Received 28 March 2022; accepted 21 June 2022. Date of publication 7 July 2022;
date of current version 26 September 2022. The review of this article was arranged by Associate Editor Cedric Richard.

Digital Object Identifier 10.1109/OJSP.2022.3189315

Graph Learning Over Partially Observed Diffusion Networks: Role of Degree Concentration

VINCENZO MATTA [1,2] (Senior Member, IEEE), AUGUSTO SANTOS [3,4], AND ALI H. SAYED [3] (Fellow, IEEE)

[1] Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, I-84084 Fisciano, SA, Italy
[2] National Inter-University Consortium for Telecommunications, I-84084 Fisciano, SA, Italy
[3] School of Engineering, École Polytechnique Fédérale de Lausanne EPFL, CH-1015 Lausanne, Switzerland
[4] Centre for Informatics and Systems, University of Coimbra, 3004-531 Coimbra, Portugal

This paper was presented in part at the 2018 Asilomar Conference on Signals, Systems, and Computers [DOI: 10.1109/ACSSC.2018.8645374], and at the 2019 IEEE International Symposium on Information Theory [DOI: 10.1109/ISIT.2019.8849234].

CORRESPONDING AUTHOR: VINCENZO MATTA (e-mail: vmatta@unisa.it)

The work of Augusto Santos was supported by the Foundation for Science and Technology, Portugal, I.P./MCTES through National Funds (PIDDAC), within the scope of CISUC R&D Unit under Grants UIDB/00326/2020 and UIDP/00326/2020. The work of Ali H. Sayed was supported by the Swiss National Science Foundation under Grant 205121-184999.

ABSTRACT This work examines the problem of learning the topology of a network from the samples of a diffusion process evolving at the network nodes, under the restriction that a limited fraction thereof is probed (*partial observability*). We provide the following main contributions. Given an estimator of the combination matrix (i.e., the matrix that quantifies the pairwise interaction between nodes), we introduce the notion of *identifiability gap*, a minimum separation between the entries of the estimated matrix that is critical to enable discrimination between connected and unconnected node pairs. Then we focus on the popular Granger estimator. First, we prove that this matrix estimator, followed by a *universal* clustering algorithm inspired by the k-means algorithm, learns faithfully the probed subgraph as the network size increases. This result is proved for the case where the network topology is obtained through an Erdős-Rényi *random* graph under *statistical concentration* of the node degrees, and the combination matrix is symmetric with nonzero entries bounded in terms of the reciprocal of the maximal and minimal degree. The analysis explores different connectivity regimes, including the *dense* regime where the probed nodes are influenced by many connections coming from the latent (hidden) part of the graph. Second, we answer a *sample complexity* question and establish that the number of samples for the Granger estimator scales almost *quadratically* with the expected graph degree. We also propose three other estimators that are proved to achieve faithful graph learning, and compare them to the Granger estimator, gaining nontrivial insights especially for the case of directed graphs.

INDEX TERMS Graph learning, network tomography, dense networks, Granger estimator, diffusion network, Erdős-Rényi graph, identifiability gap, graph concentration.

I. INTRODUCTION

Learning the graph structure that governs the evolution of a networked dynamical system from data collected at some accessible nodes is a challenging inverse problem with applications across many domains. The objective of such inferential problems is to discover the interaction profile among the network nodes since the topology has a critical effect on system behavior [3], [4], [5], [6]. Graph learning plays a central role in many applications including, among other possibilities: estimating the longevity or the source of an epidemics [7], [8]; revealing commonalities and agent influence over social networks [9], [10], [11]; discovering the routes of clandestine information flows [12], [13]; identifying defective elements [14]; addressing the fundamental issue in neuroscience that links brain functional connectivity (i.e., a "functional" topology estimated from

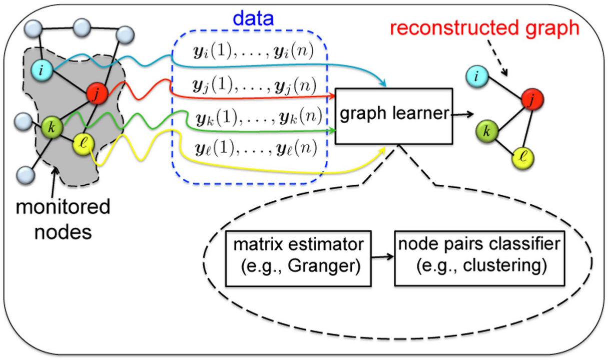
data
$\boldsymbol{y}_i(1),\ldots,\boldsymbol{y}_i(n)$
$\boldsymbol{y}_j(1),\ldots,\boldsymbol{y}_j(n)$
$\boldsymbol{y}_k(1),\ldots,\boldsymbol{y}_k(n)$
$\boldsymbol{y}_\ell(1),\ldots,\boldsymbol{y}_\ell(n)$
reconstructed graph
graph learner
monitored nodes
matrix estimator (e.g., Granger)
node pairs classifier (e.g., clustering)
i
j
k
ℓ

FIGURE 1. Illustration of the graph learning problem considered in this work.

blood-oxygenation-level-dependent signals) to brain structural connectivity (i.e., the anatomical topology of neuron interconnections) [15], [16], [17].

Depending on the particular context, the aforementioned class of problems can be referred to in different ways, including *topology inference* [18], *network tomography* [19], [20], *structure learning* [21], or *graph learning* [22]. We adopt these terminologies almost interchangeably throughout our treatment.

This article addresses the graph learning problem under the framework of *partial observability*, i.e., when only a fraction of the network nodes can be probed. This setting is particularly important in large-scale networked systems, where it is not feasible to gather data from all nodes comprising the network. We solve the learning problem for distinct regimes of network wiring, including *densely-connected* networks, a case often overlooked in the literature.

II. OVERVIEW OF THE LEARNING PROBLEM

The structure of the learning problem addressed in this work can be summarized as follows. Streams of data originating from a certain subnetwork are collected, and the goal is to estimate the (unknown) topology linking the nodes of this subnetwork from the collected data. The graph learning protocol will involve two main steps: an estimation step, where a combination matrix (i.e., a matrix quantifying the strength of the connections among the network nodes) is estimated; and a thresholding step, where node pairs linked by a strong edge (i.e., node pairs whose corresponding estimated matrix entry lies above some threshold) are deemed connected. A structurally-consistent estimator is one that ends up assigning strong ties to interacting pairs and weak ties to non-interacting pairs. In this way, at the thresholding stage, one can correctly classify the pairs as interacting and non-interacting. Fig. 1 gives a graphical summary of the aforementioned procedure.

It is useful to illustrate the learning problem in relation to some popular networked systems. To start with, let us neglect some practical limitations, in particular assume that: *i*) all nodes can be monitored (full observability); *ii*) there are no limitations in terms of computational power; and *iii*) there are no limitations on the available time samples. Then, the first inferential stage consists of finding a matrix estimator to quantify the strength of pairwise interactions in the network. One notable estimator relies on computing the (spatial) covariance matrix $R_0 = \lim_{n\to\infty} \mathbb{E}[\boldsymbol{y}_n \boldsymbol{y}_n^\top]$, where $\boldsymbol{y}_n$ denotes the vector collecting the data from all nodes at time n, and where *iii*) justifies the limit and (under an ergodic assumption) implies that the statistical average can be learned from the data. When the matrix R_0 provides a consistent estimator for the connection strengths, we talk of a *correlation network* [18]. For these networks, interactions between two nodes are *direct* and they are accordingly captured by pairwise correlations. One example of a correlation network is the ferromagnetic Ising model [23] with independent and identically distributed (i.i.d.) time samples, and under certain constraints of sparsity on the network and of regularity on the interaction weights.

Another classic model for graph learning is a *Gaussian graphical model*. In this case, R_0 is no longer the proper estimator, but its inverse R_0^{-1} (which is often referred to as the precision or concentration matrix) is a consistent estimator, in that its support coincides with the underlying graph of interactions. Over Gaussian graphical models, the pairwise interaction between adjacent nodes is affected by other nodes, and this latent influence is the reason why spatial correlation between measurements is no longer sufficient to capture the network structure.

For most standard graphical models, interactions across network nodes are described through a multivariate distribution that characterizes a collection of *dependent* random variables defined on the nodes. It is usually assumed that i.i.d. samples of these variables are available for the learning process. In other words, over graphical models the data samples do not arise from a dynamical process governing the time evolution of the nodes' outputs. In contrast, in this article we will be dealing with networked *dynamical* systems, where signals evolve at the nodes and are affected by the evolution of the signals at neighboring nodes as well. One relevant example is the diffusion or first-order Vector AutoRegressive (VAR) system described by (3) further ahead. For such graphs, the proper estimator for graph connectivity turns out to be the Granger estimator, $R_1 R_0^{-1}$, which combines in a suitable way information contained in the covariance matrix, R_0, and in the *one-lag* covariance matrix, $R_1 = \lim_{n\to\infty} \mathbb{E}[\boldsymbol{y}_n \boldsymbol{y}_{n-1}^\top]$.

A. STRUCTURAL-CONSISTENCY, HARDNESS AND SAMPLE-COMPLEXITY

We are now ready to introduce three concepts that play an important role in graph learning problems.

1) STRUCTURAL CONSISTENCY

In the previous examples, the graph structure can be retrieved from a statistical descriptor related to the measurements, i.e., R_0 for correlation networks, R_0^{-1} for Gaussian graphical models, and $R_1 R_0^{-1}$ for VAR models. Since the involved covariance matrices can be computed from the measurements, with arbitrary precision as the number of samples increases, we conclude that in the aforementioned three examples the graph can be correctly identified.

In a more general setting, consider a statistical descriptor that can be consistently estimated from the data as the number of samples goes to infinity. If, for sufficiently large networks, the statistical descriptor allows to identify the correct graph structure, we shall say that the graph learning problem is *identifiable*, and that the corresponding descriptor achieves *structural consistency*.

In this work we focus on the case in which the measurements are available from only a limited subset of nodes, with identifiability referring here to the subgraph connecting these monitored nodes. Under this setting,[1] many interesting questions arise, such as: *Does partial observability impair identifiability of the monitored subnetwork? If not, how can we design structurally-consistent estimators?*

We remark that the concepts of identifiability and structural consistency disregard complexity issues, since they assume that the necessary statistical quantities (e.g., the true covariance matrices) are available. For example, we assume that $R_1 R_0^{-1}$ can be computed exactly, which means that matrix inversion is possible whatever the size of the network, and that we have sufficient time to learn perfectly the true covariance matrices.

2) HARDNESS

How much computational complexity is required to evaluate the matrix estimator necessary for a particular graph learning problem? For example, if one is interested in the precision matrix, R_0^{-1}, the hardness is related to the complexity of the matrix inversion. Many works attempt to reduce the complexity by leveraging particular constraints such as smoothness of the node-level signals and/or sparsity of the underlying graph of interactions.

Notably, there are topology inference problems that are computationally intractable (e.g., NP-hard) [24], [25], [26], [27]. It is therefore critical to identify meaningful systems where this does not happen [28]. The present work sheds light on a relevant class of systems whereby structure learning with statistically dependent observations and under partial observability has affordable computational complexity, even for the important case of *dense* (thus, loopy and non-tree like) networks. However, we remark that the concept of hardness disregards the complexity associated with the *empirical* estimation of the pertinent matrices from the available data samples. This fundamental element of complexity is usually referred to as *sample* complexity.

3) SAMPLE COMPLEXITY

In practice, only a finite amount of data is available and, therefore, only approximate versions of the aforementioned matrix estimators can be computed. There exist several results about sample complexity in the context of high-dimensional graphical models, where the number of samples necessary to get some prescribed accuracy is related to the system parameters, e.g., to the network size and to the density of connections. Results relative to sample complexity over *dynamical* graph systems are comparably less mature [18]. Over these systems, the dependence among the time samples induced by the dynamical model complicates the theoretical analysis of the convergence of the empirical estimators, and, hence, the analysis of sample complexity. Useful results about the sample complexity of vector autoregressive models like the one addressed in this work are available in [29], [30], [31], [32]. These results do not consider the *partial observability* setting and, hence, they do not apply here.[2] Nonetheless, in [29], [30], [31], [32] we can find useful techniques to bound the errors associated to the empirical covariance matrices over vector autoregressive models. These types of bounds will be exploited in the proof of Theorem 3, where we establish the sample complexity of the estimators proposed in this work.

III. RELATED WORK

The learning problem considered in this work lies in the broad field of signal processing over graphs [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], dealing in particular with the identification of an unknown network topology from measurements gathered at the network nodes [18]. These types of inferential problems can be addressed under different settings, including the case where measurements from all network nodes are available (full observability), and the case where only a fraction of nodes is accessible (partial observability). Even if we focus on the partial observability setting, we deem it useful to start with some results pertaining to the full observability setting.

A. GRAPH LEARNING UNDER FULL OBSERVABILITY

The majority of works on graph learning over networks focuses on linear system dynamics, with nonlinear dynamics typically being tackled by variational characterizations under a small-noise assumption [47], [48], [49], or by increasing the dimensionality of the observable space [50], [51].

Topology inference for a general class of linear stochastic dynamical systems (e.g., VAR models of arbitrary order, or even non-causal linear models) is addressed in [52]. An approach based on Wiener filtering is proposed to infer the

[1] It should be remarked, however, that certain issues may arise also in the full observability case. For instance, as the network size increases, the entries of the matrix estimators might become smaller and smaller, and, hence, it is critical to identify whether a stable threshold can be found to classify connected/unconnected node pairs.

[2] In order to avoid misunderstandings, we remark that the terminology "partial samples" or "missing data" used in [31], [32] refers to a different problem. In our setting samples from only a subset of the network are available, while in [31], [32] samples from the whole network are available but, given an overall amount of data generated at each node, a certain fraction per node is randomly lost and/or corrupted.

topology, which provides exact reconstruction for self-kin networks or, in general, guarantees reconstruction of the smallest self-kin network embracing the true network.

There exist works dealing with more general dynamical systems and graphs. For example, in [53] the concept of directed information graphs is advocated to discover dependencies in networks of interacting processes linked by causal dynamics, and a metric based on *functional dependencies* is proposed in [54] to learn causal relationships over a possibly nonlinear network.

Moving closer to our setting, among linear (or linearized) systems, special attention is devoted to autoregressive diffusion models [55], [56], [57]. For instance, in [55] causal graph processes are exploited to devise a computationally tractable algorithm for graph structure recovery with performance guarantees. Recent works exploit optimization and graph signal processing techniques to feed the graph learning algorithm with proper structural constraints. In [56], [57], it is shown how to capitalize on the fact that the weighting matrix and the covariance matrix share the same eigenvectors, and how to solve the topology inverse problem through optimization methods under sparsity constraints. An account of the methods for the full observability regime can be found in [18].

We stress that most of the aforementioned methods work in the *graph spectral domain*. This has to be contrasted with the methods proposed in this paper, which rely instead on the graph edge domain. Working in the edge domain allows us to obtain a transparent relationship for the matrix estimators, which is critical to establish identifiability under the challenging partial observability setting.

In summary, while in certain cases (e.g., general linear models and/or nonlinear models) identifiability can be an issue, most of the works on diffusion models focus on reducing complexity by exploiting proper structural constraints (e.g., smoothness or sparsity of the signals defined on the graph) [18]. However, all the aforementioned results pertain to the case where node measurements from the entire network are available. We focus instead on the case in which only partial observation of the network is permitted.

B. GRAPH LEARNING UNDER PARTIAL OBSERVABILITY

A fundamental challenge of our work is performing structure learning when only *partial* observation is allowed [58]. Under this setting, results for retrieving particular network graphs (polytrees) are available in [59] and [60]. Considering instead the case of general topologies, and with focus on VAR/diffusion models like the one considered in this work, references [61], [62] establish technical conditions for exact or partial topology identifiability. However, these identifiability conditions act at a very "microscopic" level (they are formulated in terms of some precise details of the local graph structure and/or of the statistical model), and are therefore impractical over large-scale networks. In contrast, in this work we pursue a statistical approach that is genuinely tailored to the large-scale setting: an asymptotic framework is considered, where the thermodynamic limit of large networks is afforded by using *random* graphs, and the conditions on the network connection topology are summarized at a *macroscopic* level through average descriptive indicators (e.g., probability of drawing an edge). In a similar vein, detailed asymptotic analysis with performance guarantees are available for graphical models with latent variables. For example, in [63] it is shown that, under certain conditions concerning the interactions between the observed and the unobserved network nodes, the "*sparsity*+*low-rank*" framework can be exploited to estimate the amount of latent variables [63], and to reconstruct the topology of the observable subnetwork. Likewise, in [21] the graph learning problem is tackled in the context of locally-tree graphs, whereas in [64] a local separation criterion is imposed to deal with Gaussian graphical models. Still in the framework of learning graphical models with latent variables, in [65] an influence-maximization metric is proposed, to show that ferromagnetic restricted Boltzmann machines with bounded degree are an instance of graphical models that can be efficiently learned.

However, and as already explained in Section II, graphical models do not match the networked dynamical models considered in this work. For these models, results for graph learning under partial observability have been recently obtained in [19], [20], [66], [67], [68]. More specifically, *i*) in [19] the whole network graph is assumed to follow an Erdős-Rényi construction and the number of observable nodes grows with the overall network size N; whereas *ii*) in [20] the number of monitored nodes is held fixed (and, hence, the fraction of observable nodes vanishes in the limit of large N), the graph of the monitored nodes is left arbitrary, and the unobserved component continues to obey an Erdős-Rényi model. The present work focuses on the former model.[3] It is therefore necessary to explain clearly why the present work constitutes a significant progress in the context of local tomography over diffusion networks, in comparison to [19].

C. MAIN ADVANCES

The key contributions of this work in relation to [19] are as follows.

— One first advance relates to the regime of connectivity. Reference [19] addresses only the case that the network is sparsely connected, which means that the connection probability is allowed to *vanish* with N in a way that preserves network connectedness. In this work we examine also the *dense* regime, where the connection probability is *not* vanishing.

— We advance also with respect to the results currently available under the sparse regime. In [19] a consistency result is proved, for all sparsely connected networks,

[3] Even if the machinery used to prove our results can be applied to the latter model as well, we deem it useful to focus on a single model, in order to make the exposition more organic and to convey better the main message of the work.

in terms of an *average* fraction of misclassified node pairs. In the present work, we introduce in Section V a significantly stronger notion of consistency (referred to as *universal local structural consistency*), which will apply to all networks that fulfill a property of *statistical concentration* of node degrees. These networks will be shown to be all the dense networks, as well as most part of the sparse networks. Using such stronger notion of consistency answers also the following important question (posed, and only partially answered in [19], where the answer was obtained under a certain *approximation* of independence): *does the fraction of mistakes scale properly with N?*

— We are able to offer a rigorous proof that the connected and unconnected node pairs can be recovered through some *universal* clustering algorithm. In particular, we propose a variant of the k-means algorithm that is shown to be asymptotically consistent.

— We achieve the significant advance of ascertaining the *sample complexity* of the Granger estimator, i.e., we establish how the number of samples must scale with the network size to let this estimator learn the graph of probed nodes faithfully.

— Another advance relates to the topology inference algorithms. The matrix estimators available in the full-observability setting help guide the choice of a matrix estimator for the partial observability setting. For example, one can replace the Granger estimator with a version that considers only the subnetwork of observed measurements. This choice is widely adopted in causal inference from time series (when one neglects the existence of latent components), and has been adopted, e.g., in [19], [20]. In this work we characterize thoroughly the Granger estimator. Then, in Section VIII we consider three other matrix estimators, namely, the *one-three-lags estimator* (computing the difference between the one-lag and three-lags covariance matrices), the *one-lag covariance matrix* and the *covariance matrix between the residuals* (i.e., difference between subsequent time samples). We show that all the three estimators are structurally consistent.

— Finally, we examine the important case of directed graphs. Our theorems are proved under the assumption of symmetric combination matrices. However, while the three new estimators are implicitly constructed by exploiting symmetry, the construction of the Granger estimator does not assume any symmetry, and, hence, we expect that it performs well also over directed graphs. We show by numerical simulations that this is actually the case. In contrast, and remarkably, the other estimators lose their learning ability over directed graphs.

Notation: We use boldface letters to denote random variables, and normal font letters for their realizations. Matrices are denoted by capital letters, and vectors by small letters. This convention can be occasionally violated, for example, the total number of network nodes is denoted by N. The symbol $\xrightarrow{\text{p}}$ denotes convergence in probability as $N \to \infty$.

Sets and events are denoted by upper-case calligraphic letters, whereas the corresponding normal-font letter will denote the cardinality of the set. For example, the cardinality of $\mathcal{S}$ is S. The complement of $\mathcal{S}$ is denoted by $\mathcal{S}'$.

For a $K \times K$ matrix Z, the submatrix spanning the rows of Z indexed by set $\mathcal{S} \subseteq \{1, 2, \dots, K\}$ and the columns indexed by set $\mathcal{T} \subseteq \{1, 2, \dots, K\}$, is denoted by $Z_{\mathcal{S}\mathcal{T}}$, or alternatively by $[Z]_{\mathcal{S}\mathcal{T}}$. When $\mathcal{S} = \mathcal{T}$, the submatrix $Z_{\mathcal{S}\mathcal{T}}$ is abbreviated as $Z_{\mathcal{S}}$. Moreover, in the indexing of a submatrix we keep the index set of the corresponding full matrix. For example, if $\mathcal{S} = \{2, 3\}$ and $\mathcal{T} = \{2, 4, 5\}$, the submatrix $M = Z_{\mathcal{S}\mathcal{T}}$ is a 2×3 matrix, indexed as follows:

$$M = \begin{pmatrix} z_{22} & z_{24} & z_{25} \\ z_{32} & z_{34} & z_{35} \end{pmatrix} = \begin{pmatrix} m_{22} & m_{24} & m_{25} \\ m_{32} & m_{34} & m_{35} \end{pmatrix}. \tag{1}$$

For a matrix M, the symbols $\|M\|_1$, $\|M\|_2$ and $\|M\|_\infty$ denote the vector-induced ℓ_1, ℓ_2 and ℓ_∞ norms of M, respectively. The symbol $\|M\|_{\max}$ denotes instead the maximum absolute entry of M. The symbol log denotes the natural logarithm.

IV. NETWORKED DYNAMICAL SYSTEM

Let $\boldsymbol{y}_i(n)$ be the output measurement produced by node i at time n. Likewise, let $\boldsymbol{x}_i(n)$ be the input source (e.g., streaming data or noise) exciting node i at time n. It is convenient to stack the input and output variables, respectively, into the vectors:

$$\begin{aligned} \boldsymbol{x}_n &= [\boldsymbol{x}_1(n), \boldsymbol{x}_2(n), \dots, \boldsymbol{x}_N(n)]^\top, \\ \boldsymbol{y}_n &= [\boldsymbol{y}_1(n), \boldsymbol{y}_2(n), \dots, \boldsymbol{y}_N(n)]^\top. \end{aligned} \tag{2}$$

The stochastic dynamical system considered in the present work is given by the following *network diffusion* process (a.k.a. first-order Vector AutoRegressive (VAR) model):

$$\boxed{\boldsymbol{y}_n = \boldsymbol{A}\,\boldsymbol{y}_{n-1} + \sigma\,\boldsymbol{x}_n} \tag{3}$$

where $\boldsymbol{A}$ is some stable $N \times N$ matrix with nonnegative entries, and σ^2 is a variance factor. The bold notation for $\boldsymbol{A}$ is used since, as explained in the next section, we will be dealing with random graphs.

By rewriting (3) on an entrywise basis:

$$\boldsymbol{y}_i(n) = \sum_{\ell=1}^{N} \boldsymbol{a}_{i\ell}\,\boldsymbol{y}_\ell(n-1) + \sigma\,\boldsymbol{x}_i(n), \tag{4}$$

we readily see that the support-graph of $\boldsymbol{A}$ reflects the connections among the network nodes. Indeed, (4) shows that, at time n, the output of node i is updated by *combining* the outputs of other nodes from time $n-1$. In particular, node i scales the output of node ℓ by using a combination weight $\boldsymbol{a}_{i\ell}$, which implies that the output of node ℓ is *effectively used* by node i if, and only if $\boldsymbol{a}_{i\ell} \neq 0$. After the combination step, the output measurement $\boldsymbol{y}_i(n)$ is adjusted by incorporating the streaming-source value, $\boldsymbol{x}_i(n)$, which is locally available at node i at current time n.

Formulations like the one in (3) arise naturally across many application domains, e.g., in economics [69], in the variational characterization of nonlinear stochastic dynamical systems [70], or in distributed network processing applications where several useful strategies such as consensus [34], [71], [72], [73], [74], [75], [76], [77], [78], [79], [80], [81] and diffusion [82], [83], [84], [85], [86], [87] lead to data models of the form in (4).

In our *partial observability* setting, only a subset of nodes can be probed: for each node i belonging to the subset of probed nodes, $\mathcal{S}$, a stream of n measurements, $\boldsymbol{y}_i(1), \boldsymbol{y}_i(2), \ldots, \boldsymbol{y}_i(n)$ is acquired. The learning task is to reconstruct the graph of interconnections corresponding to the combination (sub)matrix $A_{\mathcal{S}}$. Since we will study the graph learning problem in the asymptotic regime where the network size N goes to infinity, it is necessary to specify how the cardinality of the probed set scales with N.

Definition 1 (Partial observability setting): The subnetwork of observable measurements, $\mathcal{S}$, has a cardinality S scaling as:

$$\frac{S}{N} \overset{N\to\infty}{\longrightarrow} \xi \in (0,1), \tag{5}$$

which means that ξ is the (asymptotic) fraction of monitored nodes. Since ξ is strictly less than one, condition (5) conforms to a *partial observability* setting.

A. RANDOM GRAPH AND COMBINATION MATRIX

In the following, we denote by $\boldsymbol{G}$ the adjacency matrix of the network graph, whose entry $\boldsymbol{g}_{ij}$ is equal to one if nodes i and j are connected, and is equal to zero otherwise. The bold notation is used because we deal with *random* graphs.

Given a realization of the random graph, a *combination matrix* is obtained by applying a certain *combination rule or policy* to this graph. The combination policy defines how the weights $\boldsymbol{a}_{ij}$ are assigned given the particular graph structure. Formally, we have that:

$$\boldsymbol{A} = \pi(\boldsymbol{G}), \tag{6}$$

where $\pi : \{0,1\}^{N\times N} \to \mathbb{R}^{N\times N}$ is a deterministic policy and the randomness of $\boldsymbol{A}$ arises from the randomness of graph $\boldsymbol{G}$.

In summary, the system in (3) contains three sources of randomness, namely, the combination matrix $\boldsymbol{A}$, the initial state $\boldsymbol{y}_0$, and the input source $\{\boldsymbol{x}_n\}$. Let us now detail how these sources of randomness interact. Once a realization $\boldsymbol{A} = A$ is given, the stochastic dynamical system in (3) evolves according to the randomness of the initial state and the input source. In the statistical physics jargon, matrix $\boldsymbol{A}$ is a *quenched* variable: once realized, it is *frozen* and process $\boldsymbol{y}_n$ evolves over the same matrix for all n.

Conditionally on $\boldsymbol{A}$, vector $\boldsymbol{y}_0$ is assumed to have finite-variance entries, possibly mutually dependent and dependent on the particular matrix realization. The random variables $\boldsymbol{x}_i(n)$ are independent of $\boldsymbol{y}_0$ and of the matrix realization. They have zero mean and unit variance, and are independent and identically distributed (i.i.d.), both spatially (i.e., w.r.t. to index i) and temporally (i.e., w.r.t. to index n).

The particular distribution of $\boldsymbol{y}_0$ and $\boldsymbol{x}_n$ is immaterial to the results stated in Theorems 1 and 2 further ahead. In comparison, Theorem 3 is proved under the assumption of Gaussian $\boldsymbol{x}_n$, and initial state $\boldsymbol{y}_0$ distributed (conditionally on $\boldsymbol{A}$) according to the stationary distribution of the VAR model.[4] These two assumptions are commonly adopted in the sample-complexity analysis of graph learning models, where the available results exploit the concentration properties of the covariance matrices R_0 and R_1 under stationary Gaussian VAR models [29], [30], [31], [32].

V. CONSISTENT GRAPH LEARNING

Let us consider a stream of n consecutive observations taken over the probed subset of nodes $\mathcal{S}$, and collected into the $S \times n$ matrix $\boldsymbol{Y}_n$ whose (ℓ, i) entry is, for $\ell \in \mathcal{S}$ and $i = 1, 2, \ldots, n$:

$$[\boldsymbol{Y}_n]_{\ell i} = \boldsymbol{y}_\ell(i). \tag{7}$$

A matrix estimator will be formally defined as some measurable function of the data $\widehat{\boldsymbol{A}}_{\mathcal{S},n} = f_n(\boldsymbol{Y}_n)$, namely,

$$f_n : \mathbb{R}^{S\times n} \to \mathbb{R}^{S\times S}. \tag{8}$$

We focus on the class of asymptotically stable estimators that converge as the number of samples increases. In particular, we consider the class of estimators that, for any realization of the combination matrix $\boldsymbol{A}$, guarantee the following convergence in probability:

$$\lim_{n\to\infty} \mathbb{P}[\|\widehat{\boldsymbol{A}}_{\mathcal{S},n} - h\|_{\max} > \epsilon | \boldsymbol{A} = A] = 0. \tag{9}$$

We remark that the limiting estimator in (9), h, is a deterministic quantity, given $\boldsymbol{A} = A$. However, this limit will be in general different for the $2^{N(N-1)/2}$ possible realizations of the random graph, i.e., we should write $h = h(A)$ in (9). In the following, we will use the following notation

$$\widehat{\boldsymbol{A}}_{\mathcal{S}} = h(\boldsymbol{A}), \tag{10}$$

where we suppressed the sample subscript n to denote the *limiting* matrix estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}}$. The terminology "limiting matrix estimator" will be generally used in our *achievability* analysis, i.e., when we disregard sample complexity and let $n = \infty$. Using (10) in (9) we have:

$$\boxed{\lim_{n\to\infty} \mathbb{P}[\|\widehat{\boldsymbol{A}}_{\mathcal{S},n} - \widehat{\boldsymbol{A}}_{\mathcal{S}}\|_{\max} > \epsilon] = 0} \tag{11}$$

Owing to dependence on $\boldsymbol{A}$, the limiting matrix estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ is still random, and its randomness is determined solely by the randomness of the underlying graph. Our main goal is to establish that, in the commonly adopted *doubly-asymptotic* framework [21], [63] where the network size N becomes large and the number of samples n increases with N, it is possible to retrieve consistently (i.e., with probability converging to

[4] The stationary distribution will depend on the particular realization A.

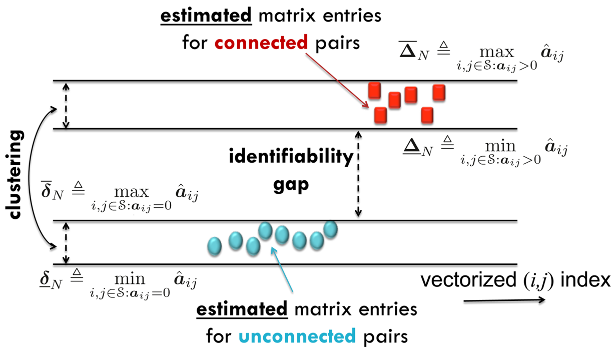
estimated matrix entries
for connected pairs
$\overline{\boldsymbol{\Delta}}_N \triangleq \max_{i,j\in\mathcal{S}:\boldsymbol{a}_{ij}>0} \hat{\boldsymbol{a}}_{ij}$
$\underline{\boldsymbol{\Delta}}_N \triangleq \min_{i,j\in\mathcal{S}:\boldsymbol{a}_{ij}>0} \hat{\boldsymbol{a}}_{ij}$
identifiability
gap
clustering
$\overline{\boldsymbol{\delta}}_N \triangleq \max_{i,j\in\mathcal{S}:\boldsymbol{a}_{ij}=0} \hat{\boldsymbol{a}}_{ij}$
$\underline{\boldsymbol{\delta}}_N \triangleq \min_{i,j\in\mathcal{S}:\boldsymbol{a}_{ij}=0} \hat{\boldsymbol{a}}_{ij}$
vectorized (i,j) index
estimated matrix entries
for unconnected pairs

FIGURE 2. **Identifiability gap.**

1) the true graph of the probed subset of nodes. In order to achieve this goal, in the next section we start by examining the asymptotic properties of the limiting estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ *as the network size N goes to infinity*.

A. ANALYSIS OF THE LIMITING ESTIMATOR AS N → ∞

In order to ascertain whether or not it is possible to discriminate interacting (i.e., connected) from non-interacting (i.e., unconnected) nodes, via observation of their output measurements, we now introduce the concept of margins and identifiability gap.

