

Estimating the Topology of Preferential Attachment Graphs Under Partial Observability

Michele Cirillo¹, Graduate Student Member, IEEE, Vincenzo Matta², Senior Member, IEEE,
and Ali H. Sayed³, Fellow, IEEE

Abstract—This work addresses the problem of learning the topology of a network from the signals emitted by the network nodes. These signals are generated over time through a linear diffusion process, where neighboring nodes exchange messages according to the underlying network topology, and aggregate them according to a certain combination matrix. We consider the demanding setting of *graph learning under partial observability*, where signals are available only from a limited fraction of nodes, and we want to establish whether the topology of these nodes can be estimated faithfully, despite the presence of possibly many latent nodes. Recent results examined this problem when the network topology is generated according to an Erdős-Rényi random model. However, Erdős-Rényi graphs feature homogeneity across nodes and independence across edges, while, over several real-world networks, significant heterogeneity is observed between very connected “hubs” and scarcely connected peripheral nodes, and the edge construction process entails significant dependencies across edges. *Preferential attachment* random graphs were conceived primarily to fill these gaps. We tackle the problem of graph learning over preferential attachment graphs by focusing on the following setting: first-order vector autoregressive models equipped with a stable *Laplacian* combination matrix, and a network topology drawn according to the popular *Bollobás-Riordan* preferential attachment model. The main result established in this work is that a combination-matrix estimator known as Granger estimator achieves graph learning under partial observability.

Index Terms—Graph learning, topology inference, preferential attachment, Bollobás-Riordan graph, partial observability.

I. INTRODUCTION

THIS work deals with the problem of learning the topology of a network, i.e., the graph, from the signals evolving across the network nodes according to discrete-time diffusion dynamics. The resulting *graph learning* problem, a.k.a. topology inference or network tomography, arises across

Manuscript received 10 December 2021; revised 28 July 2022; accepted 19 September 2022. Date of publication 6 October 2022; date of current version 20 January 2023. This work was supported in part by the Swiss National Foundation under Grant 205121-184999. An earlier version of this paper was presented in part at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2021 [DOI: 10.1109/ICASSP39728.2021.9414217]. (Corresponding author: Michele Cirillo.)

Michele Cirillo and Vincenzo Matta are with the Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, Fisciano, 84084 Salerno, Italy, and also with the National Inter-University Consortium for Telecommunications (CNIT), Italy (e-mail: micirillo@unisa.it; vmatta@unisa.it).

Ali H. Sayed is with the School of Engineering, École Polytechnique Fédérale de Lausanne EPFL, 1015 Lausanne, Switzerland (e-mail: ali.sayed@epfl.ch).

Communicated by S. Boucheron, Associate Editor for Statistical Learning. Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TIT.2022.3211078>.

Digital Object Identifier 10.1109/TIT.2022.3211078

many domains, for example, in biology, when one is interested in revealing how fish in a school coordinate their movements to escape a predator [2], [3]; in social learning, to discover how spatially dispersed network agents collaboratively share information to form their own opinions [4]; in neuroscience, when the goal is to unveil the relationships between functional and structural connectivity in the brain [5].

There exist several works addressing the graph learning problem. We refer the reader to the survey articles [6], [7] for an overview. Most works assume that signals from *all* network nodes can be gathered to estimate the topology [8], [9], [10], [11], [12], [13], [14], [15], [16], [17]. In contrast, one distinguishing feature of this work is the challenging scenario of *partial observability*, where only signals from a limited fraction of nodes are available to estimate the topology of the probed subnetwork.

Useful conditions for graph learning under partial observability have been obtained in [18], [19] for graphs having a polytree shape. More general shapes are considered in [20], [21], where, however, conditions for graph learning are formulated in terms of detailed local features of the network topology, which are not particularly suited to the large-scale setting considered here. In order to circumvent dependence on specific network structures, an asymptotic approach is considered in [22], [23], in the context of high-dimensional graphical models with latent variables. In [22], provable learning guarantees are offered under an appropriate local separation criterion, whereas in [23] graph learning is shown to be achievable under the so-called *sparsity plus low-rank* condition, namely, when the probed subnetwork is sparsely connected, and the unobserved subnetwork has suitably bounded size.

However, all the aforementioned works do not consider the time dynamics of the signals emitted by the nodes, and, hence, they are not applicable to dynamical systems like the Vector AutoRegressive (VAR) model considered here — see Eq. (2) further ahead. For such models, relevant results under full observability¹ are obtained in [24], [25], [26], [27], while partial observability is addressed in [28], relying on the aforementioned sparsity plus low-rank assumption used in [23].

Graph learning under *partial observability* with the VAR model considered in this work was recently addressed in [29], [30], [31], [32], [33], for the case where the

¹The *partial samples* setting considered in [26], [27] is different from the partial observability setting. In the latter, samples from only a subset of the network are available, while in [26], [27] samples from the *entire* network are available but, at each node, a certain fraction per node can be lost/corrupted.

dynamical system runs on top of an Erdős-Rényi random graph [34], [35]. Under reasonable technical assumptions, it was shown that graph learning is achievable under different regimes of connectivity. Erdős-Rényi graphs are by far the most common model adopted by researchers and practitioners to get insights about network problems and to validate network algorithms. The generative model of Erdős-Rényi graphs draws the edges independently one from the other, and in a completely homogeneous manner. Owing to this simple generation rule, these graphs are often unable to represent faithfully a number of phenomena emerging over real-world networks. For this reason, going beyond the Erdős-Rényi model is of critical importance.

In order to overcome the aforementioned limitations, other random graphs have been proposed in the last decades, among which a prominent role is played by *preferential attachment* graphs. The physical principles underlying the preferential attachment mechanism were originally discovered in [36], stimulating the research of several mathematical graph models, with one of the most popular and elegant being the Bollobás-Riordan model [37], [38].

These types of graphs were shown to capture several useful characteristics of real-world networks. For example, they are able to encompass the coexistence between *hubs* with many connections and peripheral, scarcely connected nodes. In contrast, as noticed, Erdős-Rényi graphs are *homogeneous* by their own nature [34], [35]. As explained in Sec. III, such enhanced descriptive power hinges basically on the statistical dependence enforced between the edges, as opposed to the edge independence implied by the Erdős-Rényi construction.

In view of this dependence, extending results valid for Erdős-Rényi graphs to preferential attachments graphs is a challenging task, which is addressed in this work providing the following main contributions. First, we formulate a sufficient condition for consistent graph learning under partial observability in terms of an *identifiability gap* (Theorem 1). This condition requires to extend existing results [6] to deal with the *randomness* of the identifiability gap that arises from the preferential attachment construction. Second, we ascertain that, when the network topology is a Bollobás-Riordan graph and the exchange of signals between neighboring nodes is ruled by the Laplacian matrix of the graph, a matrix estimator known as Granger estimator achieves consistent graph learning, assuming available an arbitrarily large number of samples (Theorem 2). Third, we establish that the *sample complexity* of the Granger estimator over Bollobás-Riordan graphs is almost linear in the network size (Theorem 3).

Notation. Matrices are denoted by upper-case letters, vectors by lower-case letters. We use boldface font to denote random variables, and normal font for their realizations. Sets and graphs are denoted by upper-case calligraphic letters. For an $N \times N$ matrix Z , the submatrix spanning the rows of Z indexed by set $\mathcal{P} \subseteq \{1, 2, \dots, N\}$ and the columns indexed by set $\mathcal{T} \subseteq \{1, 2, \dots, N\}$, is denoted by $Z_{\mathcal{P}\mathcal{T}}$, or alternatively by $[Z]_{\mathcal{P}\mathcal{T}}$. When $\mathcal{P} = \mathcal{T}$, the submatrix $Z_{\mathcal{P}\mathcal{T}}$ is abbreviated as $Z_{\mathcal{P}}$. Moreover, in the indexing of a submatrix we keep the index set of the corresponding full matrix. For example, if $\mathcal{P} = \{2, 3\}$ and $\mathcal{T} = \{2, 4, 5\}$, the submatrix $M = Z_{\mathcal{P}\mathcal{T}}$ is a 2×3 matrix,

indexed as follows:

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

For a graph \mathcal{G} , the corresponding capital letter G is used to denote its adjacency matrix, which has zero diagonal, and whose off-diagonal (k, ℓ) -entry $g_{k\ell}$ is equal to 1 if a directed edge from ℓ to k exists, and is zero otherwise. The symbol $\|\cdot\|_{\max}$ computes the maximum absolute entry of its matrix argument, whereas the symbol $\|\cdot\|_{\max\text{-off}}$ computes the maximum absolute *off-diagonal* entry of its matrix argument. The symbol $\xrightarrow{\text{p}}$ denotes convergence in probability *as the network size scales to infinity*. Likewise, the symbol $\xrightarrow{\text{a.s.}}$ denotes almost-sure convergence.

II. GRAPH LEARNING UNDER PARTIAL OBSERVABILITY

We consider a *random* graph defined over N nodes and denoted by $\mathcal{G}(N)$ (bold notation highlights graph randomness). The qualification “random” signifies that connections between nodes are drawn according to some probabilistic mechanism. Given a certain graph, the actions of the network nodes are described by a *distributed* linear dynamical system. Every node k , at time $i = 1, 2, \dots$, is driven by a random input source $\mathbf{x}_{k,i}(N)$ and produces the output signal $\mathbf{y}_{k,i}(N)$ according to the following diffusion model, a.k.a. first-order VAR model [39]:

$$\mathbf{y}_{k,i}(N) = \sum_{\ell=1}^N \mathbf{a}_{k\ell}(N) \mathbf{y}_{\ell,i-1}(N) + \mathbf{x}_{k,i}(N), \quad (2)$$

which can be conveniently recast in matrix form as:

$$\mathbf{y}_i(N) = \mathbf{A}(N) \mathbf{y}_{i-1}(N) + \mathbf{x}_i(N), \quad (3)$$

where $\mathbf{x}_i(N)$ and $\mathbf{y}_i(N)$ stack the entries $\mathbf{x}_{k,i}(N)$ and $\mathbf{y}_{k,i}(N)$ into $N \times 1$ column vectors, and where matrix $\mathbf{A}(N) = [\mathbf{a}_{k\ell}(N)]$ collects the nonnegative combination weights $\mathbf{a}_{k\ell}(N)$.² The eigenvalues of $\mathbf{A}(N)$ are assumed to lie strictly inside the unit circle to ensure that system (2) is stable. The combination matrix $\mathbf{A}(N)$ reflects the interconnections dictated by graph $\mathcal{G}(N)$. Thus, weight $\mathbf{a}_{k\ell}(N)$ is strictly positive if there is an edge from ℓ to k , and is zero otherwise. In view of (2), this structure implies that node k at time i updates its state $\mathbf{y}_{k,i}(N)$ by incorporating only previous-time signals $\mathbf{y}_{\ell,i-1}(N)$ received from nodes ℓ for which $\mathbf{a}_{k\ell}(N) > 0$. In general, $\mathbf{A}(N)$ need not be symmetric. For example, we could have $\mathbf{a}_{k\ell}(N) > 0$ and $\mathbf{a}_{\ell k}(N) = 0$. Accordingly, when we talk of “connected/disconnected pairs”, we refer to *ordered* pairs with (k, ℓ) being distinct from (ℓ, k) .

The stochastic dynamical system in (2) contains different sources of randomness. All involved random variables are assumed to lie in a common probability space (Ω, \mathcal{F}, P) . One source of randomness is given by the sequence of random graphs $\mathcal{G}(N)$, for $N = 1, 2, \dots$. We need to specify which probabilistic mechanism is used to generate such sequence. In the forthcoming sections we will focus on *preferential*

²Making explicit the dependence of the combination weights upon N is critical in our treatment, since we will need to examine the properties of these weights and related network descriptors as functions of N .

attachment random graphs, described in Sec. III-C. The combination matrix $\mathbf{A}(N)$ is a deterministic function of the graph $\mathcal{G}(N)$. Thus, once a graph realization is fixed, the combination matrix becomes deterministic, and the system in (2) evolves according to the randomness of the input signals $\mathbf{x}_{k,i}(N)$, which are independent and identically distributed (i.i.d.) w.r.t. to node index k , time index i , and network size N . These signals are statistically independent of the sequence of graphs, and, without loss of generality, are assumed to have zero mean and unit variance. The vectors $\mathbf{y}_0(N)$ that initialize the recursion (2) are assumed to be square-integrable random vectors with arbitrary distribution, independent of all input signals $\mathbf{x}_{k,i}(N)$. They are allowed to depend only on $\mathcal{G}(N)$ and, conditionally on $\mathcal{G}(N)$, they are independent of all other graphs in the sequence. The particular distribution of $\mathbf{y}_0(N)$ will be mostly immaterial for our results, since we will be dealing with the steady-state regime where the number of samples goes to infinity and the initial state does not play a role. Only when we study the sample complexity, we will assume a specific distribution for the initial state — see Theorem 3.

A. Partial Observability

The evolution of the signals $\mathbf{y}_{k,i}(N)$ is dictated by repeated interactions between neighboring nodes, and these interactions are determined by the graph topology. Thus, it is legitimate to ask whether the topology can be inferred from observing the evolution of the signals at the nodes.

However, in practice it is often impossible to collect signals from all nodes, and one can more likely observe signals emitted within a subset of nodes. Given a probed subset \mathcal{P} available under *partial* observability, and an observation time window $i = 1, 2, \dots, T$, the collection of signals available to perform graph learning will be compactly denoted by:

$$\mathbf{Y}_{\mathcal{P}}(T, N) \triangleq \{ \mathbf{y}_{k,i}(N) : k \in \mathcal{P}, i \in [1, T] \}. \quad (4)$$

Our goal is to estimate the interconnections between nodes in \mathcal{P} , namely, the topology of the partial graph $\mathcal{G}_{\mathcal{P}}(N)$ relative to \mathcal{P} , starting from the signals in (4). Formally, we need to build a *graph estimator*³:

$$\widehat{\mathcal{G}}_{\mathcal{P}}(T, N) = f(\mathbf{Y}_{\mathcal{P}}(T, N)), \quad (5)$$

where in the notation we emphasized that the properties of the graph estimator will depend on the number of samples and the network size. We will judge the goodness of a graph estimator in the classical asymptotic framework where the network size N goes to infinity, and the sample size $T = T_N$ to grow with N . The law T_N follows will characterize the so-called *sample complexity* of the estimator, namely, how the number of samples scales with the network size to provide faithful graph learning. In this analysis, we allow the probed subset to depend on N , namely, we introduce a deterministic *sequence* of subsets:

$$\mathcal{S}_N \subseteq \{1, 2, \dots, N\}, \quad (6)$$

³Formally, function f in (4) and function g in (9) are allowed to depend on T , N , and \mathcal{P} . This dependence is left implicit for ease of notation.

where \mathcal{S}_N is the probed subset of the graph of size N . For ease of notation, when a graph on N nodes or an $N \times N$ matrix is evaluated over subset \mathcal{S}_N , subscript N will be omitted, for example, we will write $\mathcal{G}_{\mathcal{S}}(N)$ in place of $\mathcal{G}_{\mathcal{S}_N}(N)$. According to this notation, a graph estimator will be said to be consistent for the family of random graphs $\mathcal{G}(N)$ and for the sequence of probed subsets \mathcal{S}_N if:

$$\lim_{N \rightarrow \infty} \mathbb{P} \left[\widehat{\mathcal{G}}_{\mathcal{S}}(T_N, N) = \mathcal{G}_{\mathcal{S}}(N) \right] = 1. \quad (7)$$

One meaningful asymptotic regime to quantify the degree of observability, studied in [29], prescribes that the cardinality of \mathcal{S}_N scales with the network size so as to converge to an asymptotic fraction ξ of probed nodes:

$$\frac{|\mathcal{S}_N|}{N} \xrightarrow{N \rightarrow \infty} \xi \in (0, 1). \quad (8)$$

We will consider this specific regime later in Sec. IV.

