linear time (1+ epsilon)-approximation algorithm*

for k-means clustering in any dimensions, in Proceedings of the 45th IEEE Annual Symposium on Foundations of Computer Science, Rome, 2004, pp. 454–462.

[25] J. LEI AND A. RINALDO, *Consistency of spectral clustering in stochastic block models*, Ann. Statist., 43 (2015), pp. 215–237.

[26] L. LJUNG, *System identification*, in Signal Analysis and Prediction, Springer, 1998, pp. 163–173.

[27] L. LJUNG, *Perspectives on system identification*, Annu. Rev. Control, 34 (2010), pp. 1–12.

[28] L. LOVÁSZ, *Large Networks and Graph Limits*, Amer. Math. Soc. Colloq. Publ. 60, American Mathematical Society, 2012.

[29] N. MASUDA, M. A. PORTER, AND R. LAMBIOTTE, *Random walks and diffusion on networks*, Phys. Rep., 716/717 (2017), pp. 1–58.

[30] G. MATEOS, S. SEGARRA, A. G. MARQUES, AND A. RIBEIRO, *Connecting the dots: Identifying network structure via graph signal processing*, IEEE Signal Process. Mag., 36 (2019), pp. 16–43.

[31] D. MATERASSI AND M. V. SALAPAKA, *On the problem of reconstructing an unknown topology via locality properties of the Wiener filter*, IEEE Trans. Automat. Control, 57 (2012), pp. 1765–1777.

[32] A. MAUROY AND J. HENDRICKX, *Spectral identification of networks using sparse measurements*, SIAM J. Appl. Dyn. Syst., 16 (2017), pp. 479–513, https://doi.org/10.1137/16M105722X.

[33] M. NITZAN, J. CASADIEGO, AND M. TIMME, *Revealing physical interaction networks from statistics of collective dynamics*, Sci. Adv., 3 (2017), e1600396.

[34] R. OLFATI-SABER, J. A. FAX, AND R. M. MURRAY, *Consensus and cooperation in networked multi-agent systems*, Proc. IEEE, 95 (2007), pp. 215–233.

[35] A. ORTEGA, P. FROSSARD, J. KOVAČEVIĆ, J. M. F. MOURA, AND P. VANDERGHEYNST, *Graph signal processing: Overview, challenges, and applications*, Proc. IEEE, 106 (2018), pp. 808–828, https://doi.org/10.1109/JPROC.2018.2820126.

[36] R. OVERBEEK, M. FONSTEIN, M. D'SOUZA, G. D. PUSCH, AND N. MALTSEV, *The use of gene clusters to infer functional coupling*, Proc. Natl. Acad. Sci. USA, 96 (1999), pp. 2896–2901.

[37] J. PEARL, *Causal inference in statistics: An overview*, Stat. Surv., 3 (2009), pp. 96–146.

[38] L. PEEL, D. B. LARREMORE, AND A. CLAUSET, *The ground truth about metadata and community detection in networks*, Sci. Adv., 3 (2017), e1602548.

[39] T. P. PEIXOTO, *Network reconstruction and community detection from dynamics*, Phys. Rev. Lett., 123 (2019), 128301.

[40] A. PERRY, J. WEED, A. S. BANDEIRA, P. RIGOLLET, AND A. SINGER, *The sample complexity of multireference alignment*, SIAM J. Math. Data Sci., 1 (2019), pp. 497–517, https://doi.org/10.1137/18M1214317.

[41] T. QIN AND K. ROHE, *Regularized spectral clustering under the degree-corrected stochastic blockmodel*, in NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems, Curran Associates, 2013, pp. 3120–3128.

[42] K. ROHE, S. CHATTERJEE, AND B. YU, *Spectral clustering and the high-dimensional stochastic blockmodel*, Ann. Statist., 39 (2011), pp. 1878–1915.

[43] A.-E. SALIBA, A. J. WESTERMANN, S. A. GORSKI, AND J. VOGEL, *Single-cell RNA-seq: Advances and future challenges*, Nucleic Acids Res., 42 (2014), pp. 8845–8860.

[44] G. SANGUINETTI, J. LAIDLER, AND N. D. LAWRENCE, *Automatic determination of the number of clusters using spectral algorithms*, in Proceedings of the 2005 IEEE Workshop on Machine Learning for Signal Processing, Mystic, CT, 2005, pp. 55–60.