Definition 2 (Margins): Let $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ be a limiting matrix estimator for $\boldsymbol{A}_{\mathcal{S}}$. The lower and upper margins corresponding to the *unconnected* pairs are defined as, respectively:[5]

$$\underline{\boldsymbol{\delta}}_N \triangleq \min_{\substack{i,j\in\mathcal{S}:\boldsymbol{a}_{ij}=0\\ i\neq j}} \widehat{\boldsymbol{a}}_{ij}, \quad \overline{\boldsymbol{\delta}}_N \triangleq \max_{\substack{i,j\in\mathcal{S}:\boldsymbol{a}_{ij}=0\\ i\neq j}} \widehat{\boldsymbol{a}}_{ij}. \tag{12}$$

Likewise, the lower and upper margins corresponding to the *connected* pairs are defined as, respectively:

$$\underline{\boldsymbol{\Delta}}_N \triangleq \min_{\substack{i,j\in\mathcal{S}:\boldsymbol{a}_{ij}>0\\ i\neq j}} \widehat{\boldsymbol{a}}_{ij}, \quad \overline{\boldsymbol{\Delta}}_N \triangleq \max_{\substack{i,j\in\mathcal{S}:\boldsymbol{a}_{ij}>0\\ i\neq j}} \widehat{\boldsymbol{a}}_{ij}. \tag{13}$$

□

The aforementioned margins are useful to examine the achievability of structural consistency for a limiting estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ — see Fig. 2 for an illustration — and lead to the concept of *identifiability gap*.

Definition 3 (Universal local structural consistency): Let $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ be a limiting matrix estimator for $\boldsymbol{A}_{\mathcal{S}}$. If there exist a positive sequence s_N, a real value η, and a positive value Γ, such that, for any $\epsilon > 0$:

$$\lim_{N\to\infty} \mathbb{P}[|s_N\, \underline{\boldsymbol{\delta}}_N - \eta| < \epsilon] = 1,$$

$$\lim_{N\to\infty} \mathbb{P}[|s_N\, \overline{\boldsymbol{\delta}}_N - \eta| < \epsilon] = 1,$$

$$\lim_{N\to\infty} \mathbb{P}[|s_N\, \underline{\boldsymbol{\Delta}}_N - (\eta+\Gamma)| < \epsilon] = 1,$$

$$\lim_{N\to\infty} \mathbb{P}[|s_N\, \overline{\boldsymbol{\Delta}}_N - (\eta+\Gamma)| < \epsilon] = 1, \tag{14}$$

we say that $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ achieves universal local structural consistency, with scaling sequence s_N, bias η, and identifiability gap Γ. □

Remark 1 (Locality): We use the qualification "local" to emphasize that the structure of the subnetwork $\mathcal{S}$ must be inferred from observations gathered *locally* in $\mathcal{S}$, even if the nodes of $\mathcal{S}$ undergo the influence of many other nodes belonging to the larger embedding network. □

Remark 2 (Bias): For the *true* combination matrix, the entries corresponding to unconnected pairs are zero. In contrast, (14) reveals that the scaled entries for unconnected pairs can be close to η, which results therefore in a *bias*. However, and remarkably, *this bias does not constitute a problem for consistent classification* of connected/unconnected node pairs, i.e., the bias does not affect in any manner identifiability. □

Remark 3 (Identifiability gap): Since we can write:

$$|s_N\, \overline{\boldsymbol{\delta}}_N - \eta| < \epsilon \Rightarrow s_N\, \overline{\boldsymbol{\delta}}_N < \eta + \epsilon, \tag{15}$$

and

$$|s_N\, \underline{\boldsymbol{\Delta}}_N - (\eta+\Gamma)| < \epsilon \Rightarrow s_N\, \underline{\boldsymbol{\Delta}}_N > \eta + \Gamma - \epsilon, \tag{16}$$

from the second and third formulas in (14) we conclude that, with high probability as N gets large:[6]

$$s_N\, \overline{\boldsymbol{\delta}}_N < \eta + \epsilon, \qquad s_N\, \underline{\boldsymbol{\Delta}}_N > \eta + \Gamma - \epsilon. \tag{18}$$

The first inequality in (18) means that the *minimum* entry of $s_N\widehat{\boldsymbol{A}}_{\mathcal{S}}$ taken over the *connected* pairs essentially stays above the value $\eta + \Gamma > \eta$. Likewise, the second inequality means that the *maximum* entry of $s_N\widehat{\boldsymbol{A}}_{\mathcal{S}}$ taken over the *unconnected* pairs essentially does not exceed η. Combining these two relationships, we conclude that the entries of the (limiting) estimated matrix corresponding to connected node pairs stand clearly separated from the entries corresponding to unconnected node pairs. The minimum amount of separation is quantified by the gap, Γ. □

The notion of structural consistency implies the existence of a threshold, comprised between η and $\eta + \Gamma$, which correctly separates (in the limit of large N) the entries of the matrix estimator, in such a way that the entries corresponding to connected pairs lie above the threshold, whereas the entries corresponding to unconnected pairs lie below the threshold.

However, an accurate determination of the separation threshold requires some prior knowledge of the monitored system. For instance, to set a detection threshold one needs to

[5] The definitions in (12) and (13) are void if the nodes in $\mathcal{S}$ are all connected or all unconnected, respectively. For these singular cases, we can formally assign arbitrary values to the margins. We will see later that, under the Erdős-Rényi model, these events are irrelevant as $N \to \infty$.

[6] Actually, the existence of a gap would be guaranteed by a weaker notion of consistency, namely, by:

$$\lim_{N\to\infty} \mathbb{P}[s_N\, \overline{\boldsymbol{\delta}}_N < \eta + \epsilon] = 1,$$

$$\lim_{N\to\infty} \mathbb{P}[s_N\, \underline{\boldsymbol{\Delta}}_N > \eta + \Gamma - \epsilon] = 1. \tag{17}$$

However, this notion would not be sufficient to guarantee the important clustering property that we discuss in Remark 4.

know the scaling sequence s_N. In the problems dealt with in this work we will see that $s_N = Np_N$, where Np_N represents the average number of neighbors in the network, and in several practical applications this number is unknown.

As a result, the threshold setting might be a critical issue, and it would be highly desirable to have a *universal* (i.e., blind and nonparametric) method to set the threshold. For example, it would be highly desirable to determine a separation threshold using machine learning tools such as a standard clustering algorithm (e.g., a k-means clustering with $k = 2$). As discussed in the next remark, this possibility is automatically enabled by the *clustering* property embodied in Definition 3.

Remark 4 (Clustering): According to the notion of *universal* structural consistency, the pair of (scaled) margins over the unconnected pairs, $s_N\,\underline{\boldsymbol{\delta}}_N$ and $s_N\,\overline{\boldsymbol{\delta}}_N$, converge to one and the same value, η, which implies that *all the entries of $s_N\widehat{\boldsymbol{A}}_{\mathcal{S}}$ corresponding to the unconnected pairs* are sandwiched between these margins — see Fig. 2. A similar behavior is observed for the scaled entries over the connected pairs, which converge altogether to $\eta + \Gamma$ since they are sandwiched between $s_N\,\underline{\boldsymbol{\Delta}}_N$ and $s_N\,\overline{\boldsymbol{\Delta}}_N$. In summary, we conclude that the connected and unconnected node pairs *cluster into well-separated classes* that can be identified, e.g., by means of a universal clustering algorithm. □

B. A CONSISTENT CLUSTERING ALGORITHM

The definition of universal local structural consistency implies that the unconnected node pairs cluster around η, whereas the connected node pairs cluster around the higher value $\eta + \Gamma$. Accordingly, *for sufficiently large N*, there is no doubt that any reasonable clustering algorithm will be able to identify properly these two clusters. For example, an asymptotically correct guess of the true clusters (i.e., of the true graph) can be obtained by simply choosing as a threshold the midpoint between the maximum and minimum matrix entries.

The simplest classification rule based on such threshold does not try to cluster the data, but it works *asymptotically* since, as $N \to \infty$, the scaled matrix entries converge to two possible values, η or $\eta + \Gamma$. For finite network and/or sample sizes, the scaled matrix entries exhibit a certain variability around these values, and it would be more appropriate to employ a clustering algorithm, like the popular k-means algorithm (in our case, we know that $k = 2$).

However, the k-means algorithm has a well-known drawback in the case of unbalanced clusters. One example of unbalanced clusters is shown in Fig. 3. We see that the k-means algorithm (top panel) tends to be highly biased by the largest cluster, resulting in a wrong configuration. Here the two centroids estimated by the k-means algorithm are both located within the largest ensemble (circles), leading to a wrong classification. Since in our model it is actually permitted that, for large N, one cluster can dominate the other one (for instance, when $p_N \to 0$, the cluster of unconnected node pairs becomes predominant), we are not guaranteed that the k-means algorithm works properly as $N \to \infty$. In summary, the k-means algorithm with $k = 2$ can mitigate finite-size issues, but it does not provide asymptotic guarantees. In order to solve this issue, we propose a simple modification of the k-means algorithm, detailed in the pseudo-code shown in Algorithm 1.

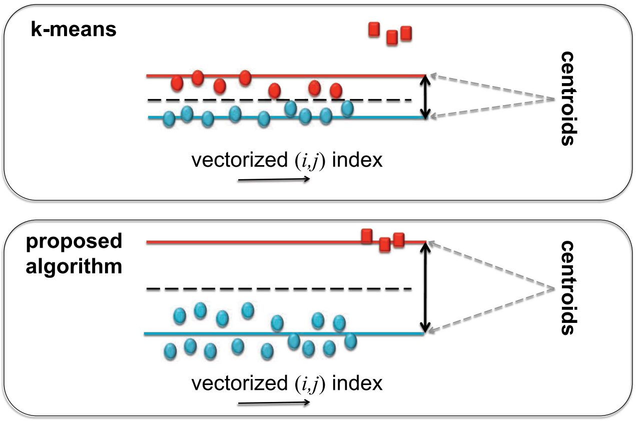

FIGURE 3. Visual comparison between the *k*-means algorithm and the clustering algorithm proposed in this work, for the case of unbalanced clusters. The true clusters are identified by different symbols (circle vs. square). The clusters produced by the algorithms are identified by different colors (blue vs. red).

Algorithm 1: $j^\star = \mathsf{clu}(v)$.

```
% v is an L × 1 vector with entries sorted in ascending order
% initialize the ensemble of admissible configurations A
$\mathcal{A} = \emptyset$;
for $j = 1 : L-1$ do
    % set 2 tentative clusters
    $\mathcal{C}_0(j) = \{v_1, v_2, \ldots, v_j\}$;
    $\mathcal{C}_1(j) = \{v_{j+1}, v_{j+2}, \ldots, v_L\}$;
    % compute the centroids of the 2 clusters
    $c_0(j) = \frac{1}{j}\sum_{i=1}^{j} v_i, \quad c_1(j) = \frac{1}{L-j}\sum_{i=j+1}^{L} v_i$;
    % check if the current configuration j is admissible, i.e., if the
    midpoint between the centroids separates the clusters
    if $v_j < \frac{c_0(j)+c_1(j)}{2} < v_{j+1}$ then
        $\mathcal{A} = \mathcal{A} \cup \{j\}$;
    end
end
% select the admissible configuration j with largest centroid distance
$j^\star = \operatorname*{argmax}_{j\in\mathcal{A}} [c_1(j) - c_0(j)]$;
```

Let v be the $L \times 1$ vector to be clustered, with entries that have been arranged in ascending order. The k-means algorithm, with $k = 2$, attempts to minimize the following cost function:

$$\sum_{v_j \in \mathcal{C}_0} (v_j - c_0)^2 + \sum_{v_j \in \mathcal{C}_1} (v_j - c_1)^2, \qquad (19)$$

over all possible clusters $\mathcal{C}_0$ and $\mathcal{C}_1$, with c_0 and c_1 being the cluster centroids, defined as:

$$c_0 = \frac{1}{|\mathcal{C}_0|}\sum_{v_j \in \mathcal{C}_0} v_j, \qquad c_1 = \frac{1}{|\mathcal{C}_1|}\sum_{v_j \in \mathcal{C}_1} v_j. \qquad (20)$$

It is useful to recall that the minimum of the cost function in (19) must fulfill the following necessary condition: the midpoint between the two centroids is a threshold that separates the two clusters [91]. In our one-dimensional case, with $k = 2$, this property implies that it suffices to consider only the cluster configurations $\mathcal{C}_0(j) = \{1, 2, \ldots, j\}$ and $\mathcal{C}_1(j) = \{j+1, j+2, \ldots, L\}$, for $j \in \{1, 2, \ldots, L-1\}$. Obviously, if two points, say v_i and v_{i+1}, are coincident, considering their possible permutations is pointless. Accordingly, we see that any possible partition is identified by an index j.

Now, for large N the true clusters of connected and unconnected pairs become almost perfectly localized, and, hence, two centroids belonging to the true clusters match the necessary condition — see the bottom panel in Fig. 3. However, as we have observed when examining Fig. 3, if the sizes of the true clusters are very different, the k-means algorithm can be in error. For this reason, we now introduce a simple modification of the k-means algorithm that overcomes this issue by taking into account also the distance between the centroids.

First, the algorithm enumerates all admissible cluster pairs through an index j spanning the set $\{1, 2, \ldots, L-1\}$. The set of indices fulfilling the necessary condition of k-means are collected in the set $\mathcal{A} = \{j_1, j_2, \ldots\}$. At this stage, the classic k-means would simply select, among these admissible configurations, the one ensuring the minimum cost. We modify this rule by selecting the index[7] $j^\star \in \mathcal{A}$ that maximizes the distance between the clusters' centroids, namely, $j^\star = \mathrm{argmax}_{j\in\mathcal{A}}[c_1(j) - c_0(j)]$, with $c_0(j)$ and $c_1(j)$ being the centroids corresponding to the clusters identified by index j. With this modified rule, we want to *i*) retain the good behavior exhibited by k-means in typical situations; and *ii*) guarantee that the algorithm achieves consistent graph learning, as we will formally establish in the next theorem.

Before stating the theorem, let us introduce the input-output relationship that relates the estimated combination matrix to the estimated graph through the clustering algorithm. The procedure is as follows. Once an estimated matrix M is computed, it is vectorized and sorted in ascending order, before feeding the clustering algorithm described in the pseudo-code reported in this page. More formally, given a matrix M with index set $\mathcal{S}$, let $\mathsf{diag}(M)$ be the diagonal matrix with same diagonal entries as M, such that $M_0 = M - \mathsf{diag}(M)$ is the matrix M with diagonal entries set to zero.

Let

$$u = \mathsf{vec}(M_0) : \mathbb{R}^{S\times S} \to \mathbb{R}^{S(S-1)} \tag{21}$$

be a one-to-one[8] mapping that transforms the *off-diagonal* entries of the $S \times S$ matrix M_0 into an $S(S-1) \times 1$ vector u. Let

$$v = \Pi\, u, \tag{22}$$

with Π being the permutation matrix that sorts the entries of u in ascending order. The vector v is given as input to the clustering algorithm $\mathsf{clu}(v)$, obtaining the clusters:

$$\mathcal{C}_0^\star = \mathcal{C}_0(j^\star), \qquad \mathcal{C}_1^\star = \mathcal{C}_1(j^\star), \tag{23}$$

as described in the pseudo-code of the algorithm. We use the reverse permutation Π^{-1} to cluster the entries of u from the corresponding clustering on the entries of v. Then, using the inverse mapping vec^{-1}, we decide which cluster a particular entry m_{ij} belongs to, and accordingly estimate an adjacency matrix $\widehat{G}$, whose main diagonal is set conventionally to 0, and whose off-diagonal entries, for all $i, j \in \mathcal{S}$ with $i \neq j$, are set as:

$$\widehat{g}_{ij} = \mathbb{I}[m_{ij} \in \mathcal{C}_1^\star], \tag{24}$$

where $\mathbb{I}[\mathcal{E}]$ denotes the indicator function, which is equal to 1 if condition $\mathcal{E}$ is true, and is equal to 0 otherwise. The overall mapping that leads from M to $\widehat{G}$, passing through the algorithm clu, will be compactly denoted by $\mathsf{graphclu}$.

Theorem 1 (Sample consistency of the proposed clustering algorithm): Let $\widehat{\boldsymbol{A}}_{\mathcal{S},n}$ be a stable matrix estimator belonging to class (11), with the limiting matrix estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ achieving universal local structural consistency according to Definition 3. Let $\boldsymbol{G}_{\mathcal{S}}$ be the (random) support graph associated to $\boldsymbol{A}_{\mathcal{S}}$ and let

$$\widehat{\boldsymbol{G}}_{\mathcal{S},n} = \mathsf{graphclu}(\widehat{\boldsymbol{A}}_{\mathcal{S},n}) \tag{25}$$

be the subgraph estimated by the proposed clustering algorithm. If the probability that $\boldsymbol{G}_{\mathcal{S}}$ is fully connected or fully disconnected vanishes as $N \to \infty$, a certain scaling law $n(N)$ exists such that:

$$\boxed{\lim_{N\to\infty} \mathbb{P}\left[\widehat{\boldsymbol{G}}_{\mathcal{S},n(N)} = \boldsymbol{G}_{\mathcal{S}}\right] = 1} \tag{26}$$

Proof: See Appendix H. ■

VI. GRANGER ESTIMATOR

Preliminarily, it is useful to examine the steady-state covariance matrix corresponding to the dynamics in (3).

According to our generative model, given a certain realization of the combination matrix $\boldsymbol{A} = A$, we let the system evolve according to (3). Exploiting (3) we see that:

$$\mathbb{E}\left[\boldsymbol{y}_n\boldsymbol{y}_n^\top \middle| \boldsymbol{A} = A\right] = A^n\mathbb{E}\left[\boldsymbol{y}_0\boldsymbol{y}_0^\top\right](A^n)^T + \sigma^2\sum_{i=0}^{n-1} A^i(A^i)^T. \tag{27}$$

Using now the stability and symmetry of A along with (27) we get the following limiting covariance (convergence of the series is guaranteed by stability):

$$\begin{aligned} R_0 = R_0(A) &= \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{y}_n\boldsymbol{y}_n^\top | \boldsymbol{A} = A\right] \\ &= \sigma^2\sum_{i=0}^{\infty} A^{2i} = \sigma^2(I - A^2)^{-1}, \end{aligned} \tag{28}$$

where I is the $N \times N$ identity matrix. The bold notation:

$$\boldsymbol{R}_0 = R_0(\boldsymbol{A}) \tag{29}$$

[7] In principle we could have multiple maximizers, but we will see later in Lemma 8 and Theorem 1 that in our case the maximizer $j^\star$ is unique with high probability.

[8] The mapping is invertible since it operates on matrices with null diagonal.

will be finally used to account for the randomness in $\boldsymbol{A}$ coming from the underlying Erdős-Rényi graph.

Likewise, we introduce the steady-state one-lag covariance matrix:

$$R_1 = R_1(A) = \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{y}_n\boldsymbol{y}_{n-1}^\top | \boldsymbol{A} = A\right], \boldsymbol{R}_1 = R_1(\boldsymbol{A}), \quad (30)$$

which exploiting the dynamics in (3) can be written as:

$$\boldsymbol{R}_1 = \boldsymbol{A}\boldsymbol{R}_0. \quad (31)$$

From (31) we obtain the following well-known relationship:

$$\boldsymbol{A} = \boldsymbol{R}_1\boldsymbol{R}_0^{-1}, \quad (32)$$

a quantity that is also referred to as the best one-step predictor or Granger estimator [52], [61]. Under the partial observability setting, it is tempting to adapt the structure in (32) by considering only the observable subnet $\mathcal{S}$ [19], [20]:

$$\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\mathrm{Gra})} = [\boldsymbol{R}_1]_{\mathcal{S}}([\boldsymbol{R}_0]_{\mathcal{S}})^{-1}, \quad (33)$$

which obviously does not allow us to incorporate well the contribution of latent nodes. In [19] it is shown that such limiting matrix estimator admits the following representation:

$$\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\mathrm{Gra})} = \boldsymbol{A}_{\mathcal{S}} + \boldsymbol{E}^{(\mathrm{Gra})}, \quad (34)$$

where (we recall that $\mathcal{S}'$ denotes the subset of unobserved nodes):

$$\boldsymbol{E}^{(\mathrm{Gra})} = \boldsymbol{A}_{\mathcal{S}\mathcal{S}'}\boldsymbol{H}[\boldsymbol{A}^2]_{\mathcal{S}'\mathcal{S}}, \quad (35)$$

with:

$$\boldsymbol{H} = (I_{\mathcal{S}'} - \boldsymbol{C})^{-1}, \qquad \boldsymbol{C} \triangleq [\boldsymbol{A}^2]_{\mathcal{S}'}. \quad (36)$$

For later use, it is also useful to rewrite (35) on an entrywise basis, for all $i, j \in \mathcal{S}$:

$$\boxed{\boldsymbol{e}_{ij}^{(\mathrm{Gra})} = \sum_{\ell,m\in\mathcal{S}'} \boldsymbol{a}_{i\ell}\boldsymbol{h}_{\ell m}\boldsymbol{a}_{mj}^{(2)}} \quad (37)$$

where the symbol $\boldsymbol{a}_{ij}^{(k)}$ denotes the (i, j) entry of the k-th matrix power $\boldsymbol{A}^k$.

VII. MAIN RESULT

In this section, we illustrate the characterization of consistency and sample complexity regarding the Granger estimator. We start by introducing the assumptions on the class of random graphs and combination matrices used to prove the results.

A. ASSUMPTIONS ON THE GRAPH

In this article we address the useful case where the network graph is generated according to the Erdős-Rényi *random graph* model, namely, an undirected graph whose edges are drawn, one independently from the other, through a sequence of Bernoulli experiments with identical probability of success (i.e., of connection) [88], [89]. In particular, the notation $\mathscr{G}(N, p_N)$ will represent an Erdős-Rényi graph over N nodes, and with connection probability p_N. Accordingly, the variables $\boldsymbol{g}_{ij}$, for $i = 1, 2, \ldots, N$ and $j > i$, are independent Bernoulli random variables with $\mathbb{P}[\boldsymbol{g}_{ij} = 1] = p_N$, and the matrix $\boldsymbol{G}$ is symmetric. As it will be clear soon, the explicit dependence of the connection probability upon N will be critical to examine the evolution of random graphs in the thermodynamic limit of large N.

As one fundamental graph descriptor, in this work we use the *degree* of a node. The degree of node i is defined as:

$$\boldsymbol{d}_i = 1 + \sum_{\ell\neq i} \boldsymbol{g}_{i\ell}, \quad (38)$$

namely, the cardinality of the i-th node neighborhood (including i itself). In particular, we shall use the minimal and maximal degrees that are defined as, respectively:

$$\boldsymbol{d}_{\min} \triangleq \min_{i=1,2,\ldots N} \boldsymbol{d}_i, \qquad \boldsymbol{d}_{\max} \triangleq \max_{i=1,2,\ldots N} \boldsymbol{d}_i. \quad (39)$$

One meaningful (and classic) way to characterize the behavior of random graphs is to examine their thermodynamic limit as the network size goes to infinity. Such an asymptotic characterization is useful because it captures average behavior that emerges with high probability over large networks.

In examining the thermodynamic behavior of random graphs, the connection probability p_N is generally allowed to scale with N. This degree of freedom allows representing different types of asymptotic graph behavior. For example, recalling that the average number of neighbors over an Erdős-Rényi graph scales as Np_N, different graph evolutions can be obtained with different choices of p_N. For example, a constant p_N will let the number of neighbors grow linearly with N. In comparison, a p_N scaling as $(\log N)/N$ would correspond to a number of neighbors growing logarithmically with N. In summary, different limiting regimes are determined by the way the connection probability evolves with N. It is useful for our purposes to list briefly the main regimes that are of interest for the forthcoming treatment.

— *Connected regime:* In this work we focus on the regime where the graph is connected with high probability. This regime prescribes that the pairwise connection probability scales as [88], [89]:

$$p_N = \frac{\log N + c_N}{N}, \qquad c_N \overset{N\to\infty}{\longrightarrow} \infty. \quad (40)$$

— *Sparse (connected) regime:* The connected regime can be obtained also when the pairwise connection probability, p_N, vanishes as N gets large. In particular, we shall refer to this scenario as the *sparse* connected regime:

$$p_N \overset{N\to\infty}{\longrightarrow} 0 \quad \text{under (40)} \quad \text{[Sparse connected regime]}. \quad (41)$$

— *Dense regime:* We call *dense* the regime where the pairwise connection probability converges to a nonzero quantity, namely $p_N \to p > 0$.

The aforementioned taxonomy basically focuses on the concepts of connectedness and sparsity. These concepts have been advocated in previous works related to topology inference under partial observability, and, in particular, some

TABLE 1. Useful Taxonomy to Illustrate the Relationships Between Concentration and Sparsity Over a Connected Erdős-Rényi Graph. The sequence ω_N goes to Infinity as $N \to \infty$

Connection probability	Concentration	Sparsity
$p_N = \omega_N \dfrac{\log N}{N} \to p > 0$	Uniform	Dense
$p_N = \omega_N \dfrac{\log N}{N} \to 0$	Uniform	Sparse

useful structural consistency results have been proved under the sparse (connected) regime.

One essential element of novelty in our analysis is exploiting a different feature, namely, the *concentration* of graph degrees. We wish to avoid confusion here: the term "concentration" does *not* refer to the number of node connections. Instead, the concept of concentration is borrowed from a common terminology in statistics, which is used to refer to statistical quantities that concentrate around some deterministic value as $N \to \infty$ [90]. In particular, we will focus our attention on the *uniform concentration properties of the minimal and maximal degrees of random graphs*.

— *Uniform concentration regime:* The uniform concentration regime is enabled by choosing the following pairwise connection probability:

$$p_N = \omega_N \frac{\log N}{N} \overset{N\to\infty}{\longrightarrow} p, \qquad \omega_N \overset{N\to\infty}{\longrightarrow} \infty, \tag{42}$$

which is tantamount to assuming that (40) holds true with the sequence c_N growing faster than $\log N$. Under this regime, the minimal and the maximal degrees of the graph both concentrate asymptotically around the expected degree $1 + (N-1)p_N \sim N p_N$, in the following precise sense:

$$\boxed{\frac{\boldsymbol{d}_{\min}}{N p_N} \overset{\mathrm{p}}{\longrightarrow} 1, \; \frac{\boldsymbol{d}_{\max}}{N p_N} \overset{\mathrm{p}}{\longrightarrow} 1 \;\; \text{[Uniform concentration]}} \tag{43}$$

The physical meaning of (43) is that both the minimal and the maximal degrees scale, asymptotically with N, as the expected degree. Indeed, (43) can be restated as: $\boldsymbol{d}_{\min} \sim N p_N + g_N$ and $\boldsymbol{d}_{\max} \sim N p_N + g'_N$, where g_N and g'_N are sequences that are asymptotically dominated by $N p_N$.

Table 1 summarizes the sparsity/concentration taxonomy arising from the previous arguments. We are now ready to extract from the above taxonomy the elements that are relevant to the forthcoming treatment.

1) Comparing (42) against (40), we see that the regime of concentration does *not* include all classes of connected Erdős-Rényi graphs. In fact, while in (40) c_N is any arbitrary divergent sequence (e.g., we can have $c_N = \log\log N$), according to (42) the sequence c_N should grow with N more than logarithmically. The regime where the graph is connected, whereas (42) is not fulfilled, will be referred to as the *very sparse* regime.

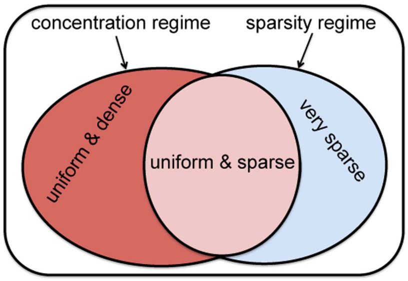

FIGURE 4. Venn diagram illustrating the relationships between concentration and sparsity over a connected Erdős-Rényi graph.

2) According to (42), the regime of concentration can be either sparse or dense. In particular, the regime is dense when $p > 0$, and is sparse when $p = 0$.

The aforementioned categorizations are illustrated in Fig. 4 by means of a Venn diagram.

B. ASSUMPTIONS ON THE COMBINATION MATRIX

The forthcoming theorems will be proved under the assumption that the combination matrix belongs to the following class.

Assumption 1 (Regular diffusion matrices): The combination matrix $\boldsymbol{A}$ is symmetric and satisfies the conditions:

$$\boxed{\sum_{\ell=1}^{N} \boldsymbol{a}_{i\ell} = \rho, \quad \frac{\kappa}{\boldsymbol{d}_{\max}} \boldsymbol{g}_{ij} \le \boldsymbol{a}_{ij} \le \frac{\kappa}{\boldsymbol{d}_{\min}} \boldsymbol{g}_{ij} \quad \forall i \neq j} \tag{44}$$

for some parameters ρ and κ, with $0 < \kappa \le \rho < 1$. □

We remark that the most common combination matrices encountered in the literature automatically satisfy Assumption 1. Some popular choices are the Laplacian and the Metropolis rules reported below, which arise naturally in many applications, for instance, they are one fundamental ingredient of *adaptive* networks [35]. The matrix entries corresponding to these combination rules are defined as follows. For $i \neq j$, $0 < \rho < 1$, and $0 < \lambda \le 1$:

$$\boldsymbol{a}_{ij} = \rho\lambda \frac{\boldsymbol{g}_{ij}}{\boldsymbol{d}_{\max}}, \qquad \text{[Laplacian rule]} \tag{45}$$

$$\boldsymbol{a}_{ij} = \rho \frac{\boldsymbol{g}_{ij}}{\max\left\{\boldsymbol{d}_i, \boldsymbol{d}_j\right\}}, \qquad \text{[Metropolis rule]} \tag{46}$$

whereas the self-weights are determined by the leftmost condition in (44), yielding $\boldsymbol{a}_{ii} = \rho - \sum_{\ell \neq i} \boldsymbol{a}_{i\ell}$. It is immediate to verify that the Laplacian rule yields a regular diffusion matrix with $\kappa = \rho\lambda$, whereas the Metropolis rule yields a regular diffusion matrix with $\kappa = \rho$.

C. GRANGER ESTIMATOR: UNIVERSAL LOCAL STRUCTURAL CONSISTENCY

The next theorem establishes the fundamental consistency properties of the Granger estimator presented in Section VI.

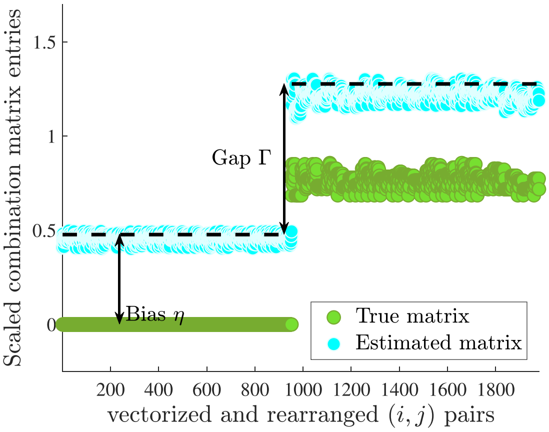

FIGURE 5. Graphical illustration of Theorem 2. In the plot, the entries of the ***true*** matrix $A_{\mathcal{S}}$ are vectorized following column-major ordering, and the (vectorized) (i, j) pairs are rearranged in such a way that the zero entries appear before the nonzero entries. The same ordering used for the true matrix is applied to the entries of the ***estimated*** matrix, $\widehat{A}_{\mathcal{S}}$. All matrices are scaled by Np_N. Broken lines display the theoretical values, computed by using (48), of the bias η and of the quantity $\eta + \Gamma$, where Γ is the identifiability gap.

Theorem 2 (Universal local structural consistency of the Granger estimator): Let $\boldsymbol{A}$ be a regular diffusion matrix with parameters ρ and κ, with the network graph drawn according to an Erdős-Rényi random graph model $\mathscr{G}(N, p_N)$ where the fraction of observable nodes, S/N, converges to some nonzero value ξ. Then, under the uniform concentration regime where:

$$p_N = \omega_N \frac{\log N}{N} \to p, \quad \text{with } \omega_N \to \infty, \tag{47}$$

the Granger estimator achieves universal local structural consistency as detailed in Definition 3, with scaling sequence $s_N = Np_N$, and with bias η and identifiability gap Γ given by:

$$\boxed{\eta = \kappa^2 p \frac{(2\rho - \kappa)(1 - \xi)}{1 - (\rho^2 - 2\rho\kappa\xi + \kappa^2\xi)}, \qquad \Gamma = \kappa} \tag{48}$$

Proof: The proof of the theorem is provided in Appendix D, and relies on a number of auxiliary lemmas and theorems reported in the appendices. In particular, the core of the proof is the following. First, Theorem 6 in Appendix C constructs *uniform* (w.r.t. N) bounds on the entries of the matrix $\boldsymbol{H}$ in (36). These bounds are useful to characterize the error associated to the Granger estimator. Then, exploiting the asymptotic concentration property of the maximal and minimal degrees, it is possible to prove the convergence of the matrix series relevant for the computation of the errors in (37). These convergence properties are finally used to compute the bias and the identifiability gap in (48). ■

The main message conveyed by Theorem 2 is illustrated in Fig. 5, where we depict: *i*) the entries of the true combination matrix, vectorized and ordered as shown in the figure, and magnified by Np_N; and *ii*) the entries of the limiting estimated combination matrix, magnified by Np_N, vectorized and ordered with the same ordering used for the true combination matrix. The essential features illustrated in Section V are clearly visible in Fig. 5. Comparing the true and estimated matrices, we can appreciate the emergence of the bias and of the gap. Both these phenomena are predicted by Theorem 2, as we can see from the theoretical values η and Γ, computed by using (48) and displayed in Fig. 5 with broken lines. We observe how the (scaled) estimated matrix entries corresponding to unconnected node pairs are clustered around the theoretical value η, whereas the entries corresponding to connected pairs are clustered around the theoretical value $\eta + \Gamma$. It is also seen how the bias does not affect separability between the groups of connected and unconnected node pairs.