B. Achievability of Consistent Graph Learning

In order to build a graph estimator, we start by building an estimator for the combination submatrix $\mathbf{A}_{\mathcal{P}}(N)$:

$$\widehat{\mathbf{A}}_{\mathcal{P}}(T, N) = g(\mathbf{Y}_{\mathcal{P}}(T, N)). \quad (9)$$

It is seen that the properties of the matrix estimator $\widehat{\mathbf{A}}_{\mathcal{P}}(T, N)$ will depend on the number of samples and the network size. We will follow the following roadmap to establish whether graph learning is possible. First, we disregard sample complexity issues and consider the limiting case where an infinite number of samples is ideally available, namely, with $T \rightarrow \infty$ and N fixed. The estimator corresponding to an infinite number of samples will be referred to as the *limiting estimator* $\widehat{\mathbf{A}}_{\mathcal{P}}(N)$, which is formally introduced in Definition 1 further ahead. Second, in Definition 2, we specify when a *limiting estimator* is successful in recovering the true graph as $N \rightarrow \infty$. Finally, in Theorem 1 we come full circle and show that graph learning is achievable when the number of samples grows together with the network size. Let us now introduce the notion of limiting matrix estimator.

Definition 1 (Limiting Matrix Estimator): We say that a matrix estimator $\widehat{\mathbf{A}}_{\mathcal{P}}(T, N)$ converges to a limiting estimator $\widehat{\mathbf{A}}_{\mathcal{P}}(N)$ if, for any $\varepsilon > 0$:

$$\lim_{T \rightarrow \infty} \mathbb{P} \left[\|\widehat{\mathbf{A}}_{\mathcal{P}}(T, N) - \widehat{\mathbf{A}}_{\mathcal{P}}(N)\|_{\max} > \varepsilon \mid \mathbf{A}(N) = \mathbf{A} \right] = 0. \quad (10)$$

□

We notice that: *i*) the conditional probability in (10) depends on the *overall* combination matrix $\mathbf{A}(N)$; and *ii*) the limiting estimator $\widehat{\mathbf{A}}_{\mathcal{P}}(N)$ inherits the randomness in $\mathbf{A}(N)$, but in (10) we used normal font to denote it since the realization of $\mathbf{A}(N)$ is fixed in the pertinent conditional space. Condition (10) is a standard condition of consistency achieved by many empirical estimators, which are convergent as the number of samples scales to infinity.

If a limiting estimator $\widehat{\mathbf{A}}_{\mathcal{P}}(N)$ exists, it makes sense to examine whether it allows us to recover faithfully the true

graph. To this end, it is useful to introduce the following notion of consistency.

Definition 2 (Universal Local Structural Consistency of the Limiting Matrix Estimator): Consider a family of random graphs $\mathcal{G}(N)$ and a deterministic sequence of probed subsets \mathcal{S}_N . If there exist a positive sequence c_N and a positive random variable γ defined on the probability space (Ω, \mathcal{F}, P) , such that:

$$\|c_N \widehat{\mathbf{A}}_{\mathcal{S}}(N) - \gamma \mathbf{G}_{\mathcal{S}}(N)\|_{\max\text{-off}} \xrightarrow{P} 0, \quad (11)$$

then we say that the limiting estimator $\widehat{\mathbf{A}}_{\mathcal{P}}(N)$ achieves universal local structural consistency for the family of graphs $\mathcal{G}(N)$ and for the sequence of probed subsets \mathcal{S}_N , with scaling sequence c_N and identifiability gap γ . \square

As a matter of terminology: *i*) the adjective *universal* is used because, as we will promptly show, Eq. (11) automatically enables the possibility of recovering the topology by means of *unsupervised* clustering, i.e., without prior knowledge as regards the network size and other system parameters; *ii*) the adjective *local* comes from the observation that the structure of the topology connecting nodes in the probed subset will be faithfully recovered by probing *only these particular nodes*; and *iii*) the adjective *structural* is used because we estimate only the structure (i.e., the support graph) underlying the combination matrix.

By making explicit the definition of $\|\cdot\|_{\max\text{-off}}$, we see that Definition 2 is equivalent to the definition of Universal Local Structural Consistency used in [6] for the case of null bias (which is of interest for the type of graphs studied here), but for one important novelty: Definition 2 encompasses the possibility that the identifiability gap is *random*. Notably, we will see that randomness in the identifiability gap does not affect graph learning achievability.

In principle, the nature of the randomness of γ looks rather abstract from the definition. It is therefore useful to explain where this randomness comes from in our specific setting. As we will see in more detail later, the Bollobás-Riordan preferential attachment procedure produces a sequence of graphs of increasing size by adding edges at each step of the procedure. In this way, we obtain a sequence of random graphs defined on the same probability space. The random variable γ will arise as a specific limiting value associated to the sequence of maximum degrees of the graphs obtained during the preferential attachment procedure.

Let us now give some insight into the practical meaning of Definition 2. Since γ is positive, from (11) we can write, for $\varepsilon > 0^4$:

$$\lim_{N \rightarrow \infty} \mathbb{P} \left[\|c_N \widehat{\mathbf{A}}_{\mathcal{S}}(N) - \gamma \mathbf{G}_{\mathcal{S}}(N)\|_{\max\text{-off}} > \varepsilon \gamma \right] = 0, \quad (13)$$

⁴Given a sequence of random variables $e(N)$ vanishing in probability with N , and a strictly positive random variable \mathbf{z} , for an arbitrary $\delta > 0$ we have:

$$\begin{aligned} \mathbb{P}[e(N) > \mathbf{z}] &\leq \mathbb{P}[e(N) > \mathbf{z}, \mathbf{z} > z_{\delta/2}] + \mathbb{P}[\mathbf{z} \leq z_{\delta/2}] \\ &\leq \mathbb{P}[e(N) > z_{\delta/2}] + \delta/2 < \delta, \end{aligned} \quad (12)$$

where $z_{\delta/2} > 0$ is chosen such that $\mathbb{P}[\mathbf{z} \leq z_{\delta/2}] \leq \delta/2$ (a condition that can be met for any δ since $\mathbb{P}[\mathbf{z} \leq 0] = 0$), and the last inequality holds for sufficiently large N since $e(N)$ vanishes in probability. The arbitrariness of δ implies that $\lim_{N \rightarrow \infty} \mathbb{P}[e(N) > \mathbf{z}] = 0$, and (13) follows from (11) by setting $e(N) = \|c_N \widehat{\mathbf{A}}_{\mathcal{S}}(N) - \gamma \mathbf{G}_{\mathcal{S}}(N)\|_{\max\text{-off}}$ and $\mathbf{z} = \varepsilon \gamma$.

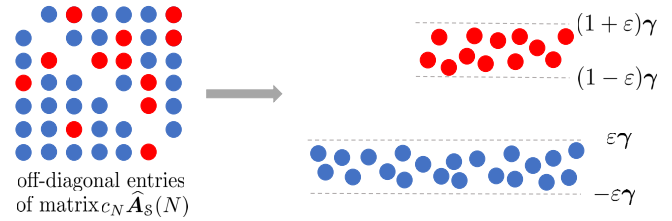


Fig. 1. Graphical illustration of Definition 2. Blue circles denote disconnected node pairs, whereas red circles denote connected pairs.

which means that, with high probability as $N \rightarrow \infty$, for all $k \neq \ell$:

$$c_N \widehat{a}_{k\ell}(N) \in \begin{cases} [(1-\varepsilon)\gamma, (1+\varepsilon)\gamma], & (k, \ell) \text{ conn.} \\ [-\varepsilon\gamma, \varepsilon\gamma], & (k, \ell) \text{ disc.} \end{cases} \quad (14)$$

We see that, for small ε , the estimated matrix entries are tightly clustered: *i*) around a positive value γ for connected node pairs, and *ii*) around zero for disconnected node pairs — see Fig. 1 for a graphical illustration. This property automatically enables the possibility of *clustering* the entries of matrix $\widehat{\mathbf{A}}_{\mathcal{S}}(N)$ so as to classify connected vs. disconnected pairs, ruling out the trivial case that the probed subgraph is either fully connected or fully disconnected, i.e., that we have only one cluster.

It is not difficult to envisage clustering algorithms that can achieve correct classification of the node pairs under condition (14). The choice of a particular clustering algorithm could matter from a practical perspective but, from a theoretical standpoint, it suffices to know that there exists at least one algorithm achieving correct clustering. To show that such an algorithm actually exists, we need to introduce first a formal definition of correct clustering.

Definition 3 (Correct Clustering): Let $\mathbf{A} = [\alpha_{k\ell}]$ be an $S \times S$ matrix with nonnegative entries and let \mathbf{G} the adjacency matrix corresponding to the support graph of \mathbf{A} . The diagonal entries of \mathbf{G} are zero by convention (since we are not interested in self-loops).

Let $\widehat{\mathbf{A}} = [\widehat{a}_{k\ell}]$ be an estimated matrix fulfilling, for certain positive values ε and a , and for all $k \neq \ell$:

$$\widehat{a}_{k\ell} \in \begin{cases} [(1-\varepsilon)a, (1+\varepsilon)a], & \text{if } \alpha_{k\ell} > 0 \\ & \text{(i.e., if } (k, \ell) \text{ is connected in } \mathbf{G}) \\ [-\varepsilon a, \varepsilon a], & \text{otherwise} \end{cases} \quad (15)$$

and let

$$\widehat{\mathbf{G}} = \text{graphclu}(\widehat{\mathbf{A}}) : \mathbb{R}^{S \times S} \rightarrow \{0, 1\}^{S \times S} \quad (16)$$

be a clustering algorithm that computes an estimated adjacency matrix $\widehat{\mathbf{G}}$. Then, algorithm $\text{graphclu}(\cdot)$ will be said to be correct if, when the sets of connected and disconnected pairs in \mathbf{G} are both non-empty, for a sufficiently small ε and independently of the value of a , we have $\widehat{\mathbf{G}} = \mathbf{G}$. \square

As an example of clustering that matches Definition 3, let us consider an algorithm that computes the midpoint between the

maximum and minimum off-diagonal entries of the estimated matrix. Using the bounds in (15) we can write:

$$\left(\frac{1}{2} - \varepsilon\right) a \leq \frac{\max_{k,\ell} \hat{\alpha}_{k\ell} - \min_{k,\ell} \hat{\alpha}_{k\ell}}{2} \leq \left(\frac{1}{2} + \varepsilon\right) a. \quad (17)$$

Accordingly, correct clustering will be surely performed if the lowest admissible value $(1 - \varepsilon)a$ for the connected pairs lies above the threshold, namely if

$$(1 - \varepsilon)a > \left(\frac{1}{2} + \varepsilon\right) a \Leftrightarrow \varepsilon < \frac{1}{4}, \quad (18)$$

and if the highest admissible value εa for the disconnected pairs lies below the threshold, namely if

$$\varepsilon a < \left(\frac{1}{2} - \varepsilon\right) a \Leftrightarrow \varepsilon < \frac{1}{4}. \quad (19)$$

In summary, the simple algorithm that employs an intermediate threshold to separate the clusters matches Definition 3, provided that $\varepsilon < 1/4$. However, it was shown in [33] that such algorithm, even if provably consistent as the network size scales to infinity, might suffer for finite network size effects. For this reason, another clustering algorithm was proposed in [33], namely, a modified k -means algorithm that: *i*) first finds all admissible k -means (with $k = 2$) cluster configurations, i.e., the configurations guaranteeing that the midpoint between the cluster centroids separate the clusters [40]); and *ii*) then selects the configuration featuring the largest distance between centroids. It was shown in [33] that this algorithm satisfies Definition 3 for any $\varepsilon < 1/6$.

From the above arguments, we conclude that a limiting estimator fulfilling Definition 2, paired with an algorithm $\text{graphclu}(\cdot)$ fulfilling Definition 3, achieves correct classification. However, in practice we need to establish that consistent graph learning is achieved by *sample* estimators, rather than by *limiting* estimators. The next theorem fills this gap, revealing that looking at the properties of the *limiting* estimator is in fact sufficient.

Theorem 1 (Universal Local Structural Consistency of the Limiting Matrix Estimator Implies Consistency of the Graph Estimator): Let $\hat{\mathbf{A}}_{\mathcal{P}}(T, N)$ be an estimator that converges to the limiting matrix estimator $\hat{\mathbf{A}}_{\mathcal{P}}(N)$ as $T \rightarrow \infty$ according to Definition 1. Consider a family of graphs $\mathcal{G}(N)$ and a sequence of probed subsets \mathcal{S}_N such that the probability that $\mathcal{G}_{\mathcal{S}}(N)$ is either fully connected or fully disconnected vanishes as $N \rightarrow \infty$. Let the limiting matrix estimator achieve universal local structural consistency according to Definition 2, and let $\text{graphclu}(\cdot)$ be a correct clustering algorithm according to Definition 3. Then, the graph estimator:

$$\hat{\mathcal{G}}_{\mathcal{P}}(T, N) = \text{graphclu}\left(\hat{\mathbf{A}}_{\mathcal{P}}(T, N)\right) \quad (20)$$

is consistent for the family of graphs $\mathcal{G}(N)$ and for the sequence of probed subsets \mathcal{S}_N , namely, there exists a scaling law T_N that ensures:

$$\lim_{N \rightarrow \infty} \mathbb{P}\left[\hat{\mathcal{G}}_{\mathcal{S}}(T_N, N) = \mathcal{G}_{\mathcal{S}}(N)\right] = 1. \quad (21)$$

Proof: See Appendix A. ■

C. Granger Estimator

Preliminarily, it is necessary to introduce the steady-state covariance matrix and the one-lag covariance matrix corresponding to model (2), which are, respectively:

$$\mathbf{R}_0(N) = \lim_{i \rightarrow \infty} \mathbb{E}\left[\mathbf{y}_i(N)\mathbf{y}_i^\top(N) \mid \mathbf{A}(N)\right], \quad (22)$$

$$\mathbf{R}_1(N) = \lim_{i \rightarrow \infty} \mathbb{E}\left[\mathbf{y}_i(N)\mathbf{y}_{i-1}^\top(N) \mid \mathbf{A}(N)\right], \quad (23)$$

where bold notation for the covariance matrices is used to encompass randomness of the underlying graph. Under model (2) (with a stable matrix $\mathbf{A}(N)$), it is known that the covariance matrix is the solution to the discrete-time Lyapunov equation $\mathbf{A}(N)\mathbf{R}_0(N)\mathbf{A}^\top(N) - \mathbf{R}_0(N) + \mathbf{I} = 0$, which is [41]:

$$\mathbf{R}_0(N) = \sum_{i=0}^{\infty} \mathbf{A}^i(N)[\mathbf{A}^i(N)]^\top. \quad (24)$$

Moreover, by exploiting (2) it is readily seen that:

$$\mathbf{R}_1(N) = \mathbf{A}(N)\mathbf{R}_0(N), \quad (25)$$

which implies the following inversion formula:

$$\mathbf{A}(N) = \mathbf{R}_1(N)\mathbf{R}_0^{-1}(N) \Rightarrow \mathbf{A}_{\mathcal{P}}(N) = [\mathbf{R}_1(N)\mathbf{R}_0^{-1}(N)]_{\mathcal{P}}. \quad (26)$$

Since covariance matrices can be faithfully estimated through *sample* covariance matrices as the number of samples increases, Eq. (26) suggests that the *true* matrix $\mathbf{A}_{\mathcal{P}}(N)$ can be actually estimated from the samples, which would imply that consistent graph learning is trivially possible.

However, under partial observability we can only compute the covariance matrices *over the probed subset*, $[\mathbf{R}_0(N)]_{\mathcal{P}}$ and $[\mathbf{R}_1(N)]_{\mathcal{P}}$. As a result, computation of the inversion formula (26) is impaired by the unavailability of signals from the latent nodes. Nevertheless, the limiting estimator (which actually depends only on the covariances over the probed subset):

$$\hat{\mathbf{A}}_{\mathcal{P}}(N) = [\mathbf{R}_1(N)]_{\mathcal{P}}([\mathbf{R}_0(N)]_{\mathcal{P}})^{-1} \quad (27)$$

is a meaningful choice that, in the context of Granger causality, is referred to as the *Granger estimator or predictor*, and attempts to provide the best linear prediction of the future samples from the past one-lag samples over the probed subset. On the other hand, from elementary matrix algebra we know that:

$$[\mathbf{R}_1(N)]_{\mathcal{P}}([\mathbf{R}_0(N)]_{\mathcal{P}})^{-1} \neq [\mathbf{R}_1(N)\mathbf{R}_0(N)^{-1}]_{\mathcal{P}}, \quad (28)$$

which gives rise to the question of whether $\hat{\mathbf{A}}_{\mathcal{P}}(N)$ can be still profitably used to estimate graph $\mathcal{G}_{\mathcal{P}}(N)$. We will answer this question in Sec. IV, with reference to our setting of *preferential attachment* Bollobás-Riordan graphs.