[45] M. T. SCHAUB, S. SEGARRA, AND H. T. WAI, *Spectral partitioning of time-varying networks with unobserved edges*, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 4938–4942.

[46] S. SEGARRA, A. G. MARQUES, G. MATEOS, AND A. RIBEIRO, *Network topology inference from spectral templates*, IEEE Trans. Signal Inform. Process. Netw., 3 (2017), pp. 467–483.

[47] S. SEGARRA, A. G. MARQUES, AND A. RIBEIRO, *Optimal graph-filter design and applications to distributed linear network operators*, IEEE Trans. Signal Process., 65 (2017), pp. 4117–4131.

[48] S. SEGARRA, M. T. SCHAUB, AND A. JADBABAIE, *Network inference from consensus dynamics*, in Proceedings of the 2017 IEEE 56th Annual Conference on Decision and Control (CDC), Melbourne, Australia, 2017, pp. 3212–3217.

[49] R. SHAFIPOUR, S. SEGARRA, A. G. MARQUES, AND G. MATEOS, *Network topology inference from*

non-stationary graph signals, in Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2017, pp. 5870–5874.

[50] S. Shahrampour and V. M. Preciado, *Reconstruction of directed networks from consensus dynamics*, in Proceedings of the 2013 American Control Conference (ACC), Washington, DC, 2013, pp. 1685–1690.

[51] S. Shahrampour and V. M. Preciado, *Topology identification of directed dynamical networks via power spectral analysis*, IEEE Trans. Automat. Control, 60 (2015), pp. 2260–2265.

[52] V. A. Smith, E. D. Jarvis, and A. J. Hartemink, *Evaluating functional network inference using simulations of complex biological systems*, Bioinformatics, 18 (2002), pp. S216–S224.

[53] P. Spirtes and K. Zhang, *Causal discovery and inference: Concepts and recent methodological advances*, Appl. Inform., 3 (2016), 3.

[54] M. Timme and J. Casadiego, *Revealing networks from dynamics: An introduction*, J. Phys. A, 47 (2014), 343001.

[55] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, *Comparing community structure to characteristics in online collegiate social networks*, SIAM Rev., 53 (2011), pp. 526–543, https://doi.org/10.1137/080734315.

[56] A. L. Traud, P. J. Mucha, and M. A. Porter, *Social structure of Facebook networks*, Phys. A, 391 (2012), pp. 4165–4180.

[57] J. N. Tsitsiklis, *Problems in Decentralized Decision Making and Computation*, Ph.D. thesis, Massachusetts Institute of Technology, 1984.

[58] U. Von Luxburg, *A tutorial on spectral clustering*, Statist. Comput., 17 (2007), pp. 395–416.

[59] H.-T. Wai, A. Scaglione, and A. Leshem, *Active sensing of social networks*, IEEE Trans. Signal Inform. Process. Netw., 2 (2016), pp. 406–419.

[60] H.-T. Wai, S. Segarra, A. E. Ozdaglar, A. Scaglione, and A. Jadbabaie, *Blind community detection from low-rank excitations of a graph filter*, IEEE Trans. Signal Process., 68 (2020), pp. 436–451.

[61] H.-T. Wai, S. Segarra, A. E. Ozdaglar, A. Scaglione, and A. Jadbabaie, *Community detection from low-rank excitations of a graph filter*, in Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada 2018, pp. 4044–4048.

[62] W.-X. Wang, Y.-C. Lai, and C. Grebogi, *Data based identification and prediction of nonlinear and complex dynamical systems*, Phys. Rep., 644 (2016), pp. 1–76.

[63] T. Xiang and S. Gong, *Spectral clustering with eigenvector selection*, Pattern Recognition, 41 (2008), pp. 1012–1029.

[64] Y. Yu, T. Wang, and R. J. Samworth, *A useful variant of the Davis-Kahan theorem for statisticians*, Biometrika, 102 (2014), pp. 315–323.

[65] Y. Yuan, G.-B. Stan, S. Warnick, and J. Goncalves, *Robust dynamical network structure reconstruction*, Automatica, 47 (2011), pp. 1230–1235.

[66] W. W. Zachary, *An information flow model for conflict and fission in small groups*, J. Anthropological Res., 33 (1977), pp. 452–473.

[67] A. Zhang and M. Wang, *Spectral state compression of Markov processes*, IEEE Trans. Inform. Theory, 66 (2020), pp. 3202–3231.