Remark 5 (Role of degree concentration): For the class of regular diffusion matrices in Assumption 1, concentration of the degrees induces concentration of the nonzero entries of the combination matrix. This creates an identifiability gap in the *true* matrix $\boldsymbol{A}_{\mathcal{S}}$. However, what is critical for graph recovery is the existence of an identifiability gap in the *estimated* matrix $\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\mathrm{Gra})}$, which is in fact proved in Theorem 2. Let us provide some intuition behind this result.

To this aim, we start by examining the useful representations of the limiting matrix estimator in (34). From this representation, we see that the existence of an identifiability gap in the limiting matrix estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\mathrm{Gra})}$ depends on the *true* matrix $\boldsymbol{A}_{\mathcal{S}}$, but will depend strongly also on the *error* matrix $\boldsymbol{E}^{(\mathrm{Gra})}$. Since each entry in $\boldsymbol{E}^{(\mathrm{Gra})}$ is a function of the entries in $\boldsymbol{A}$ (in general, also of the *latent* nodes belonging to the unobserved subset $\mathcal{S}'$), a key point is to understand how the $\boldsymbol{a}_{ij}$'s combine with each other to produce $\boldsymbol{E}^{(\mathrm{Gra})}$. As observed before, the $\boldsymbol{a}_{ij}$'s exhibit *concentration* (in their nonzero values). On the other hand, they exhibit also *randomness* (in the *location* of the nonzero entries, due to the random graph model). The attributes of concentration and randomness are critical to reveal the nontrivial result shown in Theorem 2. It will be seen that the $\boldsymbol{a}_{ij}$'s combine with each other so as to induce a concentration in $\boldsymbol{E}^{(\mathrm{Gra})}$, which in turn determines the emergence of an identifiability gap in $\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\mathrm{Gra})}$. In summary, the overall influence of latent nodes is quantified by an error matrix, whose entries converge to some deterministic quantity, equally for both connected and unconnected node pairs. In this way, the connections among the probed nodes stick out consistently from the error floor. In summary, in the limit of large networks the Granger estimator equals the ground-truth matrix plus a uniform shift of its entries and, hence, the network structure is preserved. □

We now illustrate the relevance of Theorem 2 by means of numerical experiments. In Fig. 6, we display the probability of correct graph learning for the limiting Granger estimator, with reference to the dense case (with *constant* connection probability $p_N = p = 0.1$ for all N) for different choices of combination policies and relative parameters. In Fig. 7, we consider instead one example of uniform-and-sparse regime,

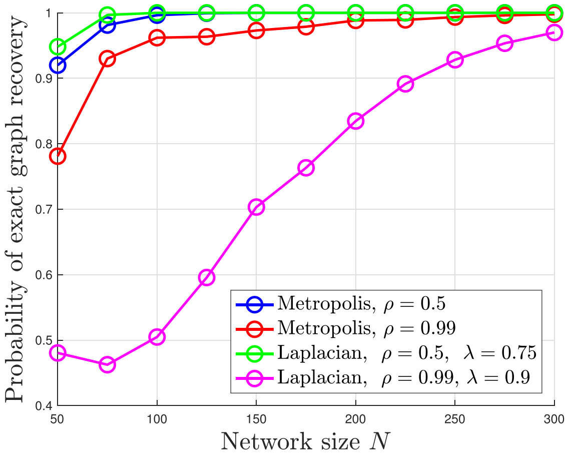

FIGURE 6. **Performance of the limiting Granger estimator as a function of N under the dense regime with connection probability $p_N = 0.1$, and fraction of probed nodes $\xi = 0.15$. In the simulations, the initial vector y_0 has all zero entries, $\sigma = 1$, and the input source samples $x_i(n)$ are i.i.d. samples from a standard Gaussian distribution. The probability of correct graph recovery is evaluated by means of 10^4 Monte Carlo runs.**

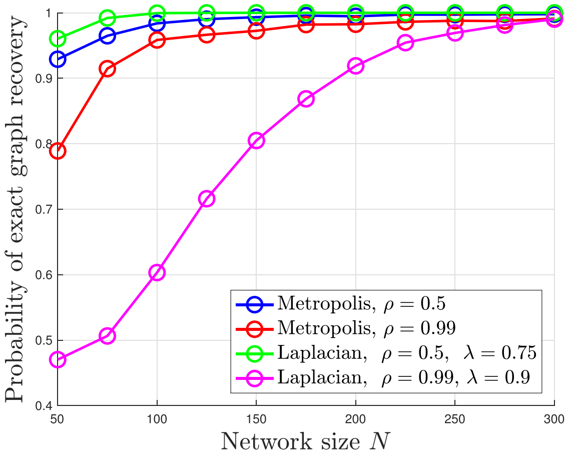

FIGURE 7. **Performance of the limiting Granger estimator as a function of N under the uniform-and-sparse regime with connection probability $p_N = (\log\log N)\frac{\log N}{N}$, and fraction of probed nodes $\xi = 0.15$. In the simulations, the initial vector y_0 has all zero entries, $\sigma = 1$, and the input source samples $x_i(n)$ are i.i.d. samples from a standard Gaussian distribution. The probability of correct graph recovery is evaluated by means of 10^4 Monte Carlo runs.**

with $\omega_N = \log\log N$, namely, with connection probability decreasing with N as

$$p_N = \log\log N \, \frac{\log N}{N}. \tag{49}$$

In both plots, we see that the simulations match well the theoretical predictions obtained from Theorem 2, since the probability of correct graph learning approaches 1 as the network size grows.

D. GRANGER ESTIMATOR: SAMPLE COMPLEXITY

In light of Theorem 1, the universal local structural consistency of the *limiting* estimators $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ implies consistency of the matrix estimators $\widehat{\boldsymbol{A}}_{\mathcal{S},n} = f_n(\boldsymbol{Y}_n)$ (i.e., of the *real* estimators based on the measured samples) as the sample size n grows with the network size N with some law $n(N)$. This section is devoted to establishing which law $n(N)$ is sufficient to achieve consistent learning for the Granger estimator.

First of all, it is necessary to introduce the following estimator:

$$\widehat{\boldsymbol{A}}^{(\mathrm{Gra})}_{\mathcal{S},n} = [\widehat{\boldsymbol{R}}_{1,n}]_{\mathcal{S}}([\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}})^{-1}, \tag{50}$$

which is nothing but the finite-sample version of (32) obtained by replacing the true covariance matrices with their empirical counterparts. The estimator in (50) is assumed to be unspecified when $[\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}}$ is singular (which must happen when $n < N$). In order to bypass the instability related to matrix inversion, we introduce also the following regularized version of the Granger estimator.

For $i \in \mathcal{S}$, the i-th row of the regularized Granger estimator $\widehat{\boldsymbol{A}}^{(\mathrm{reGra})}_{\mathcal{S},n}$ is a solution to the constrained optimization problem (here x is a row vector, and by $[M]_{i\mathcal{S}}$ we denote the i-th row of submatrix $M_{\mathcal{S}}$):

$$\min_{x\in\mathbb{R}^S} \left\| x\,[\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [\widehat{\boldsymbol{R}}_{1,n}]_{i\mathcal{S}} \right\|_\infty \quad \text{s.t. } \|x\|_1 \le 1. \tag{51}$$

The ℓ_1-constraint on the admissible row vectors x in (51) arises from the fact that the rows of our target estimator, the *limiting* Granger estimator, have ℓ_1-norm bounded by 1 — see (315) in Appendix I.

It is also useful to observe that, in view of (50), when the sample covariance matrix is invertible, the non-regularized Granger estimator is the only matrix that yields a zero residual in (51). As a result, whenever $\|\widehat{\boldsymbol{A}}^{(\mathrm{Gra})}_{\mathcal{S},n}\|_\infty \le 1$ the plain and regularized Granger estimators coincide.

Now, assuming that the empirical covariance matrices converge to the true covariance matrices as $n \to \infty$, it is easily seen that both the plain and regularized Granger estimators converge to their limiting counterparts and, hence, guarantee condition (11). Since in Theorem 2 we established that the *limiting* Granger estimator achieves universal local structural consistency, in view of Theorem 1 this property implies that estimators (50) and (51) are able to learn well the underlying subgraph of probed nodes, provided that the number of samples n grows as the network size N diverges. The goal of a sample complexity analysis is to establish how this number of samples should scale with N in order to grant correct graph learning. The forthcoming theorem provides an answer for the class of models (3) with Gaussian input source.

Theorem 3 (Sample complexity of the Granger estimator): Assume that model (3) holds with i.i.d. standard Gaussian source data $\{\boldsymbol{x}_n\}$, and with initial state $\boldsymbol{y}_0$ distributed (conditionally on $\boldsymbol{A}$) according to the stationary distribution of the VAR process in (3). Let $\boldsymbol{A}$ be a regular diffusion matrix with parameters ρ and κ, with the network graph drawn according to an Erdős-Rényi random graph model $\mathscr{G}(N, p_N)$ where the fraction of observable nodes, S/N, converges to some nonzero value ξ. Let $\widehat{\boldsymbol{A}}_{\mathcal{S},n}$ be the regularized Granger estimator in (51). Then, under the uniform concentration regime where:

$$p_N = \omega_N \frac{\log N}{N} \overset{N\to\infty}{\longrightarrow} p, \quad \text{with } \omega_N \to \infty, \qquad (52)$$

we have that:

$$\lim_{N\to\infty} \mathbb{P}[\mathsf{graphclu}(\widehat{\boldsymbol{A}}_{\mathcal{S},n(N)}) = \boldsymbol{G}_{\mathcal{S}}] = 1, \qquad (53)$$

provided that the number of samples is on the order of:[9]

$$\boxed{n(N) = \Omega\left((Np_N)^2 \log S\right)} \qquad (54)$$

Moreover, in the dense regime where $p > 0$, the same result holds for the non-regularized Granger estimator in (50).

Proof: See Appendix I. ■

We see from (54) that under the dense regime the growth is essentially quadratic in N, since p_N converges to some positive constant p. In order to see what happens under the sparse regime, it is useful to apply (52) and rewrite the sample size in (54) as (we recall that S grows linearly with N in view of (5)):

$$n(N) \sim (\omega_N \log N)^2 \log S \sim \omega_N^2 (\log N)^3, \qquad (55)$$

revealing that the specific sample complexity under the sparse regime depends on the specific speed of growth of the sequence ω_N, which in fact regulates the sparsity of the problem.

The scaling law found in (54) is significant since it matches well with the scaling laws that have been found in the literature in relation to other graphical models.

For linear dynamical systems like (3) operating under *full observability*, sample complexity has been examined for a fixed error in estimating the combination matrix and/or the covariance matrices. It has been shown that the sample complexity grows logarithmically with N, but proportionally to the inverse square of the error [29], [30], [31], [32]. In our case, such growth would imply the $(Np_N)^2$-scaling, since we need the error to be bounded by $\epsilon/(Np_N)$.

Under the regime of *partial observability*, an algorithm is proposed in [92] whose sample complexity scales as $\log N$ when the node degree is kept fixed. When the degree grows with N (as in our setting), then the sample complexity in [92] contains an additional $(Np_N)^3$ factor. However, the authors of [92] indicate that they suspect the exponent could be reduced to 2, which would then match our result. In addition, to grant graph recovery a lower bound on the minimum combination-matrix entry is assumed in [92], but this bound is not allowed to scale with N.

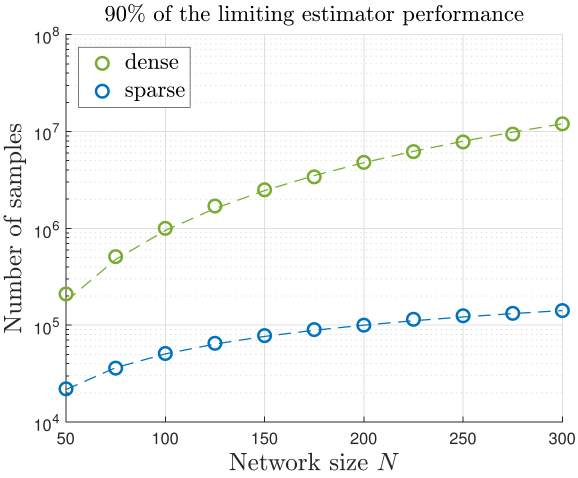

FIGURE 8. Sample complexity. For every N, we evaluated numerically (circles) the number of samples necessary to reach 90% of the performance reached by the limiting estimator for the same value of N. Curves displayed with broken line are $\propto (Np_N)^2 \log S$, i.e., they refer to the theoretical sample complexity predicted by Theorem 3. In the dense case, $p_N = 0.1$, whereas in the sparse case $p_N = (\log\log N)\frac{\log N}{N}$. The combination matrix is obtained through a Laplacian rule with $\rho = 0.5$ and $\lambda = 0.75$, and the fraction of probed nodes is $\xi = 0.15$. In the simulations, the initial vector y_0 has all zero entries, $\sigma = 1$, and the input source samples $x_i(n)$ are i.i.d. samples from a standard Gaussian distribution.

Remark 6 (Factors affecting sample complexity): The sample complexity found in Theorem 3 is primarily determined by the error in estimating the empirical covariance matrices, which is examined in Lemma 9. The limitations of the empirical covariance matrices in high-dimensional settings are well known. For example, the covariance matrix is singular when $n < N$. Even for larger n the number of parameters to be estimated (i.e., the matrix entries) grows quadratically, thus requiring a high number of samples to get good performance. On the other hand, when focusing on the $\|\cdot\|_{\max}$ error norm, it is known that a maximum error up to an ϵ can be obtained when the number of samples grows *logarithmically* with the dimension [29], [30], [31], [32]. This effect is summarized by the $\log S$ factor appearing in (54). Unfortunately, this is not the main factor determining the sample complexity. The main factor is instead the Np_N term. This term arises because, in order to guarantee that the matrix entries are well classifiable, we need to guarantee a maximum error up to $\epsilon/(Np_N)$, which is related to the scaling law of the smallest combination-matrix entry. □

We now complement the asymptotic result in Theorem 3 with an *empirical* evaluation of the sample complexity. This is a cumbersome task from a computational viewpoint, which is however critical to establish the practical relevance of Theorem 3. In Fig. 8, we considered different values of the network size N. For every N, we evaluated numerically (circles) the number of samples necessary to reach 90% of the performance reached by the limiting estimator for the same value of N. We display with broken line the theoretical curves, which are proportional to $(Np_N)^2 \log S$, according to the theoretical scaling law in (54). Under both the dense and sparse regimes, we see that the match with the theoretical curves is excellent. Therefore, under the *dense* regime (green) the scaling law of

[9] With the $\Omega(\cdot)$ notation we mean that there exist a constant $C > 0$ and a value $N_0 > 0$ such that, for all $N \geq N_0$, any sample complexity scaling as $n(N) \geq C(Np_N)^2 \log S$ guarantees consistent learning.

the sample complexity is almost quadratic in N. In comparison, under the *sparse* uniform concentration regime (blue) the sample complexity is mitigated, and scales as:

$$(\log\log N)^2(\log N)^3, \tag{56}$$

since in the considered example p_N was chosen as in (49).

VIII. OTHER ESTIMATORS

So far we have shown that, over uniformly-concentrated Erdős-Rényi graphs, the Granger estimator operating under partial observability achieves consistent graph learning. The proof has been carried out assuming symmetric combination matrices, albeit the Granger estimator construction does not require symmetry. Symmetry has been assumed for technical reasons, since it leads to a closed-form expression for the covariance matrix that helps in the proof.

In this section, we present three other estimators that provably achieve universal local structural consistency under Assumption 1. However, differently from the Granger estimator, the construction itself of these estimators relies on symmetry. For this reason, we will show later that these estimators are sensitive to the symmetry assumption and their performance deteriorates over asymmetric combination matrices, whereas the performance of the Granger estimator does not.

A. ONE-THREE-LAGS ESTIMATOR

We have already observed that (3) can be exploited to obtain the formula $\boldsymbol{R}_1 = \boldsymbol{A}\boldsymbol{R}_0$. Actually, the formula can be generalized to covariance matrices corresponding to any lag, namely, for $m = 1, 2, \ldots$ we have:

$$R_m = R_m(A) = \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{y}_n\boldsymbol{y}_{n-m}^\top|\boldsymbol{A}=A\right] = A^m R_0(A),$$
$$\boldsymbol{R}_m = R_m(\boldsymbol{A}) = \boldsymbol{A}^m\boldsymbol{R}_0. \tag{57}$$

Considering now the series expansion in (28), from (57) we get:

$$\boldsymbol{R}_1 - \boldsymbol{R}_3 = \sigma^2\boldsymbol{A}\left(I + \boldsymbol{A}^2 + \boldsymbol{A}^4 + \cdots\right) - \sigma^2\boldsymbol{A}^3\left(I + \boldsymbol{A}^2 + \boldsymbol{A}^4 + \cdots\right) = \sigma^2\boldsymbol{A}, \tag{58}$$

which motivates the *one-three-lags* limiting estimator proposed in [93]:

$$\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\text{1-3-lags})} = [\boldsymbol{R}_1]_{\mathcal{S}} - [\boldsymbol{R}_3]_{\mathcal{S}} = \sigma^2\boldsymbol{A}_{\mathcal{S}}, \tag{59}$$

revealing that *in the symmetric case*, the desired submatrix over the probed subset $\mathcal{S}$ is equal, up to a scaling factor, to the difference between the one-lag and the three-lags covariance matrices.

B. LIMITING ONE-LAG ESTIMATOR

In this section, we use the one-lag covariance matrix to estimate the combination matrix.[10] The reason behind such choice is the following series expansion of the one-lag covariance matrix:

$$\boldsymbol{R}_1 = \boldsymbol{A}\boldsymbol{R}_0 = \sigma^2\boldsymbol{A}(I-\boldsymbol{A}^2)^{-1} = \sigma^2\big(\boldsymbol{A} + \boldsymbol{A}^3 + \boldsymbol{A}^5 + \cdots\big). \tag{60}$$

When applied only to the submatrix corresponding to $\mathcal{S}$, (60) yields:

$$\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\text{1-lag})} = [\boldsymbol{R}_1]_{\mathcal{S}} = \sigma^2\big(\boldsymbol{A}_{\mathcal{S}} + [\boldsymbol{A}^3]_{\mathcal{S}} + [\boldsymbol{A}^5]_{\mathcal{S}} + \cdots\big). \tag{61}$$

It is convenient to rewrite (61) as:

$$\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\text{1-lag})} = \sigma^2\big(\boldsymbol{A}_{\mathcal{S}} + \boldsymbol{E}^{(\text{1-lag})}\big), \tag{62}$$

where:

$$\boldsymbol{E}^{(\text{1-lag})} = \sum_{h=1}^{\infty}[\boldsymbol{A}^{2h+1}]_{\mathcal{S}}, \tag{63}$$

or, for all $i, j \in \mathcal{S}$:

$$\boxed{\boldsymbol{e}_{ij}^{(\text{1-lag})} = \sum_{h=1}^{\infty}\boldsymbol{a}_{ij}^{(2h+1)}} \tag{64}$$

C. LIMITING RESIDUAL ESTIMATOR

Let us introduce the residual vector that computes the (scaled) difference between consecutive time samples:

$$\boldsymbol{r}_n \triangleq \frac{\boldsymbol{y}_n - \boldsymbol{y}_{n-1}}{\sqrt{2}}. \tag{65}$$

We observe that:

$$\lim_{n\to\infty}\mathbb{E}[\boldsymbol{r}_n\boldsymbol{r}_n^\top|\boldsymbol{A}=A] = R_0 - R_1 = \sigma^2(I+A)^{-1}. \tag{66}$$

Accordingly, it makes sense to introduce the following limiting estimator:

$$\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\text{res})} = [\boldsymbol{R}_1]_{\mathcal{S}} - [\boldsymbol{R}_0]_{\mathcal{S}} = -\sigma^2\left[(I+\boldsymbol{A})^{-1}\right]_{\mathcal{S}} = \sigma^2\left(\boldsymbol{A}_{\mathcal{S}} - I_{\mathcal{S}} - [\boldsymbol{A}^2]_{\mathcal{S}} + [\boldsymbol{A}^3]_{\mathcal{S}} + \cdots\right). \tag{67}$$

The structure of (67) motivates the introduction of the matrix:

$$\boldsymbol{E}^{(\text{res})} = -I_{\mathcal{S}} + \sum_{h=1}^{\infty}([\boldsymbol{A}^{2h+1}]_{\mathcal{S}} - [\boldsymbol{A}^{2h}]_{\mathcal{S}}), \tag{68}$$

yielding:

$$\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\text{res})} = \sigma^2\big(\boldsymbol{A}_{\mathcal{S}} + \boldsymbol{E}^{(\text{res})}\big). \tag{69}$$

Equation (68) implies that, for all $i, j \in \mathcal{S}$, with $i \neq j$:

$$\boxed{\boldsymbol{e}_{ij}^{(\text{res})} = \sum_{h=1}^{\infty}\left(\boldsymbol{a}_{ij}^{(2h+1)} - \boldsymbol{a}_{ij}^{(2h)}\right)} \tag{70}$$

[10] When σ^2 is known, the relationship between $\boldsymbol{R}_1 = \sigma^2\boldsymbol{A}(I-\boldsymbol{A}^2)^{-1}$ and $\boldsymbol{A}$ is invertible under the full observability regime. This can be easily shown by resorting to the spectral decomposition of $\boldsymbol{R}_1/\sigma^2$ and $\boldsymbol{A}$ (which share the same eigenvectors), and by finding the eigenvalues of $\boldsymbol{A}$ from those of $\boldsymbol{R}_1/\sigma^2$; this inversion operation can be successfully realized because $\boldsymbol{A}$ is symmetric and has spectral radius less than one. However, we remark that σ^2 is assumed unknown and, more importantly, that we work under a partial observability regime, and, hence, the aforementioned inversion procedure does not apply.

TABLE 2. Biases and Identifiability Gaps of the Estimators Listed in the Leftmost Column. In all Cases, the Scaling Sequence is $s_N = Np_N$. In the Formulas we set $\zeta = \rho - \kappa$

Estimator	Bias η	Identifiability gap Γ
Granger	$\kappa^2 p \, \dfrac{(2\rho - \kappa)\,(1-\xi)}{1-(\rho^2 - 2\rho\kappa\xi + \kappa^2\xi)}$	κ
one-three-lags	0	$\sigma^2\kappa$
one-lag	$\sigma^2\kappa^2 p \, \dfrac{\rho + \rho\zeta^2 + 2\zeta}{(1-\rho^2)(1-\zeta^2)^2}$	$\dfrac{1+\zeta^2}{(1-\zeta^2)^2} \times \sigma^2\kappa$
residual	$-\dfrac{\sigma^2\kappa^2 p}{(1+\rho)(1+\zeta)^2}$	$\dfrac{\sigma^2\kappa}{(1+\zeta)^2}$

We are ready to state the theorem that characterizes the consistency of the aforementioned three estimators.

Theorem 4 (Universal local structural consistency of the estimators in Section VIII): Let $\boldsymbol{A}$ be a regular diffusion matrix with parameters ρ and κ, with the network graph drawn according to an Erdős-Rényi random graph model $\mathscr{G}(N, p_N)$ where the fraction of observable nodes, S/N, converges to some nonzero value ξ. Then, under the uniform concentration regime where:

$$p_N = \omega_N \frac{\log N}{N} \to p, \qquad \text{with } \omega_N \to \infty, \tag{71}$$

the one-three-lags, one-lag, and residual estimators achieve universal local structural consistency as detailed in Definition 3, with scaling sequence $s_N = Np_N$, and with the biases and identifiability gaps listed in Table 2.

Proof: See Appendix E. ∎

It is instructive to compare the different estimators reported in Table 2. First of all, we notice that the proper scaling sequence for all estimators is $s_N = Np_N$. Consider now the identifiability gap. The gap of the Granger estimator is equal to the bounding constant κ that characterizes the regular diffusion matrix (44). The other estimators exhibit two main differences with respect to the Granger estimator. First, they depend also on the variance of the input process, σ^2. This behavior should be expected, since in the Granger estimator the one-lag covariance matrix multiplies the inverse of the covariance matrix, and, hence, the effect of σ^2 disappears. In contrast, the other estimators do not cancel this effect out. Second, when $\kappa \neq \rho$ (a condition that occurs, for instance, for the Laplacian rule), the term $\sigma^2\kappa$ multiplies a factor that is a function of $\rho - \kappa$. This factor is greater than one for the one-lag estimator, whereas it is smaller than one for the residual estimator. Note that the dependence alone upon σ^2, as well as a magnified/reduced gap, do not imply any conclusion about the structure-learning performance of the pertinent estimators for finite network and/or sample sizes. What plays a role in this case is the spread of the matrix entries around their limiting values, and the identifiability gap does not contain information about such spread.

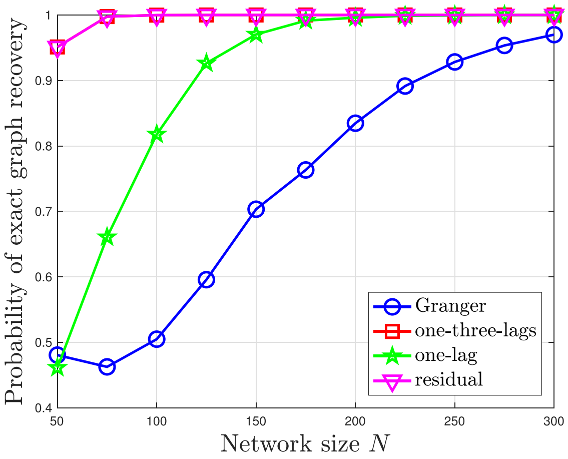

FIGURE 9. Performance of the limiting estimators proposed in this work, as a function of N under the dense regime with connection probability $p_N = 0.1$, and fraction of probed nodes $\xi = 0.15$. The combination matrix is obtained through a Laplacian rule with $\rho = 0.99$ and $\lambda = 0.9$. In the simulations, the initial vector y_0 has all zero entries, $\sigma = 1$, and the input source samples $x_i(n)$ are i.i.d. samples from a standard Gaussian distribution. The probability of correct graph recovery is evaluated by means of 10^4 Monte Carlo runs.

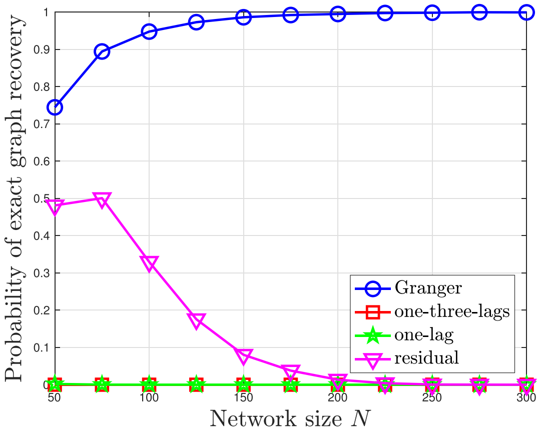

FIGURE 10. Performance of the limiting estimators proposed in this work as a function of N, under asymmetric combination matrices. Specifically, the random graph is generated as a *binomial graph* with connection probability equal to 0.1, and the combination matrix is obtained through a uniform averaging rule with $\rho = 0.99$ and $\lambda = 0.2$. The fraction of probed nodes $\xi = 0.15$. In the simulations, the initial vector y_0 has all zero entries, $\sigma = 1$, and the input source samples $x_i(n)$ are i.i.d. samples from a standard Gaussian distribution. The probability of correct graph recovery is evaluated by means of 10^4 Monte Carlo runs.

Let us switch to the analysis of the bias. First, we notice that $\eta = 0$ for the one-three-lags estimator,[11] which is obvious since (the limiting version of) this estimator is equal

[11] In order to avoid misunderstandings, we point out that in our treatment we use the term "bias" to quantify the distance from zero of the estimated matrix entries corresponding to unconnected pairs. In this respect, the one-three-lags estimator is unbiased. However, it is not an unbiased estimator of the combination matrix $\boldsymbol{A}_{\mathcal{S}}$, due to the presence of the scaling factor.

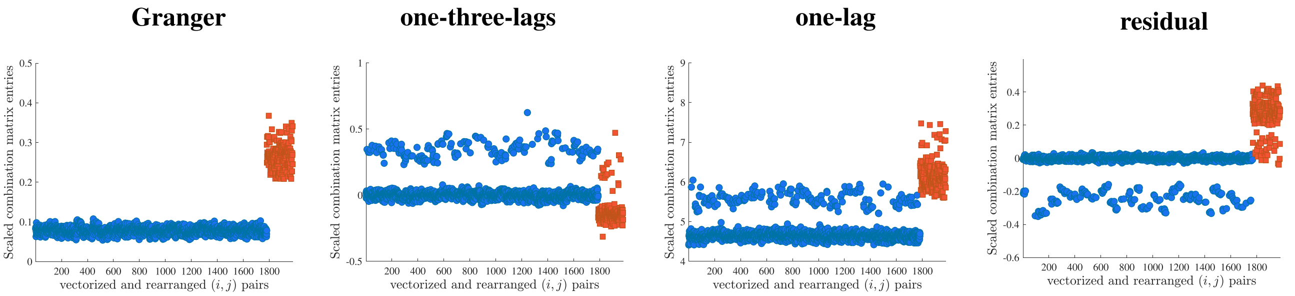

FIGURE 11. **Directed graphs. The four panels refer to the limiting matrix estimators considered in this work. The entries of the *estimated* matrix, $\widehat{A}_{\mathcal{S}}$ are displayed according to the following rule. First, the entries of the *true* matrix $A_{\mathcal{S}}$ are vectorized following column-major ordering, and the (vectorized) (i, j) pairs are rearranged in such a way that the zero entries appear before the nonzero entries. Such ordering is then applied to the entries of the *estimated* matrix, $\widehat{A}_{\mathcal{S}}$, which are scaled by Np_N, and displayed with blue circles if they correspond to unconnected pairs, and with red squares otherwise. The underlying graph is a *binomial graph* with connection probability equal to 0.1. The combination matrix is obtained through the uniform averaging rule in (72), with parameters $\rho = 0.99$ and $\lambda = 0.2$. We see that the Granger estimator preserves the identifiability gap, whereas the other estimators do not.**

to the true matrix scaled by σ^2. The biases of the other estimators are proportional to the limiting connection probability p and, hence, we have always $\eta = 0$ in the sparse case, where $p = 0$. Furthermore, we observe that only the bias of the Granger estimator depends upon the fraction of monitored nodes ξ. This finding makes sense, since the Granger estimator is based upon inversion of a partial matrix (which clearly varies with the number of latent variables), whereas the other estimators are natively determined by pairwise correlations. In comparison, only the bias of the Granger estimator does not depend upon σ^2, and this behavior is easily grasped in light of the explanation in the previous paragraph.

We complement our theoretical findings with some numerical experiments. First, we test the limiting estimators over one of the settings adopted in Fig. 6. The results are reported in Fig. 9. We see that, in agreement with Theorem 4, all the limiting estimators have probability of correct graph recovery converging to 1 as N grows. In particular, the one-three-lags estimator (which, as we know, coincides with the true matrix) has the same performance as the residual estimator. These estimators perform better than the one-lag estimator, which is in turn better than the Granger estimator. This ordering is relative to this particular example, and we have observed various behavior in other experiments, depending on different factors, such as the degree of observability or the type of combination matrix. However, there is a more important comparison that should be made between the Granger estimator and the other estimators, which pertains the case of directed graphs.

To this end, we generate a directed topology by using a *binomial graph*, namely, a random graph where all directed edges are drawn as i.i.d. realizations of Bernoulli variables, with the success corresponding to the edge presence. We remark that, over a binomial graph, the event of a directed edge $i \to j$ is independent of the event of a directed edge $j \to i$. Then, we consider the following *uniform averaging* combination policy, with $0 < \lambda \leq 1$:

$$\boldsymbol{a}_{ij} = \frac{\rho\lambda}{\boldsymbol{d}_i}\boldsymbol{g}_{ij}, \tag{72}$$

and $\boldsymbol{a}_{ii} = \rho - \sum_{\ell \neq i}^{N} \boldsymbol{a}_{i\ell}$.

The performance of the different estimators is reported in Fig. 10. We see that all but the Granger estimator fail in learning the true graph. This happens also for the one-three-lags estimator, which was used because it reproduces (up to the scaling factor σ^2) the true combination matrix. However, it must be remarked that this property relies exclusively on the symmetry assumption, and is lost in the asymmetric case. Also the other two estimators are sensitive to symmetry. More information is gained by examining the matrix-entries realizations shown in Fig. 11, where we see clearly that, in this example with a directed graph, *the identifiability gap is preserved by the Granger estimator, but is lost by the other proposed estimators*, a behavior that explains clearly the performance obtained in Fig. 10. In particular, the estimator that seems to suffer more from the lack of symmetry is the one-three-lags estimator, with a kind of reversed pattern exhibited in the second plot of Fig. 11. This behavior confirms a classical mismatched-model trade-off: the more one method is matched to a nominal scenario (i.e., the symmetric scenario), the more sensitivity it exhibits in a mismatched scenario.