Finally, we introduce the *sample* Granger estimator:

$$\hat{\mathbf{A}}_{\mathcal{P}}(T, N) = [\mathbf{R}_1(T, N)]_{\mathcal{P}}([\mathbf{R}_0(T, N)]_{\mathcal{P}})^{-1}, \quad (29)$$

which replaces the true covariance matrices appearing in (27) with the sample covariance matrices $\mathbf{R}_0(T, N)$ and $\mathbf{R}_1(T, N)$,

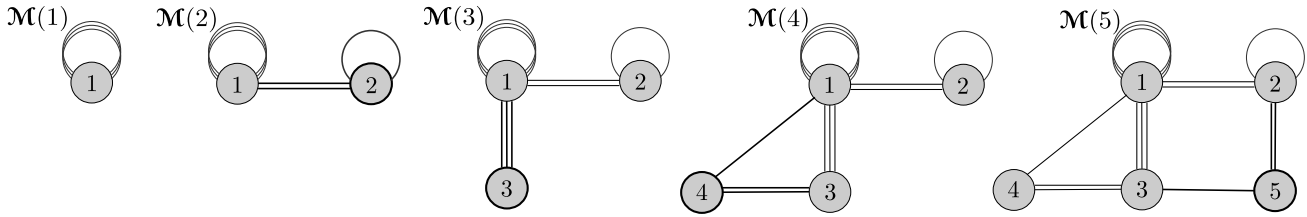


Fig. 2. One example of iterative construction of a Bollobás-Riordan multigraph with parameter $\eta = 3$.

whose (k, ℓ) -entries are defined as:

$$[\mathbf{R}_j(T, N)]_{k\ell} = \frac{1}{T-j} \sum_{i=1}^T \mathbf{y}_{k,i}(N) \mathbf{y}_{\ell,i-j}^\top(N), \quad j = 0, 1. \quad (30)$$

By ergodicity, the sample Granger estimator converges to the limiting estimator in (27) in the sense of Definition 1.

D. Laplacian Combination Matrix

Once a network graph $\mathcal{G}(N)$ is constructed, it is necessary to define a policy to assign the combination matrix $\mathbf{A}(N)$. Combination matrices arise across several domains, including distributed optimization, adaptation and learning over networks, social learning, network stochastic control. In all these domains, the system designer is called to run an algorithm (e.g., for optimization, learning, control) over a certain network topology. To this end, he/she devises a distributed procedure where the network nodes exchange locally information according to a certain combination matrix. One popular choice in these contexts is the Laplacian policy [42], [43], [44], [45], [46], [47]. The graph Laplacian $\mathbf{L}(N)$ is a matrix whose off-diagonal (k, ℓ) -entry is -1 for connected pairs (k, ℓ) and 0 otherwise, and whose k -th main diagonal entry is the degree of node k . Starting from $\mathbf{L}(N)$, the Laplacian combination matrix is defined as:

$$\mathbf{A}(N) = \rho \times (\mathbf{I} - c \mathbf{L}(N)), \quad c \triangleq \frac{\lambda}{1 + \mu_{\mathcal{G}}(N)}, \quad (31)$$

where $\mu_{\mathcal{G}}(N)$ is the maximal degree of $\mathcal{G}(N)$, $\lambda \leq 1$ is a positive parameter that tunes the relative importance of the self-weights, and $\rho < 1$ is a positive parameter that grants stability of the dynamical system in (2) — see [41]. The Laplacian combination rule can be conveniently described in terms of the individual entries as follows, for $k \neq \ell$:

$$\begin{cases} \mathbf{a}_{k\ell}(N) = 0, & (k, \ell) \text{ disconnected,} \\ \mathbf{a}_{k\ell}(N) = \frac{\rho\lambda}{1 + \mu_{\mathcal{G}}(N)}, & (k, \ell) \text{ connected,} \\ \mathbf{a}_{kk}(N) = \rho - \sum_{\substack{\ell=1 \\ \ell \neq k}}^N \mathbf{a}_{k\ell}(N). \end{cases} \quad (32)$$

III. BOLLOBÁS-RIORDAN MODEL

Preferential attachment graphs are typically obtained through an iterative process that goes as follows. Starting from a graph with a certain structure, at each subsequent iteration one node is added, along with some edges connecting this node to the graph constructed until that iteration. The term “preferential attachment” is used because the probability that

the new node is connected to an existing node is proportional to the degree of the latter. Therefore, nodes that have already experienced a large amount of connections are favored, giving rise to a dichotomy in the network, where some nodes emerge as hubs with most of the connections, whereas the remaining nodes become peripheral and feature few connections.

The way to build a preferential attachment model is not unique. Since the pioneering work [36], several preferential attachment models have been proposed. One of the most popular variants is the Bollobás-Riordan random graph, which is the model examined in this work [37], [38]. The Bollobás-Riordan model provides an elegant mathematical formulation that allows to capture many features of real-world networks and to obtain clean analytical results for useful graph descriptors (e.g., node degrees, minimum and maximal degrees, centrality measures). Let us delve into the mathematical description of the Bollobás-Riordan model [37], [38].

First of all, a Bollobás-Riordan graph is a *multigraph*, which means that multiple self-loops and multiple edges are permitted. A random multigraph of size n will be denoted by $\mathcal{M}(n)$ and its *adjacency matrix* by $\mathbf{M}(n)$. Matrix $\mathbf{M}(n)$ is the symmetric (since Bollobás-Riordan graphs are undirected) $n \times n$ matrix whose off-diagonal (k, ℓ) -entry $m_{k\ell}(n)$ is the number of edges between nodes k and ℓ , and whose diagonal entry $m_{kk}(n)$ is the number of self-loops of node k .

Then, the Bollobás-Riordan preferential-attachment model with parameter $\eta \in \mathbb{N}$ generates iteratively a random sequence of multigraphs $\mathcal{M}(n)$, for $n = 1, 2, \dots$, according to the following procedure — see Fig. 2 for a graphical illustration. The initial multigraph $\mathcal{M}(1)$ is a deterministic multigraph with one node and η self-loops. Multigraph $\mathcal{M}(n)$ is constructed starting from $\mathcal{M}(n-1)$ by adding a new node n and η new connections (edges or self-loops). Specifically, η steps are performed, and at each step node n is connected to a node randomly chosen from the set $\{1, 2, \dots, n\}$. The intermediate multigraph obtained at steps $s = 1, 2, \dots, \eta$, is denoted by $\mathcal{M}(n; s)$. Accordingly, since after η steps we obtain the updated multigraph $\mathcal{M}(n)$, we have the identity $\mathcal{M}(n; \eta) = \mathcal{M}(n)$. Likewise, we have $\mathcal{M}(n; 0) = \mathcal{M}(n-1)$.

Exploiting the procedure shown in Fig. 2 we can argue that the adjacency matrices possess the following structure:

$$\mathbf{M}(n) = \left(\begin{array}{c|c} & \begin{matrix} m_{1,n}(n) \\ \vdots \\ m_{n-1,n}(n) \end{matrix} \\ \hline \mathbf{M}(n-1) & \\ \hline m_{n,1}(n) \cdots m_{n,n-1}(n) & m_{n,n}(n) \end{array} \right), \quad (33)$$

with $\mathcal{M}(1) = \eta$. In fact, when passing from $\mathcal{M}(n-1)$ to $\mathcal{M}(n)$ we simply attach η new edges to the new node n , so that the number of edges $\mathbf{m}_{k\ell}(n-1)$ between any pair of nodes $k, \ell < n$ remains unaltered. In comparison, the adjacency matrix entries relative to the fresh node n evolve according to the following rule:

$$\mathbf{m}_{nk}(n) = \mathbf{m}_{kn}(n) = \sum_{s=1}^{\eta} \mathbb{I}(v(n; s) = k), \quad k \in \{1, \dots, n\} \quad (34)$$

where $\mathbb{I}(\cdot)$ is the indicator function (which is equal to 1 if its argument is true and is zero otherwise) and we denote by $v(n; s)$ the particular node that becomes connected to n through the edge introduced at step s . For this reason, for any $k \leq \ell \leq n$, the (k, ℓ) -entry of the adjacency matrix is actually determined only at iteration ℓ , and, hence, it makes sense to drop the dependence on n and write:

$$\mathbf{m}_{k\ell}(n) = \mathbf{m}_{k\ell}(\ell) \triangleq \mathbf{m}_{k\ell} = \mathbf{m}_{\ell k}. \quad (35)$$

A. Node Degrees and Preferential Attachment Rule

We adopt the standard convention that the degree of node k , denoted by $\mathbf{d}_{\mathcal{M},k}(n)$, is the number of edges connected to k plus twice⁵ the number of self-loops [48]:

$$\mathbf{d}_{\mathcal{M},k}(n) = \sum_{\substack{\ell=1 \\ \ell \neq k}}^n \mathbf{m}_{k\ell} + 2\mathbf{m}_{kk}. \quad (36)$$

Likewise, we denote by $\mathbf{d}_{\mathcal{M},k}(n; s)$ the degree of node k in the *intermediate* multigraph $\mathcal{M}(n; s)$.

At each step s , the degree of a node $k \neq n$ in the intermediate multigraph $\mathcal{M}(n; s)$ increases by 1 if the node picked at step s is equal to k , namely,

$$\mathbf{d}_{\mathcal{M},k}(n; s) = \mathbf{d}_{\mathcal{M},k}(n; s-1) + \mathbb{I}(v(n; s) = k). \quad (37)$$

In comparison, the degree of the new node n increases by 1 if the node picked at step s is equal to $k < n$, while it increases by 2 if node n itself is picked, since each self-loop is counted twice in the degree, with the initialization $\mathbf{d}_{\mathcal{M},n}(n; 0) = 0$.

$$\mathbf{d}_{\mathcal{M},n}(n; s) = \mathbf{d}_{\mathcal{M},n}(n; s-1) + 1 + \mathbb{I}(v(n; s) = n). \quad (38)$$

The description of the multigraph construction is now completed by assigning the probability that a particular node is picked. Consider first the probability that the new node n is attached to an existing node $k < n$, namely,

$$\mathbb{P}[v(n; s) = k | \mathcal{M}(n; s-1)] = \frac{\mathbf{d}_{\mathcal{M},k}(n; s-1)}{1 + \sum_{\ell=1}^n \mathbf{d}_{\mathcal{M},\ell}(n; s-1)}. \quad (39)$$

Let us ignore for now the term 1 appearing in the denominator. We see that the probability mass function in (39) matches well the preferential attachment paradigm, since we see that nodes

⁵If we sum all degrees over index k in (36), each edge is counted twice (because $\mathbf{m}_{k\ell} = \mathbf{m}_{\ell k}$), and each self-loop is counted twice (because of the factor 2). As a result, with the adopted convention the *half*-sum of the degrees in the multigraph is exactly equal to the total number of edges and self-loops.

with higher degrees in $\mathcal{M}(n; s-1)$ are more likely to be connected to the incoming node n , and so their degrees are more likely to increase further as the multigraph construction proceeds, according to “*the rich get richer*” philosophy.

We switch to the probability that a self-loop is created on the new node n :

$$\mathbb{P}[v(n; s) = n | \mathcal{M}(n; s-1)] = \frac{1 + \mathbf{d}_{\mathcal{M},n}(n; s-1)}{1 + \sum_{\ell=1}^n \mathbf{d}_{\mathcal{M},\ell}(n; s-1)}. \quad (40)$$

The term 1 in the numerator corresponds to first attaching one end of a new edge to the new node and updating the degree of that node before attaching the other end of the edge [37], [38]. Note that this term grants a nonzero probability of self-loops (we recall that $\mathbf{d}_{\mathcal{M},n}(n; 0) = 0$) when the new node enters the system. The term 1 in the denominator is necessary to get an admissible probability mass function, i.e., to let the sum of the probabilities in (39) and (40) be equal to 1.

It is useful to provide a more explicit representation for the denominator in (39) and (40). Since we know (see footnote 5) that the half-sum of all degrees is equal to the total number of edges and self-loops, and since at each step the Bollobás-Riordan construction adds exactly η new connections, we get the following equality:

$$1 + \sum_{k=1}^n \mathbf{d}_{\mathcal{M},k}(n; s-1) = 1 + 2\eta(n-1) + 2(s-1), \quad (41)$$

which reveals that the denominator of the preferential attachment probability is a deterministic quantity. Finally, by merging (39) and (40) in a single equation, and using (41) to represent the denominator, we get, for all $k = 1, 2, \dots, n$, the compact representation:

$$\mathbb{P}[v(n; s) = k | \mathcal{M}(n; s-1)] = \frac{\delta_{kn} + \mathbf{d}_{\mathcal{M},k}(n; s-1)}{1 + 2\eta(n-1) + 2(s-1)}, \quad (42)$$

where δ_{kn} is the Kronecker delta.

B. Maximal Degree

One fundamental graph descriptor that will play a critical role in our analysis is the *maximal degree* $\mu_{\mathcal{M}}(N)$. In particular, we will rely on the asymptotic growth of the maximal degree with the network size N , which, as formally stated in Appendix B, was found to be on the order of \sqrt{N} , in the following rigorous sense [49, Th. 8.14, p. 280]:

$$\frac{\mu_{\mathcal{M}}(N)}{\sqrt{N}} \xrightarrow{\text{a.s.}} \mu, \quad (43)$$

where $\xrightarrow{\text{a.s.}}$ denotes almost-sure convergence as $N \rightarrow \infty$, and μ is a certain *positive* random variable.

The square-root growth of the maximal degree in a Bollobás-Riordan graph can be related to the well known *power-law* or *scale-free* behavior of these graphs. The power-law decay refers to the average number of nodes with degree equal to d , which was shown to scale as an inverse power of d , precisely as d^{-3} for Bollobás-Riordan graphs. It was shown in [38] that such heavy-tailed behavior,

as opposed, for instance, to the exponential tail corresponding to an Erdős-Rényi graph, reflects into a faster growth of the maximal degree, namely, the \sqrt{N} growth prescribed by (43).

C. From Multigraph $\mathcal{M}(N)$ to Graph $\mathcal{G}(N)$

The *multigraph* structure was chosen by Bollobás and Riordan because it was instrumental to prove a number of theoretical results [37], [38]. The final goal of their model, however, was to construct a standard (i.e., simple) graph, with single edges and no self-loops. Actually, the multigraphs generated according to the Bollobás-Riordan model are approximately similar to simple graphs, since it is possible to show that the fraction of edges that are either repetitions or self-loops vanishes as N grows, as formally stated in the following lemma. Before stating the lemma, it is useful to notice that, by construction, the number of edges in $\mathcal{M}(n)$ is equal to ηn since we start with η loops in $\mathcal{M}(1)$ and add η new edges at a time.

Lemma 1 (Equivalence Between Bollobás-Riordan Multigraphs and Simple Graphs): Let $\mathcal{M}(N)$ be a Bollobás-Riordan multigraph of size N , and let $\hat{m}(N)$ and $\check{m}(N)$ be the number of self-loops and redundant edges, respectively. Then, the number of self-loops and redundant edges are asymptotically negligible w.r.t. the total number of connections ηN , namely,

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}[\hat{m}(N) + \check{m}(N)]}{\eta N} = 0. \quad (44)$$

Proof: See Appendix C. ■

According to Lemma 1, it makes sense to introduce the *simple*⁶ graph $\mathcal{G}(N)$ associated to a multigraph $\mathcal{M}(N)$, obtained by uprooting all self-loops and redundant edges from $\mathcal{M}(N)$. The entries of the adjacency matrix $\mathbf{G}(N)$ of graph $\mathcal{G}(N)$ are:

$$g_{kk} = 0, \quad g_{k\ell} = \min\{m_{k\ell}, 1\} \text{ for } k \neq \ell. \quad (45)$$

Likewise, the degree of node k in $\mathcal{G}(n)$ and the corresponding maximal degree are, respectively:

$$d_{\mathcal{G},k}(N) = \sum_{\ell=1}^N g_{k\ell}, \quad \mu_{\mathcal{G}}(N) = \max_{k \in [1, N]} d_{\mathcal{G},k}(N). \quad (46)$$

The equivalence between $\mathcal{M}(N)$ and $\mathcal{G}(N)$ holds also in terms of maximal degrees, as stated in the following lemma.