IX. CONCLUDING REMARKS AND OPEN ISSUES

This work examined the problem of graph learning when data can be collected from a limited subset of nodes. The goal is to learn the topology of the subgraph of probed nodes. We showed that an estimator of the combination matrix known as Granger estimator, followed by a universal clustering algorithm, is able to achieve faithful graph learning, requiring a number of samples that grows almost quadratically with the expected node degree. We explored various regimes of connectivity, including the often overlooked regime of *dense* connectivity. Several works in the literature of graph learning assume in fact sparsity in the graph of connections. The role of sparsity in graph learning can be (at least) twofold. On one hand, sparsity can be leveraged to reduce the complexity associated to the estimators (we have seen this effect also in our analysis of the sample complexity). On the other hand, in the presence of latent variables, sparsity can be leveraged since it reduces the (unseen) effect of the latent unobserved nodes on the probed nodes [20], [63]. One revealing conclusion stemming from our analysis is that, under the setting considered in this work (Erdős-Rényi graphs and regular diffusion matrices) an important role is played by another structural property of the graph, namely, the *statistical concentration of node degrees*. Finally, we proposed three structurally consistent estimators and compare them against the Granger estimator, obtaining nontrivial insights, especially over directed graphs.

There are several open issues that may deserve attention. This work focused on the Erdős-Rényi model, and used certain regularity assumptions on the diffusion matrix. A useful extension would be to examine structural consistency for other graph models, and/or under different regularity assumptions. For example, one might have network heterogeneity (e.g., different connectivity across nodes) and/or dependence in the edge formation process, and an interesting question is whether or not consistency can be achieved under these conditions. Preliminary results along this direction are available in [94].

Another open issue concerns *directed* graphs, which are relevant, e.g., in the context of causal inference. In this work we have examined this issue through numerical experiments, showing that the Granger estimator seems to preserve its learning capability over directed graphs, while the other proposed estimators are much more sensitive to asymmetries in the graph construction. From the theoretical side, the graph-edge-domain approach developed in this work could be exploited and generalized to get insights about the directed graph setting.

APPENDIX A
USEFUL PROPERTIES OF MAXIMAL AND MINIMAL DEGREES

We denote by $\mathcal{B}_i(N, q)$, with $i = 1, 2, \ldots, K$, a sequence of K binomial random variables (not necessarily independent) with success probability q over N independent Bernoulli trials. Moreover, we denote by $\mathcal{B}_{\max}(N, q, K)$ and $\mathcal{B}_{\min}(N, q, K)$ the maximum and the minimum over this sequence, respectively. The following two relationships are standard inequalities arising from the application of the Chernoff bounding technique, and will be the fundamental building blocks to characterize the asymptotic behavior of several random quantities arising in our problem. The inequalities are as follows.[12] For any $t > 0$:

$$\mathbb{P}[\mathcal{B}_{\max}(N, q, K) \geq x] \leq K e^{-xt + Nq(e^t - 1)}, \tag{75}$$

$$\mathbb{P}[\mathcal{B}_{\min}(N, q, K) \leq x] \leq K e^{xt - Nq(1 - e^{-t})}. \tag{76}$$

We now apply these fundamental bounds to some specific random variables that are of interest in our setting.

We start by characterizing the behavior of the variables:

$$\mathcal{B}_{\max}(N, p_N, N), \qquad \mathcal{B}_{\min}(N, p_N, N), \tag{77}$$

which, as we will see, are useful to characterize the behavior of the maximal and minimal degree of the graphs that we use in this work. The forthcoming lemma contains fundamental (classic) results about the asymptotic behavior of $\mathcal{B}_{\max}(N, p_N, N)$ and $\mathcal{B}_{\min}(N, p_N, N)$ under the different regimes for the probability p_N.

Lemma 1 (Asymptotic scaling of $\mathcal{B}_{\max}(N, p_N, N)$ and $\mathcal{B}_{\min}(N, p_N, N)$): Let the probability p_N scale with N according to (42). Then:

$$\boxed{\frac{\mathcal{B}_{\max}(N, p_N, N)}{N p_N} \xrightarrow{\text{p}} 1, \qquad \frac{\mathcal{B}_{\min}(N, p_N, N)}{N p_N} \xrightarrow{\text{p}} 1} \tag{78}$$

Proof: The following inequality, holding for all $\epsilon > 0$, is easily obtained from (75) by setting $K = N$, $q = p_N$, $x = (1 + \epsilon) N p_N$, $t = \log(1 + \epsilon)$, and $g_\epsilon \triangleq 1 + (1 + \epsilon)(\log(1 + \epsilon) - 1)$:

$$\mathbb{P}[\mathcal{B}_{\max}(N, p_N, N) \geq (1 + \epsilon) N p_N] \leq N e^{-N p_N g_\epsilon}. \tag{79}$$

Using now (42) in (79) we get:

$$\mathbb{P}[\mathcal{B}_{\max}(N, p_N, N) \geq (1 + \epsilon) N p_N] \leq N^{1 - \omega_N g_\epsilon} \overset{N \to \infty}{\longrightarrow} 0, \tag{80}$$

which follows because $g_\epsilon > 0$ for all $\epsilon > 0$, as $g_0 = 0$ and $d g_\epsilon / d\epsilon > 0$ for all $\epsilon > 0$.

Likewise, the following inequality, holding for all $0 < \epsilon < 1$, is easily obtained from (76) by setting $K = N$, $q = p_N$, $x = (1 - \epsilon) N p_N$, $t = -\log(1 - \epsilon)$, and $h_\epsilon \triangleq 1 - (1 - \epsilon)(1 - \log(1 - \epsilon))$:

$$\mathbb{P}[\mathcal{B}_{\min}(N, p_N, N) \leq (1 - \epsilon) N p_N] \leq N^{1 - \omega_N h_\epsilon} \overset{N \to \infty}{\longrightarrow} 0, \tag{81}$$

which follows because $h_\epsilon > 0$ for all $0 < \epsilon < 1$, as $h_0 = 0$ and $d h_\epsilon / d\epsilon > 0$ for all $0 < \epsilon < 1$. By joining (80) with (81), and observing that $\mathcal{B}_{\max}(N, p_N, N) \geq \mathcal{B}_{\min}(N, p_N, N)$, we conclude that (78) holds true. ■

As a simple corollary to Lemma 1, we can now obtain the characterization of the maximal and minimal degrees.

Corollary 1 (Behavior of $\boldsymbol{d}_{\max}$ and $\boldsymbol{d}_{\min}$): If the connection probability of the Erdős-Rényi model obeys (42), then we have:

$$\boxed{\frac{\boldsymbol{d}_{\max}}{N p_N} \xrightarrow{\text{p}} 1, \qquad \frac{\boldsymbol{d}_{\min}}{N p_N} \xrightarrow{\text{p}} 1[\text{Uniform concentration}]} \tag{82}$$

[12] For any $t > 0$, we can write:

$$\mathbb{P}[\mathcal{B}_i(N, q) \geq x] = \mathbb{P}[e^{\mathcal{B}_i(N,q)t} \geq e^{xt}] \leq e^{-xt}\, \mathbb{E}[e^{\mathcal{B}_i(N,q)t}], \tag{73}$$

where the latter inequality is an application of Markov's inequality. Since a binomial variable of parameters N and q is the sum of N independent Bernoulli variables with success probability equal to q, we can further write:

$$\mathbb{E}[e^{\mathcal{B}_i(N,q)t}] = \left(q e^t + 1 - q\right)^N = \left(1 + q(e^t - 1)\right)^N \leq e^{Nq(e^t - 1)}, \tag{74}$$

where the latter inequality follows by observing that, for $z > 0$, one has $(1 + z)^N \leq e^{Nz}$. Combining (74) with (73) yields (75). (76) is worked out with a similar technique.

Proof: The degree of a single node is equal to 1 plus (because in our setting the degree counts also the node itself) a binomial random variable with parameters $N-1$ and p_N. Therefore, we have the following representation:

$$\boldsymbol{d}_{\max} = 1 + \mathfrak{B}_{\max}(N-1, p_N, N), \tag{83}$$

$$\boldsymbol{d}_{\min} = 1 + \mathfrak{B}_{\min}(N-1, p_N, N). \tag{84}$$

In order to obtain useful bounds involving $\boldsymbol{d}_{\max}$ and $\boldsymbol{d}_{\min}$, let us introduce a modified sequence of binomial variables, obtained by adding one more (independent) Bernoulli trial to each binomial variable $\mathfrak{B}_i(N-1, p_N)$, with $i = 1, 2, \ldots, N$. The corresponding maximum and minimum taken over the modified sequence will be denoted by $\widetilde{\mathfrak{B}}_{\max}(N, p_N, N)$ and $\widetilde{\mathfrak{B}}_{\min}(N, p_N, N)$, respectively. Since a Bernoulli variable can be either zero or one, from (83) and (84) we get readily the following bounds:

$$\boldsymbol{d}_{\max} \le 1 + \widetilde{\mathfrak{B}}_{\max}(N, p_N, N), \tag{85}$$

$$\boldsymbol{d}_{\min} \ge \widetilde{\mathfrak{B}}_{\min}(N, p_N, N), \tag{86}$$

and, hence, the claims of the corollary follow readily from Lemma 1, with the factor 1 playing no role as $N \to \infty$. ∎

A. ANOTHER USEFUL CONCENTRATION RESULT

Lemma 2 (Maximum and minimum of N^2 binomial variables with success probability p_N^2): Assume that the success probability obeys (42). Then we have that:

$$\boxed{\frac{\mathfrak{B}_{\max}(N, p_N^2, N^2)}{N p_N} \xrightarrow{\text{p}} p, \qquad \frac{\mathfrak{B}_{\min}(N, p_N^2, N^2)}{N p_N} \xrightarrow{\text{p}} p} \tag{87}$$

Proof: If $p_N \to p > 0$, we can set $p_N' = p_N^2$, and obviously p_N' converges to $p^2 > 0$, implying that the binomial variables of parameters N and p_N' are generated under the uniform concentration regime. The result in (87) then readily follows by exploiting (75) and (76) as done in the proof of Lemma 1.

If $p_N \to p = 0$, it suffices to prove the claim for the maximum. Applying (75) we can write:

$$\begin{aligned}\mathbb{P}[\mathfrak{B}_{\max}(N, p_N^2, N^2) \ge \epsilon N p_N] &\le N^2 e^{-N p_N [\epsilon t - p_N(e^t - 1)]} \\ &= N^2 N^{-\omega_N [\epsilon t - p_N(e^t - 1)]}.\end{aligned} \tag{88}$$

where, in the last step, we used the equality $N p_N = \omega_N \log N$ that follows from (42). Moreover, since we are considering the case where $p_N \to 0$ as $N \to \infty$, for any $\epsilon' > 0$ and for sufficiently large N we will have $p_N < \epsilon'$, so that, asymptotically, it is legitimate to replace (88) with:

$$\mathbb{P}[\mathfrak{B}_{\max}(N, p_N^2, N^2) \ge \epsilon N p_N] \le N^2 N^{-\omega_N [\epsilon t - \epsilon'(e^t - 1)]}. \tag{89}$$

Now, choosing ϵ' small enough so that $\epsilon'(e^t - 1) < \epsilon t$, we finally get:

$$\mathbb{P}[\mathfrak{B}_{\max}(N, p_N^2, N^2) \ge \epsilon N p_N] \overset{N \to \infty}{\longrightarrow} 0, \tag{90}$$

which completes the proof of the lemma. ∎

APPENDIX B
BOUNDS ON THE POWERS OF A

In the following, the symbol $\boldsymbol{a}_{ij}^{(k)}$ denotes the (i, j) entry of the k-th matrix power $\boldsymbol{A}^k$, and $\mathcal{N}$ denotes the set of all nodes. Table 3 lists some random variables that are necessary to state and prove the theorems.

We start with a technical lemma that will be used to prove Theorem 2.

Lemma 3 (Bounds on the entries of $\boldsymbol{A}^k$): The entries of the combination matrix power $\boldsymbol{A}^k$ are bounded as follows:

$$\underline{\boldsymbol{\alpha}}_k \le \boldsymbol{a}_{ii}^{(k)} \le \overline{\boldsymbol{\alpha}}_k, \tag{91}$$

and, for $i \neq j$:

$$\underline{\boldsymbol{\beta}}_k \boldsymbol{a}_{ij} + \underline{\boldsymbol{\gamma}}_k \mathfrak{m} \le \boldsymbol{a}_{ij}^{(k)} \le \overline{\boldsymbol{\beta}}_k \boldsymbol{a}_{ij} + \overline{\boldsymbol{\gamma}}_k \mathfrak{M}, \tag{92}$$

where, for $k \ge 2$, the (random) sequences $\overline{\boldsymbol{\alpha}}_k$, $\underline{\boldsymbol{\alpha}}_k$, $\overline{\boldsymbol{\beta}}_k$, $\underline{\boldsymbol{\beta}}_k$, $\overline{\boldsymbol{\gamma}}_k$, and $\underline{\boldsymbol{\gamma}}_k$, are determined by the following recursions:

$$\overline{\boldsymbol{\alpha}}_{k+1} = \mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\alpha}}_k + \mathfrak{M}_a\, \rho^k, \tag{93}$$

$$\overline{\boldsymbol{\beta}}_{k+1} = \overline{\boldsymbol{\alpha}}_k + \mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\beta}}_k, \tag{94}$$

$$\overline{\boldsymbol{\gamma}}_{k+1} = \overline{\boldsymbol{\beta}}_k + \mathfrak{M}_{a,\text{sum}}\, \overline{\boldsymbol{\gamma}}_k, \tag{95}$$

with the initialization choices:

$$\overline{\boldsymbol{\alpha}}_2 = \mathfrak{M}_{a_2,\text{self}}, \quad \overline{\boldsymbol{\beta}}_2 = 2\,\mathfrak{M}_{a,\text{self}}, \quad \overline{\boldsymbol{\gamma}}_2 = 1, \tag{96}$$

and

$$\underline{\boldsymbol{\alpha}}_{k+1} = \mathfrak{m}_{a,\text{self}}\, \underline{\boldsymbol{\alpha}}_k, \tag{97}$$

$$\underline{\boldsymbol{\beta}}_{k+1} = \underline{\boldsymbol{\alpha}}_k + \mathfrak{m}_{a,\text{self}}\, \underline{\boldsymbol{\beta}}_k, \tag{98}$$

$$\underline{\boldsymbol{\gamma}}_{k+1} = \underline{\boldsymbol{\beta}}_k + \mathfrak{m}_{a,\text{sum}}\, \underline{\boldsymbol{\gamma}}_k, \tag{99}$$

with the initialization choices:

$$\underline{\boldsymbol{\alpha}}_2 = \mathfrak{m}_{a_2,\text{self}}, \quad \underline{\boldsymbol{\beta}}_2 = 2\,\mathfrak{m}_{a,\text{self}}, \quad \underline{\boldsymbol{\gamma}}_2 = 1. \tag{100}$$

Proof: We start by examining the relationships pertaining to the main diagonal terms, namely, (91). For $k = 2$ the claim is trivially true with the values of $\overline{\boldsymbol{\alpha}}_2$ and $\underline{\boldsymbol{\alpha}}_2$ in (96) and (100), because of the definitions of $\mathfrak{M}_{a_2,\text{self}}$ and $\mathfrak{m}_{a_2,\text{self}}$ reported on line 3) of Table 3. We shall therefore reason by induction to prove that (91) holds for an arbitrary k. In particular, we assume the claim verified for k, and manage to prove that it is verified for $k+1$. To this aim, we start by writing the diagonal terms of the $(k+1)$-th matrix power as:

$$\boldsymbol{a}_{ii}^{(k+1)} = \sum_{\ell \in \mathcal{N}} \boldsymbol{a}_{i\ell} \boldsymbol{a}_{\ell i}^{(k)} = \boldsymbol{a}_{ii} \boldsymbol{a}_{ii}^{(k)} + \sum_{\substack{\ell \in \mathcal{N} \\ \ell \neq i}} \boldsymbol{a}_{i\ell} \boldsymbol{a}_{\ell i}^{(k)}, \tag{101}$$

and, hence:

$$\boldsymbol{a}_{ii} \boldsymbol{a}_{ii}^{(k)} \le \boldsymbol{a}_{ii}^{(k+1)} \le \boldsymbol{a}_{ii} \boldsymbol{a}_{ii}^{(k)} + \mathfrak{M}_a\, \rho^k, \tag{102}$$

where we have used the fact that $\sum_{\ell \in \mathcal{N}} \boldsymbol{a}_{\ell i}^{(k)} = \rho^k$ along with the definition of $\mathfrak{M}_a$ appearing on line 1) of Table 3. In (102), we can bound the term $\boldsymbol{a}_{ii}$ by using the definitions on line 2)

TABLE 3. **Random Variables and Convergences Relevant for the Proofs of the Theorems**

	Random variable	Random variable	Limit (in probability)
1)	$\mathfrak{M}_a \triangleq \max\limits_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \boldsymbol{a}_{ij}$	$\mathfrak{m}_a \triangleq \min\limits_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \boldsymbol{a}_{ij}$	0
2)	$\mathfrak{M}_{a,\text{self}} \triangleq \max\limits_{i\in\mathcal{N}} \boldsymbol{a}_{ii}$	$\mathfrak{m}_{a,\text{self}} \triangleq \min\limits_{i\in\mathcal{N}} \boldsymbol{a}_{ii}$	$\rho-\kappa$
3)	$\mathfrak{M}_{a_2,\text{self}} \triangleq \max\limits_{i\in\mathcal{N}} \boldsymbol{a}_{ii}^{(2)}$	$\mathfrak{m}_{a_2,\text{self}} \triangleq \min\limits_{i\in\mathcal{N}} \boldsymbol{a}_{ii}^{(2)}$	$(\rho-\kappa)^2$
4)	$\mathfrak{M}_{a,\text{sum}} \triangleq \max\limits_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \sum\limits_{\substack{\ell\in\mathcal{N}\\ \ell\neq j}} \boldsymbol{a}_{i\ell}$	$\mathfrak{m}_{a,\text{sum}} \triangleq \min\limits_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \sum\limits_{\substack{\ell\in\mathcal{N}\\ \ell\neq j}} \boldsymbol{a}_{i\ell}$	ρ
5)	$\mathfrak{M}_{c,\text{self}} \triangleq \max\limits_{\ell\in\mathcal{S}'} \boldsymbol{c}_{\ell\ell}$	$\mathfrak{m}_{c,\text{self}} \triangleq \min\limits_{\ell\in\mathcal{S}'} \boldsymbol{c}_{\ell\ell}$	$(\rho-\kappa)^2$
6)	$\mathfrak{M}_{a,\text{sum}}^{(\mathcal{S}')} \triangleq \max\limits_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \sum\limits_{\substack{h\in\mathcal{S}'\\ h\neq\ell,m}} \boldsymbol{a}_{hm}$	$\mathfrak{m}_{a,\text{sum}}^{(\mathcal{S}')} \triangleq \min\limits_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \sum\limits_{\substack{h\in\mathcal{S}'\\ h\neq\ell,m}} \boldsymbol{a}_{hm}$	$\kappa(1-\xi)$
7)	$\widetilde{\mathfrak{M}}_{a,\text{sum}}^{(\mathcal{S}')} \triangleq \max\limits_{i\in\mathcal{S}} \sum\limits_{\ell\in\mathcal{S}'} \boldsymbol{a}_{i\ell}$	$\widetilde{\mathfrak{m}}_{a,\text{sum}}^{(\mathcal{S}')} \triangleq \min\limits_{i\in\mathcal{S}} \sum\limits_{\ell\in\mathcal{S}'} \boldsymbol{a}_{i\ell}$	$\kappa(1-\xi)$
8)	$\widetilde{\widetilde{\mathfrak{M}}}_{a,\text{sum}}^{(\mathcal{S}')} \triangleq \max\limits_{i,j\in\mathcal{S}} \sum\limits_{\ell\in\mathcal{S}'} \boldsymbol{a}_{i\ell} \sum\limits_{\substack{h\in\mathcal{N}\\ h\neq j}} \boldsymbol{a}_{hj} \sum\limits_{\substack{m\in\mathcal{S}'\\ m\neq\ell,h}} \boldsymbol{a}_{mh}$	$\widetilde{\widetilde{\mathfrak{m}}}_{a,\text{sum}}^{(\mathcal{S}')} \triangleq \min\limits_{i,j\in\mathcal{S}} \sum\limits_{\ell\in\mathcal{S}'} \boldsymbol{a}_{i\ell} \sum\limits_{\substack{h\in\mathcal{N}\\ h\neq j}} \boldsymbol{a}_{hj} \sum\limits_{\substack{m\in\mathcal{S}'\\ m\neq\ell,h}} \boldsymbol{a}_{mh}$	$\kappa^3(1-\xi)^2$
9)	$\widetilde{\mathfrak{M}}^{(\mathcal{S}')} \triangleq \max\limits_{i\in\mathcal{S}} \sum\limits_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell m}$	$\widetilde{\mathfrak{m}}^{(\mathcal{S}')} \triangleq \min\limits_{i\in\mathcal{S}} \sum\limits_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell m}$	$\kappa^2(1-\xi)^2$
10)	$\widetilde{\widetilde{\mathfrak{M}}}^{(\mathcal{S}')} \triangleq \max\limits_{\substack{i,j\in\mathcal{S}\\ i\neq j}} \sum\limits_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{mj}$	$\widetilde{\widetilde{\mathfrak{m}}}^{(\mathcal{S}')} \triangleq \min\limits_{\substack{i,j\in\mathcal{S}\\ i\neq j}} \sum\limits_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{mj}$	$\kappa^2(1-\xi)^2$
11)	$\mathfrak{M}_{c,\text{sum}} \triangleq \max\limits_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \sum\limits_{\substack{h\in\mathcal{S}'\\ h\neq m}} \boldsymbol{c}_{\ell h}$	$\mathfrak{m}_{c,\text{sum}} \triangleq \min\limits_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \sum\limits_{\substack{h\in\mathcal{S}'\\ h\neq m}} \boldsymbol{c}_{\ell h}$	$\rho^2-2\rho\kappa\xi+\kappa^2\xi$
12)	$\mathfrak{M} \triangleq \max\limits_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \sum\limits_{\substack{\ell\in\mathcal{N}\\ \ell\neq i,j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j}$	$\mathfrak{m} \triangleq \min\limits_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \sum\limits_{\substack{\ell\in\mathcal{N}\\ \ell\neq i,j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j}$	$\left.\begin{array}{c} Np_N\,\mathfrak{M} \\ Np_N\,\mathfrak{m} \end{array}\right\} \xrightarrow{\text{p}} \kappa^2 p$
13)	$\mathfrak{M}^{(\mathcal{S}')} \triangleq \max\limits_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \sum\limits_{\substack{\ell\in\mathcal{S}'\\ \ell\neq i,j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j}$	$\mathfrak{m}^{(\mathcal{S}')} \triangleq \min\limits_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \sum\limits_{\substack{\ell\in\mathcal{S}'\\ \ell\neq i,j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j}$	$\left.\begin{array}{c} Np_N\,\mathfrak{M}^{(\mathcal{S}')} \\ Np_N\,\mathfrak{m}^{(\mathcal{S}')} \end{array}\right\} \xrightarrow{\text{p}} \kappa^2 p(1-\xi)$
14)	$\mathfrak{M}_{a_3,\text{sum}}^{(\mathcal{S}')} \triangleq \max\limits_{\substack{i,j\in\mathcal{S}\\ i\neq j}} \sum\limits_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell m}\boldsymbol{a}_{mj}$	$\mathfrak{m}_{a_3,\text{sum}}^{(\mathcal{S}')} \triangleq \min\limits_{\substack{i,j\in\mathcal{S}\\ i\neq j}} \sum\limits_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell m}\boldsymbol{a}_{mj}$	$\left.\begin{array}{c} Np_N\,\mathfrak{M}_{a_3,\text{sum}}^{(\mathcal{S}')} \\ Np_N\,\mathfrak{m}_{a_3,\text{sum}}^{(\mathcal{S}')} \end{array}\right\} \xrightarrow{\text{p}} \kappa^3 p(1-\xi)^2$

of Table 3, and the term $\boldsymbol{a}_{ii}^{(k)}$ by using (91) (which is true for k by the induction hypothesis), yielding:

$$\mathfrak{m}_{a,\text{self}}\,\underline{\boldsymbol{\alpha}}_k \le \boldsymbol{a}_{ii}^{(k+1)} \le \mathfrak{M}_{a,\text{self}}\,\overline{\boldsymbol{\alpha}}_k + \mathfrak{M}_a\,\rho^k, \tag{103}$$

from which we conclude that (91) holds true for $k+1$, with the sequences $\overline{\boldsymbol{\alpha}}_k$ and $\underline{\boldsymbol{\alpha}}_k$ obeying the recursions in (93) and (97), respectively.

We switch to the proof of (92). In particular, we will focus on the upper bound in (92), since the proof for the lower bound is similar. For all $i, j \in \mathcal{N}$, with $i \neq j$, we have:

$$\begin{aligned}
\boldsymbol{a}_{ij}^{(2)} &= \sum_{\ell\in\mathcal{N}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j} = (\boldsymbol{a}_{ii}+\boldsymbol{a}_{jj})\boldsymbol{a}_{ij} + \sum_{\substack{\ell\in\mathcal{N}\\ \ell\neq i,j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j} \\
&\le 2\,\mathfrak{M}_{a,\text{self}}\,\boldsymbol{a}_{ij} + \sum_{\substack{\ell\in\mathcal{N}\\ \ell\neq i,j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j} &\quad (104)\\
&\le 2\,\mathfrak{M}_{a,\text{self}}\,\boldsymbol{a}_{ij} + \mathfrak{M} &\quad (105)\\
&\le 2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}_a + \mathfrak{M}, &\quad (106)
\end{aligned}$$

where we have applied the definitions listed on lines 1), 2) and 8) of Table 3. First, we observe that (105) implies that the right inequality in (92) holds true in the case $k = 2$, with the choices detailed in (96). Moreover, we have:

$$\boldsymbol{a}_{ij}^{(k+1)} = \sum_{\ell\in\mathcal{N}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j}^{(k)} = \boldsymbol{a}_{ij}\boldsymbol{a}_{jj}^{(k)} + \sum_{\substack{\ell\in\mathcal{N}\\ \ell\neq j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j}^{(k)}. \tag{107}$$

Therefore, since (91) holds true, and assuming that (92) holds for an arbitrary $k \ge 2$, we have:

$$\begin{aligned}
\boldsymbol{a}_{ij}^{(k+1)} &\le \overline{\boldsymbol{\alpha}}_k \boldsymbol{a}_{ij} + \overline{\boldsymbol{\beta}}_k \sum_{\substack{\ell\in\mathcal{N}\\ \ell\neq j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j} + \overline{\boldsymbol{\gamma}}_k\,\mathfrak{M} \sum_{\substack{\ell\in\mathcal{N}\\ \ell\neq j}} \boldsymbol{a}_{i\ell} \\
&\le (\overline{\boldsymbol{\alpha}}_k + \overline{\boldsymbol{\beta}}_k\,\mathfrak{M}_{a,\text{self}})\boldsymbol{a}_{ij} + (\overline{\boldsymbol{\beta}}_k + \overline{\boldsymbol{\gamma}}_k\,\mathfrak{M}_{a,\text{sum}})\,\mathfrak{M},
\end{aligned} \tag{108}$$

which shows that the right inequality in (92) holds with the sequences $\overline{\boldsymbol{\beta}}_k$ and $\overline{\boldsymbol{\gamma}}_k$ obeying (94) and (95), respectively, with the initialization choices in (96). ■

Let us introduce the following series (all the infinite summations will be shown to be bounded), defined in terms of the variables introduced in Lemma 3:

$$\overline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{even})} \triangleq \sum_{h=1}^{\infty} \overline{\boldsymbol{\alpha}}_{2h}, \quad \underline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{even})} \triangleq \sum_{h=1}^{\infty} \underline{\boldsymbol{\alpha}}_{2h}, \tag{109}$$

$$\overline{\boldsymbol{\Sigma}}_{\beta}^{(\text{even})} \triangleq \sum_{h=1}^{\infty} \overline{\boldsymbol{\beta}}_{2h}, \quad \underline{\boldsymbol{\Sigma}}_{\beta}^{(\text{even})} \triangleq \sum_{h=1}^{\infty} \underline{\boldsymbol{\beta}}_{2h}, \tag{110}$$

$$\overline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{even})} \triangleq \sum_{h=1}^{\infty} \overline{\boldsymbol{\gamma}}_{2h}, \quad \underline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{even})} \triangleq \sum_{h=1}^{\infty} \underline{\boldsymbol{\gamma}}_{2h}, \tag{111}$$

and

$$\overline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{odd})} \triangleq \sum_{h=1}^{\infty} \overline{\boldsymbol{\alpha}}_{2h+1}, \quad \underline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{odd})} \triangleq \sum_{h=1}^{\infty} \underline{\boldsymbol{\alpha}}_{2h+1}, \tag{112}$$

$$\overline{\boldsymbol{\Sigma}}_{\beta}^{(\text{odd})} \triangleq \sum_{h=1}^{\infty} \overline{\boldsymbol{\beta}}_{2h+1}, \quad \underline{\boldsymbol{\Sigma}}_{\beta}^{(\text{odd})} \triangleq \sum_{h=1}^{\infty} \underline{\boldsymbol{\beta}}_{2h+1}, \tag{113}$$

$$\overline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{odd})} \triangleq \sum_{h=1}^{\infty} \overline{\boldsymbol{\gamma}}_{2h+1}, \quad \underline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{odd})} \triangleq \sum_{h=1}^{\infty} \underline{\boldsymbol{\gamma}}_{2h+1}. \tag{114}$$

We are now ready to state and prove the first theorem, which provides upper and lower bounds on sums of powers of the matrix $\boldsymbol{A}$. These bounds will be critical to examine the concentration behavior of the one-lag and residual estimators.

The main message conveyed by the theorem is that, for all $i, j \in \mathbb{N}$, the individual (i, j) entries of sums of powers of $\boldsymbol{A}$ can be upper and lower bounded in terms of the elements $\boldsymbol{a}_{ij}$, and in terms of suitable bounding random variables that do not depend on the node indices i and j. These bounding random variables are $\mathfrak{m}$ and $\mathfrak{M}$ on line 8) of Table 3, and the "$\boldsymbol{\Sigma}$" variables appearing in (109)–(114). We remark that, in order to prove our main results in Theorem 2, we will not need the explicit expression of these "$\boldsymbol{\Sigma}$", but only their limiting properties, detailed in (118) and (119).

Theorem 5 (Bounds on the sum of powers of $\boldsymbol{A}$*):* The combination matrix $\boldsymbol{A}$ fulfills the following bounds for all $i, j \in \mathbb{N}$:

$$\underline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{even})} \leq \sum_{h=1}^{\infty} \boldsymbol{a}_{ii}^{(2h)} \leq \overline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{even})}, \underline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{odd})} \leq \sum_{h=1}^{\infty} \boldsymbol{a}_{ii}^{(2h+1)} \leq \overline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{odd})}, \tag{115}$$

and for $i \neq j$:

$$\boldsymbol{a}_{ij}\, \underline{\boldsymbol{\Sigma}}_{\beta}^{(\text{even})} + \mathfrak{m}\, \underline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{even})} \leq \sum_{h=1}^{\infty} \boldsymbol{a}_{ij}^{(2h)} \leq \boldsymbol{a}_{ij}\, \overline{\boldsymbol{\Sigma}}_{\beta}^{(\text{even})} + \mathfrak{M}\, \overline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{even})}, \tag{116}$$

$$\boldsymbol{a}_{ij}\, \underline{\boldsymbol{\Sigma}}_{\beta}^{(\text{odd})} + \mathfrak{m}\, \underline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{odd})} \leq \sum_{h=1}^{\infty} \boldsymbol{a}_{ij}^{(2h+1)} \leq \boldsymbol{a}_{ij}\, \overline{\boldsymbol{\Sigma}}_{\beta}^{(\text{odd})} + \mathfrak{M}\, \overline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{odd})}. \tag{117}$$

If $\boldsymbol{A}$ is a regular diffusion matrix of parameters ρ and κ according to Assumption 1, then under the uniform concentration regime the bounding variables "$\boldsymbol{\Sigma}$" converge in probability, as $N \to \infty$, to deterministic quantities, namely,

$$\begin{aligned}
&\overline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{even})} \text{ and } \underline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{even})} \xrightarrow{\text{p}} \frac{\zeta^2}{1-\zeta^2},\\
&\overline{\boldsymbol{\Sigma}}_{\beta}^{(\text{even})} \text{ and } \underline{\boldsymbol{\Sigma}}_{\beta}^{(\text{even})} \xrightarrow{\text{p}} \frac{2\zeta}{(1-\zeta^2)^2},\\
&\overline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{even})} \text{ and } \underline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{even})} \xrightarrow{\text{p}} \frac{1+\zeta^2+2\rho\,\zeta}{(1-\rho^2)(1-\zeta^2)^2},
\end{aligned} \tag{118}$$

$$\begin{aligned}
&\overline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{odd})} \text{ and } \underline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{odd})} \xrightarrow{\text{p}} \frac{\zeta^3}{1-\zeta^2},\\
&\overline{\boldsymbol{\Sigma}}_{\beta}^{(\text{odd})} \text{ and } \underline{\boldsymbol{\Sigma}}_{\beta}^{(\text{odd})} \xrightarrow{\text{p}} \frac{3\,\zeta^2-\zeta^4}{(1-\zeta^2)^2},\\
&\overline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{odd})} \text{ and } \underline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{odd})} \xrightarrow{\text{p}} \frac{\rho+\rho\,\zeta^2+2\,\zeta}{(1-\rho^2)(1-\zeta^2)^2}.
\end{aligned} \tag{119}$$

where we set $\zeta = \rho - \kappa$.

Proof: Calling upon Lemma 3, and using in (91) and (92) the definitions (109)–(114), we immediately get (115), (116), and (117). However, it is necessary to show that the "$\boldsymbol{\Sigma}$" random variables are proper random variables, i.e., that all the infinite summations in (109)–(114) are bounded. We will explain this conclusion with reference to the upper bounding sequences, with the case of the lower bounding sequences being dealt with similarly. The system of recursions in (93)–(95) can be solved by calculating first $\overline{\boldsymbol{\alpha}}_k$, then $\overline{\boldsymbol{\beta}}_k$ (after substituting $\overline{\boldsymbol{\alpha}}_k$) and finally $\overline{\boldsymbol{\gamma}}_k$ (after substituting $\overline{\boldsymbol{\beta}}_k$). Applying Lemma 7, it can be verified that all the obtained solutions are linear combinations of geometric sequences with ratio strictly smaller than one, from which finiteness of the summations $\overline{\boldsymbol{\Sigma}}_{\alpha}$, $\overline{\boldsymbol{\Sigma}}_{\beta}$ and $\overline{\boldsymbol{\Sigma}}_{\gamma}$ follows.

Next we focus on proving the convergence in probability in (118) and (119). To this end, it is instrumental to introduce the following auxiliary series:

$$\begin{aligned}
\overline{\boldsymbol{\Sigma}}_{\alpha} &\triangleq \overline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{even})} + \overline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{odd})}, & \underline{\boldsymbol{\Sigma}}_{\alpha} &\triangleq \underline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{even})} + \underline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{odd})},\\
\overline{\boldsymbol{\Sigma}}_{\beta} &\triangleq \overline{\boldsymbol{\Sigma}}_{\beta}^{(\text{even})} + \overline{\boldsymbol{\Sigma}}_{\beta}^{(\text{odd})}, & \underline{\boldsymbol{\Sigma}}_{\beta} &\triangleq \underline{\boldsymbol{\Sigma}}_{\beta}^{(\text{even})} + \underline{\boldsymbol{\Sigma}}_{\beta}^{(\text{odd})},\\
\overline{\boldsymbol{\Sigma}}_{\gamma} &\triangleq \overline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{even})} + \overline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{odd})}, & \underline{\boldsymbol{\Sigma}}_{\gamma} &\triangleq \underline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{even})} + \underline{\boldsymbol{\Sigma}}_{\gamma}^{(\text{odd})}.
\end{aligned} \tag{120}$$

We start by showing that the following convergence takes place:

$$\begin{aligned}
&\overline{\boldsymbol{\Sigma}}_{\alpha} \text{ and } \underline{\boldsymbol{\Sigma}}_{\alpha} \xrightarrow{\text{p}} \frac{\zeta^2}{1-\zeta},\\
&\overline{\boldsymbol{\Sigma}}_{\beta} \text{ and } \underline{\boldsymbol{\Sigma}}_{\beta} \xrightarrow{\text{p}} \frac{1-(1-\zeta)^2}{(1-\zeta)^2},\\
&\overline{\boldsymbol{\Sigma}}_{\gamma} \text{ and } \underline{\boldsymbol{\Sigma}}_{\gamma} \xrightarrow{\text{p}} \frac{1}{(1-\rho)(1-\zeta)^2}.
\end{aligned} \tag{121}$$

First of all, it is convenient to rewrite the series in (120) in the following more explicit form:

$$\overline{\boldsymbol{\Sigma}}_{\alpha} \triangleq \sum_{k=2}^{\infty} \overline{\boldsymbol{\alpha}}_{k}, \quad \underline{\boldsymbol{\Sigma}}_{\alpha} \triangleq \sum_{k=2}^{\infty} \underline{\boldsymbol{\alpha}}_{k}, \tag{122}$$

$$\overline{\boldsymbol{\Sigma}}_\beta \triangleq \sum_{k=2}^{\infty} \overline{\boldsymbol{\beta}}_k, \quad \underline{\boldsymbol{\Sigma}}_\beta \triangleq \sum_{k=2}^{\infty} \underline{\boldsymbol{\beta}}_k, \tag{123}$$

$$\overline{\boldsymbol{\Sigma}}_\gamma \triangleq \sum_{k=2}^{\infty} \overline{\boldsymbol{\gamma}}_k, \quad \underline{\boldsymbol{\Sigma}}_\gamma \triangleq \sum_{k=2}^{\infty} \underline{\boldsymbol{\gamma}}_k. \tag{124}$$

Let us consider (93). By summing over index k and using (122), we can write:

$$\overline{\boldsymbol{\Sigma}}_\alpha = \overline{\boldsymbol{\alpha}}_2 + \mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\Sigma}}_\alpha + \frac{\mathfrak{M}_a\, \rho^2}{1-\rho} \Rightarrow \overline{\boldsymbol{\Sigma}}_\alpha = \frac{\mathfrak{M}_{a_2,\text{self}} + \boldsymbol{\epsilon}}{1 - \mathfrak{M}_{a,\text{self}}}, \tag{125}$$

where we have set $\boldsymbol{\epsilon} = \frac{\mathfrak{M}_a\, \rho^2}{1-\rho}$. Likewise, operating on (94) and using (123), we get:

$$\overline{\boldsymbol{\Sigma}}_\beta = \overline{\boldsymbol{\beta}}_2 + \overline{\boldsymbol{\Sigma}}_\alpha + \mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\Sigma}}_\beta \Rightarrow \overline{\boldsymbol{\Sigma}}_\beta = \frac{2\,\mathfrak{M}_{a,\text{self}} + \overline{\boldsymbol{\Sigma}}_\alpha}{1 - \mathfrak{M}_{a,\text{self}}}. \tag{126}$$

Finally, applying the above procedure to (95) and using (124), we obtain:

$$\overline{\boldsymbol{\Sigma}}_\gamma = \overline{\boldsymbol{\gamma}}_2 + \overline{\boldsymbol{\Sigma}}_\beta + \mathfrak{M}_{a,\text{sum}}\, \overline{\boldsymbol{\Sigma}}_\gamma \Rightarrow \overline{\boldsymbol{\Sigma}}_\gamma = \frac{1 + \overline{\boldsymbol{\Sigma}}_\beta}{1 - \mathfrak{M}_{a,\text{sum}}}, \tag{127}$$

which, using (125) and (126), after straightforward algebra yields:

$$\overline{\boldsymbol{\Sigma}}_\gamma = \frac{1 - \mathfrak{M}^2_{a,\text{self}} + \mathfrak{M}_{a_2,\text{self}} + \boldsymbol{\epsilon}}{(1 - \mathfrak{M}_{a,\text{sum}})(1 - \mathfrak{M}_{a,\text{self}})^2}. \tag{128}$$

Now, in view of the convergences in probability proved in Lemma 6 — specifically, in view of (206), (209), (212) — it is legitimate to replace the pertinent variables in (125), (126), and (128), with their limits (that are reported in Table 3). After some lengthy, though straightforward, algebra, this replacement leads to (121).

We now move on to prove (119). Using the definitions of $\overline{\boldsymbol{\Sigma}}^{(\text{even})}_\alpha$ and $\overline{\boldsymbol{\Sigma}}^{(\text{odd})}_\alpha$ in (109) and (112), and summing over k the terms in (93), we see that:

$$\begin{aligned}
\overline{\boldsymbol{\Sigma}}^{(\text{odd})}_\alpha &= \sum_{\substack{k=3\\ k \text{ odd}}}^{\infty} \overline{\boldsymbol{\alpha}}_k = \sum_{\substack{k=2\\ k \text{ even}}}^{\infty} \overline{\boldsymbol{\alpha}}_{k+1} \\
&= \mathfrak{M}_{a,\text{self}} \sum_{\substack{k=2\\ k \text{ even}}}^{\infty} \overline{\boldsymbol{\alpha}}_k + \mathfrak{M}_a \sum_{\substack{k=2\\ k \text{ even}}}^{\infty} \rho^k \\
&= \mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\Sigma}}^{(\text{even})}_\alpha + \mathfrak{M}_a \sum_{k=1}^{\infty} \rho^{2k} \\
&= \mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\Sigma}}_\alpha - \mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\Sigma}}^{(\text{odd})}_\alpha + \mathfrak{M}_a \frac{\rho^2}{1-\rho^2},
\end{aligned} \tag{129}$$

yielding:

$$\overline{\boldsymbol{\Sigma}}^{(\text{odd})}_\alpha = \frac{\mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\Sigma}}_\alpha + \boldsymbol{\epsilon}'}{1 + \mathfrak{M}_{a,\text{self}}}, \tag{130}$$

where we defined $\boldsymbol{\epsilon}' = \mathfrak{M}_a \frac{\rho^2}{1-\rho^2}$. Using now (120) and (130), we further obtain:

$$\overline{\boldsymbol{\Sigma}}^{(\text{even})}_\alpha = \frac{\overline{\boldsymbol{\Sigma}}_\alpha - \boldsymbol{\epsilon}'}{1 + \mathfrak{M}_{a,\text{self}}}. \tag{131}$$

Proceeding in a similar way, from (94) we obtain:

$$\begin{aligned}
\overline{\boldsymbol{\Sigma}}^{(\text{odd})}_\beta &= \overline{\boldsymbol{\Sigma}}^{(\text{even})}_\alpha + \mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\Sigma}}^{(\text{even})}_\beta \\
&= \overline{\boldsymbol{\Sigma}}^{(\text{even})}_\alpha + \mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\Sigma}}_\beta - \mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\Sigma}}^{(\text{odd})}_\beta,
\end{aligned} \tag{132}$$

yielding:

$$\overline{\boldsymbol{\Sigma}}^{(\text{odd})}_\beta = \frac{\mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\Sigma}}_\beta + \overline{\boldsymbol{\Sigma}}^{(\text{even})}_\alpha}{1 + \mathfrak{M}_{a,\text{self}}}, \tag{133}$$

and

$$\overline{\boldsymbol{\Sigma}}^{(\text{even})}_\beta = \frac{\overline{\boldsymbol{\Sigma}}_\beta - \overline{\boldsymbol{\Sigma}}^{(\text{even})}_\alpha}{1 + \mathfrak{M}_{a,\text{self}}}. \tag{134}$$

Finally, from (95), we can write:

$$\begin{aligned}
\overline{\boldsymbol{\Sigma}}^{(\text{odd})}_\gamma &= \overline{\boldsymbol{\Sigma}}^{(\text{even})}_\beta + \mathfrak{M}_{a,\text{sum}}\, \overline{\boldsymbol{\Sigma}}^{(\text{even})}_\gamma \\
&= \overline{\boldsymbol{\Sigma}}^{(\text{even})}_\beta + \mathfrak{M}_{a,\text{sum}}\, \overline{\boldsymbol{\Sigma}}_\gamma - \mathfrak{M}_{a,\text{sum}}\, \overline{\boldsymbol{\Sigma}}^{(\text{odd})}_\beta,
\end{aligned} \tag{135}$$

yielding:

$$\overline{\boldsymbol{\Sigma}}^{(\text{odd})}_\gamma = \frac{\mathfrak{M}_{a,\text{sum}}\, \overline{\boldsymbol{\Sigma}}_\gamma + \overline{\boldsymbol{\Sigma}}^{(\text{even})}_\beta}{1 + \mathfrak{M}_{a,\text{sum}}}, \tag{136}$$

and

$$\overline{\boldsymbol{\Sigma}}^{(\text{even})}_\gamma = \frac{\overline{\boldsymbol{\Sigma}}_\gamma - \overline{\boldsymbol{\Sigma}}^{(\text{even})}_\beta}{1 + \mathfrak{M}_{a,\text{sum}}}. \tag{137}$$

The limiting results in (118) and (119) are now obtained by replacing, in (130), (131), (133), (134), (136), and (137), the pertinent random variables with their limiting counterparts shown in Table 3. ■

APPENDIX C
BOUNDS ON THE MATRIX H IN (36)

The following technical lemma is instrumental to prove Theorem 6. We recall that, from (36), we have the definition $\boldsymbol{C} \triangleq [\boldsymbol{A}^2]_{\mathcal{S}'}$.

Lemma 4 (Bounds on the entries of $\boldsymbol{C}^k$*):* The entries of the matrix power $\boldsymbol{C}^k$ are bounded as follows:

$$\underline{\boldsymbol{\alpha}}'_k \le \boldsymbol{c}^{(k)}_{\ell\ell} \le \overline{\boldsymbol{\alpha}}'_k, \tag{138}$$

and, for $\ell \neq m$:

$$\underline{\boldsymbol{\beta}}'_k\, \boldsymbol{a}_{\ell m} + \underline{\boldsymbol{\gamma}}'_k \le \boldsymbol{c}^{(k)}_{\ell m} \le \overline{\boldsymbol{\beta}}'_k\, \boldsymbol{a}_{\ell m} + \overline{\boldsymbol{\gamma}}'_k, \tag{139}$$

where, for $k \ge 1$, the (random) sequences $\overline{\boldsymbol{\alpha}}'_k$, $\underline{\boldsymbol{\alpha}}'_k$, $\overline{\boldsymbol{\beta}}'_k$, $\underline{\boldsymbol{\beta}}'_k$, $\overline{\boldsymbol{\gamma}}'_k$, and $\underline{\boldsymbol{\gamma}}'_k$, are determined by the following recursions:

$$\overline{\boldsymbol{\alpha}}'_{k+1} = \mathfrak{M}_{c,\text{self}}\, \overline{\boldsymbol{\alpha}}'_k + (2\,\mathfrak{M}_{a,\text{self}}\, \mathfrak{M}_a + \mathfrak{M})\, \rho^{2k}, \tag{140}$$

$$\overline{\boldsymbol{\beta}}'_{k+1} = 2\,\mathfrak{M}_{a,\text{self}}\, \overline{\boldsymbol{\alpha}}'_k + \mathfrak{M}_{c,\text{self}}\, \overline{\boldsymbol{\beta}}'_k, \tag{141}$$

$$\overline{\boldsymbol{\gamma}}'_{k+1} = \mathfrak{M}\,\overline{\boldsymbol{\alpha}}'_k + (2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}^{(\mathcal{S}')} + \mathfrak{M}\,\mathfrak{M}^{(\mathcal{S}')}_{a,\text{sum}})\overline{\boldsymbol{\beta}}'_k + \mathfrak{M}_{c,\text{sum}}\,\overline{\boldsymbol{\gamma}}'_k, \tag{142}$$

with the initialization choices:

$$\overline{\boldsymbol{\alpha}}'_1 = \mathfrak{M}_{c,\text{self}}, \quad \overline{\boldsymbol{\beta}}'_1 = 2\,\mathfrak{M}_{a,\text{self}}, \quad \overline{\boldsymbol{\gamma}}'_1 = \mathfrak{M}, \tag{143}$$

and

$$\underline{\boldsymbol{\alpha}}'_{k+1} = \mathfrak{m}_{c,\text{self}}\,\underline{\boldsymbol{\alpha}}'_k, \tag{144}$$

$$\underline{\boldsymbol{\beta}}'_{k+1} = 2\,\mathfrak{m}_{a,\text{self}}\,\underline{\boldsymbol{\alpha}}'_k + \mathfrak{m}_{c,\text{self}}\,\underline{\boldsymbol{\beta}}'_k, \tag{145}$$

$$\underline{\boldsymbol{\gamma}}'_{k+1} = \mathfrak{m}\,\underline{\boldsymbol{\alpha}}'_k + (2\,\mathfrak{m}_{a,\text{self}}\,\mathfrak{m}^{(\mathcal{S}')} + \mathfrak{m}\,\mathfrak{m}^{(\mathcal{S}')}_{a,\text{sum}})\underline{\boldsymbol{\beta}}'_k + \mathfrak{m}_{c,\text{sum}}\,\underline{\boldsymbol{\gamma}}'_k, \tag{146}$$

with the initialization choices:

$$\underline{\boldsymbol{\alpha}}'_1 = \mathfrak{m}_{c,\text{self}}, \quad \underline{\boldsymbol{\beta}}'_1 = 2\,\mathfrak{m}_{a,\text{self}}, \quad \underline{\boldsymbol{\gamma}}'_1 = \mathfrak{m}. \tag{147}$$

Proof: We start with the inequalities pertaining to the main diagonal of matrices $\boldsymbol{C}^k$, namely, with (138). For $k = 1$ we use the definitions of $\overline{\boldsymbol{\alpha}}'_1$ and $\underline{\boldsymbol{\alpha}}'_1$ in (143) and (147), respectively, to see that (138) is trivially satisfied in view of the definitions appearing on line 6) of Table 3. We shall now prove that (138) holds for an arbitrary k by induction. Assume thus that (138) is true for k, we need to show that it is true for $k+1$. From the definition of matrix $\boldsymbol{C}$ in (36), we can write its terms on the main diagonal as:

$$\boldsymbol{c}^{(k+1)}_{\ell\ell} = \sum_{h\in\mathcal{S}'} \boldsymbol{c}_{\ell h}\boldsymbol{c}^{(k)}_{h\ell} = \boldsymbol{c}_{\ell\ell}\boldsymbol{c}^{(k)}_{\ell\ell} + \sum_{\substack{h\in\mathcal{S}'\\h\neq\ell}} \boldsymbol{c}_{\ell h}\boldsymbol{c}^{(k)}_{h\ell}. \tag{148}$$

In view of (106) we can write:

$$\sum_{\substack{h\in\mathcal{S}'\\h\neq\ell}} \boldsymbol{c}_{\ell h}\boldsymbol{c}^{(k)}_{h\ell} = \sum_{\substack{h\in\mathcal{S}'\\h\neq\ell}} \boldsymbol{a}^{(2)}_{\ell h}\boldsymbol{c}^{(k)}_{h\ell} \leq (2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}_a + \mathfrak{M}) \sum_{\substack{h\in\mathcal{S}'\\h\neq\ell}} \boldsymbol{c}^{(k)}_{h\ell}. \tag{149}$$

Moreover, since $\boldsymbol{C} = [\boldsymbol{A}^2]_{\mathcal{S}'}$, the first relationship in (44) can be recursively applied to show that:

$$\sum_{\substack{h\in\mathcal{S}'\\h\neq\ell}} \boldsymbol{c}^{(k)}_{h\ell} \leq \rho^{2k}. \tag{150}$$

Therefore, from (148), (149) and (150) we conclude that:

$$\mathfrak{m}_{c,\text{self}}\,\boldsymbol{c}^{(k)}_{\ell\ell} \leq \boldsymbol{c}^{(k+1)}_{\ell\ell} \leq \mathfrak{M}_{c,\text{self}}\,\boldsymbol{c}^{(k)}_{\ell\ell} + (2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}_a + \mathfrak{M})\,\rho^{2k}, \tag{151}$$

having further used the definitions on line 6) of Table 3 to bound the term $\boldsymbol{c}_{\ell\ell}$. Since we have assumed that (138) holds true for k, we can further apply (138) into (151), yielding:

$$\mathfrak{m}_{c,\text{self}}\,\underline{\boldsymbol{\alpha}}'_k \leq \boldsymbol{c}^{(k+1)}_{\ell\ell} \leq \mathfrak{M}_{c,\text{self}}\,\overline{\boldsymbol{\alpha}}'_k + (2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}_a + \mathfrak{M})\,\rho^{2k}, \tag{152}$$

which reveals that (138) holds true for $k+1$, with the sequences $\overline{\boldsymbol{\alpha}}'_k$ and $\underline{\boldsymbol{\alpha}}'_k$ obeying the recursions in (140) and (144), respectively.

Let us move on to examine the case $\ell \neq m$. For all $\ell, m \in \mathcal{S}'$, with $\ell \neq m$, we can use (105) to conclude that:

$$\boldsymbol{c}_{\ell m} \leq 2\,\mathfrak{M}_{a,\text{self}}\,\boldsymbol{a}_{\ell m} + \mathfrak{M}. \tag{153}$$

Equation (153) shows that the upper bound in (139) holds for $k = 1$, with the choices in (143). Now, rewriting the relationship $\boldsymbol{C}^{k+1} = \boldsymbol{C}\boldsymbol{C}^k$ on an entrywise basis, we have:

$$\boldsymbol{c}^{(k+1)}_{\ell m} = \sum_{h\in\mathcal{S}'} \boldsymbol{c}_{\ell h}\boldsymbol{c}^{(k)}_{hm} = \boldsymbol{c}_{\ell m}\boldsymbol{c}^{(k)}_{mm} + \sum_{\substack{h\in\mathcal{S}'\\h\neq m}} \boldsymbol{c}_{\ell h}\boldsymbol{c}^{(k)}_{hm}. \tag{154}$$

Now, we can bound $\boldsymbol{c}^{(k)}_{mm}$ by applying (138) and $\boldsymbol{c}_{\ell m}$ by applying (153). Moreover, assuming as induction hypothesis that (139) holds true for an arbitrary k, we can bound $\boldsymbol{c}^{(k)}_{hm}$, finally obtaining from (154):

$$\begin{aligned}\boldsymbol{c}^{(k+1)}_{\ell m} &\leq \overline{\boldsymbol{\alpha}}'_k(2\,\mathfrak{M}_{a,\text{self}}\,\boldsymbol{a}_{\ell m} + \mathfrak{M}) \\ &\quad + \overline{\boldsymbol{\beta}}'_k \sum_{\substack{h\in\mathcal{S}'\\h\neq m}} \boldsymbol{c}_{\ell h}\boldsymbol{a}_{hm} + \overline{\boldsymbol{\gamma}}'_k \sum_{\substack{h\in\mathcal{S}'\\h\neq m}} \boldsymbol{c}_{\ell h} \\ &\leq \overline{\boldsymbol{\alpha}}'_k(2\,\mathfrak{M}_{a,\text{self}}\,\boldsymbol{a}_{\ell m} + \mathfrak{M}) \\ &\quad + \overline{\boldsymbol{\beta}}'_k \sum_{\substack{h\in\mathcal{S}'\\h\neq m}} \boldsymbol{c}_{\ell h}\boldsymbol{a}_{hm} + \mathfrak{M}_{c,\text{sum}}\,\overline{\boldsymbol{\gamma}}'_k,\end{aligned} \tag{155}$$

where in the last step we applied the definition of $\mathfrak{M}_{c,\text{sum}}$ appearing on line 7) of Table 3 to bound the sum of the $\boldsymbol{c}_{\ell h}$. Let us focus on the last summation in (155). Using the definitions on line 6) of Table 3 to bound the term $\boldsymbol{c}_{\ell\ell}$, and (153) to bound the term $\boldsymbol{c}_{\ell h}$ for $h \neq \ell$, we get:

$$\begin{aligned}\sum_{\substack{h\in\mathcal{S}'\\h\neq m}} \boldsymbol{c}_{\ell h}\boldsymbol{a}_{hm} &= \boldsymbol{c}_{\ell\ell}\boldsymbol{a}_{\ell m} + \sum_{\substack{h\in\mathcal{S}'\\h\neq\ell,m}} \boldsymbol{c}_{\ell h}\boldsymbol{a}_{hm} \\ &\leq \mathfrak{M}_{c,\text{self}}\,\boldsymbol{a}_{\ell m} + 2\,\mathfrak{M}_{a,\text{self}} \sum_{\substack{h\in\mathcal{S}'\\h\neq\ell,m}} \boldsymbol{a}_{\ell h}\boldsymbol{a}_{hm} \\ &\quad + \mathfrak{M} \sum_{\substack{h\in\mathcal{S}'\\h\neq\ell,m}} \boldsymbol{a}_{hm} \\ &\leq \mathfrak{M}_{c,\text{self}}\,\boldsymbol{a}_{\ell m} + 2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}^{(\mathcal{S}')} + \mathfrak{M}\,\mathfrak{M}^{(\mathcal{S}')}_{a,\text{sum}},\end{aligned} \tag{156}$$

where, in the latter inequality, we have further applied the definitions listed on lines 5) and 9) of Table 3. Using now (156) into (155) we get:

$$\boldsymbol{c}^{(k+1)}_{\ell m} \leq \underbrace{(2\,\mathfrak{M}_{a,\text{self}}\,\overline{\boldsymbol{\alpha}}'_k + \mathfrak{M}_{c,\text{self}}\,\overline{\boldsymbol{\beta}}'_k)}_{\overline{\boldsymbol{\beta}}'_{k+1}}\boldsymbol{a}_{\ell m}$$

$$+\underbrace{[\mathfrak{M}\,\overline{\boldsymbol{\alpha}}'_k + (2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}^{(\mathcal{S}')} + \mathfrak{M}\,\mathfrak{M}^{(\mathcal{S}')}_{a,\text{sum}})\overline{\boldsymbol{\beta}}'_k + \mathfrak{M}_{c,\text{sum}}\,\overline{\boldsymbol{\gamma}}'_k]}_{\overline{\gamma}'_{k+1}}, \tag{157}$$

which shows that the right inequality in (139) holds also for $k+1$, with $\overline{\boldsymbol{\beta}}'_k$, and $\overline{\boldsymbol{\gamma}}'_k$ obeying the recursions in (141) and (142), respectively, with the initialization choices in (143). The proof of the left inequality in (139) is similar. ■

Let us now introduce the following series (all the infinite summations will be shown to be bounded):

$$\overline{\boldsymbol{\Phi}}_\alpha \triangleq \sum_{k=1}^{\infty} \overline{\boldsymbol{\alpha}}'_k, \quad \underline{\boldsymbol{\Phi}}_\alpha \triangleq \sum_{k=1}^{\infty} \underline{\boldsymbol{\alpha}}'_k, \tag{158}$$

$$\overline{\boldsymbol{\Phi}}_\beta \triangleq \sum_{k=1}^{\infty} \overline{\boldsymbol{\beta}}'_k, \quad \underline{\boldsymbol{\Phi}}_\beta \triangleq \sum_{k=1}^{\infty} \underline{\boldsymbol{\beta}}'_k, \tag{159}$$

$$\overline{\boldsymbol{\Phi}}_\gamma \triangleq \sum_{k=1}^{\infty} \overline{\boldsymbol{\gamma}}'_k, \quad \underline{\boldsymbol{\Phi}}_\gamma \triangleq \sum_{k=1}^{\infty} \underline{\boldsymbol{\gamma}}'_k. \tag{160}$$

The next theorem provides upper and lower bounds on the matrix $\boldsymbol{H}$ introduced in (36). These bounds will be critical to examine the concentration behavior of the Granger estimator.

Theorem 6 (Bounds on the matrix $\boldsymbol{H}$): The matrix $\boldsymbol{H}$ in (36) fulfills the following bounds, for all $\ell, m \in \mathcal{S}'$:

$$1 + \underline{\boldsymbol{\Phi}}_\alpha \le \boldsymbol{h}_{\ell\ell} \le 1 + \overline{\boldsymbol{\Phi}}_\alpha, \tag{161}$$

and for $\ell \ne m$:

$$\underline{\boldsymbol{\Phi}}_\beta\, \boldsymbol{a}_{\ell m} + \underline{\boldsymbol{\Phi}}_\gamma \le \boldsymbol{h}_{\ell m} \le \overline{\boldsymbol{\Phi}}_\beta\, \boldsymbol{a}_{\ell m} + \overline{\boldsymbol{\Phi}}_\gamma. \tag{162}$$

If $\boldsymbol{A}$ is a regular diffusion matrix of parameters ρ and κ according to Assumption 1, then under the uniform concentration regime the bounding variables "$\boldsymbol{\Phi}$" converge in probability, as $N \to \infty$, to deterministic quantities, namely,

$$\begin{aligned}
&\overline{\boldsymbol{\Phi}}_\alpha \text{ and } \underline{\boldsymbol{\Phi}}_\alpha \xrightarrow{\text{p}} \frac{\zeta^2}{1-\zeta^2},\\
&\overline{\boldsymbol{\Phi}}_\beta \text{ and } \underline{\boldsymbol{\Phi}}_\beta \xrightarrow{\text{p}} \frac{2\,\zeta}{(1-\zeta^2)^2},\\
&N p_N \overline{\boldsymbol{\Phi}}_\gamma \text{ and } N p_N \underline{\boldsymbol{\Phi}}_\gamma \xrightarrow{\text{p}} \phi,
\end{aligned} \tag{163}$$

where we set:

$$\phi \triangleq \kappa^2 p \frac{1-\zeta^2 + 2\,\zeta[2\,\zeta\,(1-\xi) + \kappa(1-\xi)]}{[1-(\rho^2 - 2\rho\kappa\xi + \kappa^2\xi)][1-\zeta^2]^2}, \quad \zeta = \rho - \kappa. \tag{164}$$

Proof: The matrix $\boldsymbol{H}$ defined in (36) can be expressed as:

$$\boldsymbol{H} = (I_{\mathcal{S}'} - \boldsymbol{C})^{-1} = I_{\mathcal{S}'} + \boldsymbol{C} + \boldsymbol{C}^2 + \cdots \tag{165}$$

Calling upon Lemma 4 and summing the inequalities in (138) over index k, from (158) we immediately get (161). Likewise, summing over k the inequalities in (139), from (159) and (160) we obtain (162).

However, it is necessary to show that the "$\boldsymbol{\Phi}$" random variables are proper random variables, i.e., that all the infinite summations in (158)–(160) are bounded. We will explain this conclusion with reference to the upper bounding sequences, with the case of the lower bounding sequences being dealt with similarly. The system of recursions in (140)–(142) can be solved by calculating first $\overline{\boldsymbol{\alpha}}'_k$, then $\overline{\boldsymbol{\beta}}'_k$ (after substituting $\overline{\boldsymbol{\alpha}}'_k$) and finally $\overline{\boldsymbol{\gamma}}'_k$ (after substituting $\overline{\boldsymbol{\alpha}}'_k$ and $\overline{\boldsymbol{\beta}}'_k$). Applying Lemma 7, it can be verified that all the obtained solutions are linear combinations of geometric sequences with ratio strictly smaller than one, from which convergence of the series $\overline{\boldsymbol{\Phi}}_\alpha$, $\overline{\boldsymbol{\Phi}}_\beta$ and $\overline{\boldsymbol{\Phi}}_\gamma$ in (158)–(160) follows.

Next we focus on proving the convergence in probability in (163). Let us consider (140). By summing over index k and using (158), we can write:

$$\overline{\boldsymbol{\Phi}}_\alpha = \overline{\boldsymbol{\alpha}}'_1 + \mathfrak{M}_{c,\text{self}}\,\overline{\boldsymbol{\Phi}}_\alpha + \boldsymbol{\epsilon} \Rightarrow \overline{\boldsymbol{\Phi}}_\alpha = \frac{\mathfrak{M}_{c,\text{self}} + \boldsymbol{\epsilon}}{1-\mathfrak{M}_{c,\text{self}}}, \tag{166}$$

where we have set:

$$\boldsymbol{\epsilon} = (2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}_a + \mathfrak{M})\,\frac{\rho^2}{1-\rho^2}. \tag{167}$$

Likewise, from (141), summing over index k and using (159), we can write:

$$\overline{\boldsymbol{\Phi}}_\beta = \overline{\boldsymbol{\beta}}'_1 + 2\,\mathfrak{M}_{a,\text{self}}\,\overline{\boldsymbol{\Phi}}_\alpha + \mathfrak{M}_{c,\text{self}}\,\overline{\boldsymbol{\Phi}}_\beta, \tag{168}$$

yielding:

$$\overline{\boldsymbol{\Phi}}_\beta = \frac{2\,\mathfrak{M}_{a,\text{self}}\,(1+\overline{\boldsymbol{\Phi}}_\alpha)}{1-\mathfrak{M}_{c,\text{self}}}. \tag{169}$$

Using now (166) into (169), we get:

$$\overline{\boldsymbol{\Phi}}_\beta = \frac{2\,\mathfrak{M}_{a,\text{self}}}{(1-\mathfrak{M}_{c,\text{self}})^2}(1+\boldsymbol{\epsilon}). \tag{170}$$

Finally, applying the same procedure to (142) and using (160), we obtain:

$$\begin{aligned}
\overline{\boldsymbol{\Phi}}_\gamma &= \overline{\boldsymbol{\gamma}}'_1 + \mathfrak{M}\,\overline{\boldsymbol{\Phi}}_\alpha + (2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}^{(\mathcal{S}')} + \mathfrak{M}\,\mathfrak{M}^{(\mathcal{S}')}_{a,\text{sum}})\overline{\boldsymbol{\Phi}}_\beta\\
&\quad + \mathfrak{M}_{c,\text{sum}}\,\overline{\boldsymbol{\Phi}}_\gamma,
\end{aligned} \tag{171}$$

which yields the solution:

$$\overline{\boldsymbol{\Phi}}_\gamma = \frac{\mathfrak{M}(1+\overline{\boldsymbol{\Phi}}_\alpha) + (2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}^{(\mathcal{S}')} + \mathfrak{M}\,\mathfrak{M}^{(\mathcal{S}')}_{a,\text{sum}})\overline{\boldsymbol{\Phi}}_\beta}{1-\mathfrak{M}_{c,\text{sum}}}. \tag{172}$$

Now, in view of the convergences in probability proved in Lemma 6 — specifically, in view of (206), (214), (223), (224) and (230) — it is legitimate to replace the pertinent variables with their limits (that are reported in Table 3) in (166), (169), and (172), which, after some lengthy, though straightforward, algebra, leads to (163). ■

APPENDIX D PROOF OF THEOREM 2

We start by proving an auxiliary lemma.