Lemma 2: Let $\mu_{\mathcal{M}}(N)$ and $\mu_{\mathcal{G}}(N)$ be the maximal degrees of the multigraph $\mathcal{M}(N)$ and of the associated simple graph $\mathcal{G}(N)$, respectively. We have that:

$$\frac{\mu_{\mathcal{M}}(N) - \mu_{\mathcal{G}}(N)}{\sqrt{N}} \xrightarrow{p} 0, \quad (47)$$

which further implies:

$$\frac{\mu_{\mathcal{G}}(N)}{\sqrt{N}} \xrightarrow{p} \mu, \quad (48)$$

where μ is the same limiting variable introduced in (43).

Proof: See Appendix C. ■

In the following, we will refer to graph $\mathcal{G}(N)$ as Bollobás-Riordan simple graph, or simply as Bollobás-Riordan graph.

⁶In graph theory, the qualification “simple” is used to stress that the graph has no self-loops and no multiple edges.

IV. STRUCTURAL CONSISTENCY AND SAMPLE COMPLEXITY OF THE GRANGER ESTIMATOR

In this section we collect the main results established in this work. First, we ascertain that the *limiting* Granger estimator achieves consistent graph learning under partial observability.

Theorem 2 (Universal Local Structural Consistency of the Limiting Granger Estimator): Let us consider the dynamical system (2), with Laplacian combination matrix as in (32), and with network graph $\mathcal{G}(N)$ being a simple graph obtained from a Bollobás-Riordan multigraph $\mathcal{M}(N)$ with step parameter η . Then, the limiting Granger estimator in (27) achieves universal local structural consistency according to Definition 2 for any sequence of probed subsets \mathcal{S}_N , with scaling sequence \sqrt{N} and identifiability gap:

$$\gamma \triangleq \frac{\rho\lambda}{\mu}, \quad (49)$$

where μ is given by (43).

Proof: See Appendix F. ■

Equation (43) reveals that the *limiting* (scaled) maximal degree μ is an *intrinsic* property of the specific Bollobás-Riordan multigraph instance. In other words, as the Bollobás-Riordan multigraph construction progresses, the maximal degree, scaled by \sqrt{N} , tends to become stable, and converges to a certain value μ . However, this value is *random*, implying that if we repeat the Bollobás-Riordan multigraph construction with the same parameters, we would obtain a *different* value for μ . In view of (49), this implies that the value of the identifiability gap that is critical for graph learning purposes is random as well, i.e., it depends on the particular graph sequence. This is a fundamental difference that distinguishes the behavior of Bollobás-Riordan graphs from the behavior of Erdős-Rényi graphs, where the identifiability gap is instead deterministic and independent of the particular graph realizations. However, and remarkably, we already know from Theorem 1 that randomness of the gap does not impair the possibility of consistent graph recovery.⁷

We observe that the coupling between Bollobás-Riordan graphs (which are undirected) and the Laplacian rule gives rise to a symmetric combination matrix. Under this assumption, the series for the covariance matrix in (24) can be computed as (I denotes the $N \times N$ identity matrix):

$$\mathbf{R}_0(N) = \sum_{i=0}^{\infty} \mathbf{A}^{2i}(N) = (\mathbf{I} - \mathbf{A}^2(N))^{-1}, \quad (50)$$

whose structure is exploited in the proofs of our results. While symmetry is useful to develop these technical arguments, we remark that the Granger estimator is based upon relation (25), which does *not* rely on symmetry at all. This observation, along with the series structure in (24) that is similar to the structure exploited in Appendix F for the symmetric case, suggests that the Granger estimator can work also with non-symmetric matrices, as we will show in Sec. V by examining a *directed* version of Bollobás-Riordan graphs.

⁷The result in Theorem 1 is asymptotic in N . On the other hand, for finite network sizes, the graph learning performance will depend on the particular realization, e.g., larger values of γ will imply a better performance.

In order to conclude the proof of graph learning achievability, we need to establish that the *sample* Granger estimator in (29) achieves consistent learning. To this end, we notice that the sample Granger estimator converges to the limiting Granger estimator by ergodicity. Then, in view of Theorem 1, the claim in Theorem 2 implies that the graph estimator in (20) is consistent for any sequence \mathcal{S}_N such that the sequence of graphs $\mathcal{G}_{\mathcal{S}}(N)$ does not end up being trivial (i.e., fully connected or fully disconnected) as $N \rightarrow \infty$. As shown in Lemma 6 (Appendix B), when the cardinality of \mathcal{S}_N scales linearly with N as in (8), these pathological situations are ruled with high probability as $N \rightarrow \infty$, for any nonzero fraction ξ of probed nodes.

However, the question of how the number of samples T must scale with N to achieve consistent learning is still open. To answer this question we need to focus on sample complexity. First of all, when dealing with covariance-based estimators, one source of sample complexity comes from how well-conditioned these matrices are. For this reason, it is useful to replace (29) with its *regularized* counterpart, namely, a matrix $\widehat{\mathbf{A}}_{\mathcal{P}}(T, N)$ constructed as follows. For $k \in \mathcal{P}$, the k -th row of $\widehat{\mathbf{A}}_{\mathcal{P}}(T, N)$ is a solution to the constrained optimization problem (here $x \in \mathbb{R}^{|\mathcal{P}|}$ is a row vector, and, for a matrix M , the notation $[M]_{k\mathcal{P}}$ denotes the k -th row of submatrix $M_{\mathcal{P}}$):

$$\min_{x \in \mathbb{R}^{|\mathcal{P}|}} \|x [\mathbf{R}_0(T, N)]_{\mathcal{P}} - [\mathbf{R}_1(T, N)]_{k\mathcal{P}}\|_{\infty} \quad \text{s.t.} \quad \|x\|_1 \leq 1. \quad (51)$$

We remark that, when the sample covariance matrix is invertible, the non-regularized Granger estimator in (29) is the only matrix that yields a zero residual in (51). As a result, when the non-regularized Granger estimator fulfills the constraint in (51), the two estimators coincide.

We are ready to state our sample complexity result. Following [24], [25], [26], [27], the analysis is performed under the following classical assumption on the initialization of system (2).

Assumption 1 (Stationary Gaussian VAR): For the sample complexity analysis, we assume that the input signals $\mathbf{x}_{k,i}(N)$ are standard Gaussian variables (independent w.r.t. to node index k , time index i , and network size N). Under these conditions, given a realization $\mathbf{A}(N) = A$ of the combination matrix, the VAR process in (2) admits a Gaussian stationary distribution (which is a function of A) [24], [39]. We assume that, given $\mathbf{A}(N) = A$, the initial vector $\mathbf{y}_0(N)$ is distributed according to the stationary distribution.

Theorem 3 (Sample Complexity of Regularized Granger Estimator): Let us consider the dynamical system (2) operating under Assumption 1, with Laplacian combination matrix as in (32), and with network graph $\mathcal{G}(N)$ being a simple graph obtained from a Bollobás-Riordan multigraph $\mathcal{M}(N)$ with step parameter η . Let the number of samples T_N grow with N as:

$$T_N = \omega_N N \log N, \quad (52)$$

where ω_N can be chosen as a positive sequence diverging in an arbitrarily slow fashion. Then, the graph estimator (20)

obtained by using the regularized Granger estimator in (51) is consistent for any sequence of probed subsets \mathcal{S}_N fulfilling (8).

Proof: See Appendix G. ■

Let us comment on the main ramifications of Theorem 3. First of all, since the sequence ω_N can grow in an arbitrarily slow fashion, any sample complexity that scales slightly faster than $N \log N$ achieves consistency. Therefore, the bottom line of Theorem 3 is that the sample complexity of the proposed estimator is *essentially linear*. Let us now see where this linear law originates from.

According to the Laplacian matrix structure in (32), the growth of the maximal degree determines the way the nonzero entries of the combination matrix shrink down as $N \rightarrow \infty$. The smaller they are, the higher is the precision required by the sample estimators to distinguish the nonzero entries from the zero entries. For this reason, faster scaling laws of the maximal degree become more demanding in terms of number of samples. This argument can be made rigorous, and is in fact exploited in the proof of Theorem 3 to show that the sample complexity goes essentially (i.e., up to a $\log N$ factor) as $\mu_{\mathcal{G}}^2(N)$. As a result, the \sqrt{N} -growth of the maximal degree over Bollobás-Riordan graphs reflects into a final sample complexity that is essentially linear in N .

In summary, from a technical viewpoint we conclude that the main factor influencing sample complexity is the maximal degree of the graph. However, it is useful to relate this behavior to more “physical” attributes of the system, to capture the factors that play a domineering role on sample complexity. One important attribute of Bollobás-Riordan graphs is *sparsity*. Bollobás-Riordan graphs are *very sparsely* connected, since, over a total number of possible $N(N-1)/2$ edges, only ηN edges are drawn, which results into a sparsity ratio (no. of connected edges over total no. of possible edges) scaling as $1/N$. The sporadic presence of connections might suggest that the nodes have small degree, which, in the light of the previous discussion, would suggest a slow growth of the maximal degree. However, this conclusion is not precise. To understand why, it is useful to contrast Bollobás-Riordan graphs against Erdős-Rényi graphs. We consider in particular Erdős-Rényi graphs under the uniform concentration regime because under this regime results on sample complexity are available [6], [33]. Erdős-Rényi graphs are built homogeneously (i.e., presence/absence of edges is established in an i.i.d. manner). This homogeneity implies in particular that the maximal and average degree of an Erdős-Rényi graph scale comparably, i.e., $\mu_{\mathcal{G}}(N) \sim Np$, where p is the connection probability. Accordingly, over *sparse* Erdős-Rényi graphs where $p \approx (\log N)/N$, the sample complexity is *polylogarithmic* in N , whereas over *dense* graphs with constant p it is *almost-quadratic* in N .

Let us see what happens over Bollobás-Riordan graphs. Notably, the latter graphs are *sparser* than the sparsest connected Erdős-Rényi graphs! In fact, we observed that the sparsity ratio of Bollobás-Riordan graphs is $1/N$, whereas for sparse connected Erdős-Rényi graphs we have a sparsity ratio given by the connection probability $\approx (\log N)/N$. However, despite such increased sparsity, the maximal degree of Bollobás-Riordan graphs grows as \sqrt{N} , namely, faster than

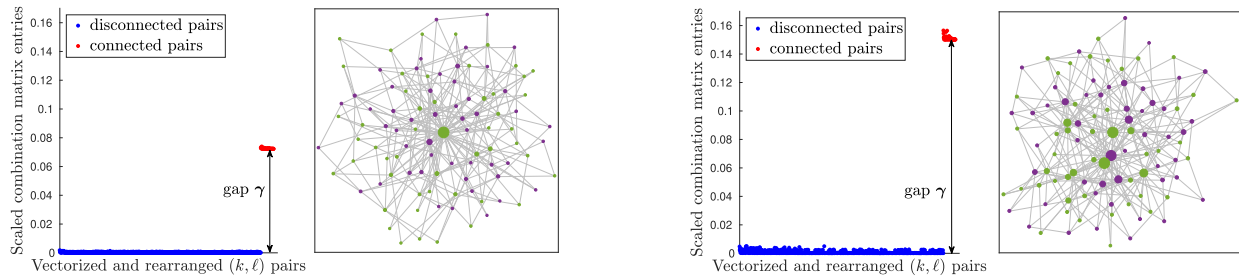


Fig. 3. Illustration of Theorem 2, for two realizations of a Bollobás-Riordan graph with number of nodes $N = 100$ and parameter $\eta = 3$. The plots show the entries of the limiting Granger estimator in (27), scaled by \sqrt{N} , vectorized and rearranged so that entries corresponding to disconnected nodes come first. The vertical arrow displays the gap γ . In the shown network topologies, probed nodes are displayed in green, while latent nodes in purple, with the circle radius being proportional to the node degree. The probed subset has cardinality $N/2 = 50$, and its nodes are randomly picked from $\{1, 2, \dots, N\}$ without replacement. The parameters of the Laplacian matrix are $\rho = 0.5$ and $\lambda = 0.75$.

the logarithmic law characterizing sparse Erdős-Rényi graphs. This difference must be ascribed to the fact that Bollobás-Riordan graphs are highly *inhomogeneous* and, hence, even with a small number of overall connections, there are nodes with a very large number of neighbors, inducing a faster growth of the *maximal* degree.

V. ILLUSTRATIVE EXAMPLES

A. Synthetic Data

We start by illustrating the significance of Theorem 2 by examining the behavior of the limiting Granger estimator on synthetic data.

The two panels in Fig. 3 display the pattern exhibited by the entries of the *limiting* Granger estimator, $\hat{A}_{\mathcal{P}}(N)$, for two realizations of the random graph $\mathcal{G}(N)$ with a probed subset \mathcal{P} containing half the nodes of the entire network. For both realizations, the entries of $\hat{A}_{\mathcal{P}}(N)$ are scaled by \sqrt{N} , and, for clarity of visualization, they are vectorized and rearranged so that the entries corresponding to disconnected nodes come first. The vertical arrow displays the gap γ , which was estimated using (43), (47) and (49), with reference to the pertinent graph topology shown in the figure. The following notable effects are observed. First, in perfect accordance with Theorem 2, we see the emergence of an *identifiability gap* that separates clearly the entries corresponding to disconnected node pairs from the entries corresponding to connected node pairs. Second, *clustering* is definitely visible: the entries pertaining to disconnected nodes cluster around zero, whereas the entries corresponding to connected nodes around γ , the limiting value displayed by the vertical arrow. Last but not least, by comparing side-by-side the panels in Fig. 3, we see that the two different realizations correspond to different values of the gap γ , which confirms that this gap is in fact random.

We see from (49) that the limiting random variable μ is one fundamental ingredient of the identifiability gap. It is therefore useful to examine the statistical distribution of μ , and in particular its behavior in comparison to the finite-size (scaled) maximal degree $\mu_{\mathcal{G}}(N)/\sqrt{N}$ — see (48). To this end, in Fig. 4 we display: *i*) the empirical histograms of the scaled maximal degree, for three values of N (first three panels from the left); and *ii*) the empirical histogram corresponding

to the limiting variable μ (rightmost panel). To obtain the latter histogram, we exploit the following result.

Theorem 17 in [37]: Let p_1, p_2, \dots be the points of a Poisson process with rate η , i.e., equal to the number of new edges added at each iteration of the Bollobás-Riordan procedure (see Sec. III). The limiting random variable in (64) is equal to:

$$\mu = \max_{n=1,2,\dots} z_n - z_{n-1}. \quad (53)$$

where, for $n = 0, 1, \dots$, we defined:

$$z_0 \triangleq 0, \quad z_n \triangleq 2\eta\sqrt{p_{\eta n}}. \quad (54)$$

□

According to this theorem, we simulate a Poisson process with rate η and use the expression of μ provided in (53) and (54). By comparing the different panels in Fig. 4, we see that the distribution of the scaled maximal degree approaches the distribution of μ as N increases, and that the result is stable yet for the values $N = 100$ and $N = 250$. These are interesting values since, in the range $[100, 250]$, the probability of correct graph learning is close to 1, as we can appreciate from the quantitative performance analysis reported in Fig. 5.

More specifically, in Fig. 5 we show the probability of correct graph learning evaluated empirically over 10^3 Monte Carlo runs, as a function of the network size N . Specifically, the dynamical evolution in (2) is simulated over a network of increasing size N ranging from 50 to 250, and we consider a subset of probed nodes having cardinality $\lfloor \xi N \rfloor$, with $\xi = 0.15$. The curves displayed with continuous line refer to the *limiting* Granger estimator in (27), which is obtained by using the *true* covariance matrices. Markers refer to the regularized Granger estimator in (51), which is instead computed over the samples. The take-away messages from Fig. 5 are that: *i*) for sufficiently large number of samples, the learning curve of the empirical Granger estimator reaches the curve of the limiting Granger estimator; and *ii*) consistent learning is progressively achieved as N grows.

In Fig. 5, a relatively large number of samples is considered, and kept constant across all values of the network size N . Another useful analysis pertains to the effective number of samples necessary to achieve a target learning probability. In Fig. 6 we evaluate empirically the number of samples needed to get a probability equal to 90% of the probability of

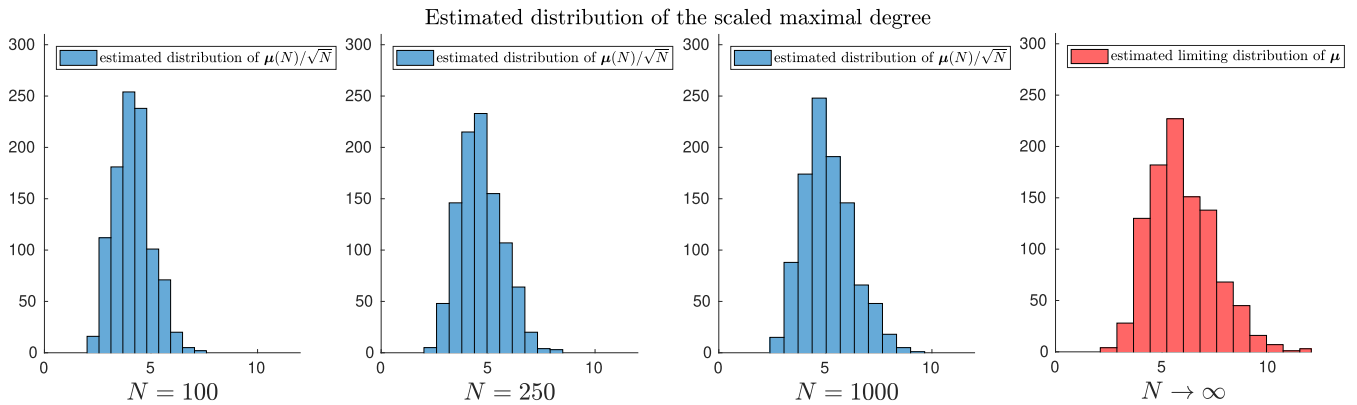


Fig. 4. *First three panels.* Empirical histograms, obtained over 10^3 Monte Carlo runs, relative to the scaled maximal degree $\mu_g(N)/\sqrt{N}$. *Rightmost panel.* Empirical histogram relative to the limiting random variable μ , obtained by simulating, over 10^3 Monte Carlo runs, a Poisson process of rate η , and by exploiting relations (53) and (54).

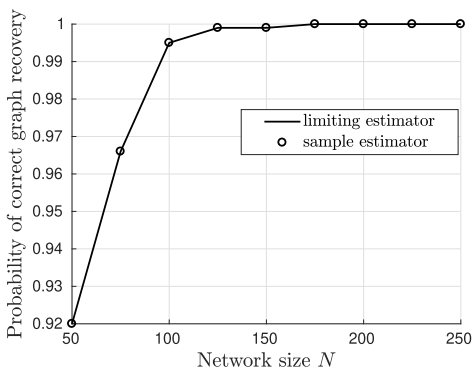


Fig. 5. Probability of correct graph recovery, computed over 10^3 Monte Carlo runs, for different values of the network size N . We remark that *correct* graph recovery here means that graphs with even a single wrong edge are counted as an erroneous experiment. The graphs are generated according to a Bollobás-Riordan model with parameter $\eta = 3$. The sequence of probed subsets fulfills (8) with $\xi = 0.15$. We consider: the *limiting* estimator (27) obtained by using the true covariances (solid line); and the *empirical* estimator (51) obtained by using the sample covariances (markers) evaluated over $T = 3 \cdot 10^6$ samples. The clustering algorithm applied to the Granger estimator is the modified k -means algorithm proposed in [33]. The parameters of the Laplacian matrix are $\rho = 0.5$ and $\lambda = 0.75$.

correct learning achieved by the limiting Granger estimator. The blue curve corresponds to Bollobás-Riordan graphs, and shows a growth that matches well the almost-linear growth prescribed by Theorem 3.

It is useful to compare the observed behavior against the behavior of Erdős-Rényi graphs. The sample complexity laws relative to Erdős-Rényi graphs mentioned in the previous section are confirmed by the curves in Fig. 6, revealing in particular that: *i*) the intermediate growth rate is given by Bollobás-Riordan graphs (blue curve), with *almost-linear* sample complexity; *ii*) the highest sample complexity is *quadratic*, and is required by *dense* Erdős-Rényi graphs (green curve); and *iii*) the lowest sample complexity is achieved by *sparse* Erdős-Rényi graphs (red curve), and depends on the specific law chosen for the vanishing connection probability p .

The bottom line is that: *i*) *sparsity* of preferential attachment graphs makes them easier to learn than *dense* graphs; whereas

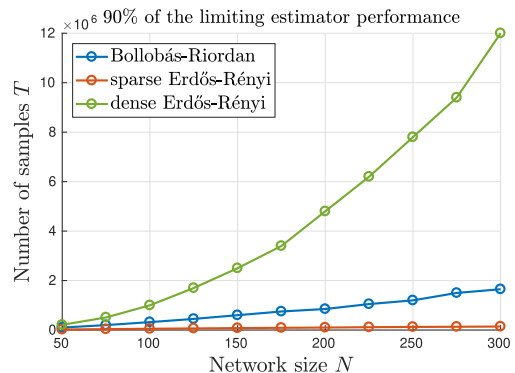


Fig. 6. Sample complexity of preferential attachment graphs, compared against sparse and dense Erdős-Rényi graphs. The curves depict the number of samples needed by the *empirical* estimator to attain 90% of the performance (i.e., probability of correct graph recovery) of the *limiting* estimator, for different values of N . The preferential attachment graphs are generated as Bollobás-Riordan graphs with parameter $\eta = 3$. The sparse Erdős-Rényi graphs are generated with a connection probability $p = \frac{\log N}{N} \cdot \log \log N$, whereas for the dense Erdős-Rényi graphs we have $p = 0.5$. The underlying probability of correct graph recovery is evaluated over 10^3 Monte Carlo runs. The parameters of the Laplacian matrix are $\rho = 0.5$ and $\lambda = 0.75$.

ii) *heterogeneity* of preferential attachment graphs implies a power-law behavior that reflects into a \sqrt{N} -growth of the maximal degree, making them harder to learn than *sparse homogeneous* graphs.

Finally, we provide some quantitative data as regards the computational complexity of the graph learning strategy in the considered examples. To this end, we now report the run times relative to the XPS 7390 laptop of Dell Inc.[®], equipped with an i7 Intel[®] processor and a 16GB RAM. The graph learning algorithm can be decoupled in two steps: *i*) computing the Granger estimator; and *ii*) performing the clustering algorithm on its entries. The cost associated to the clustering algorithm is negligible. In the first step, if we use the regularized Granger estimator, we need an optimization algorithm to solve numerically (51). In our simulations, we employed the MATLAB[®] package CVX [50], [51], which exhibited a run time ranging from ≈ 3 s to ≈ 8 s when N ranges from 100 to 250, with $\xi = 0.15$. The run time reduces to less than

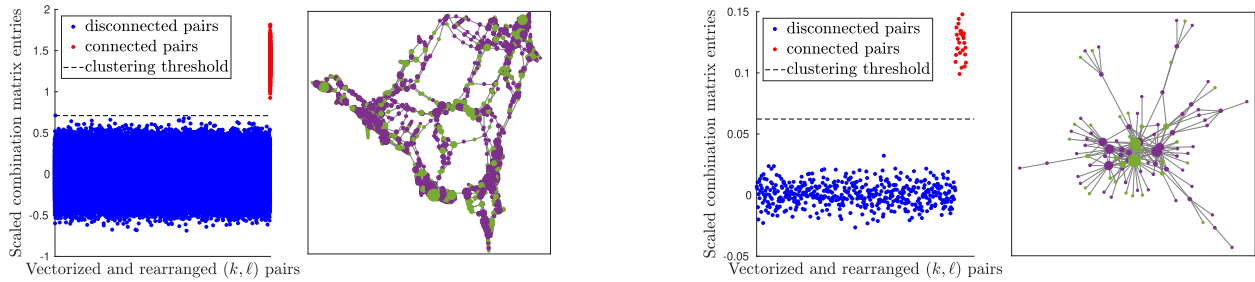


Fig. 7. Experiments over real-world topologies. The parameters of the Laplacian matrix are $\rho = 0.5$ and $\lambda = 0.75$. *Left.* A simulation run of (2) over a power-grid network of $N = 4941$ nodes taken from the web-repository [54]. The plot shows the entries of the regularized Granger estimator in (51), scaled by \sqrt{N} , vectorized and rearranged so that entries corresponding to disconnected nodes come first. In this run we set $T = 2 \cdot 10^5$, and consider a subset of probed nodes having cardinality $N/3 = 1647$. In the network topology, probed nodes are displayed in green, while latent nodes in purple, with the circle radius being proportional to the node degree. *Right.* The same general setting as in the left panel, with reference to a network of $N = 100$ real-world routers, whose connection topology was extracted from a bigger network available in the web-repository [54]. In this run we set $T = 10^6$, and consider a subset of probed nodes having cardinality $N/2 = 50$.

1 ms if we use instead the non-regularized Granger estimator⁸ in (29), which in the considered examples was found to coincide with its regularized counterpart — see the discussion following (51).

B. Real Networks and Directed Graphs

So far, we tested our results over *synthetic* network topologies generated according to the Bollobás-Riordan procedure described in Sec. III. Since the main motivation behind the challenging study of these graphs is their similarity to real-world graphs, in this section we examine some topologies of existing networks. The examples that we are going to illustrate should be intended in the following way. We are given the topology of a real-world network, such as, e.g., a power-grid network, a network of routers, or a social network, which can support the implementation of distributed learning algorithms for different useful purposes. We therefore use the assigned network topology to build/run on top of it a distributed algorithm, for example, an adaptive distributed detection algorithm, or a social learning algorithm, which are examples matching well the considered model — see [6, Sec. V-A], [52], [53]. Then, the focus of topology inference is to solve the reverse learning problem of retrieving the network graph from partial observation of the nodes' output.

We are now ready to illustrate the tests conducted over two real-world networks provided by a popular web-repository [54], corresponding to the topologies shown in Fig. 7. As a preliminary comment, we can see that these topologies exhibit a dichotomous structure with “hubs” featuring many connections as opposed to “peripheral” nodes with few connections. Similarly shaped structures match well the heterogeneity guaranteed by Bollobás-Riordan models, while they are impossible to mimic through the independent/homogeneous Erdős-Rényi generation.

The example in the left panel of Fig. 7 refers to a power-grid network composed of $N = 4941$ nodes, connected according to the displayed topology. Over this topology, we let the VAR

⁸We remark that no theoretical proof on the sample complexity of the non-regularized Granger estimator is available.

system (2) run with a Laplacian combination policy, and then applied our inference algorithm under the case that only one third of the nodes are probed, with $T = 2 \cdot 10^5$ samples. The results of the test are shown in the left plot in Fig. 7. We see that the clustering algorithm is able to separate correctly the disconnected/connected nodes, therefore providing faithful graph learning. Moreover, we see that the spread exhibited by the matrix entries of the *sample* estimator is larger than the spread exhibited by the limiting estimator, which makes perfect sense.

The second example, illustrated in the right panel of Fig. 7, refers to a network of 100 routers connected according to the shown topology, which was extracted from a bigger network reported in the web-repository [54]. In this case, 50% of the nodes are probed, and we have $T = 10^6$ available samples. The results of the test are shown in the right plot in Fig. 7, where we can appreciate that the graph learning algorithm successfully classifies the node connections within the probed subnetwork. In comparison to the left panel, we see that in the right panel the spread of the sample estimators is reduced, which is a consequence of the fact that we have a larger number of samples and a smaller network size.

Finally, we test whether the Granger estimator can achieve faithful graph learning over *directed* graphs. While Bollobás-Riordan graphs are naturally undirected, there are of course several straightforward ways to devise directed variants thereof, see, e.g., [37]. Perhaps the simplest way is to perform, at each step of the preferential attachment construction: *i*) the insertion of a *directed* edge from the new node n to an existing node, based on an attachment probability ruled by the *in-degree* of the existing nodes; and *ii*) the insertion of a *directed* edge from an existing node to the new node n , based on an attachment probability ruled by the *out-degree* of the existing nodes. Such construction is used in Fig. 8, where we report two realizations relative to the parameters described in the caption. Regarding the Laplacian matrix, in the directed case we use definition (32) with the maximal *in-degree*. We see that, even in this non-symmetric case, the limiting Granger estimator is still able to separate well connected from disconnected pairs.

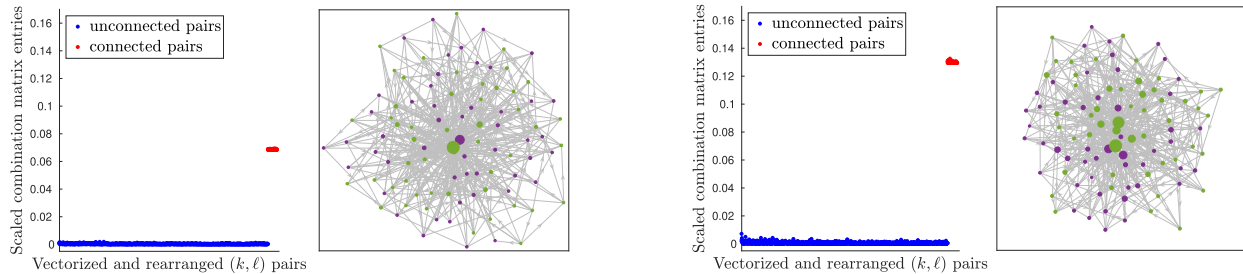


Fig. 8. Two realizations of a *directed* Bollobás-Riordan graph, whose construction is detailed in the main text. We set $N = 100$ and $\eta = 3$. The plots show the entries of the limiting Granger estimator in (27), scaled by \sqrt{N} , vectorized and rearranged so that entries corresponding to disconnected nodes come first. In the shown network topologies, probed nodes are displayed in green, while latent nodes in purple, with the circle radius being proportional to the node *in-degree*. The probed subset has cardinality $N/2 = 50$, and its nodes are randomly picked from $\{1, 2, \dots, N\}$ without replacement. The parameters of the Laplacian matrix are $\rho = 0.5$ and $\lambda = 0.75$.

VI. CONCLUSION

We examined the problem of learning a network graph from the signals diffusing across the network according to the VAR model in (2). The distinguishing features of our work are: *i*) the network topology is modeled as a preferential attachment random graph; and *ii*) only part of the network is monitored (partial observability). We established that the Granger estimator under partial observability achieves faithful graph learning (with high probability as the network grows) for the class of Bollobás-Riordan graphs when the signals of neighboring nodes are combined according to the Laplacian matrix of the graph.

Previous results on consistent graph learning under partial observability, for diffusion models like (2), were relative to Erdős-Rényi graphs, which cannot reproduce faithfully the behavior of several useful real-world networks. In contrast, preferential attachment graphs were shown to be powerful in capturing useful real-world effects such as node heterogeneity and statistical dependence across graph edges. Accordingly, moving from Erdős-Rényi to Bollobás-Riordan graphs constitutes a useful research advance, which was rather demanding, especially because the multigraph construction relies on a preferential attachment mechanism, which introduces significant dependence across the edges, thus preventing from application of the simpler i.i.d. models adopted for the former graph models. Exploiting statistical concentration results for dependent processes, we are able to examine in detail the limiting properties of these graphs, and to ascertain (Theorem 2) that the entries of the Granger matrix estimator computed over the probed subnetwork split into two classes, separated by an *identifiability gap*. As a distinguishing feature of the Bollobás-Riordan model, this gap is a *random* variable, which depends on the particular instance of the multigraph generation, as opposed to the deterministic gap observed for Erdős-Rényi graphs. We proved that the emergence of the gap is critical to enable *achievable graph learning* (Theorem 1), and characterized the scaling law that relates how the number of samples must grow with the network size (Theorem 3), finding that the sample complexity is slightly larger than $N \log N$.