Lemma 5 (Sufficient conditions for universal local structural consistency): Let the network graph be drawn according to an Erdős-Rényi random graph model, and let $\boldsymbol{A}$ be a regular diffusion matrix with parameters ρ and κ. Let $\mathcal{S}$ be the set of observable nodes and consider then a limiting estimator:

$$\widehat{\boldsymbol{A}}_{\mathcal{S}} = \boldsymbol{A}_{\mathcal{S}} + \boldsymbol{E}. \tag{173}$$

Assume that, for all $i, j \in \mathcal{S}$, with $i \neq j$:

$$\underline{\boldsymbol{w}}_N \boldsymbol{a}_{ij} + \underline{z}_N \le \boldsymbol{e}_{ij} \le \overline{\boldsymbol{w}}_N \boldsymbol{a}_{ij} + \overline{z}_N, \tag{174}$$

where the quantities $\underline{\boldsymbol{w}}_N$, $\overline{\boldsymbol{w}}_N$, $\underline{z}_N$ and $\overline{z}_N$ do depend on the network size, N, but they do not depend on (i, j), and fulfill the following convergences:

$$\underline{\boldsymbol{w}}_N \xrightarrow{\text{p}} w, \qquad \overline{\boldsymbol{w}}_N \xrightarrow{\text{p}} w,$$
$$N p_N \underline{z}_N \xrightarrow{\text{p}} z, \qquad N p_N \overline{z}_N \xrightarrow{\text{p}} z. \tag{175}$$

Then, under the uniform concentration regime, the limiting estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ achieves universal local structural consistency, with scaling sequence $s_N = N p_N$, bias $\eta = z$, and identifiability gap $\Gamma = \kappa\,(1 + w)$.

Proof: Using (12) and (173) we can write:

$$\underline{\boldsymbol{\delta}}_N = \min_{\substack{i,j \in \mathcal{S}: \boldsymbol{a}_{ij} = 0 \\ i \neq j}} \boldsymbol{e}_{ij}, \qquad \overline{\boldsymbol{\delta}}_N = \max_{\substack{i,j \in \mathcal{S}: \boldsymbol{a}_{ij} = 0 \\ i \neq j}} \boldsymbol{e}_{ij}, \tag{176}$$

and, hence, from (174) we get:

$$N p_N \, \underline{z}_N \le N p_N \, \underline{\boldsymbol{\delta}}_N \le N p_N \, \overline{\boldsymbol{\delta}}_N \le N p_N \, \overline{z}_N. \tag{177}$$

Using (175), we conclude that:

$$N p_N \, \overline{\boldsymbol{\delta}}_N \xrightarrow{\text{p}} z, \qquad N p_N \, \underline{\boldsymbol{\delta}}_N \xrightarrow{\text{p}} z. \tag{178}$$

Let us now examine the connected pairs. From (174) we know that:

$$(1 + \underline{\boldsymbol{w}}_N) \boldsymbol{a}_{ij} + \underline{z}_N \le \boldsymbol{a}_{ij} + \boldsymbol{e}_{ij} \le (1 + \overline{\boldsymbol{w}}_N) \boldsymbol{a}_{ij} + \overline{z}_N, \tag{179}$$

which, used along with (13) gives:

$$N p_N \underline{\boldsymbol{\Delta}}_N \ge N p_N (1 + \underline{\boldsymbol{w}}_N) \min_{\substack{i,j \in \mathcal{S}: \boldsymbol{a}_{ij} > 0 \\ i \neq j}} \boldsymbol{a}_{ij} + N p_N \underline{z}_N, \tag{180}$$

and

$$N p_N \overline{\boldsymbol{\Delta}}_N \le N p_N (1 + \overline{\boldsymbol{w}}_N) \max_{\substack{i,j \in \mathcal{S}: \boldsymbol{a}_{ij} > 0 \\ i \neq j}} \boldsymbol{a}_{ij} + N p_N \overline{z}_N. \tag{181}$$

In view of Assumption 1 we can write, for all pairs (i, j) where $\boldsymbol{a}_{ij} > 0$:

$$\kappa \frac{N p_N}{\boldsymbol{d}_{\max}} \le N p_N \boldsymbol{a}_{ij} \le \kappa \frac{N p_N}{\boldsymbol{d}_{\min}}. \tag{182}$$

Using (182) in (180) and (181), calling upon Corollary 1, and exploiting the convergences in (175), we conclude that:

$$N p_N \, \underline{\boldsymbol{\Delta}}_N \xrightarrow{\text{p}} \kappa\,(1 + w) + z, N p_N \, \overline{\boldsymbol{\Delta}}_N \xrightarrow{\text{p}} \kappa\,(1 + w) + z. \tag{183}$$

It remains to apply the definition of bias and identifiability gap in (14) to get the claim of the lemma. ■

Proof of Theorem 2: The proof boils down to combining Theorem 6 with Lemma 5.

From (37) we can write, for $i, j \in \mathcal{S}$, with $i \neq j$:

$$\begin{aligned}
\boldsymbol{e}_{ij}^{(\mathrm{Gra})} &= \sum_{\ell, m \in \mathcal{S}'} \boldsymbol{a}_{i\ell} \boldsymbol{h}_{\ell m} \boldsymbol{a}_{mj}^{(2)} \\
&= \sum_{\ell \in \mathcal{S}'} \boldsymbol{a}_{i\ell} \boldsymbol{h}_{\ell\ell} \boldsymbol{a}_{\ell j}^{(2)} + \sum_{\substack{\ell, m \in \mathcal{S}' \\ \ell \neq m}} \boldsymbol{a}_{i\ell} \boldsymbol{h}_{\ell m} \boldsymbol{a}_{mj}^{(2)} \\
&\le (1 + \overline{\boldsymbol{\Phi}}_{\alpha}) \sum_{\ell \in \mathcal{S}'} \boldsymbol{a}_{i\ell} \boldsymbol{a}_{\ell j}^{(2)} && (184) \\
&\quad + \overline{\boldsymbol{\Phi}}_{\beta} \sum_{\substack{\ell, m \in \mathcal{S}' \\ \ell \neq m}} \boldsymbol{a}_{i\ell} \boldsymbol{a}_{\ell m} \boldsymbol{a}_{mj}^{(2)} && (185) \\
&\quad + \overline{\boldsymbol{\Phi}}_{\gamma} \sum_{\substack{\ell, m \in \mathcal{S}' \\ \ell \neq m}} \boldsymbol{a}_{i\ell} \boldsymbol{a}_{mj}^{(2)}, && (186)
\end{aligned}$$

where the inequality is obtained by bounding the entries of the matrix $\boldsymbol{H}$, specifically, we have that (184) follows from (161), whereas (185) and (186) follow from (162).

Let us focus on (186). Using (104) to bound the term $\boldsymbol{a}_{mj}^{(2)}$, we can write:

$$\begin{aligned}
\sum_{\substack{\ell, m \in \mathcal{S}' \\ \ell \neq m}} \boldsymbol{a}_{i\ell} \boldsymbol{a}_{mj}^{(2)} &\le 2\, \mathfrak{M}_{a,\mathrm{self}} \sum_{\substack{\ell, m \in \mathcal{S}' \\ \ell \neq m}} \boldsymbol{a}_{i\ell} \boldsymbol{a}_{mj} \\
&\quad + \sum_{\substack{\ell, m \in \mathcal{S}' \\ \ell \neq m}} \boldsymbol{a}_{i\ell} \sum_{\substack{h \in \mathcal{N} \\ h \neq m, j}} \boldsymbol{a}_{mh} \boldsymbol{a}_{hj}. && (187)
\end{aligned}$$

We now use (105) to bound the term $\boldsymbol{a}_{\ell j}^{(2)}$ in (184) and the term $\boldsymbol{a}_{mj}^{(2)}$ in (185), whereas we use (187) to bound (186), finally obtaining:

$$\begin{aligned}
\boldsymbol{e}_{ij}^{(\mathrm{Gra})} &\le (1 + \overline{\boldsymbol{\Phi}}_{\alpha}) \sum_{\ell \in \mathcal{S}'} \boldsymbol{a}_{i\ell} (2\, \mathfrak{M}_{a,\mathrm{self}} \boldsymbol{a}_{\ell j} + \mathfrak{M}) \\
&\quad + \overline{\boldsymbol{\Phi}}_{\beta} \sum_{\substack{\ell, m \in \mathcal{S}' \\ \ell \neq m}} \boldsymbol{a}_{i\ell} \boldsymbol{a}_{\ell m} (2\, \mathfrak{M}_{a,\mathrm{self}} \boldsymbol{a}_{mj} + \mathfrak{M}) \\
&\quad + 2 \overline{\boldsymbol{\Phi}}_{\gamma} \mathfrak{M}_{a,\mathrm{self}} \sum_{\substack{\ell, m \in \mathcal{S}' \\ \ell \neq m}} \boldsymbol{a}_{i\ell} \boldsymbol{a}_{mj} \\
&\quad + \overline{\boldsymbol{\Phi}}_{\gamma} \sum_{\ell \in \mathcal{S}'} \boldsymbol{a}_{i\ell} \sum_{\substack{h \in \mathcal{N} \\ h \neq j}} \boldsymbol{a}_{hj} \sum_{\substack{m \in \mathcal{S}' \\ m \neq \ell, h}} \boldsymbol{a}_{mh}, && (188)
\end{aligned}$$

which can be recast in the following convenient form:

$$\begin{aligned}
\boldsymbol{e}_{ij}^{(\mathrm{Gra})} &\le (1 + \overline{\boldsymbol{\Phi}}_{\alpha}) \left[2\, \mathfrak{M}_{a,\mathrm{self}} \mathfrak{M}^{(\mathcal{S}')} + \mathfrak{M} \widetilde{\mathfrak{M}}_{a,\mathrm{sum}}^{(\mathcal{S}')} \right] \\
&\quad + \overline{\boldsymbol{\Phi}}_{\beta} \left[2\, \mathfrak{M}_{a,\mathrm{self}} \mathfrak{M}_{a_3,\mathrm{sum}}^{(\mathcal{S}')} + \mathfrak{M} \widetilde{\mathfrak{M}}^{(\mathcal{S}')} \right] \\
&\quad + \overline{\boldsymbol{\Phi}}_{\gamma} \left[2\, \mathfrak{M}_{a,\mathrm{self}} \overset{\approx}{\mathfrak{M}}{}^{(\mathcal{S}')} + \overset{\approx}{\mathfrak{M}}{}_{a,\mathrm{sum}}^{(\mathcal{S}')} \right] \triangleq \overline{z}_N, && (189)
\end{aligned}$$

where we have used the definitions listed on lines 9), 10), 11), 12), 13) and 14) of Table 3. The arguments leading to this result can be repeated by replacing upper bounds with lower bounds, and maxima with minima (e.g., $\mathfrak{M}$ replaced by $\mathfrak{m}$, or $\overline{\boldsymbol{\Phi}}_\alpha$ replaced by $\underline{\boldsymbol{\Phi}}_\alpha$), yielding:

$$\begin{aligned} \boldsymbol{e}_{ij}^{(\text{Gra})} \geq (1+\underline{\boldsymbol{\Phi}}_\alpha)\left[2\,\mathfrak{m}_{a,\text{self}}\,\mathfrak{m}^{(\mathcal{S}')}\,+\mathfrak{m}\,\widetilde{\mathfrak{m}}_{a,\text{sum}}^{(\mathcal{S}')}\right] \\ +\underline{\boldsymbol{\Phi}}_\beta\left[2\,\mathfrak{m}_{a,\text{self}}\,\mathfrak{m}_{a_3,\text{sum}}^{(\mathcal{S}')}+\mathfrak{m}\,\widetilde{\mathfrak{m}}^{(\mathcal{S}')}\right] \\ +\underline{\boldsymbol{\Phi}}_\gamma\left[2\,\mathfrak{m}_{a,\text{self}}\,\overset{\approx}{\mathfrak{m}}{}^{(\mathcal{S}')}+\overset{\approx}{\mathfrak{m}}{}_{a,\text{sum}}^{(\mathcal{S}')}\right] \triangleq \underline{z}_N. \end{aligned} \tag{190}$$

Now, under the uniform concentration regime we can use the pertinent convergences in probability listed in Table 3, in conjunction with the convergences in (163), which, after some tedious but straightforward algebra, lead to:

$$N p_N\,\bar{z}_N \xrightarrow{\text{p}} \eta, \qquad N p_N\,\underline{z}_N \xrightarrow{\text{p}} \eta \tag{191}$$

where η is the bias corresponding to the Granger estimator in Table 2. Accordingly, we can conclude that the error for the Granger estimator fulfills the hypotheses of Lemma 5, with the choice $\overline{\boldsymbol{w}}_N=\underline{\boldsymbol{w}}_N=0$, and with the quantities $\bar{z}_N$ and $\underline{z}_N$ defined in (189) and (190), respectively. This concludes the proof for the claim pertaining to the behavior of the limiting Granger estimator. ■

APPENDIX E
PROOF OF THEOREM 4

Proof: Since the limiting one-three-lags estimator is equal to the true matrix $\boldsymbol{A}_\mathcal{S}$, the proof for this estimator comes from Assumption 1 and Corollary 1.

The proof for the one-lag and residual estimators boils down to combining Theorem 5 with Lemma 5. In light of the definitions of bias and gap, it suffices to prove the claim with $\sigma^2=1$, and then scale the values obtained for the bias and the gap by an arbitrary σ^2.

Using (64), the error corresponding to the one-lag estimator can be written as (for all $i,j\in\mathcal{S}$ with $i\neq j$):

$$\boldsymbol{e}_{ij}^{(\text{1-lag})}=\sum_{h=1}^{\infty}\boldsymbol{a}_{ij}^{(2h+1)}, \tag{192}$$

and, hence, using the bounds in (117), we can write:

$$\underline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}\,\boldsymbol{a}_{ij}+\underline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{m}\leq \boldsymbol{e}_{ij}^{(\text{1-lag})}\leq \overline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}\,\boldsymbol{a}_{ij}+\overline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{M}. \tag{193}$$

Examining the bounds on $\boldsymbol{e}_{ij}^{(\text{1-lag})}$, we see that there are two contributions in the error. For example, let us consider the upper bound. The first contribution, $\overline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}$, multiplies the entries of the combination matrix, $\boldsymbol{a}_{ij}$, thus playing a role for connected node pairs. The second contribution, $\overline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{M}$, plays a role for *all* nodes, whether or note they are connected.

Now, using the convergence results in (119) and in (230), simple algebraic calculations lead to:

$$N p_N\,\underline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{m}\xrightarrow{\text{p}}\eta, \quad N p_N\,\overline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{M}\xrightarrow{\text{p}}\eta, \tag{194}$$

where η is equal to the bias of the one-lag estimator as defined in the pertinent row of Table 2. For what concerns $\underline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}$ and $\overline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}$, from (119) we see that:

$$\underline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}\xrightarrow{\text{p}}\Gamma/\kappa-1, \quad \overline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}\xrightarrow{\text{p}}\Gamma/\kappa-1, \tag{195}$$

where Γ is equal to the bias of the one-lag estimator as defined in the pertinent row of Table 2 (recall that we are working with $\sigma^2=1$). It remains to apply Lemma 5, with the choices:

$$\begin{aligned} \overline{\boldsymbol{w}}_N&=\overline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}, & \underline{\boldsymbol{w}}_N&=\underline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}, \\ \bar{z}_N&=\overline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{M}, & \underline{z}_N&=\underline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{m}, \end{aligned} \tag{196}$$

which concludes the proof for the limiting one-lag estimator.

Let us switch to the analysis of the residual estimator. Using (70), the error corresponding to the residual estimator can be written as (for all $i,j\in\mathcal{S}$ with $i\neq j$):

$$\boldsymbol{e}_{ij}^{(\text{res})}=\sum_{h=1}^{\infty}\boldsymbol{a}_{ij}^{(2h+1)}-\sum_{h=1}^{\infty}\boldsymbol{a}_{ij}^{(2h)}, \tag{197}$$

and, hence, using the bounds in (116) and (117), we can write:

$$\boldsymbol{e}_{ij}^{(\text{res})}\leq(\overline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}-\underline{\boldsymbol{\Sigma}}_\beta^{(\text{even})})\,\boldsymbol{a}_{ij}+(\overline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{M}-\underline{\boldsymbol{\Sigma}}_\gamma^{(\text{even})}\,\mathfrak{m}), \tag{198}$$

and

$$\boldsymbol{e}_{ij}^{(\text{res})}\geq(\underline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}-\overline{\boldsymbol{\Sigma}}_\beta^{(\text{even})})\,\boldsymbol{a}_{ij}+(\underline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{m}-\overline{\boldsymbol{\Sigma}}_\gamma^{(\text{even})}\,\mathfrak{M}). \tag{199}$$

Now, using the convergence results in (118), (119) and (230), simple algebraic calculations lead to:

$$\begin{aligned} N p_N\,(\overline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{M}-\underline{\boldsymbol{\Sigma}}_\gamma^{(\text{even})}\,\mathfrak{m})\xrightarrow{\text{p}}\eta, \\ N p_N\,(\underline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{m}-\overline{\boldsymbol{\Sigma}}_\gamma^{(\text{even})}\,\mathfrak{M})\xrightarrow{\text{p}}\eta, \end{aligned} \tag{200}$$

where η corresponds to the bias for the residual estimator listed in the pertinent row of Table 2. Likewise, exploiting (118), (119) and the representation of the gap Γ of the residual estimator in Table 2, we can prove that:

$$\begin{aligned} (\overline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}-\underline{\boldsymbol{\Sigma}}_\beta^{(\text{even})})\xrightarrow{\text{p}}\Gamma/\kappa-1, \\ (\underline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}-\overline{\boldsymbol{\Sigma}}_\beta^{(\text{even})})\xrightarrow{\text{p}}\Gamma/\kappa-1. \end{aligned} \tag{201}$$

It remains to apply Lemma 5, with the choices:

$$\begin{aligned} \overline{\boldsymbol{w}}_N&=\overline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}-\underline{\boldsymbol{\Sigma}}_\beta^{(\text{even})}, \\ \underline{\boldsymbol{w}}_N&=\underline{\boldsymbol{\Sigma}}_\beta^{(\text{odd})}-\overline{\boldsymbol{\Sigma}}_\beta^{(\text{even})}, \end{aligned} \tag{202}$$

and

$$\begin{aligned} \bar{z}_N&=\overline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{M}-\underline{\boldsymbol{\Sigma}}_\gamma^{(\text{even})}\,\mathfrak{m}, \\ \underline{z}_N&=\underline{\boldsymbol{\Sigma}}_\gamma^{(\text{odd})}\,\mathfrak{m}-\overline{\boldsymbol{\Sigma}}_\gamma^{(\text{even})}\,\mathfrak{M}, \end{aligned} \tag{203}$$

which concludes the proof of the theorem. ■

APPENDIX F
USEFUL CONVERGENCE RESULTS

Lemma 6 (List of Convergences under Uniform Concentration): If the connection probability fulfills (42), the convergences listed in Table 3 hold true.

Proof:

1)
$$\boxed{\mathfrak{M}_a \xrightarrow{\text{p}} 0, \qquad \mathfrak{m}_a \xrightarrow{\text{p}} 0} \tag{204}$$

From the inequalities in (44) we know that, for $i \neq j$:

$$\boldsymbol{a}_{ij} \leq \frac{\kappa}{\boldsymbol{d}_{\min}}, \tag{205}$$

which implies (204) in view of Corollary 1.

2)
$$\boxed{\mathfrak{M}_{a,\text{self}} \xrightarrow{\text{p}} \rho - \kappa, \qquad \mathfrak{m}_{a,\text{self}} \xrightarrow{\text{p}} \rho - \kappa} \tag{206}$$

Since $\boldsymbol{a}_{ii} = \rho - \sum_{\substack{\ell \in \mathcal{N} \\ \ell \neq i}} \boldsymbol{a}_{i\ell}$, from (44) we can write:

$$\boldsymbol{a}_{ii} \leq \rho - \frac{\kappa}{\boldsymbol{d}_{\max}} \sum_{\ell \in \mathcal{N} \ell \neq i} \boldsymbol{g}_{i\ell} = \rho - \kappa \frac{\boldsymbol{d}_i - 1}{\boldsymbol{d}_{\max}}$$
$$\leq \rho - \kappa \frac{\boldsymbol{d}_{\min} - 1}{\boldsymbol{d}_{\max}}. \tag{207}$$

Therefore, recalling that $\mathfrak{M}_{a,\text{self}} \triangleq \max_{i=1,2,\ldots,N} \boldsymbol{a}_{ii}$, we can write:

$$\mathfrak{M}_{a,\text{self}} \leq \rho - \kappa \frac{\boldsymbol{d}_{\min} - 1}{\boldsymbol{d}_{\max}}. \tag{208}$$

In view of Corollary 1, we have that the ratio $\frac{\boldsymbol{d}_{\min}-1}{\boldsymbol{d}_{\max}}$ converges to 1 in probability. Repeating the same reasoning with lower bounds in place of upper bounds, and with minima in place of maxima, yields the same result, and, hence, (206) follows.

3)
$$\boxed{\mathfrak{M}_{a_2,\text{self}} \xrightarrow{\text{p}} (\rho - \kappa)^2, \quad \mathfrak{m}_{a_2,\text{self}} \xrightarrow{\text{p}} (\rho - \kappa)^2} \tag{209}$$

We can write:

$$\boldsymbol{a}_{ii}^{(2)} = \sum_{\ell \in \mathcal{N}} \boldsymbol{a}_{i\ell} \boldsymbol{a}_{\ell i} = \boldsymbol{a}_{ii}^2 + \sum_{\substack{\ell \in \mathcal{N} \\ \ell \neq i}}^{N} \boldsymbol{a}_{i\ell} \boldsymbol{a}_{\ell i}$$
$$\leq \boldsymbol{a}_{ii}^2 + \frac{\kappa^2}{\boldsymbol{d}_{\min}^2} \sum_{\substack{\ell \in \mathcal{N} \\ \ell \neq i}}^{N} \boldsymbol{g}_{i\ell}$$
$$\leq \boldsymbol{a}_{ii}^2 + \kappa^2 \frac{\boldsymbol{d}_{\max} - 1}{\boldsymbol{d}_{\min}^2}, \tag{210}$$

where the intermediate inequality follows by (44). Therefore, recalling that $\mathfrak{M}_{a_2,\text{self}} \triangleq \max_{i \in \mathcal{N}} \boldsymbol{a}_{ii}^{(2)}$, we can write:

$$\mathfrak{M}_{a_2,\text{self}} \leq \mathfrak{M}_{a,\text{self}}^2 + \kappa^2 \frac{\boldsymbol{d}_{\max} - 1}{\boldsymbol{d}_{\min}^2}, \tag{211}$$

where the last term vanishes in probability in view of Corollary 1. Using now (206), and repeating the same reasoning with lower bounds in place of upper bounds, and with minima in place of maxima, the result in (209) follows.

4)
$$\boxed{\mathfrak{M}_{a,\text{sum}} \xrightarrow{\text{p}} \rho, \qquad \mathfrak{m}_{a,\text{sum}} \xrightarrow{\text{p}} \rho} \tag{212}$$

From the first relationship in (44) we can write:

$$\sum_{\substack{\ell \in \mathcal{N} \\ \ell \neq j}} \boldsymbol{a}_{i\ell} = \rho - \boldsymbol{a}_{ij}, \tag{213}$$

and the claim in (212) follows readily by (204).

5)
$$\boxed{\mathfrak{M}_{a,\text{sum}}^{(\mathcal{S}')} \xrightarrow{\text{p}} \kappa(1 - \xi), \quad \mathfrak{m}_{a,\text{sum}}^{(\mathcal{S}')} \xrightarrow{\text{p}} \kappa(1 - \xi)} \tag{214}$$

In view of (44) we can write:

$$\sum_{\substack{h \in \mathcal{S}' \\ h \neq \ell, m}} \boldsymbol{a}_{hm} \leq \frac{\kappa}{\boldsymbol{d}_{\min}} \sum_{\substack{h \in \mathcal{S}' \\ h \neq \ell, m}} \boldsymbol{g}_{hm}. \tag{215}$$

Now we observe that the random variable:

$$\sum_{\substack{h \in \mathcal{S}' \\ h \neq \ell, m}} \boldsymbol{g}_{hm} \tag{216}$$

is a binomial random variable with number of trials equal to $S' - 2$, and success probability equal to p_N. In other words, we get the following representation:

$$\max_{\substack{\ell, m \in \mathcal{S}' \\ \ell \neq m}} \sum_{\substack{h \in \mathcal{S}' \\ h \neq \ell, m}} \boldsymbol{g}_{hm} = \mathcal{B}_{\max}(S' - 2, p_N, (S' - 1)S'), \tag{217}$$

because maximization is carried over all pairs $\ell, m \in \mathcal{S}'$, with $\ell \neq m$. Moreover, since in the uniform concentration regime we have:

$$p_N = \omega_N \frac{\log N}{N}, \qquad \omega_N \overset{N \to \infty}{\longrightarrow} \infty, \tag{218}$$

and since $S'/N \to 1 - \xi$ as $N \to \infty$, we can regard the connection probability p_N as a connection probability scaling with respect to S', namely,

$$p_N = \omega_N \frac{\log S'}{S'} \frac{S'}{N} \frac{\log N}{\log(S'/N) + \log N}$$
$$= \omega_{S'} \frac{\log S'}{S'} \triangleq p_{S'}, \tag{219}$$

where:

$$\omega_{S'} = \omega_N \frac{S'}{N} \frac{\log N}{\log(S'/N) + \log N} \overset{N \to \infty}{\longrightarrow} \infty. \tag{220}$$

This shows that the uniform concentration regime can be referred also to the scaling of the involved quantities w.r.t. S' (in place of N). Accordingly, applying (75) with the choices $K = (S' - 1)S'$, $N = S' - 2$, and $p_{\mathcal{S}'} = \omega_{\mathcal{S}'} \log(S')/S'$, and reasoning as done in the proof of of Lemma 2, we get:

$$\frac{\mathcal{B}_{\max}(S'-2, p_{\mathcal{S}'}, (S'-1)S')}{S' p_{\mathcal{S}'}} \xrightarrow{\mathrm{p}} 1. \tag{221}$$

It remains to use (217) in (215) to get:

$$\sum_{\substack{h\in\mathcal{S}'\\ h\neq \ell,m}} \boldsymbol{a}_{hm} \le \kappa \underbrace{\frac{N p_N}{\boldsymbol{d}_{\min}}}_{\xrightarrow{\mathrm{p}} 1} \underbrace{\frac{S'}{N}}_{\to 1-\xi} \times \underbrace{\frac{\mathcal{B}_{\max}(S'-2, p_{\mathcal{S}'}, (S'-1)S')}{S' p_{\mathcal{S}'}}}_{\xrightarrow{\mathrm{p}} 1}. \tag{222}$$

Repeating the same reasoning with lower bounds in place of upper bounds, and with minima in place of maxima, yields the same result, and, hence, (214) follow.

6)

$$\boxed{\mathfrak{M}_{c,\text{self}} \xrightarrow{\mathrm{p}} (\rho-\kappa)^2, \quad \mathfrak{m}_{c,\text{self}} \xrightarrow{\mathrm{p}} (\rho-\kappa)^2} \tag{223}$$

This result follows readily by repeating the same steps used to prove (209).

7)

$$\boxed{\begin{aligned} &\mathfrak{M}_{c,\text{sum}} \xrightarrow{\mathrm{p}} \rho^2 - 2\rho\kappa\xi + \kappa^2\xi \\ &\mathfrak{m}_{c,\text{sum}} \xrightarrow{\mathrm{p}} \rho^2 - 2\rho\kappa\xi + \kappa^2\xi \end{aligned}} \tag{224}$$

Using the definition of $\boldsymbol{C}$ in (36), we note that we can write:

$$\begin{aligned} \sum_{\substack{h\in\mathcal{S}'\\ h\neq m}} \boldsymbol{c}_{\ell h} &= \sum_{\substack{h\in\mathcal{S}'\\ h\neq m}} \sum_{j\in\mathcal{N}} \boldsymbol{a}_{\ell j}\boldsymbol{a}_{jh} \\ &= \sum_{j\in\mathcal{S}} \boldsymbol{a}_{\ell j} \sum_{\substack{h\in\mathcal{S}'\\ h\neq m}} \boldsymbol{a}_{jh} + \sum_{j\in\mathcal{S}'} \boldsymbol{a}_{\ell j} \sum_{\substack{h\in\mathcal{S}'\\ h\neq m}} \boldsymbol{a}_{jh}. \end{aligned} \tag{225}$$

Applying the same procedure used in the previous items of this section, it is readily proved that, if $j \in \mathcal{S}$ (and, hence the self-term $\boldsymbol{a}_{\ell\ell}$ is not present, because $\ell \in \mathcal{S}'$):

$$\max_{j\in\mathcal{S}} \sum_{\substack{h\in\mathcal{S}'\\ h\neq m}} \boldsymbol{a}_{jh} \xrightarrow{\mathrm{p}} \kappa(1-\xi), \min_{j\in\mathcal{S}} \sum_{\substack{h\in\mathcal{S}'\\ h\neq m}} \boldsymbol{a}_{jh} \xrightarrow{\mathrm{p}} \kappa(1-\xi), \tag{226}$$

whereas, if $j \in \mathcal{S}'$:

$$\max_{j\in\mathcal{S}} \sum_{\substack{h\in\mathcal{S}'\\ h\neq m}} \boldsymbol{a}_{jh} \xrightarrow{\mathrm{p}} \rho-\kappa\xi, \min_{j\in\mathcal{S}} \sum_{\substack{h\in\mathcal{S}'\\ h\neq m}} \boldsymbol{a}_{jh} \xrightarrow{\mathrm{p}} \rho-\kappa\xi. \tag{227}$$

Likewise, we can show that:

$$\max_{\ell\in\mathcal{S}'} \sum_{j\in\mathcal{S}} \boldsymbol{a}_{\ell j} \xrightarrow{\mathrm{p}} \kappa\xi, \min_{\ell\in\mathcal{S}'} \sum_{j\in\mathcal{S}} \boldsymbol{a}_{\ell j} \xrightarrow{\mathrm{p}} \kappa\xi, \tag{228}$$

and that:

$$\max_{\ell\in\mathcal{S}'} \sum_{j\in\mathcal{S}'} \boldsymbol{a}_{\ell j} \xrightarrow{\mathrm{p}} \rho-\kappa\xi, \min_{\ell\in\mathcal{S}'} \sum_{j\in\mathcal{S}'} \boldsymbol{a}_{\ell j} \xrightarrow{\mathrm{p}} \rho-\kappa\xi. \tag{229}$$

Plugging (226)–(229) into (225) finally yields (224).

8)

$$\boxed{N p_N\, \mathfrak{M} \xrightarrow{\mathrm{p}} \kappa^2 p, \qquad N p_N\, \mathfrak{m} \xrightarrow{\mathrm{p}} \kappa^2 p} \tag{230}$$

In view of (44), we can write:

$$\sum_{\substack{\ell\in\mathcal{N}\\ \ell\neq i,j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j} \le \frac{\kappa^2}{\boldsymbol{d}_{\min}^2} \sum_{\substack{\ell\in\mathcal{N}\\ \ell\neq i,j}} \boldsymbol{g}_{i\ell}\boldsymbol{g}_{\ell j}. \tag{231}$$

Now we see that the quantity:

$$\sum_{\substack{\ell\in\mathcal{N}\\ \ell\neq i,j}} \boldsymbol{g}_{i\ell}\boldsymbol{g}_{\ell j} \tag{232}$$

is a binomial random variable with number of trials equal to $N-2$, and success probability equal to p_N^2, since when $\ell \neq i, j$, the product variable $\boldsymbol{g}_{i\ell}\boldsymbol{g}_{\ell j}$ is a Bernoulli variable with success probability p_N^2. Therefore, we are allowed to introduce the definition:

$$\mathcal{B}_{\max}(N-2, p_N^2, (N-1)N) = \max_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \sum_{\substack{\ell\in\mathcal{N}\\ \ell\neq i,j}} \boldsymbol{g}_{i\ell}\boldsymbol{g}_{\ell j}, \tag{233}$$

which, in view of Lemma 2, yields:

$$\frac{1}{N p_N} \max_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \sum_{\ell\neq i,j} \boldsymbol{g}_{i\ell}\boldsymbol{g}_{\ell j} \xrightarrow{\mathrm{p}} p. \tag{234}$$

Now, from the definition of $\mathfrak{M}$ in Table 3, line 8), we get:

$$N p_N \mathfrak{M} \le \kappa^2 \underbrace{\frac{\mathcal{B}_{\max}(N-2, p_N^2, (N-1)N)}{N p_N}}_{\xrightarrow{\mathrm{p}} p} \underbrace{\frac{N^2 p_N^2}{\boldsymbol{d}_{\min}^2}}_{\xrightarrow{\mathrm{p}} 1}. \tag{235}$$

Repeating the same reasoning with lower bounds in place of upper bounds, and with minima in place of maxima, we get the claim in (230).

9)

$$\boxed{\begin{aligned} &N p_N\, \mathfrak{M}^{(\mathcal{S}')} \xrightarrow{\mathrm{p}} \kappa^2 p(1-\xi) \\ &N p_N\, \mathfrak{m}^{(\mathcal{S}')} \xrightarrow{\mathrm{p}} \kappa^2 p(1-\xi) \end{aligned}} \tag{236}$$

The proof for the case where $p = 0$ comes from (230) because, from the definitions listed on lines 8) and 9) of Table 3, we see that $\mathfrak{M}^{(\mathcal{S}')} \le \mathfrak{M}$. The proof for the case where $p > 0$ is readily obtained by using the same arguments leading to (230).

10)

$$\boxed{\begin{aligned} &N p_N\, \mathfrak{M}_{a_3,\text{sum}}^{(\mathcal{S}')} \xrightarrow{\mathrm{p}} \kappa^3 p(1-\xi)^2 \\ &N p_N\, \mathfrak{m}_{a_3,\text{sum}}^{(\mathcal{S}')} \xrightarrow{\mathrm{p}} \kappa^3 p(1-\xi)^2 \end{aligned}} \tag{237}$$

We have that:

$$\sum_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell m}\boldsymbol{a}_{mj} = \sum_{\ell\in\mathcal{S}'}\boldsymbol{a}_{i\ell}\sum_{\substack{m\in\mathcal{S}'\\ m\neq\ell}}\boldsymbol{a}_{\ell m}\boldsymbol{a}_{mj}$$
$$\leq \max_{i\in\mathcal{S}}\sum_{\ell\in\mathcal{S}'}\boldsymbol{a}_{i\ell}\max_{j\in\mathcal{S},\ell\in\mathcal{S}'}\sum_{\substack{m\in\mathcal{S}'\\ m\neq\ell}}\boldsymbol{a}_{\ell m}\boldsymbol{a}_{mj}. \tag{238}$$

Reasoning as done for proving (236), we can show that:

$$N p_N \max_{j\in\mathcal{S},\ell\in\mathcal{S}'}\sum_{\substack{m\in\mathcal{S}'\\ m\neq\ell}}\boldsymbol{a}_{\ell m}\boldsymbol{a}_{mj} \xrightarrow{\mathrm{p}} \kappa^2 p(1-\xi). \tag{239}$$

Likewise, reasoning as done for proving (214), we can show that:

$$\max_{i\in\mathcal{S}}\sum_{\ell\in\mathcal{S}'}\boldsymbol{a}_{i\ell} \xrightarrow{\mathrm{p}} \kappa(1-\xi). \tag{240}$$

Finally, using (239) and (240) into (238), repeating the same reasoning with lower bounds in place of upper bounds, and with minima in place of maxima, we get (237).

11) The following list of convergences is obtained by trivial variations on the previous proofs.

$$\boxed{\begin{array}{ll}
\widetilde{\mathfrak{M}}^{(\mathcal{S}')}_{a,\mathrm{sum}} \xrightarrow{\mathrm{p}} \kappa(1-\xi), & \widetilde{\mathfrak{m}}^{(\mathcal{S}')}_{a,\mathrm{sum}} \xrightarrow{\mathrm{p}} \kappa(1-\xi)\\
\widetilde{\widetilde{\mathfrak{M}}}^{(\mathcal{S}')}_{a,\mathrm{sum}} \xrightarrow{\mathrm{p}} \kappa^3(1-\xi)^2, & \widetilde{\widetilde{\mathfrak{m}}}^{(\mathcal{S}')}_{a,\mathrm{sum}} \xrightarrow{\mathrm{p}} \kappa^3(1-\xi)^2\\
\widetilde{\mathfrak{M}}^{(\mathcal{S}')} \xrightarrow{\mathrm{p}} \kappa^2(1-\xi)^2, & \widetilde{\mathfrak{m}}^{(\mathcal{S}')} \xrightarrow{\mathrm{p}} \kappa^2(1-\xi)^2\\
\widetilde{\widetilde{\mathfrak{M}}}^{(\mathcal{S}')} \xrightarrow{\mathrm{p}} \kappa^2(1-\xi)^2, & \widetilde{\widetilde{\mathfrak{m}}}^{(\mathcal{S}')} \xrightarrow{\mathrm{p}} \kappa^2(1-\xi)^2
\end{array}} \tag{241}$$

■

APPENDIX G
USEFUL LEMMA

Lemma 7: Let $0<\alpha<1$, $0<\rho_\ell<1$ and $\beta_\ell\in\mathbb{R}$ for all $\ell=1,2,\ldots,L$, and introduce the following recursion:

$$f_{k+1} = \alpha f_k + \sum_{\ell=1}^{L}\beta_\ell\rho_\ell^k. \tag{242}$$

Then, f_k is equal to:

$$\left(f_0 + \sum_{\ell=1}^{L}\frac{\beta_\ell}{\alpha-\rho_\ell}\right)\alpha^k - \sum_{\ell=1}^{L}\frac{\beta_\ell}{\alpha-\rho_\ell}\rho_\ell^k, \tag{243}$$

namely, f_k can be represented as a linear combination of geometric sequences with ratio strictly smaller than one, a structure that will be particularly convenient in the proofs of Theorems 5 and 6.

Proof: Exploiting (242), we can write:

$$f_1 = \alpha f_0 + \sum_{\ell=1}^{L}\beta_\ell,$$

$$f_2 = \alpha^2 f_0 + \alpha\sum_{\ell=1}^{L}\beta_\ell + \sum_{\ell=1}^{L}\beta_\ell\rho_\ell,$$

$$\vdots$$

$$f_k = \alpha^k f_0 + \sum_{\ell=1}^{L}\beta_\ell\sum_{j=0}^{k-1}\alpha^{k-1-j}\rho_\ell^j. \tag{244}$$

The last equation can be manipulated as follows:

$$f_k = \alpha^k f_0 + \sum_{\ell=1}^{L}\beta_\ell\alpha^{k-1}\sum_{j=0}^{k-1}\left(\frac{\rho_\ell}{\alpha}\right)^j$$
$$= \alpha^k f_0 + \sum_{\ell=1}^{L}\beta_\ell\alpha^{k-1}\frac{1-(\rho_\ell/\alpha)^k}{1-\rho_\ell/\alpha}$$
$$= \alpha^k f_0 + \sum_{\ell=1}^{L}\beta_\ell\frac{\alpha^k-\rho_\ell^k}{\alpha-\rho_\ell}, \tag{245}$$

which corresponds to (243). ■

APPENDIX H
SAMPLE CONSISTENCY

We start with a useful lemma that characterizes the output of the clustering algorithm proposed in Section V-B, when its input is constituted by two well-separated classes.

Lemma 8 (Consistency of the clustering algorithm): Let

$$v_1 \leq v_2 \leq \cdots \leq v_L \tag{246}$$

be a set of real numbers such that, for $\nu_0,\nu_1\in\mathbb{R}$, $\epsilon>0$, and $k\in\{1,2,\ldots,L\}$:

$$\nu_0-\epsilon < v_i < \nu_0+\epsilon, \quad i=1,2,\ldots,k$$
$$\nu_1-\epsilon < v_i < \nu_1+\epsilon, \quad i=k+1,k+2,\ldots,L. \tag{247}$$

Then, for all $\epsilon>0$ such that:

$$\epsilon < \frac{\nu_1-\nu_0}{6}, \tag{248}$$

the clustering algorithm $\mathsf{clu}(v)$ described in Section V-B, produces as output a *unique* pair of clusters corresponding to $j^\star = k$, namely:

$$\mathcal{C}_0^\star = \mathcal{C}_0(k) = \{v_1,v_2,\ldots,v_k\},$$
$$\mathcal{C}_1^\star = \mathcal{C}_1(k) = \{v_{k+1},v_{k+2},\ldots,v_L\}. \tag{249}$$

Proof: We recall that the algorithm $\mathsf{clu}(v)$ considers a pair of clusters as admissible if the midpoint between the centroids of the clusters separates them. Accordingly, we start by showing that the cluster configuration in (249) is admissible. In view of (247), the centroids of $\mathcal{C}_0(k)$ and $\mathcal{C}_1(k)$ fulfill the following bounds:

$$\nu_0-\epsilon < c_0(k) < \nu_0+\epsilon, \qquad \nu_1-\epsilon < c_1(k) < \nu_1+\epsilon, \tag{250}$$

and, hence, their midpoint fulfills the condition:

$$\frac{\nu_0+\nu_1}{2}-\epsilon<\frac{c_0(k)+c_1(k)}{2}<\frac{\nu_0+\nu_1}{2}+\epsilon. \tag{251}$$

In view of (247) and (251), a sufficient condition for the clusters in (249) to be admissible is the following:

$$\nu_0+\epsilon<\frac{\nu_0+\nu_1}{2}-\epsilon,\qquad \frac{\nu_0+\nu_1}{2}+\epsilon<\nu_1-\epsilon, \tag{252}$$

which amounts to:

$$\epsilon<\frac{\nu_1-\nu_0}{4}, \tag{253}$$

and the admissibility of the configuration in (249) follows by (248).

In principle, other admissible configurations could exist. Next we show that if another admissible configuration distinct from (249) exists, then this configuration must necessarily exhibit a smaller inter-cluster distance. First, we note that, in view of (250), the distance between the centroids of the clusters in (249) fulfills the bound:

$$c_1(k)-c_0(k)>\nu_1-\nu_0-2\epsilon. \tag{254}$$

Let us assume that another admissible configuration j exists, for a certain $j>k$. Since now the point $v_j>\nu_1-\epsilon$ belongs to $\mathcal{C}_0(j)$, and since we assumed that configuration j is admissible, then v_j must be smaller than the midpoint between the centroids, yielding, in view of (247):

$$\begin{aligned}\nu_1-\epsilon<v_j&<\frac{c_0(j)+c_1(j)}{2}\\ &\Rightarrow -c_0(j)<c_1(j)-2(\nu_1-\epsilon)\\ &\Rightarrow c_1(j)-c_0(j)<2c_1(j)-2(\nu_1-\epsilon).\end{aligned} \tag{255}$$

On the other hand, in view of (247) we have:

$$c_1(j)<\nu_1+\epsilon, \tag{256}$$

which used along with (255) yields:

$$c_1(j)-c_0(j)<4\epsilon. \tag{257}$$

Examining the bound on the distance in (254), we see that the condition:

$$\nu_1-\nu_0-2\epsilon>4\epsilon\Leftrightarrow\epsilon<\frac{\nu_1-\nu_0}{6}, \tag{258}$$

is sufficient for the configuration in (249) to maximize the inter-cluster distance. Finally, since the situation is similar for $j<k$, the proof is complete. ■

Proof of Theorem 1: In the following, for $i,j\in\mathcal{S}$, the (i,j) entry of the estimated matrix $\widehat{\boldsymbol{A}}_{\mathcal{S},n}$ is denoted by $\widehat{\boldsymbol{a}}_{ij}(n)$. Likewise, the (i,j) entry of the limiting estimated matrix $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ is denoted by $\widehat{\boldsymbol{a}}_{ij}$. Let us introduce the following events (we recall that bold notation refers to random variables):

$$\mathcal{E}_0(n,N)=\left\{\max_{\substack{i,j\in\mathcal{S}:a_{ij}=0\\ i\neq j}}\left|s_N\widehat{\boldsymbol{a}}_{ij}(n)-\eta\right|<\epsilon\right\},$$

$$\mathcal{E}_1(n,N)=\left\{\max_{\substack{i,j\in\mathcal{S}:a_{ij}>0\\ i\neq j}}\left|s_N\widehat{\boldsymbol{a}}_{ij}(n)-\eta-\Gamma\right|<\epsilon\right\}. \tag{259}$$

In view of Lemma 8, assuming a sufficiently small ϵ, the clustering algorithm proposed in Section V-B reconstructs the exact graph whenever $\mathcal{E}_0(n,N)\cap\mathcal{E}_1(n,N)$ occurs. Accordingly, the theorem will be proved if we show that for some $n(N)$:

$$\lim_{N\to\infty}\mathbb{P}[\mathcal{E}_0(n(N),N)\cap\mathcal{E}_1(n(N),N)]=1. \tag{260}$$

To this aim, let us introduce the events:

$$\mathcal{E}_0(N)=\left\{\max_{\substack{i,j\in\mathcal{S}:a_{ij}=0\\ i\neq j}}\left|s_N\widehat{\boldsymbol{a}}_{ij}-\eta\right|<\epsilon/2\right\},$$

$$\mathcal{E}_1(N)=\left\{\max_{\substack{i,j\in\mathcal{S}:a_{ij}>0\\ i\neq j}}\left|s_N\widehat{\boldsymbol{a}}_{ij}-\eta-\Gamma\right|<\epsilon/2\right\}, \tag{261}$$

and

$$\mathcal{F}(n,N)=\left\{s_N\,\|\widehat{\boldsymbol{A}}_{\mathcal{S},n}-\widehat{\boldsymbol{A}}_{\mathcal{S}}\|_{\max}<\epsilon/2\right\}. \tag{262}$$

By triangle inequality, we have the implication:

$$\mathcal{E}_0(N)\cap\mathcal{E}_1(N)\cap\mathcal{F}(n,N)\Rightarrow\mathcal{E}_0(n,N)\cap\mathcal{E}_1(n,N), \tag{263}$$

yielding:

$$\mathbb{P}\left[\mathcal{E}_0(N)\cap\mathcal{E}_1(N)\cap\mathcal{F}(n,N)\right]\leq\mathbb{P}\left[\mathcal{E}_0(n,N)\cap\mathcal{E}_1(n,N)\right]. \tag{264}$$

Since by assumption the limiting matrix estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ achieves universal local structural consistency we know that:

$$\lim_{N\to\infty}\mathbb{P}\left[\mathcal{E}_0(N)\cap\mathcal{E}_1(N)\right]=1. \tag{265}$$

Consider now a sequence $\epsilon_N>0$ that vanishes as $N\to\infty$. Since $\widehat{\boldsymbol{A}}_{\mathcal{S},n}$ is in the class defined by (11), for any fixed N, there exists $n(N)$ such that, for all $n\geq n(N)$ we have:

$$\mathbb{P}\left[\mathcal{F}(n,N)\right]>1-\epsilon_N, \tag{266}$$

or, by the sandwich theorem:

$$\lim_{N\to\infty}\mathbb{P}\left[\mathcal{F}(n(N),N)\right]=1. \tag{267}$$

Using now (265) and (267) in (264) implies (260) and, hence, the claim of the theorem. ■

APPENDIX I
SAMPLE COMPLEXITY

In the following we will make repeated use of the following inequality, holding for any two matrices A,B:

$$\|AB\|_{\max}\leq\min(\|A\|_{\max}\|B\|_1,\|A\|_\infty\|B\|_{\max}). \tag{268}$$

Given a regular diffusion matrix fulfilling Assumption 1 with parameters ρ and κ, let $\zeta=\rho-\kappa$. We introduce, for a small

δ, with $0 < \delta < 1 - \zeta^2$, the auxiliary quantities:

$$\varphi_1 \triangleq \frac{(1-\zeta^2-\delta)(1-\rho)}{16\sqrt{2}\sigma^2}, \varphi_2 \triangleq \frac{(1-\rho^2)^2(1-\rho)}{16\sqrt{2}\sigma^2}. \quad (269)$$

We notice that when δ is sufficiently small, $\varphi_1 > \varphi_2$. Next we introduce some auxiliary functions. Let

$$\mathsf{b}_n(x) = \begin{cases} \min\left\{1\,,\; S^2\left(e^{-n/2} + e^{-\left[\sqrt{n}x - \sqrt{2}\right]^2}\right)\right\}, & \sqrt{n}x > \sqrt{2} \\ 1, & \sqrt{n}x \le \sqrt{2} \end{cases} \quad (270)$$

where x is a positive quantity, n is the sample size and S is the number of probed nodes. It is immediate to verify that $\mathsf{b}_n(x)$ is non-increasing in x, and that:

$$\lim_{n\to\infty} \mathsf{b}_n(x) = 0 \qquad \text{for all } x > 0. \quad (271)$$

The second auxiliary function is:

$$\mathsf{b}_N(\delta) = \mathbb{P}\left[\frac{\min_{i\in\mathcal{N}}[\boldsymbol{R}_0]_{ii}}{(\max_{i\in\mathcal{N}}[\boldsymbol{R}_0]_{ii})^2} < \frac{1-\zeta^2-\delta}{\sigma^2}\right]. \quad (272)$$

We will now examine the behavior of $b_N(\delta)$ as $N \to \infty$. We start by showing that, as $N \to \infty$, the minimum and maximum diagonal entry of $\boldsymbol{R}_0$ concentrate (in probability) around the same value. In view of (29) we have, for all $i \in \mathcal{N}$:

$$[\boldsymbol{R}_0]_{ii} = \sigma^2[(I - \boldsymbol{A}^2)^{-1}]_{ii} = \sigma^2\left(1 + \sum_{h=1}^{\infty} \boldsymbol{a}_{ii}^{(2h)}\right), \quad (273)$$

which, using (115), yields:

$$\sigma^2\left(1 + \underline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{even})}\right) \le [\boldsymbol{R}_0]_{ii} \le \sigma^2\left(1 + \overline{\boldsymbol{\Sigma}}_{\alpha}^{(\text{even})}\right). \quad (274)$$

Applying to the lower and upper bounds in (274) the convergence shown in the first line of (118), by the sandwich theorem we get:

$$[\boldsymbol{R}_0]_{ii} \xrightarrow{\text{p}} \sigma^2\left(1 + \frac{\zeta^2}{1-\zeta^2}\right) = \frac{\sigma^2}{1-\zeta^2}, \qquad \forall i \in \mathcal{N}. \quad (275)$$

We note from (274) that the convergence in (275) is uniform across the index i. In other words,

$$\min_{i\in\mathcal{N}}[\boldsymbol{R}_0]_{ii} \xrightarrow{\text{p}} \frac{\sigma^2}{1-\zeta^2},$$
$$\max_{i\in\mathcal{N}}[\boldsymbol{R}_0]_{ii} \xrightarrow{\text{p}} \frac{\sigma^2}{1-\zeta^2}, \quad (276)$$

further implying that:

$$\frac{\min_{i\in\mathcal{N}}[\boldsymbol{R}_0]_{ii}}{(\max_{i\in\mathcal{N}}[\boldsymbol{R}_0]_{ii})^2} \xrightarrow{\text{p}} \frac{1-\zeta^2}{\sigma^2}. \quad (277)$$

Since for all $\delta > 0$,

$$\frac{1-\zeta^2-\delta}{\sigma^2} < \frac{1-\zeta^2}{\sigma^2}, \quad (278)$$

from (272) we conclude that:

$$\lim_{N\to\infty} \mathsf{b}_N(\delta) = 0. \quad (279)$$

The following lemma characterizes the rate of convergence of the sample covariance estimators. This lemma adapts Lemmas 1 and 2 in [30] to exploit the peculiarities of our setting.

Lemma 9 (Convergence rate of the sample covariance estimators): Let us consider the VAR model in (3) with i.i.d. standard Gaussian input source and initial state distributed (conditionally on $\boldsymbol{A}$) according to the stationary distribution of the VAR process. Let the combination matrix $\boldsymbol{A}$ fulfill Assumption 1 with parameters ρ and κ, and let the underlying graph fulfill the Erdős-Rényi model under the uniform concentration regime. Then, for all $\epsilon > 0$ we have:

$$\mathbb{P}\left[\|[\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [\boldsymbol{R}_0]_{\mathcal{S}}\|_{\max} > \epsilon\right]$$
$$\le 3\left[\mathsf{b}_n(\epsilon\,\varphi_1) + \mathsf{b}_n(\epsilon\,\varphi_2)\,\mathsf{b}_N(\delta)\right], \quad (280)$$

and

$$\mathbb{P}\left[\|[\widehat{\boldsymbol{R}}_{1,n}]_{\mathcal{S}} - [\boldsymbol{R}_1]_{\mathcal{S}}\|_{\max} > \epsilon\right]$$
$$\le 4\left[\mathsf{b}_{n-1}(\epsilon\,\varphi_1) + \mathsf{b}_{n-1}(\epsilon\,\varphi_2)\,\mathsf{b}_N(\delta)\right]. \quad (281)$$

Proof: Preliminarily, it is worth noticing that the assumption of stationary initial state is adopted because some results about covariance matrices that will be used to prove our theorems are obtained under this assumption, which is classically adopted in sample complexity analysis [29], [30], [31], [32], [95]. As observed, e.g., in [29], such assumption entails a mild restriction since the (stable) system in (3) converges exponentially to the stationary regime.

Let us start by examining the probability in (280) for a fixed realization of the combination matrix $\boldsymbol{A} = A$, i.e., we consider the quantity, for $i, j \in \mathcal{N}$:

$$\mathbb{P}\left[|[\widehat{\boldsymbol{R}}_{0,n}]_{ij} - [\boldsymbol{R}_0]_{ij}| > \epsilon \middle| \boldsymbol{A} = A\right]$$
$$= \mathbb{P}\left[|[\widehat{\boldsymbol{R}}_{0,n}]_{ij} - [R_0]_{ij}| > \epsilon \middle| \boldsymbol{A} = A\right], \quad (282)$$

where

$$R_0 = \sigma^2(I - A^2)^{-1}, \quad (283)$$

and where, in the considered conditional space, $\widehat{\boldsymbol{R}}_{0,n}$ is an empirical covariance matrix estimated over a VAR model (3) with *deterministic* combination matrix A. In fact, given the graph generation (i.e., given a certain $\boldsymbol{A} = A$), the random sequence $\boldsymbol{y}_n$ evolves according to (3), with $\boldsymbol{x}_n$ being the assigned i.i.d. zero-mean, unit-variance source, and with $\boldsymbol{y}_0$ following the stationary distribution of the VAR model corresponding to A. For these types of models, an upper bound on the probability appearing in (282) is derived in [30] relying on a concentration result for the covariance matrix entries shown in [95]. More precisely, to obtain an upper bound on the probability in (282) we now modify the proof of Lemma 1 in [30] by exploiting the peculiarities of our model. In order to avoid redundancy, we do not report here the complete arguments and proofs in [30]. We limit ourselves to modify the specific part of the proof that is necessary to obtain our bounds.

Let us consider, for $m, m' = 1, 2, \ldots, n$, the following known relationship that is obtained exploiting (3):[13]

$$R_{m-m'} = \mathbb{E}[\boldsymbol{y}_m \boldsymbol{y}_{m'}^\top | \boldsymbol{A} = A] = A^{|m-m'|} R_0. \tag{284}$$

In [30] the following inequality is used:

$$\|R_{m-m'}\|_{\max} \le \|R_{m-m'}\|_2 \le \rho^{|m-m'|} \|R_0\|_2. \tag{285}$$

In our case we can replace (285) by the following inequality, which exploits additional constraints on A:

$$\begin{aligned}\|R_{m-m'}\|_{\max} &\le \|A^{|m-m'|}\|_\infty \|R_0\|_{\max} \\ &= \rho^{|m-m'|} \max_{i \in \mathcal{N}} [R_0]_{ii},\end{aligned} \tag{286}$$

where the last equality comes from noticing that: i) the matrix A/ρ is doubly stochastic; and ii) the off-diagonal entries of the covariance matrix are upper bounded as $|[R_0]_{ij}| \le \sqrt{[R_0]_{ii}[R_0]_{jj}}$ by Cauchy-Schwarz inequality.

Applying (286) in the proof of Lemma 1 in [30] (with every other detail of the proof being unaltered) we get, for $\sqrt{n}\epsilon\varphi(R_0) > \sqrt{2}$ and all $i, j \in \mathcal{N}$ with $i \ne j$:

$$\begin{aligned}&\mathbb{P}\left[|[\widehat{\boldsymbol{R}}_{0,n}]_{ij} - [R_0]_{ij}| > \epsilon | \boldsymbol{A} = A\right] \\ &\le 3\left(e^{-n/2} + e^{-\left[\sqrt{n}\epsilon\,\varphi(R_0) - \sqrt{2}\right]^2}\right),\end{aligned} \tag{287}$$

where we introduced the definition:

$$\varphi(R_0) = \frac{(1-\rho)}{16\sqrt{2}} \frac{\min_{i\in\mathcal{N}}[R_0]_{ii}}{(\max_{i\in\mathcal{N}}[R_0]_{ii})^2}. \tag{288}$$

Using the union bound over the set of probed nodes $\mathcal{S}$ and the definition in (270), from (287) and (288) we get, for all n:

$$\mathbb{P}\left[\|[\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [R_0]_{\mathcal{S}}\|_{\max} > \epsilon | \boldsymbol{A} = A\right] \le 3\,\mathsf{b}_n(\epsilon\,\varphi(R_0)). \tag{289}$$

We now complete the proof by taking into account the randomness of the combination matrix $\boldsymbol{A}$. To this end, let $\mathcal{R}$ the set of all possible realizations of $\boldsymbol{A}$, and introduce the following set:

$$\mathcal{T} \triangleq \left\{A \in \mathcal{R} : \frac{\min_{i\in\mathcal{N}}[R_0]_{ii}}{(\max_{i\in\mathcal{N}}[R_0]_{ii})^2} \ge \frac{1-\zeta^2-\delta}{\sigma^2}\right\}, \tag{290}$$

where we recall that R_0 is a function of A — see (283). We notice that $\mathcal{T}$ is a high-probability set, since in view of (272) and (279) we have:

$$\mathbb{P}\left[\boldsymbol{A} \in \mathcal{T}'\right] = \mathsf{b}_N(\delta) \overset{N\to\infty}{\longrightarrow} 0. \tag{291}$$

By the law of total probability we can write:

$$\begin{aligned}&\mathbb{P}\left[\|[\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [\boldsymbol{R}_0]_{\mathcal{S}}\|_{\max} > \epsilon\right] \\ &= \sum_{A\in\mathcal{R}} \mathbb{P}\left[\|[\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [R_0]_{\mathcal{S}}\|_{\max} > \epsilon \middle| \boldsymbol{A} = A\right] \mathbb{P}[\boldsymbol{A} = A] \\ &\le 3 \sum_{A\in\mathcal{R}} \mathsf{b}_n(\epsilon\,\varphi(R_0))\, \mathbb{P}[\boldsymbol{A} = A],\end{aligned} \tag{292}$$

where we have used (289). Recalling now the definition of φ_1 in (269) and of $\varphi(R_0)$ in (288), we see that when $A \in \mathcal{T}$ we have $\varphi(R_0) > \varphi_1$, yielding, since $\mathsf{b}_n(x)$ is non-increasing in x:

$$\mathsf{b}_n(\epsilon\,\varphi(R_0)) \le \mathsf{b}_n(\epsilon\,\varphi_1), \quad \forall A \in \mathcal{T}. \tag{293}$$

Moreover, from (273) we see that:

$$\sigma^2 \le [R_0]_{ii} \le \frac{\sigma^2}{1-\rho^2}, \tag{294}$$

yielding:

$$\frac{\min_{i\in\mathcal{N}}[R_0]_{ii}}{(\max_{i\in\mathcal{N}}[R_0]_{ii})^2} \ge \frac{(1-\rho^2)^2}{\sigma^2}, \tag{295}$$

and in view of the definition of φ_2 in (269) we conclude that $\varphi(R_0) \ge \varphi_2$ for all R_0, implying then:

$$\mathsf{b}_n(\epsilon\,\varphi(R_0)) \le \mathsf{b}_n(\epsilon\,\varphi_2). \tag{296}$$

Applying (293) and (296) in (292), we can write:

$$\begin{aligned}&\mathbb{P}\left[\|[\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [\boldsymbol{R}_0]_{\mathcal{S}}\|_{\max} > \epsilon\right] \\ &\le 3\sum_{A\in\mathcal{T}} \mathsf{b}_n(\epsilon\,\varphi(R_0))\,\mathbb{P}[\boldsymbol{A} = A] \\ &\quad + 3\sum_{A\in\mathcal{T}'} \mathsf{b}_n(\epsilon\,\varphi(R_0))\,\mathbb{P}[\boldsymbol{A} = A] \\ &\le 3\mathsf{b}_n(\epsilon\,\varphi_1) + 3\mathsf{b}_n(\epsilon\,\varphi_2)\,\mathbb{P}[\boldsymbol{A} \in \mathcal{T}'],\end{aligned} \tag{297}$$

and (280) follows by observing that $\mathbb{P}[\boldsymbol{A} \in \mathcal{T}'] = \mathsf{b}_N(\delta)$ — see (291). In order to obtain (281), we must repeat the same steps shown above to the proof of Lemma 2 in [30]. ■

Let us comment briefly on the structure of the bounds in (280) and (281). For clarity of presentation, we focus on (280), since similar considerations would apply to (281). We notice that the bound in (296) could be used for all possible realizations of $\boldsymbol{A}$, while in (297) we used it only for $A \in \mathcal{T}'$. This is because for A belonging to the high-probability set $\mathcal{T}$, the covariance matrix $\boldsymbol{R}_0$ exhibits a concentration property that allowed us to obtain the bound $\mathsf{b}_n(\epsilon\,\varphi_1)$ in (293), which goes to zero with n faster than $\mathsf{b}_n(\epsilon\,\varphi_2)$, since $\varphi_1 > \varphi_2$. On the other hand, examining (280) we see that the other bound $\mathsf{b}_n(\epsilon\,\varphi_2)$ is further multiplied by the quantity $\mathsf{b}_N(\delta)$, which is independent on the sample size n and vanishes as the network size N goes to infinity as a consequence of the aforementioned concentration properties. As a possible extension, one could try to refine also the term $\mathsf{b}_n(\epsilon\,\varphi_2)$ by exploring alternative bounding techniques — see, e.g., [29], [32].

Corollary 2 (Scaling law useful for sample complexity): Assume the same conditions used in Lemma 9. If

$$n(N) = \Omega\left((Np_N)^2 \log S\right), \tag{298}$$

then we have that:

$$\lim_{N\to\infty} \mathbb{P}\left[\|[\widehat{\boldsymbol{R}}_{0,n(N)}]_{\mathcal{S}} - [\boldsymbol{R}_0]_{\mathcal{S}}\|_{\max} > \frac{\epsilon}{Np_N}\right] = 0, \tag{299}$$

[13] The specific representation $R_{m-m'} = A^{|m'-m|}R_0$ holds due to the symmetry of A.

$$\lim_{N\to\infty} \mathbb{P}\left[\|[\widehat{\boldsymbol{R}}_{1,n(N)}]_{\mathcal{S}} - [\boldsymbol{R}_1]_{\mathcal{S}}\|_{\max} > \frac{\epsilon}{Np_N}\right] = 0. \tag{300}$$

Proof: We will prove the claim with reference to (299), with the proof being identical for (300). Examining (280), we see that in order to get (299) it suffices to guarantee that:

$$\lim_{N\to\infty} b_{n(N)}\left(\frac{\epsilon\,\varphi_1}{Np_N}\right) = 0, \tag{301}$$

since the second term in (280) goes to zero automatically as $N\to\infty$ due to the presence of $\mathrm{b}_N(\delta)$, whatever law $n(N)$ is chosen. Let us first focus on the second exponential term in (270), which can be written as:

$$\exp\left\{-\left(\frac{\sqrt{n}\,\epsilon\,\varphi_1}{Np_N} - \sqrt{2}\right)^2 + 2\log S\right\}. \tag{302}$$

After some straightforward algebra, the exponent in (302) can be more conveniently rewritten as:

$$-\left[\left(\sqrt{\frac{n\epsilon^2\,\varphi_1^2}{2(Np_N)^2\log S}} - \frac{1}{\sqrt{\log S}}\right)^2 - 1\right]\log S^2. \tag{303}$$

Now, since $S/N \to \xi > 0$ as $N\to\infty$, we see that $S\to\infty$. Therefore, the second exponential term in (270) converges to zero if the quantity under brackets in (303) becomes asymptotically positive. Noticing that the term $(\sqrt{\log S})^{-1}$ vanishes as $N\to\infty$, we see that this condition is verified if we have, for $\epsilon' > 0$:

$$\frac{n\epsilon^2\,\varphi_1^2}{2(Np_N)^2\log S} > 1 + \epsilon', \tag{304}$$

from some N onward, which corresponds to (298). Finally, since with the found scaling law $n(N)$ diverges faster than $\log S$, the first exponential term in (270) converges to zero as well, and the desired claim in (301) is obtained. ■

The next two lemmas are auxiliary to the proof of Theorem 3.