There are many other open questions that might deserve attention. For example, achievability is expressed in a worst-case perspective where *perfect* graph reconstruction

is required. It would be useful to relax the criterion to encompass a possible fraction of misclassified edges, and see how this impacts the performance and the requested number of samples. Moreover, since classification is performed by using an unsupervised clustering algorithm, it would be useful to see whether one could explore side-information to set a classification threshold, and see how this changes the fraction of misclassified nodes.

We established that the identifiability gap over Bollobás-Riordan graphs is *random*, as opposed to the deterministic nature that was proved over Erdős-Rényi graphs. This difference stimulates an open question that concerns the connections between the nature (deterministic or random) of the identifiability gap and the generative mechanism of the graph.

Also considering that the ubiquity of scale-free networks is a debated point [55], it would be interesting to consider other useful graph models, such as the stochastic block model, Chung-Lu graphs, random dot product or random geometric graphs. In particular, it would be interesting to ascertain whether randomness of the identifiability gap is related to the scale-free property or the preferential attachment mechanism.

The conducted analysis does not depend on a specific selection rule for the subset of probed nodes. Indeed, some of our results hold for any (deterministic) subset of probed nodes. For Theorem 3, we just need to impose that the cardinality of this subset scales linearly with N so as to avoid trivial (i.e., fully connected or fully disconnected) subgraphs as $N \rightarrow \infty$. Regarding the choice of the probed subset, it is interesting to consider an adversarial perspective where a malicious entity wants to select it to impair the topology inference algorithm. In the traditional setting, attacks to network graphs have the goal of impairing the connectivity properties the network. One classical attack is the *deletion* attack, where the attacker has the freedom of deleting some nodes [56]. The goal is to reduce the connectivity of the selected subgraph surviving after the deletion process. In particular, when the attacker leverages knowledge of the preferential attachment construction, he/she can delete all the “oldest” nodes to minimize connectivity. In our context, connectivity of probed nodes is not of interest, but one way to impair the clustering algorithm would be to select a subset that is fully connected or fully disconnected. However, when the cardinality of the probed subset grows linearly with N , we know that both these extreme situations

occur with vanishing probability as N grows. The picture changes if the attack does not rely only on the graph model, and the adversary has the power of choosing the subset based on the actual realizations of the graph and/or of the nodes' output. In this case, the subset becomes statistically dependent on other random variables, and the statistical properties of the combination matrix and its estimated counterpart change due to the dependence introduced between the subset and the matrix entries. Carrying out the analysis under this scenario requires a different analysis that is not covered by the results of this work.

Devising matrix estimators different from (51), exploiting in particular some other structural constraints such as sparsity or smoothness can be useful to reduce sample complexity.

Another important aspect pertains to *online* graph learning algorithms, for example, one could implement an online version of (51) that learns from streaming data, encompassing the possibility that the graph topology changes due to the evolutionary mechanism of the Bollobás-Riordan graph itself.

Finally, other useful research advances concern the generalization to higher-order VAR processes, nonlinear models, or other combination matrices.

APPENDICES

The proofs reported in these appendices are organized as follows. In Appendix A we prove Theorem 1. In Appendices B and C we collect several properties on the Bollobás-Riordan model useful for our analysis. By exploiting these results, in Appendix D we show how the underlying graph structure influences the behavior of the Laplacian combination matrix. Next, in Appendix E we establish some useful properties of the limiting Granger estimator. Our main results, namely, Theorems 2 and 3, are proved in Appendices F and G, respectively. Finally, in Appendix H we collect two auxiliary results used in our analysis.

APPENDIX A PROOF OF THEOREM 1

By application of the triangle inequality we can write:

$$\begin{aligned} & \|c_N \widehat{\mathbf{A}}_S(T, N) - \gamma \mathbf{G}_S(N)\|_{\max\text{-off}} \\ & \leq c_N \|\widehat{\mathbf{A}}_S(T, N) - \widehat{\mathbf{A}}_S(N)\|_{\max\text{-off}} \\ & + \|c_N \widehat{\mathbf{A}}_S(N) - \gamma \mathbf{G}_S(N)\|_{\max\text{-off}}. \end{aligned} \quad (55)$$

Let us focus on the first term on the RHS of (55). From the definition of limiting estimator in (10), we have, for any $N \in \mathbb{N}$ and $\varepsilon > 0$:

$$\lim_{T \rightarrow \infty} \mathbb{P} \left[c_N \|\widehat{\mathbf{A}}_S(T, N) - \widehat{\mathbf{A}}_S(N)\|_{\max\text{-off}} > \varepsilon \right] = 0. \quad (56)$$

By definition of limit, from (56) we conclude that, for any $N \in \mathbb{N}$ and any $\delta, \varepsilon > 0$, there exists always a value $T_0(N, \delta, \varepsilon)$ such that for all $T \geq T_0(N, \delta, \varepsilon)$:

$$\mathbb{P} \left[c_N \|\widehat{\mathbf{A}}_S(T, N) - \widehat{\mathbf{A}}_S(N)\|_{\max\text{-off}} > \varepsilon \right] \leq \delta. \quad (57)$$

Let now f_N and g_N be two positive sequences vanishing with N with arbitrary laws. Since the reduction of δ and/or ε is a more demanding condition, the function $T_0(N, \delta, \varepsilon)$ can be

always chosen to be non-increasing w.r.t. both δ and ε , which implies that for sufficiently large N :

$$T_0(N, f_N, g_N) \geq T_0(N, \delta, \varepsilon), \quad (58)$$

further implying, in view of (57):

$$\mathbb{P} \left[c_N \|\widehat{\mathbf{A}}_S(T_0(N, f_N, g_N), N) - \widehat{\mathbf{A}}_S(N)\|_{\max\text{-off}} > \varepsilon \right] \leq \delta. \quad (59)$$

In other words, if the number of samples scales with N as $T_N = T_0(N, f_N, g_N)$, we can write:

$$c_N \|\widehat{\mathbf{A}}_S(T_N, N) - \widehat{\mathbf{A}}_S(N)\|_{\max\text{-off}} \xrightarrow{\mathbb{P}} 0. \quad (60)$$

Plugging this result into (55) and noticing that the second term on the RHS of (55) vanishes in probability in view of Definition 2 (since by assumption the limiting estimator achieves universal local structural consistency), we conclude that:

$$\|c_N \widehat{\mathbf{A}}_S(T_N, N) - \gamma \mathbf{G}_S(N)\|_{\max\text{-off}} \xrightarrow{\mathbb{P}} 0, \quad (61)$$

which implies that, for any $\varepsilon > 0$ and all $k \neq \ell$ we have (see footnote 4):

$$c_N \widehat{\mathbf{a}}_{k\ell}(T_N, N) \in \begin{cases} [(1 - \varepsilon)\gamma, (1 + \varepsilon)\gamma], & (k, \ell) \text{ conn.} \\ [-\varepsilon\gamma, \varepsilon\gamma], & (k, \ell) \text{ disc.} \end{cases} \quad (62)$$

with high probability as $N \rightarrow \infty$. By assumption, there exists a certain $\varepsilon > 0$ such that the algorithm `graphclu`(\cdot) achieves successful classification for all configurations fulfilling (62), provided that $\mathcal{G}_S(N)$ is neither fully connected nor fully disconnected. Thus, the proof is complete because we have just shown that configurations fulfilling (62) occur with high probability as $N \rightarrow \infty$, and by assumption the probability that $\mathcal{G}_S(N)$ is fully connected or fully disconnected vanishes as $N \rightarrow \infty$. ■

APPENDIX B

USEFUL RESULTS ON BOLLOBÁS-RIORDAN MULTIGRAPHS

We start by enunciating two properties of the maximal degree $\mu_{\mathcal{M}}(N)$ that will be critical in our development.

Theorem 8.14 in [49, p. 280]: For $N = 1, 2, \dots$, let $\mu_{\mathcal{M}}(N)$ be the maximal degree sequence defined over the multigraph sequence $\mathcal{M}(N)$. For any $m \in \mathbb{N}$ we have that:

$$\limsup_{N \rightarrow \infty} \mathbb{E} \left[\left(\frac{\mu_{\mathcal{M}}(N)}{\sqrt{N}} \right)^m \right] < \infty. \quad (63)$$

Moreover, there exists a strictly positive random variable μ such that:

$$\frac{\mu_{\mathcal{M}}(N)}{\sqrt{N}} \xrightarrow{\text{a.s.}} \mu, \quad (64)$$

where $\xrightarrow{\text{a.s.}}$ denotes almost-sure convergence. □

We continue by proving a lemma that provides a uniform upper bound on the *preferential attachment probability*.

Lemma 3 (Bounds on the Preferential Attachment Probability): The preferential attachment probability defined in (42) obeys the following bound:

$$\mathbb{P}[\mathbf{v}(n; s) = k | \mathcal{M}(n; s - 1)] < \frac{\mathbf{d}_{\mathcal{M}, k}(n - 1) + 2\eta}{2\eta(n - 1)}. \quad (65)$$

Moreover, for any set $\mathcal{J} \subseteq \{1, 2, \dots, s-1\}$, we have that:

$$\mathbb{P}[\mathbf{v}(n; s) = k | \{\mathbf{v}(n; \tau)\}_{\tau \in \mathcal{J}}, \mathcal{M}(n-1)] < \frac{\mathbf{d}_{\mathcal{M},k}(n-1) + 2\eta}{2\eta(n-1)}. \quad (66)$$

Proof: First we focus on the numerator in (42). Joining (37) and (38), we can write the degree of node k in multigraph $\mathcal{M}(n, s-1)$ as:

$$\mathbf{d}_{\mathcal{M},k}(n; s-1) = \mathbf{d}_{\mathcal{M},k}(n; s-2) + \mathbb{I}(\mathbf{v}(n; s-1) = k) + \delta_{kn}, \quad (67)$$

where δ_{kn} is the Kronecker delta. Developing the recursion in (67) over index s , we get:

$$\begin{aligned} \mathbf{d}_{\mathcal{M},k}(n; s-1) &= \mathbf{d}_{\mathcal{M},k}(n-1) + \sum_{t=1}^{s-1} \mathbb{I}(\mathbf{v}(n; t) = k) \\ &+ (s-1)\delta_{kn} \leq \mathbf{d}_{\mathcal{M},k}(n-1) + 2(s-1), \end{aligned} \quad (68)$$

which implies that the numerator in (42) is upper bounded as:

$$\begin{aligned} \delta_{kn} + \mathbf{d}_{\mathcal{M},k}(n; s-1) &\leq \mathbf{d}_{\mathcal{M},k}(n-1) + 2s-1 \\ &< \mathbf{d}_{\mathcal{M},k}(n-1) + 2\eta, \end{aligned} \quad (69)$$

where the last inequality follows by observing that $s \leq \eta$.

We switch to the analysis of the denominator in (42), which is lower bounded as:

$$1 + 2\eta(n-1) + 2(s-1) > 2\eta(n-1). \quad (70)$$

Using (69) and (70) in (42), we get (65). It remains to prove (66). By applying the law of total probability, we can write:

$$\begin{aligned} &\mathbb{P}[\mathbf{v}(n; s) = k | \{\mathbf{v}(n; \tau)\}_{\tau \in \mathcal{J}}, \mathcal{M}(n-1)] \\ &\stackrel{(a)}{=} \sum_{\mathcal{M}} \mathbb{P}[\mathbf{v}(n; s) = k | \mathcal{M}(n, s-1) = \mathcal{M}] \\ &\times \mathbb{P}[\mathcal{M}(n, s-1) = \mathcal{M} | \{\mathbf{v}(n; \tau)\}_{\tau \in \mathcal{J}}, \mathcal{M}(n-1)] \\ &\stackrel{(b)}{<} \frac{\mathbf{d}_{\mathcal{M},k}(n-1) + 2\eta}{2\eta(n-1)}. \end{aligned} \quad (71)$$

In the summation, \mathcal{M} spans the space of possible multigraphs $\mathcal{M}(n, s-1)$ compatible with the multigraph $\mathcal{M}(n-1)$ and the collection of selected nodes $\{\mathbf{v}(n; \tau)\}_{\tau \in \mathcal{J}}$, and where: (a) comes from the fact that the preferential attachment rule is Markovian, namely, the selection at step s depends only on the previous multigraph $\mathcal{M}(n; s-1)$ (regardless of any details about the previous multigraph evolution); while (b) comes from (65). ■

Lemma 4 (Sum of Maximal Degree Powers): For all $m \geq 2$ we have that:

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \sum_{n=1}^N \mathbb{E} \left[\left(\frac{\mu_{\mathcal{M}}(n) + 2\eta}{n} \right)^m \right] = 0. \quad (72)$$

Proof: First we observe that:

$$\begin{aligned} &\frac{1}{\sqrt{N}} \sum_{n=1}^N \mathbb{E} \left[\left(\frac{\mu_{\mathcal{M}}(n) + 2\eta}{n} \right)^m \right] \\ &= \frac{\sum_{n=1}^N n^{-m/2}}{\sqrt{N}} \frac{\sum_{n=1}^N \mathbb{E} \left[\left(\frac{\mu_{\mathcal{M}}(n) + 2\eta}{n} \right)^m \right]}{\sum_{n=1}^N n^{-m/2}}. \end{aligned} \quad (73)$$

From the Stolz-Cesàro theorem, for any two positive sequences f_N and g_N with $g_N \rightarrow \infty$ as $N \rightarrow \infty$, we have that [57], [58]:

$$\limsup_{N \rightarrow \infty} \frac{\sum_{n=1}^N f_n}{\sum_{n=1}^N g_n} \leq \limsup_{N \rightarrow \infty} \frac{f_N}{g_N}, \quad (74)$$

which applied to the last fraction in (73) yields:

$$\begin{aligned} &\limsup_{N \rightarrow \infty} \frac{\sum_{n=1}^N \mathbb{E} \left[\left(\frac{\mu_{\mathcal{M}}(n) + 2\eta}{n} \right)^m \right]}{\sum_{n=1}^N n^{-m/2}} \\ &\leq \limsup_{N \rightarrow \infty} \frac{\mathbb{E} \left[\left(\frac{\mu_{\mathcal{M}}(N) + 2\eta}{N} \right)^m \right]}{N^{-m/2}} \\ &= \limsup_{N \rightarrow \infty} \mathbb{E} \left[\left(\frac{\mu_{\mathcal{M}}(N) + 2\eta}{\sqrt{N}} \right)^m \right] < \infty, \end{aligned} \quad (75)$$

where the last inequality follows by (63).

Focusing on the first fraction in (73) we have that, for $m > 2$:

$$\sum_{n=1}^{\infty} n^{-m/2} = \zeta(m/2), \quad (76)$$

where $\zeta(\cdot)$ is the Riemann zeta function (which is finite), while for $m = 2$ the summation in (76) is the harmonic number, which diverges logarithmically as $N \rightarrow \infty$. Accordingly, we conclude that for all $m \geq 2$:

$$\lim_{N \rightarrow \infty} \frac{\sum_{n=1}^N n^{-m/2}}{\sqrt{N}} = 0. \quad (77)$$

The claim of the lemma now follows by applying (75) and (77). ■

Lemma 5 (Correlation Between Degrees): For any $1 \leq k < \ell \leq N$, we have that:

$$\mathbb{E}[\mathbf{m}_{kN} \mathbf{m}_{\ell N} | \mathcal{M}(N-1)] < \left(\frac{\mu_{\mathcal{M}}(N-1) + 2\eta}{N-1} \right)^2. \quad (78)$$

Proof: It is convenient to introduce the following binary random variables:

$$\beta_{ks}(N) \triangleq \mathbb{I}(\mathbf{v}(N; s) = k), \quad (79)$$

for $1 \leq k \leq N$ and $1 \leq s \leq \eta$. Using (34), we can write:

$$\begin{aligned} &\mathbb{E}[\mathbf{m}_{kN} \mathbf{m}_{\ell N} | \mathcal{M}(N-1)] \\ &= \mathbb{E} \left[\sum_{s=1}^{\eta} \beta_{ks}(N) \sum_{s=1}^{\eta} \beta_{\ell s}(N) \middle| \mathcal{M}(N-1) \right] \\ &= \sum_{\substack{1 \leq s, t \leq \eta \\ s \neq t}} \mathbb{E}[\beta_{ks}(N) \beta_{\ell t}(N) | \mathcal{M}(N-1)], \end{aligned} \quad (80)$$

where the last step holds true since $k \neq \ell$ by assumption. Let us examine the behavior of the individual term in (80) and focus, without loss of generality, on the case $s > t$. Since

$\beta_{ks}(N)$ and $\beta_{\ell t}(N)$ are indicator variables, we have that:

$$\begin{aligned} & \mathbb{E} [\beta_{ks}(N)\beta_{\ell t}(N)|\mathcal{M}(N-1)] \\ &= \mathbb{P}[\mathbf{v}(N; s) = k, \mathbf{v}(N; t) = \ell | \mathcal{M}(N-1)] \\ &= \mathbb{P}[\mathbf{v}(N; s) = k | \mathbf{v}(N; t) = \ell, \mathcal{M}(N-1)] \\ &\times \mathbb{P}[\mathbf{v}(N; t) = \ell | \mathcal{M}(N-1)] \\ &< \frac{\eta(\eta-1)}{2} \left(\frac{\mu_{\mathcal{M}}(N-1) + 2\eta}{2\eta(N-1)} \right)^2, \end{aligned} \quad (81)$$

where the last inequality follows from (66), and the claim follows by observing that $\eta(\eta-1)/(8\eta^2) < 1$. ■

APPENDIX C

EQUIVALENCE BETWEEN BOLLOBÁS-RIORDAN MULTIGRAPHS AND SIMPLE GRAPHS

Proof of Lemma 1: We start with the analysis of the number of self-loops in $\mathcal{M}(N)$, which is equal to:

$$\mathring{m}(N) \triangleq \sum_{k=1}^N m_{kk}. \quad (82)$$

Given a multigraph $\mathcal{M}(k-1)$, let us consider the η steps, $s = 1, 2, \dots, \eta$, necessary to build the multigraph $\mathcal{M}(k)$. Consider a sequence of nodes v_1, v_2, \dots, v_η selected during the η steps, with the prescription that *exactly* m out of the η nodes are equal to k , i.e., we have m self-loops attached to the new node k . We denote by \mathcal{V}_m the ensemble of configurations v_1, v_2, \dots, v_η that match such prescription. We note in passing that the cardinality of \mathcal{V}_m is equal to $\binom{\eta}{m}$. According to the adopted notation, the probability of having exactly m self-loops attached to k admits the following representation:

$$\begin{aligned} & \mathbb{P}[m_{kk} = m | \mathcal{M}(k-1)] \\ &\stackrel{(a)}{=} \sum_{\mathcal{V}_m} \prod_{s=1}^{\eta} \mathbb{P}[\mathbf{v}(k; s) = v_s | \{\mathbf{v}(k; \tau) = v_\tau\}_{\tau=1}^{s-1}, \mathcal{M}(k-1)] \\ &\leq \sum_{\mathcal{V}_m} \prod_{s: v_s=k} \mathbb{P}[\mathbf{v}(k; s) = k | \{\mathbf{v}(k; \tau) = v_\tau\}_{\tau=1}^{s-1}, \mathcal{M}(k-1)] \\ &\stackrel{(b)}{\leq} \binom{\eta}{m} \frac{1}{(k-1)^m}, \end{aligned} \quad (83)$$

where (a) follows by the chain rule and (b) follows from (66), once noticing that $d_{\mathcal{M},k}(k-1) = 0$. Now, by exploiting the definition of $\mathcal{M}(1)$, the multigraph made by a single node with η self-loops, we can write:

$$\mathring{m}(N) = \eta + \sum_{k=2}^N m_{kk}, \quad (84)$$

which, in view of (83) yields:

$$\begin{aligned} & \mathbb{E}[\mathring{m}(N) | \mathcal{M}(k-1)] = \\ & \eta + \sum_{k=2}^N \sum_{m=1}^{\eta} m \mathbb{P}[m_{kk} = m | \mathcal{M}(k-1)] \\ & \leq \eta + \sum_{k=2}^N \sum_{m=1}^{\eta} m \binom{\eta}{m} \left(\frac{2\eta}{k-1} \right)^m \\ & \leq \eta + C(\eta) \sum_{k=2}^N \frac{1}{k-1}, \end{aligned} \quad (85)$$

where the finite constant $C(\eta)$ is implicitly defined in the last step of (85). We immediately see from (85) that:

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}[\mathring{m}(N)]}{N} = 0, \quad (86)$$

since the last summation in (85) is the harmonic number, which grows logarithmically with N .

We continue by examining the expected number of redundant edges in $\mathcal{M}(N)$:

$$\mathbb{E}[\mathring{m}(N)] = \sum_{\ell=1}^N \sum_{k=1}^{\ell-1} \mathbb{E}[\mathring{m}_{k\ell}], \quad (87)$$

where the number of redundant edges between two distinct nodes k and ℓ in the multigraph $\mathcal{M}(\ell)$ can be conveniently represented as:

$$\mathring{m}_{k\ell} \triangleq \max\{m_{k\ell} - 1, 0\}, \quad (88)$$

We want to show that:

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}[\mathring{m}(N)]}{N} = 0. \quad (89)$$

In order to prove (89), we can call upon the Stolz-Cesàro theorem, and apply (74) with the choices $f_\ell = \sum_{k=1}^{\ell-1} \mathbb{E}[\mathring{m}_{k\ell}]$ and $g_\ell = 1$ (which corresponds to apply the Cesàro-mean theorem), implying that it would suffice to show that:

$$\lim_{\ell \rightarrow \infty} \sum_{k=1}^{\ell-1} \mathbb{E}[\mathring{m}_{k\ell}] = 0. \quad (90)$$

Reasoning as done to obtain (83), we have the following representation:

$$\begin{aligned} & \mathbb{P}[m_{k\ell} = m | \mathcal{M}(\ell-1)] \\ & \leq \sum_{\mathcal{V}_m} \prod_{s: v_s=k} \mathbb{P}[\mathbf{v}(\ell; s) = k | \{\mathbf{v}(\ell; \tau) = v_\tau\}_{\tau=1}^{s-1}, \mathcal{M}(\ell-1)] \\ & \leq \binom{\eta}{m} \left(\frac{d_{\mathcal{M},k}(\ell-1) + 2\eta}{2\eta(\ell-1)} \right)^m, \end{aligned} \quad (91)$$

where in the last step we applied (66). Exploiting (88), from (91) we can compute the conditional expected value of $\mathring{m}_{k\ell}$, obtaining:

$$\begin{aligned} & \mathbb{E}[\mathring{m}_{k\ell} | \mathcal{M}(\ell-1)] = \sum_{m=1}^{\eta-1} m \mathbb{P}[\mathring{m}_{k\ell} = m | \mathcal{M}(\ell-1)] \\ & = \sum_{m=2}^{\eta} (m-1) \mathbb{P}[m_{k\ell} = m | \mathcal{M}(\ell-1)] \\ & \leq \sum_{m=2}^{\eta} (m-1) \binom{\eta}{m} \left(\frac{d_{\mathcal{M},k}(\ell-1) + 2\eta}{2\eta(\ell-1)} \right)^m \\ & \leq \left(\frac{d_{\mathcal{M},k}(\ell-1) + 2\eta}{2\eta(\ell-1)} \right) \\ & \times \sum_{m=2}^{\eta} (m-1) \binom{\eta}{m} \left(\frac{\mu_{\mathcal{M}}(\ell-1) + 2\eta}{2\eta(\ell-1)} \right)^{m-1}, \end{aligned} \quad (92)$$

where, in the last step the degree of node k has been upper bounded $m-1$ times by the maximal degree. Summing over

index k , from (93) we get:

$$\begin{aligned} \sum_{k=1}^{\ell-1} \mathbb{E}[\ddot{m}_{k\ell} | \mathcal{M}(\ell-1)] &\leq \frac{\sum_{k=1}^{\ell-1} (\mathbf{d}_{\mathcal{M},k}(\ell-1) + 2\eta)}{2\eta(\ell-1)} \\ &\times \sum_{m=2}^{\eta} (m-1) \binom{\eta}{m} \left(\frac{\mu_{\mathcal{M}}(\ell-1) + 2\eta}{2\eta(\ell-1)} \right)^{m-1}. \end{aligned} \quad (94)$$

Using (41) and recalling that $\mathbf{d}_{\mathcal{M},k}(\ell-1) = \mathbf{d}_{\mathcal{M},k}(\ell-1; \eta)$, we have that:

$$\sum_{k=1}^{\ell-1} \mathbf{d}_{\mathcal{M},k}(\ell-1) = 2\eta(\ell-1), \quad (95)$$

which, taking expectation w.r.t. to $\mathcal{M}(\ell-1)$ in (94), yields:

$$\begin{aligned} \sum_{k=1}^{\ell-1} \mathbb{E}[\ddot{m}_{k\ell}] \\ \leq 2 \sum_{m=2}^{\eta} (m-1) \binom{\eta}{m} \mathbb{E} \left[\left(\frac{\mu_{\mathcal{M}}(\ell-1) + 2\eta}{2\eta(\ell-1)} \right)^{m-1} \right], \end{aligned} \quad (96)$$

and the claim in (90) follows by (63), which, further applying (86), completes the proof of the lemma. ■

Next we prove the asymptotic equivalence between the *maximal* degrees of the multigraph and the associated simple graph, claimed in Lemma 2.

Proof of Lemma 2: Considering the definition of the multigraph degree in (36), and separating the contribution of the redundant edges in (88) from the contribution of the simple graph term in (45), we can write:

$$\begin{aligned} \mathbf{d}_{\mathcal{M},k}(N) &= 2\mathbf{m}_{kk} + \sum_{\substack{\ell=1 \\ \ell \neq k}}^N \mathbf{m}_{k\ell} \\ &\stackrel{(a)}{=} 2\mathbf{m}_{kk} + \sum_{\substack{\ell=1 \\ \ell \neq k}}^N \mathbf{g}_{k\ell} + \sum_{\ell=1}^{k-1} \ddot{m}_{k\ell} + \sum_{\ell=k+1}^N \ddot{m}_{k\ell} \\ &\stackrel{(b)}{<} 3\eta + \mathbf{d}_{\mathcal{G},k}(N) + \sum_{\ell=k+1}^N \ddot{m}_{k\ell}, \end{aligned} \quad (97)$$

where in (a) we adopt the convention that the second summation is equal to zero when $k=1$, and that the third summation is equal to zero when $k=N$; and in (b) the inequality follows because the number of self-loops attached to node k at cycle k , as well as the number of edges connecting node k to a node $\ell < k$, are at most equal to the number of steps η , namely, $\mathbf{m}_{kk} \leq \eta$ and $\sum_{\ell=1}^{k-1} \ddot{m}_{k\ell} < \eta$. Taking the maximum over $k \in [1, N]$ in (97) we can write:

$$\begin{aligned} 0 \leq \mu_{\mathcal{M}}(N) - \mu_{\mathcal{G}}(N) &< 3\eta + \max_{k \in [1, N]} \sum_{\ell=k+1}^N \ddot{m}_{k\ell} \\ &= 3\eta + \max_{k \in [1, N-1]} \sum_{\ell=k+1}^N \ddot{m}_{k\ell}, \end{aligned} \quad (98)$$

where the equality follows because of the summation is zero when $k=N$. For any $k \geq 1$, let us introduce the sequence:

$$\mathbf{u}_k(\ell) \triangleq \begin{cases} \ddot{m}_{k\ell}, & k < \ell, \\ 0, & \text{otherwise.} \end{cases} \quad (99)$$

It is readily verified that the family of sequences $\{\mathbf{u}_k(\ell)\}_{\ell \geq 1}$ indexed by the parameter k , matches the hypotheses of Lemma 13, with the choices $\Theta = \mathbb{N} \setminus \{0\}$, $b = \eta$, and with the filtration $\{\mathcal{F}(\ell)\}_{\ell \geq 1}$ generated by the random sequence $\{\mathcal{M}(\ell)\}_{\ell \geq 1}$. In particular, for any $\varepsilon > 0$, and for any $N \geq 1$, by choosing $\mathcal{J} = [1, N]$ and $u = \varepsilon\sqrt{N}$, we can apply Lemma 13 and write:

$$\begin{aligned} \mathbb{P} \left[\max_{k \in [1, N-1]} \sum_{\ell=k+1}^N \ddot{m}_{k\ell} > \varepsilon\sqrt{N} \right] \\ \leq N e^{-\frac{3\varepsilon}{16\eta}\sqrt{N}} + \mathbb{P} \left[\max_{k \in [1, N-1]} \mathbf{C}_k(N) > \frac{\varepsilon}{2}\sqrt{N} \right], \end{aligned} \quad (100)$$

where

$$\mathbf{C}_k(N) = \sum_{\ell=k+1}^N \mathbb{E}[\ddot{m}_{k\ell} | \mathcal{M}(\ell-1)]. \quad (101)$$

On the other hand, for $1 \leq k \leq N-1$ we have that:

$$\begin{aligned} \mathbf{C}_k(N) &\leq \sum_{\ell=2}^N \mathbb{E}[\ddot{m}_{k\ell} | \mathcal{M}(\ell-1)] \\ &\leq \sum_{\ell=2}^N \sum_{m=2}^{\eta} (m-1) \binom{\eta}{m} \left(\frac{\mathbf{d}_{\mathcal{M},k}(\ell-1) + 2\eta}{2\eta(\ell-1)} \right)^m \\ &\leq \sum_{\ell=2}^N \sum_{m=2}^{\eta} (m-1) \binom{\eta}{m} \left(\frac{\mu_{\mathcal{M}}(\ell-1) + 2\eta}{2\eta(\ell-1)} \right)^m, \end{aligned} \quad (102)$$

where in the second inequality we applied (92), whereas the third inequality follows from the definition of maximal degree. Applying Markov's inequality and exploiting (102) we obtain:

$$\begin{aligned} \mathbb{P} \left[\max_{k \in [1, N-1]} \mathbf{C}_k(N) \geq \frac{\varepsilon}{2}\sqrt{N} \right] \\ \leq \frac{2}{\varepsilon} \sum_{m=2}^{\eta} (m-1) \binom{\eta}{m} \frac{1}{\sqrt{N}} \sum_{\ell=1}^N \mathbb{E} \left[\left(\frac{\mu_{\mathcal{M}}(\ell) + 2\eta}{\ell} \right)^m \right]. \end{aligned} \quad (103)$$

Substituting (103) into (100) and applying Lemma 4, from (98) we obtain the convergence in (47). Then, the convergence in (48) comes directly from (43), and the proof is complete. ■

Lemma 6 (Subgraphs $\mathcal{G}_{\mathcal{P}}(N)$ Are Nontrivial): Let $\mathcal{G}(N)$ be the simple graph associated to a Bollobás-Riordan multigraph $\mathcal{M}(N)$, and let $\mathcal{G}_{\mathcal{S}}(N)$ be the subgraph over a subset sequence \mathcal{S}_N fulfilling (8). Then: *i*) for sufficiently large N , subgraph $\mathcal{G}_{\mathcal{S}}(N)$ is not fully connected; and *ii*) the probability that $\mathcal{G}_{\mathcal{S}}(N)$ is fully disconnected vanishes as $N \rightarrow \infty$.

Proof: The sum of all degrees in multigraph $\mathcal{M}(N)$ grows linearly with N . In order to have a fully connected subgraph $\mathcal{G}_{\mathcal{S}}(N)$, the sum of all degrees of this subgraph should be equal to $|\mathcal{S}_N|^2$, which scales as N^2 in view of (8). Therefore, subgraph $\mathcal{G}_{\mathcal{S}}(N)$ cannot be fully connected as N grows.