Lemma 10: Let A, R_0, R_1, $\widehat{\mathsf{A}}$, $\widehat{\mathsf{R}}_0$ and $\widehat{\mathsf{R}}_1$ be square matrices of equal size, with:

$$\|\mathsf{A}\|_\infty \le 1, \quad \mathsf{R}_1 = \mathsf{A}\mathsf{R}_0. \tag{305}$$

Assume that $\widehat{\mathsf{R}}_0$ is invertible and let

$$\widehat{\mathsf{A}} = \widehat{\mathsf{R}}_1\widehat{\mathsf{R}}_0^{-1}. \tag{306}$$

Then we have that:

$$\|\widehat{\mathsf{A}} - \mathsf{A}\|_{\max} \le \|\widehat{\mathsf{R}}_0^{-1}\|_1\left(\|\widehat{\mathsf{R}}_0 - \mathsf{R}_0\|_{\max} + \|\widehat{\mathsf{R}}_1 - \mathsf{R}_1\|_{\max}\right). \tag{307}$$

Proof: From (306) we can write:

$$\begin{aligned}
\|\widehat{\mathsf{A}} - \mathsf{A}\|_{\max} &= \|(\widehat{\mathsf{R}}_1 - \mathsf{A}\widehat{\mathsf{R}}_0)\widehat{\mathsf{R}}_0^{-1}\|_{\max}\\
&\le \|\widehat{\mathsf{R}}_0^{-1}\|_1\,\|\widehat{\mathsf{R}}_1 - \mathsf{A}\widehat{\mathsf{R}}_0\|_{\max}\\
&= \|\widehat{\mathsf{R}}_0^{-1}\|_1\,\|\widehat{\mathsf{R}}_1 - \mathsf{R}_1\\
&\quad + \underbrace{\mathsf{R}_1 - \mathsf{A}\mathsf{R}_0}_{=0 \text{ from } (305)} + \mathsf{A}\mathsf{R}_0 - \mathsf{A}\widehat{\mathsf{R}}_0\|_{\max}\\
&\le \|\widehat{\mathsf{R}}_0^{-1}\|_1\,\|\widehat{\mathsf{R}}_1 - \mathsf{R}_1\|_{\max}\\
&\quad + \|\widehat{\mathsf{R}}_0^{-1}\|_1\,\|\mathsf{A}\|_\infty\|\widehat{\mathsf{R}}_0 - \mathsf{R}_0\|_{\max},
\end{aligned} \tag{308}$$

and the result in (307) follows because $\|\mathsf{A}\|_\infty \le 1$ in view of (305). ■

Lemma 11: Let A, R_0, R_1, $\widehat{\mathsf{A}}$, $\widehat{\mathsf{R}}_0$ and $\widehat{\mathsf{R}}_1$ be square matrices defined over an index set $\mathcal{S}$, with:

$$\|\mathsf{A}\|_\infty \le 1, \mathsf{R}_1 = \mathsf{A}\mathsf{R}_0. \tag{309}$$

Let the i-th row of $\widehat{\mathsf{A}}$ be a solution to the optimization problem:

$$\min_x \left\|x\,\widehat{\mathsf{R}}_0 - [\widehat{\mathsf{R}}_1]_{i\mathcal{S}}\right\|_\infty \quad \text{s.t. } \|x\|_1 \le 1, \tag{310}$$

where x is a row vector and $[\widehat{\mathsf{R}}_1]_{i\mathcal{S}}$ is the i-th row of $[\widehat{\mathsf{R}}_1]_{\mathcal{S}}$.

If R_0 is invertible, then we have that:

$$\|\widehat{\mathsf{A}} - \mathsf{A}\|_{\max} \le 2\|\mathsf{R}_0^{-1}\|_1\left(\|\widehat{\mathsf{R}}_0 - \mathsf{R}_0\|_{\max} + \|\widehat{\mathsf{R}}_1 - \mathsf{R}_1\|_{\max}\right). \tag{311}$$

Proof: We recall that for a vector the $\|\cdot\|_\infty$ norm amounts to the maximum absolute value of its entries, whereas for matrices it is the maximum absolute row sum. Accordingly, we see that the matrix A introduced in the statement of the lemma is a candidate solution to (310) since by assumption $\|\mathsf{A}\|_\infty \le 1$. As a result, a solution to the optimization problem in (310) fulfills:

$$\begin{aligned}
\|\widehat{\mathsf{A}}\widehat{\mathsf{R}}_0 - \widehat{\mathsf{R}}_1\|_{\max} &\le \|\mathsf{A}\widehat{\mathsf{R}}_0 - \widehat{\mathsf{R}}_1\|_{\max}\\
&= \|\mathsf{A}\widehat{\mathsf{R}}_0 - \mathsf{A}\mathsf{R}_0 + \mathsf{R}_1 - \widehat{\mathsf{R}}_1\|_{\max}\\
&\le \|\mathsf{A}\|_\infty\,\|\widehat{\mathsf{R}}_0 - \mathsf{R}_0\|_{\max} + \|\mathsf{R}_1 - \widehat{\mathsf{R}}_1\|_{\max}.
\end{aligned} \tag{312}$$

On the other hand, we can write:

$$\begin{aligned}
\|\widehat{\mathsf{A}} - \mathsf{A}\|_{\max} &= \|(\widehat{\mathsf{A}}\mathsf{R}_0 - \mathsf{R}_1)\mathsf{R}_0^{-1}\|_{\max}\\
&\le \|\mathsf{R}_0^{-1}\|_1\,\|\widehat{\mathsf{A}}\mathsf{R}_0 - \mathsf{R}_1\|_{\max}\\
&= \|\mathsf{R}_0^{-1}\|_1\,\|\widehat{\mathsf{A}}\mathsf{R}_0 - \widehat{\mathsf{A}}\widehat{\mathsf{R}}_0\\
&\quad + \widehat{\mathsf{A}}\widehat{\mathsf{R}}_0 - \widehat{\mathsf{R}}_1 + \widehat{\mathsf{R}}_1 - \mathsf{R}_1\|_{\max}\\
&\le \|\mathsf{R}_0^{-1}\|_1\,\|\widehat{\mathsf{A}}\|_\infty\,\|\widehat{\mathsf{R}}_0 - \mathsf{R}_0\|_{\max}\\
&\quad + \|\mathsf{R}_0^{-1}\|_1\,\|\widehat{\mathsf{A}}\widehat{\mathsf{R}}_0 - \widehat{\mathsf{R}}_1\|_{\max}\\
&\quad + \|\mathsf{R}_0^{-1}\|_1\,\|\widehat{\mathsf{R}}_1 - \mathsf{R}_1\|_{\max},
\end{aligned} \tag{313}$$

and the claim in (311) follows from (312). ■

Proof of Theorem 3: Examining the proof of Theorem 1, and in particular (262), we see that the sample complexity of

$\widehat{\boldsymbol{A}}_{\mathcal{S},n}$ can be determined by finding, for a sufficiently small $\epsilon > 0$, a law $n(N)$ that ensures:

$$\lim_{N\to\infty} \mathbb{P}\left[\|\widehat{\boldsymbol{A}}_{\mathcal{S},n(N)} - \widehat{\boldsymbol{A}}_{\mathcal{S}}\|_{\max} > \frac{\epsilon}{Np_N}\right] = 0. \tag{314}$$

Let us consider the regularized sample Granger estimator defined by (51). First we show that the *limiting* Granger estimator obeys the following bound:

$$\|\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\mathrm{Gra})}\|_{\infty} \le 1. \tag{315}$$

In view of (35) we can write:

$$\widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\mathrm{Gra})} = \boldsymbol{A}_{\mathcal{S}} + \boldsymbol{E}^{(\mathrm{Gra})} = \boldsymbol{A}_{\mathcal{S}} + \boldsymbol{A}_{\mathcal{S},\mathcal{S}'}\underbrace{\boldsymbol{H}[\boldsymbol{A}^2]_{\mathcal{S}'\mathcal{S}}}_{\boldsymbol{F}}, \tag{316}$$

where we remark that the matrix $\boldsymbol{F}$ is nonnegative by construction. Exploiting (316) we have:

$$\sum_{j\in\mathcal{S}} \widehat{\boldsymbol{a}}_{ij}^{(\mathrm{Gra})} = \sum_{j\in\mathcal{S}} \boldsymbol{a}_{ij} + \sum_{\ell\in\mathcal{S}'} \boldsymbol{a}_{i\ell} \sum_{j\in\mathcal{S}} \boldsymbol{f}_{\ell j} \le \rho \le 1, \tag{317}$$

where the last inequality follows since from (75) in [19] we know that:

$$\sum_{j\in\mathcal{S}} \boldsymbol{f}_{\ell j} \le 1. \tag{318}$$

Equation (317) reveals that (315) holds true. Accordingly, we can call upon Lemma 11 and use (311) to write:

$$\begin{aligned}&\|\widehat{\boldsymbol{A}}_{\mathcal{S},n}^{(\mathrm{reGra})} - \widehat{\boldsymbol{A}}_{\mathcal{S}}^{(\mathrm{Gra})}\|_{\max}\\ &\le 2\|([\boldsymbol{R}_0]_{\mathcal{S}})^{-1}\|_1 \left(\|[\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [\boldsymbol{R}_0]_{\mathcal{S}}\|_{\max}\right.\\ &\quad \left. + \|[\widehat{\boldsymbol{R}}_{1,n}]_{\mathcal{S}} - [\boldsymbol{R}_1]_{\mathcal{S}}\|_{\max}\right).\end{aligned} \tag{319}$$

We conclude that the scaling law in (298) will be sufficient for the regularized Granger estimator if we show that $\|([\boldsymbol{R}_0]_{\mathcal{S}})^{-1}\|_1$ is bounded by some constant. To this end, let

$$\boldsymbol{Z} = \frac{I - \boldsymbol{A}^2}{\sigma^2} \Leftrightarrow \boldsymbol{R}_0 = \boldsymbol{Z}^{-1}. \tag{320}$$

From one of the block matrix representations for the inverse matrix we have that [96, p. 18]:

$$\begin{aligned}([\boldsymbol{R}_0]_{\mathcal{S}})^{-1} &= \boldsymbol{Z}_{\mathcal{S}} - \boldsymbol{Z}_{\mathcal{S}\mathcal{S}'}(\boldsymbol{Z}_{\mathcal{S}'})^{-1}\boldsymbol{Z}_{\mathcal{S}'\mathcal{S}}\\ &= \frac{I_{\mathcal{S}} - [\boldsymbol{A}^2]_{\mathcal{S}} - [\boldsymbol{A}^2]_{\mathcal{S}\mathcal{S}'}\boldsymbol{H}[\boldsymbol{A}^2]_{\mathcal{S}'\mathcal{S}}}{\sigma^2}\\ &= \frac{I_{\mathcal{S}} - [\boldsymbol{A}^2]_{\mathcal{S}} - [\boldsymbol{A}^2]_{\mathcal{S}\mathcal{S}'}\boldsymbol{F}}{\sigma^2}.\end{aligned} \tag{321}$$

Since $\boldsymbol{A}$ is nonnegative with $\|\boldsymbol{A}^2\|_{\infty} = \rho^2$, while $\boldsymbol{F}$ is nonnegative and fulfills (318), we conclude that:

$$\|[\boldsymbol{A}^2]_{\mathcal{S}} + [\boldsymbol{A}^2]_{\mathcal{S}\mathcal{S}'}\boldsymbol{F}\|_{\infty} \le \rho^2, \tag{322}$$

which, in view of (321) and the symmetry of $\boldsymbol{R}_0$, yields:

$$\|\left([\boldsymbol{R}_0]_{\mathcal{S}}\right)^{-1}\|_1 \le \frac{1+\rho^2}{\sigma^2}, \tag{323}$$

which completes the proof for the regularized Granger estimator.

Let us finally prove that in the dense case where $p_N \to p > 0$, the convergence in (314) with the scaling law in (54) holds for the non-regularized Granger estimator in (50). To this end, we expand $[\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}}$ as:

$$\begin{aligned}[\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} &= [\boldsymbol{R}_0]_{\mathcal{S}} + \left([\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [\boldsymbol{R}_0]_{\mathcal{S}}\right)\\ &= [\boldsymbol{R}_0]_{\mathcal{S}}\left(I_{\mathcal{S}} + \underbrace{([\boldsymbol{R}_0]_{\mathcal{S}})^{-1}\left([\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [\boldsymbol{R}_0]_{\mathcal{S}}\right)}_{\boldsymbol{D}}\right).\end{aligned} \tag{324}$$

Now we observe that:

$$\begin{aligned}\|\boldsymbol{D}\|_1 &\le \|([\boldsymbol{R}_0]_{\mathcal{S}})^{-1}\|_1 \; \|([\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [\boldsymbol{R}_0]_{\mathcal{S}})\|_1\\ &\le \frac{1+\rho^2}{\sigma^2} S \; \|([\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [\boldsymbol{R}_0]_{\mathcal{S}})\|_{\max}\\ &= \frac{1+\rho^2}{\sigma^2 p_N}\frac{S}{N}\left(Np_N \; \|([\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}} - [\boldsymbol{R}_0]_{\mathcal{S}})\|_{\max}\right)\\ &\le \left(\frac{(1+\rho^2)\xi}{\sigma^2 p} + \epsilon\right)\epsilon = \epsilon',\end{aligned} \tag{325}$$

where: *i*) in the second-to-last inequality we used (323) and the fact that for an $S \times S$ matrix the $\|\cdot\|_1$ norm is upper bounded by S times the $\|\cdot\|_{\max}$ norm; *ii*) we used the fact that $S/N \to \xi$ and $p_N \to p > 0$ as $N \to \infty$; and *iii*) in view of Corollary 2, the last inequality holds (and, hence, $[\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}}$ is invertible) with probability converging to 1 as $N \to \infty$. Finally, applying matrix inversion to (324), and upper bounding the norm of $(I_{\mathcal{S}} + \boldsymbol{D})^{-1}$ as in [96, p. 301], we get:

$$\begin{aligned}\|([\widehat{\boldsymbol{R}}_{0,n}]_{\mathcal{S}})^{-1}\|_1 &\le \|([\boldsymbol{R}_0]_{\mathcal{S}})^{-1}\|_1 \; \|(I_{\mathcal{S}} + \boldsymbol{D})^{-1}\|_1\\ &\le \frac{(1+\rho^2)(1 - \|\boldsymbol{D}\|_1)^{-1}}{\sigma^2}\\ &\le \frac{1+\rho^2}{\sigma^2(1-\epsilon')},\end{aligned} \tag{326}$$

and (314) follows by using (326) and Corollary 2 in Lemma 10. ■

REFERENCES

[1] V. Matta, A. Santos, and A. H. Sayed, “Tomography of large adaptive networks under the dense latent regime,” in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, 2018, pp. 2144–2148.

[2] V. Matta, A. Santos, and A. H. Sayed, “Graph learning with partial observations: Role of degree concentration,” in *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, 2019, pp. 1312–1316.

[3] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[4] T. Liggett, *Interacting Particle Systems*. Berlin Heidelberg, Germany: Springer, 2005.

[5] P. Robert, *Stochastic Networks and Queues*. Berlin Heidelberg, Germany: Springer, 2003.

[6] M. Porter and J. Gleeson, *Dynamical Systems on Networks: A Tutorial*. Berlin, Germany: Springer, 2016.

[7] A. Ganesh, L. Massoulié, and D. Towsley, "The effect of network topology on the spread of epidemics," in *Proc. IEEE 24th Annu. Joint Conf. IEEE Comput. Commun. Societies*, 2005, vol. 2, pp. 1455–1466.

[8] P. C. Pinto, P. Thiran, and M. Vetterli, "Locating the source of diffusion in large-scale networks," *Phys. Rev. Lett.*, vol. 109, pp. 068702-1–068702-5, Aug. 2012.

[9] S. Mahdizadehaghdam, H. Wang, H. Krim, and L. Dai, "Information diffusion of topic propagation in social media," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 4, pp. 569–581, Dec. 2016.

[10] S. Marano, V. Matta, and P. Willett, "The importance of being earnest: Social sensing with unknown agent quality," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 3, pp. 306–320, Sep. 2016.

[11] H. Salami, B. Ying, and A. H. Sayed, "Social learning over weakly connected graphs," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 2, pp. 222–238, Jun. 2017.

[12] P. Venkitasubramaniam, T. He, and L. Tong, "Anonymous networking amidst eavesdroppers," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2770–2784, Jun. 2008.

[13] S. Marano, V. Matta, T. He, and L. Tong, "The embedding capacity of information flows under renewal traffic," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1724–1739, Mar. 2013.

[14] M. Cheraghchi, A. Karbasi, S. Mohajer, and V. Saligrama, "Graph-constrained group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 248–262, Jan. 2012.

[15] C. J. Honey et al., "Predicting human resting-state functional connectivity from structural connectivity," *Proc. Nat. Acad. Sci.*, vol. 106, no. 6, pp. 2035–2040, Feb. 2009.

[16] B. Mišić et al., "Network-level structure-function relationships in human neocortex," *Cereb. Cortex*, vol. 26, no. 7, pp. 3285–3296, Jul. 2016.

[17] R. Liégeois, A. Santos, V. Matta, D. Van de Ville, and A. H. Sayed, "Revisiting correlation-based functional connectivity and its relationship with structural connectivity," *Netw. Neurosci.*, vol. 4, no. 4, pp. 1235–1251, 2020.

[18] G. Mateos, S. Segarra, A. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 16–43, May 2019.

[19] V. Matta and A. H. Sayed, "Consistent tomography under partial observations over adaptive networks," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 622–646, Jan. 2019.

[20] A. Santos, V. Matta, and A. H. Sayed, "Local tomography of large networks under the low-observability regime," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 587–613, Jan. 2020.

[21] A. Anandkumar and R. Valluvan, "Learning loopy graphical models with latent variables: Efficient methods and guarantees," *Ann. Statist.*, vol. 41, no. 2, pp. 401–435, Apr. 2013.

[22] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from data under Laplacian and structural constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 825–841, Sep. 2017.

[23] J. Bento and A. Montanari, "Which graphical models are difficult to learn?," in *Proc. Neural Inf. Process. Syst.*, Vancouver, Canada, 2009, pp. 1303–1311.

[24] D. M. Chickering, D. Heckerman, and C. Meek, "Large-sample learning of Bayesian networks is NP-hard," *J. Mach. Learn. Res.*, vol. 5, pp. 1287–1330, Dec. 2004.

[25] S. E. Shimony, "Finding MAPs for belief networks is NP-hard," *Artif. Intell.*, vol. 68, no. 2, pp. 399–410, Aug. 1994.

[26] G. Bresler, D. Gamarnik, and D. Shah, "Hardness of parameter estimation in graphical models," in *Proc. Neural Inf. Process. Syst.*, Montréal, Canada, 2014, pp. 1062–1070.

[27] A. Bogdanov, E. Mossel, and S. Vadhan, "The complexity of distinguishing Markov random fields," in *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, A. Goel, K. Jansen, J. D. P. Rolim, and R. Rubinfeld, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 331–342.

[28] G. Bresler, "Efficiently learning Ising models on arbitrary graphs," in *Proc. ACM Symp. Theory Comput.*, Portland, OR, USA, 2015, pp. 771–782.

[29] J. Bento, M. Ibrahimi, and A. Montanari, "Learning networks of stochastic differential equations," in *Proc. Neural Inf. Process. Syst.*, Vancouver, Canada, 2010, pp. 172–180.

[30] F. Han, H. Lu, and H. Liu, "A direct estimation of high dimensional stationary vector autoregressions," *J. Mach. Learn. Res.*, vol. 16, pp. 3115–3150, Dec. 2015.

[31] P.-L. Loh and M. J. Wainwright, "High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity," *Ann. Statist.*, vol. 40, no. 3, pp. 1637–1664, Apr. 2012.

[32] M. Rao, A. Kipnis, M. Javidi, Y. Eldar, and A. Goldsmith, "System identification with partial samples: Non-asymptotic analysis," in *Proc. IEEE Conf. Decis. Control*, Las Vegas, NV, USA, 2016, pp. 2938–2944.

[33] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Found. Trends Mach. Learn.*, vol. 7, no. 4/5, pp. 311–801, 2014.

[34] A. H. Sayed, "Adaptive networks," *Proc. IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.

[35] V. Matta and A. H. Sayed, "Estimation and detection over adaptive networks," in *Cooperative and Graph Signal Processing*, P. Djuric and C. Richard, Eds. Amsterdam, The Netherlands: Elsevier, 2018, pp. 69–106.

[36] A. H. Sayed and X. Zhao, "Asynchronous adaptive networks," in *Cooperative and Graph Signal Processing*, P. Djuric and C. Richard, Eds. Amsterdam, The Netherlands: Elsevier, 2018, pp. 3–68.

[37] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[38] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.

[39] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.

[40] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[41] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3775–3789, Jul. 2016.

[42] M. Tsitsvero, S. Barbarossa, and P. D. Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4845–4860, Sep. 2016.

[43] P. Di Lorenzo, P. Banelli, E. Isufi, S. Barbarossa, and G. Leus, "Adaptive graph signal processing: Algorithms and optimal sampling strategies," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3584–3598, Jul. 2018.

[44] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5911–5926, Nov. 2017.

[45] S. P. Chepuri and G. Leus, "Graph sampling for covariance estimation," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 451–466, Sep. 2017.

[46] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3462–3477, Jul. 2017.

[47] E. S. C. Ching and H. C. Tam, "Reconstructing links in directed networks from noisy dynamics," *Phys. Rev. E*, vol. 95, no. 1, pp. 010301-1–010301-5, Jan. 2017.

[48] D. Napoletani and T. D. Sauer, "Reconstructing the topology of sparsely connected dynamical networks," *Phys. Rev. E*, vol. 77, no. 2, pp. 026103-1–026103-5, Feb. 2008.

[49] J. Ren, W.-X. Wang, B. Li, and Y.-C. Lai, "Noise bridges dynamical correlation and topology in coupled oscillator networks," *Phys. Rev. Lett.*, vol. 104, no. 5, pp. 058701-1–058701-4, Feb. 2010.

[50] Y. Yang, T. Luo, Z. Li, X. Zhang, and P. S. Yu, "A robust method for inferring network structures," in *Sci. Rep.*, vol. 7, no. 5221, pp. 1–12, Jul. 2017.

[51] A. Mauroy and J. Goncalves, "Linear identification of nonlinear systems: A lifting technique based on the Koopman operator," in *Proc. IEEE Conf. Decis. Control*, Las Vegas, NV, USA, 2016, pp. 6500–6505.

[52] D. Materassi and M. V. Salapaka, “On the problem of reconstructing an unknown topology via locality properties of the Wiener filter,” *IEEE Trans. Autom. Control*, vol. 57, no. 7, pp. 1765–1777, Jul. 2012.

[53] C. J. Quinn, N. Kiyavash, and T. P. Coleman, “Directed information graphs,” *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6887–6909, Dec. 2015.

[54] J. Etesami and N. Kiyavash, “Measuring causal relationships in dynamical systems through recovery of functional dependencies,” *IEEE Trans. Signal Inf. Process. Netw*., vol. 3, no. 4, pp. 650–659, Dec. 2017.

[55] J. Mei and J. Moura, “Signal processing on graphs: Causal modeling of unstructured data,” *IEEE Trans. Signal Process*., vol. 65, no. 8, pp. 2077–2092, Apr. 2017.

[56] B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and M. G. Rabbat, “Characterization and inference of graph diffusion processes from observations of stationary signals,” *IEEE Trans. Signal Inf. Process. Netw*., vol. 4, no. 3, pp. 481–496, Sep. 2018.

[57] S. Segarra, M. T. Schaub, and A. Jadbabaie, “Network inference from consensus dynamics,” in *Proc. IEEE Conf. Decis. Control*, 2017, pp. 3212–3217.

[58] V. Matta, A. Santos, and A. H. Sayed, “Graph learning under partial observability,” *Proc. IEEE*, vol. 108, no. 11, pp. 2049–2066, Nov. 2020.

[59] D. Materassi and M. V. Salapaka, “Network reconstruction of dynamical polytrees with unobserved nodes,” in *Proc. IEEE Conf. Decis. Control*, Maui, HI, USA, 2012, pp. 4629–4634.

[60] J. Etesami, N. Kiyavash, and T. Coleman, “Learning minimal latent directed information polytrees,” *Neural Comput.*, vol. 28, no. 9, pp. 1723–1768, Aug. 2016.

[61] P. Geiger, K. Zhang, B. Schölkopf, M. Gong, and D. Janzing, “Causal inference by identification of vector autoregressive processes with hidden components,” in *Proc. Int. Conf. Mach. Learn.*, Lille, France, 2015, vol. 37, pp. 1917–1925.

[62] D. Materassi and M. V. Salapaka, “Identification of network components in presence of unobserved nodes,” in *Proc. IEEE Conf. Decis. Control*, Osaka, Japan, 2015, pp. 1563–1568.

[63] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, “Latent variable graphical model selection via convex optimization,” *Ann. Statist.*, vol. 40, no. 4, pp. 1935–1967, Aug. 2012.

[64] A. Anandkumar, V. Y. F. Tan, F. Huang, and A. S. Willsky, “High-dimensional Gaussian graphical model selection: Walk summability and local separation criterion,” *J. Mach. Learn. Res.*, vol. 13, pp. 2293–2337, Jan. 2012.

[65] G. Bresler, F. Koehler, A. Moitra, and E. Mossel, “Learning restricted Boltzmann machines via influence maximization,” in *Proc. ACM Symp. Theory Comput.*, Phoenix, AZ, USA, Jun. 2019, pp. 828–839.

[66] V. Matta and A. H. Sayed, “Tomography of adaptive multi-agent networks under limited observation,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Calgary, Canada, 2018, pp. 6638–6642.

[67] A. Santos, V. Matta, and A. H. Sayed, “Divide-and-conquer tomography for large-scale networks,” in *Proc. IEEE Data Sci. Workshop*, Lausanne, Switzerland, 2018, pp. 170–174.

[68] A. Santos, V. Matta, and A. H. Sayed, “Consistent tomography over diffusion networks under the low-observability regime,” in *Proc. IEEE Int. Symp. Inf. Theory*, Vail, CO, USA, 2018, pp. 1839–1843.

[69] A. Moneta, N. Chlaß, D. Entner, and P. Hoyer, “Causal search in structural vector autoregressive models,” in *Proc. Neural Inf. Process. Syst.*, Vancouver, Canada, 2009, pp. 95–118.

[70] P.-Y. Lai, “Reconstructing network topology and coupling strengths in directed networks of discrete-time dynamics,” *Phys. Rev. E*, vol. 95, no. 2, pp. 022311-1–022311-13, Feb. 2017.

[71] L. Xiao and S. Boyd, “Fast linear iterations for distributed averaging,” *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, Sep. 2004.

[72] J. Tsitsiklis, D. Bertsekas, and M. Athans, “Distributed asynchronous deterministic and stochastic gradient optimization algorithms,” *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.

[73] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[74] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, “Gossip algorithms for distributed signal processing,” *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[75] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, “Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis,” *IEEE Trans. Signal Process*., vol. 59, no. 9, pp. 4381–4396, Sep. 2011.

[76] D. Bajovic, D. Jakovetic, J. M. F. Moura, J. Xavier, and B. Sinopoli, “Large deviations performance of consensus innovations distributed detection with non-Gaussian observations,” *IEEE Trans. Signal Process*., vol. 60, no. 11, pp. 5987–6002, Nov. 2012.

[77] S. Kar and J. M. F. Moura, “Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs,” *IEEE J. Sel. Topics Signal Process*., vol. 5, no. 4, pp. 674–690, Aug. 2011.

[78] P. Braca, S. Marano, and V. Matta, “Enforcing consensus while monitoring the environment in wireless sensor networks,” *IEEE Trans. Signal Process*., vol. 56, no. 7, pp. 3375–3380, Jul. 2008.

[79] P. Braca, S. Marano, V. Matta, and P. Willett, “Asymptotic optimality of running consensus in testing binary hypotheses,” *IEEE Trans. Signal Process*., vol. 58, no. 2, pp. 814–825, Feb. 2010.

[80] J. Chen and A. H. Sayed, “On the learning behavior of adaptive networks—Part I: Transient analysis,” *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3487–3517, Jun. 2015.

[81] J. Chen and A. H. Sayed, “On the learning behavior of adaptive networks—Part II: Performance analysis,” *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3518–3548, Jun. 2015.

[82] C. G. Lopes and A. H. Sayed, “Diffusion least-mean squares over adaptive networks: Formulation and performance analysis,” *IEEE Trans. Signal Process*., vol. 56, no. 7, pp. 3122–3136, Jul. 2008.

[83] F. S. Cattivelli and A. H. Sayed, “Diffusion LMS strategies for distributed estimation,” *IEEE Trans. Signal Process*., vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[84] F. S. Cattivelli and A. H. Sayed, “Distributed detection over adaptive networks using diffusion adaptation,” *IEEE Trans. Signal Process*., vol. 59, no. 5, pp. 1917–1932, May 2011.

[85] A. H. Sayed, S. Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, “Diffusion strategies for adaptation and learning over networks,” *IEEE Signal Process. Mag*., vol. 30, no. 3, pp. 155–171, May 2013.

[86] V. Matta, P. Braca, S. Marano, and A. H. Sayed, “Diffusion-based adaptive distributed detection: Steady-state performance in the slow adaptation regime,” *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4710–4732, Aug. 2016.

[87] V. Matta, P. Braca, S. Marano, and A. H. Sayed, “Distributed detection over adaptive networks: Refined asymptotics and the role of connectivity,” *IEEE Trans. Signal Inf. Process. Netw*., vol. 2, no. 4, pp. 442–460, Dec. 2016.

[88] P. Erdõs and A. Rényi, “On random graphs I,” *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959.

[89] B. Bollobás, *Random Graphs*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[90] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A. Nonasymptotic Theory of Independence*. London, U.K.: Oxford Univ. Press, 2013.

[91] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. New York, NY, USA: Springer, 2001.

[92] A. Jalali and S. Sanghavi, “Learning the dependence graph of time series with latent factors,” in *Proc. Int. Conf. Mach. Learn.*, Edinburgh, Scotland, U.K., 2012, pp. 619–626.

[93] Y. Chen, Z. Wang, and X. Shen, “An unbiased symmetric matrix estimator for topology inference under partial observability,” *IEEE Signal Process. Lett*., vol. 29, pp. 1257–1261, 2022.

[94] M. Cirillo, V. Matta, and A. H. Sayed, “Learning Bollobás-Riordan graphs under partial observability,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Jun. 2021, pp. 5360–5364.

[95] S. Negahban and M. J. Wainwright, “Estimation of (near) low-rank matrices with noise and high-dimensional scaling,” *Ann. Statist.*, vol. 39, no. 2, pp. 1069–1097, Apr. 2011.

[96] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

VINCENZO MATTA (Senior Member, IEEE) received the Laurea degree in electronic engineering and the Ph.D. degree in information engineering from the University of Salerno, Fisciano, Italy, in 2001 and 2005, respectively. He is currently a Full Professor of telecommunications with the Department of Information and Electrical Engineering and Applied Mathematics, the University of Salerno, Fisciano, Italy. He is an author of more than 130 articles published on international journals and proceedings of international conferences. His research interests include signal processing, network theory and statistical learning, focusing on: adaptation and learning over networks, distributed inference, communications and security, multiobject-multisensor tracking and data fusion, detection of gravitational waves. Dr. Matta is an Associate Editor for the IEEE Open Journal of Signal Processing. He was an Associate Editor for the IEEE Transactions on Aerospace and Electronic Systems, the IEEE Signal Processing Letters, the IEEE Transactions on Signal and Information Processing over Networks, and as a Senior Area Editor for the IEEE Signal Processing Letters. He is Member of the Sensor Array and Multichannel Technical Committee of the Signal Processing Society (SPS), and was IEEE SPS Representative to the IEEE Transactions on Signal and Information Processing over Networks.

AUGUSTO SANTOS received the B.Sc. and M.Sc. degrees in electrical and computer engineering from Instituto Superior Técnico (IST), Lisbon-Portugal, and the Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, and IST. He held a two year Postdoctoral Scholar Position with Carnegie Mellon University and another two year with the École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. His research interests include high-dimensional statistical inference, especially on the problem of graph learning with latent variables, qualitative analysis of complex networked dynamical systems. He is currently a Researcher with the Centre for Informatics and Systems, University of Coimbra, Coimbra, Portugal.

ALI H. SAYED (Fellow, IEEE) is the Dean of Engineering at the École Polytechnique Fédérale de Lausanne Switzerland, where he also leads the Adaptive Systems Laboratory. He is a Member of the U.S. National Academy of Engineering, and The World Academy of Sciences. He was the President of the IEEE Signal Processing Society in 2018 and 2019. His research interests include adaptation and learning theories, data and network sciences, and statistical inference. His work has been recognized with several awards including the 2022 IEEE Fourier Award, the 2020 Norbert Wiener Society Award, the 2015 Education Award, and the 2012 Technical Achievement Award from the IEEE Signal Processing Society. He also was the recipient of the 2014 Papoulis Award from the European Association for Signal Processing, 2005 Terman Award from the American Society for Engineering Education, and several best paper awards. He is a Fellow of EURASIP and the American Association for the Advancement of Science.