We move on to examine the probability that $\mathcal{G}_{\mathcal{S}}(N)$ is fully disconnected. Let

$$\mathcal{S}_N = \{n_1, n_2, \dots, n_{|\mathcal{S}_N|}\}, \quad \mathcal{V}_{\ell} \triangleq \{n_1, n_2, \dots, n_{\ell}\}, \quad (104)$$

with:

$$n_1 < n_2 < \dots < n_{|\mathcal{S}_N|}. \quad (105)$$

We can write:

$$\begin{aligned} & \mathbb{P}[\mathcal{G}_S(N) \text{ is fully disconnected}] \\ & \leq \mathbb{P}[\mathbf{v}(n_2; 1) \notin \mathcal{V}_1, \mathbf{v}(n_3; 1) \notin \mathcal{V}_2, \dots, \mathbf{v}(n_{|\mathcal{S}_N|}; 1) \notin \mathcal{V}_{|\mathcal{S}_N|-1}] \\ & = \prod_{\ell=2}^{|\mathcal{S}_N|} \mathbb{P}[\mathbf{v}(n_\ell; 1) \notin \mathcal{V}_{\ell-1} | \{\mathbf{v}(n_k; 1) \notin \mathcal{V}_{k-1}\}_{k=2}^{\ell-1}], \quad (106) \end{aligned}$$

where the inequality follows by considering only the connections at step $s = 1$ of the preferential attachment construction, whereas the equality follows by the chain rule. Now, from (42), for any $k < n$ we obtain the following lower bound:

$$\mathbb{P}[\mathbf{v}(n; 1) = k | \mathcal{M}(n-1)] \geq \frac{\eta}{1 + 2\eta(n-1)} \geq \frac{1}{2n}. \quad (107)$$

Accordingly, the individual term in (106) can be upper bounded as follows:

$$\begin{aligned} & \mathbb{P}[\mathbf{v}(n_\ell; 1) \notin \mathcal{V}_{\ell-1} | \{\mathbf{v}(n_k; 1) \notin \mathcal{V}_{k-1}\}_{k=2}^{\ell-1}] \\ & = 1 - \sum_{k=1}^{\ell-1} \mathbb{P}[\mathbf{v}(n_\ell; 1) = n_k | \{\mathbf{v}(n_k; 1) \notin \mathcal{V}_{k-1}\}_{k=2}^{\ell-1}] \\ & \leq 1 - \frac{\ell-1}{2n_\ell} \leq 1 - \frac{\ell-1}{2N}. \quad (108) \end{aligned}$$

Using (108) in (106), we finally get⁹:

$$\begin{aligned} & \mathbb{P}[\mathcal{G}_{\mathcal{S}_N}(N) \text{ is fully disconnected}] \\ & \leq \prod_{\ell=2}^{|\mathcal{S}_N|} \left(1 - \frac{\ell-1}{2N}\right) \leq \prod_{\ell=|\mathcal{S}_N|/2+1}^{|\mathcal{S}_N|} \left(1 - \frac{\ell-1}{2N}\right) \\ & \leq \left(1 - \frac{|\mathcal{S}_N|}{4N}\right)^{|\mathcal{S}_N|/2} \xrightarrow{N \rightarrow \infty} 0, \quad (110) \end{aligned}$$

where the second inequality holds because all terms in the product are smaller than 1, while convergence holds in view of (8). ■

APPENDIX D USEFUL RESULTS ON MATRIX $\mathbf{A}(N)$

Theorem 4 (Correlation Between Matrix Entries): Let

$$\mathfrak{M}(N) \triangleq \max_{\substack{k, \ell \in [1, N] \\ k < \ell}} \sum_{\substack{j=1 \\ j \neq k, \ell}}^N \mathbf{a}_{jk}(N) \mathbf{a}_{j\ell}(N). \quad (111)$$

Then we have that:

$$\sqrt{N} \mathfrak{M}(N) \xrightarrow{p} 0. \quad (112)$$

⁹For each multigraph $\mathcal{M}(n_\ell-1)$ such that $\{\mathbf{v}(n_k; 1) \notin \mathcal{V}_{k-1}\}_{k=2}^{\ell-1}$, we can write:

$$\begin{aligned} & \mathbb{P}[\mathbf{v}(n_\ell; 1) = n_k | \{\mathbf{v}(n_k; 1) \notin \mathcal{V}_{k-1}\}_{k=2}^{\ell-1}, \mathcal{M}(n_\ell-1)] \\ & = \mathbb{P}[\mathbf{v}(n_\ell; 1) = n_k | \mathcal{M}(n_\ell-1)]. \quad (109) \end{aligned}$$

Since the bound in (107) does not depend on $\mathcal{M}(n-1)$, by applying the law of total probability, we can use (107) to bound also the probability $\mathbb{P}[\mathbf{v}(n_\ell; 1) = n_k | \{\mathbf{v}(n_k; 1) \notin \mathcal{V}_{k-1}\}_{k=2}^{\ell-1}]$.

Proof: Using (32) in (111) we get:

$$\sqrt{N} \mathfrak{M}(N) = \left(\frac{\rho\lambda\sqrt{N}}{1 + \mu_{\mathcal{G}}(N)} \right)^2 \frac{1}{\sqrt{N}} \underbrace{\max_{\substack{k, \ell \in [1, N] \\ k < \ell}} \sum_{\substack{j=1 \\ j \neq k, \ell}}^N \mathbf{g}_{jk} \mathbf{g}_{j\ell}}_{t_N}. \quad (113)$$

Since the term $\frac{\sqrt{N}}{1 + \mu_{\mathcal{G}}(N)}$ converges almost surely to $1/\mu$, it is sufficient to show that the term t_N in (113) vanishes in probability as $N \rightarrow \infty$. To this aim, it is expedient to work in terms of the original multigraph $\mathcal{M}(N)$ that originates the simple graph $\mathcal{G}(N)$. We have that:

$$\begin{aligned} t_N & \leq \frac{1}{\sqrt{N}} \max_{\substack{k, \ell \in [1, N] \\ k < \ell}} \sum_{\substack{j=1 \\ j \neq k, \ell}}^N \mathbf{m}_{jk} \mathbf{m}_{j\ell} \\ & = \frac{1}{\sqrt{N}} \max_{\substack{k, \ell \in [1, N] \\ k < \ell}} \underbrace{\sum_{\substack{j=1 \\ j \neq k, \ell}}^{\ell} \mathbf{m}_{jk} \mathbf{m}_{j\ell}}_{t'_N} \\ & \quad + \underbrace{\frac{1}{\sqrt{N}} \max_{\substack{k, \ell \in [1, N] \\ k < \ell}} \sum_{j=\ell+1}^N \mathbf{m}_{jk} \mathbf{m}_{j\ell}}_{t''_N}. \quad (114) \end{aligned}$$

By construction, $\mathbf{m}_{k\ell} \in \{1, \dots, \eta\}$, and using (36), for any $k < \ell$ we have:

$$\sum_{\substack{j=1 \\ j \neq k, \ell}}^{\ell} \mathbf{m}_{jk} \mathbf{m}_{j\ell} \leq \eta \sum_{\substack{j=1 \\ j \neq k, \ell}}^{\ell} \mathbf{m}_{j\ell} \leq \eta d_{\mathcal{M}, \ell}(\ell) \leq 2\eta^2, \quad (115)$$

where the last inequality holds because, in the multigraph $\mathcal{M}(\ell)$, node ℓ has only η edges, and so its degree $d_{\mathcal{M}, \ell}(\ell)$ is upper bounded by 2η . Applying (115) to the random sequence t'_N in (114), we conclude that t'_N vanishes almost surely as $N \rightarrow \infty$. It remains to show that t''_N in (114) vanishes in probability. To this end, we call upon Lemma 13, by introducing the following family of sequences, for any $k, \ell \in \mathbb{N}$ with $1 \leq k < \ell$:

$$\mathbf{u}_{k\ell}(j) \triangleq \begin{cases} \mathbf{m}_{jk} \mathbf{m}_{j\ell}, & j > \ell, \\ 0, & \text{otherwise.} \end{cases} \quad (116)$$

Following the notation adopted in Lemma 13, we have a family of sequences $\{\mathbf{u}_{k\ell}(j)\}_{j \geq 1}$ parameterized by the set:

$$\Theta = \{(k, \ell) \in \mathbb{N}^2 : 1 \leq k < \ell\}. \quad (117)$$

Moreover, we introduce the aggregate variable:

$$\mathbf{U}_{k\ell}(N) \triangleq \sum_{j=1}^N \mathbf{u}_{k\ell}(j). \quad (118)$$

Convergence to zero of the random variable t''_N in (114) is equivalent to the following statement:

$$\forall \varepsilon > 0, \quad \lim_{N \rightarrow \infty} \mathbb{P} \left[\max_{\substack{k, \ell \in [1, N] \\ k < \ell}} \mathbf{U}_{k\ell}(N) > \varepsilon \sqrt{N} \right] = 0. \quad (119)$$

Now, for any $(k, \ell) \in \Theta$:

$$\mathbf{u}_{k\ell}(1) = 0, \quad \text{and} \quad \forall j > 1, \quad 0 \leq \mathbf{u}_{k\ell}(j) \leq \eta^2, \quad (120)$$

where the upper bound follows because, in view of (34) and (35), both \mathbf{m}_{jk} and $\mathbf{m}_{j\ell}$ cannot exceed the number of steps η . From (120) we see that the family of sequences in (116) meets the hypotheses of Lemma 13 with the filtration $\{\mathcal{F}(n)\}_{n \geq 1}$ generated by the random sequence $\{\mathcal{M}(n)\}_{n \geq 1}$. We conclude that the probability in (119) is upper bounded by:

$$\frac{N(N-1)}{2} e^{-\frac{3}{16\eta^2} \varepsilon \sqrt{N}} + \mathbb{P} \left[\max_{\substack{k, \ell \in [1, N] \\ k < \ell}} \mathbf{C}_{k\ell}(N) > \frac{\varepsilon}{2} \sqrt{N} \right], \quad (121)$$

where:

$$\mathbf{C}_{k\ell}(N) \triangleq \sum_{j=1}^N \mathbb{E}[\mathbf{u}_{k\ell}(j) | \mathcal{M}(j-1)]. \quad (122)$$

On the other hand, from Lemma 5 we have that:

$$\max_{\substack{k, \ell \in [1, N] \\ k < \ell}} \mathbf{C}_{k\ell}(N) < \sum_{j=1}^N \left(\frac{\mu_{\mathcal{M}(j-1)} + 2\eta}{j-1} \right)^2. \quad (123)$$

Therefore, applying Markov's inequality and (123) to the second term in (121), we conclude that this term is upper bounded by:

$$\frac{2}{\varepsilon \sqrt{N}} \sum_{j=1}^N \mathbb{E} \left[\left(\frac{\mu_{\mathcal{M}(j-1)} + 2\eta}{j-1} \right)^2 \right], \quad (124)$$

which vanishes as $N \rightarrow \infty$ in view of Lemma 4, concluding the proof of the theorem. \blacksquare

APPENDIX E DETERMINISTIC PROPERTIES OF THE LIMITING GRANGER ESTIMATOR

In this section we obtain an upper bound on the error of the limiting Granger estimator. To this aim, we start by proving two auxiliary lemmas that hold for any $N \times N$ scaled left-stochastic matrix $A = [a_{k\ell}]$, namely, for any matrix whose entries satisfy the conditions:

$$a_{k\ell} \geq 0, \quad \sum_{\ell=1}^N a_{k\ell} = \rho. \quad (125)$$

In the following analysis we denote by $a_{k\ell}^{(i)}$ the (k, ℓ) -entry of the matrix power A^i , and we use the following quantity:

$$\mathbf{M} \triangleq \max_{\substack{k, \ell \in [1, N] \\ k \neq \ell}} \sum_{\substack{j=1 \\ j \neq k, \ell}}^N a_{kj} a_{j\ell}. \quad (126)$$

Lemma 7 (Bounds on Matrix Powers): Let A be an $N \times N$ scaled left-stochastic matrix as in (125). For $i = 1, 2, \dots$, we have that:

- The main diagonal entries of A^{2i} satisfy the inequalities:

$$a_{kk}^{(2i)} \leq \alpha_i, \quad (127)$$

where the sequence α_i is recursively defined as:

$$\alpha_1 = \rho^2, \quad \alpha_{i+1} = \rho^2 \alpha_i + \rho^{2(i+1)}. \quad (128)$$

- The off-diagonal entries of A^{2i} satisfy the inequalities:

$$a_{k\ell}^{(2i)} \leq \beta_i a_{k\ell} + \gamma_i \quad (129)$$

where β_i and γ_i are two sequences recursively defined as:

$$\beta_1 = 2\rho, \quad \beta_{i+1} = 2\rho \alpha_i + \rho^2 \beta_i, \quad (130)$$

$$\gamma_1 = \mathbf{M}, \quad \gamma_{i+1} = \mathbf{M} \alpha_i + 3\rho \mathbf{M} \beta_i + \rho^2 \gamma_i. \quad (131)$$

Proof: Preliminary, it is useful to observe that¹⁰:

$$\sum_{\ell=1}^N a_{k\ell}^{(i)} = \rho^i. \quad (133)$$

We start by proving (127) by induction. For $i = 1$, the claim follows directly from (133). We shall therefore prove that (127) holds for $i + 1$, assuming that it holds for i . To this aim, let us write the diagonal terms of matrix $A^{2(i+1)}$ as:

$$a_{kk}^{(2i+2)} = \sum_{\ell=1}^N a_{k\ell}^{(2i)} a_{\ell k}^{(2)} = a_{kk}^{(2i)} a_{kk}^{(2)} + \sum_{\substack{\ell=1 \\ \ell \neq k}}^N a_{k\ell}^{(2i)} a_{\ell k}^{(2)}. \quad (134)$$

We observe that (133) implies in particular the following inequalities:

$$a_{kk}^{(2)} \leq \rho^2, \quad a_{\ell k}^{(2)} \leq \rho^2, \quad \sum_{\substack{\ell=1 \\ \ell \neq k}}^N a_{k\ell}^{(2i)} \leq \rho^{2i}, \quad (135)$$

which, applied in (134), yield:

$$a_{kk}^{(2i+2)} \leq \rho^2 a_{kk}^{(2i)} + \rho^{2(i+1)}. \quad (136)$$

Since $a_{kk}^{(2i)} \leq \alpha_i$ by the induction hypothesis, from (136) we get:

$$a_{kk}^{(2i+2)} \leq \rho^2 \alpha_i + \rho^{2(i+1)} = \alpha_{i+1}, \quad (137)$$

which corresponds to (127) for the case $i + 1$, and the claim for the diagonal entries is proved.

We continue by proving (129) by induction. For any $k, \ell = 1, 2, \dots, N$, with $k \neq \ell$, we have:

$$\begin{aligned} a_{k\ell}^{(2)} &= \sum_{h=1}^N a_{kh} a_{h\ell} = (a_{kk} + a_{\ell\ell}) a_{k\ell} + \sum_{\substack{h=1 \\ h \neq k, \ell}}^N a_{kh} a_{h\ell} \\ &\leq 2\rho a_{k\ell} + \mathbf{M} \triangleq \beta_1 a_{k\ell} + \gamma_1, \end{aligned} \quad (138)$$

where: *i*) in the inequality we exploited the fact that the diagonal entries of A are upper bounded by ρ in view of (125), and we used the definition of \mathbf{M} in (126); and *ii*) in the last equality we applied the definitions of β_1 and γ_1 appearing in (130) and (131), respectively. We conclude from (138) that the claim in (129) holds for $i = 1$. Let us now show that, if the

¹⁰We can prove this property by induction as follows. For $i = 1$ the property is exactly (125), while the induction step comes from:

$$\sum_{\ell=1}^N a_{k\ell}^{(i+1)} = \sum_{\ell=1}^N \sum_{h=1}^N a_{kh}^{(i)} a_{h\ell} = \sum_{h=1}^N a_{kh}^{(i)} \sum_{\ell=1}^N a_{h\ell} = \rho^i \rho. \quad (132)$$

claim holds for a generic i , then it holds for $i + 1$. To this aim, we observe that:

$$\begin{aligned}
a_{k\ell}^{(2i+2)} &= \sum_{h=1}^N a_{kh}^{(2)} a_{h\ell}^{(2i)} = a_{k\ell}^{(2)} a_{\ell\ell}^{(2i)} + \sum_{\substack{h=1 \\ h \neq \ell}}^N a_{kh}^{(2)} a_{h\ell}^{(2i)} \\
&\leq a_{k\ell}^{(2)} a_{\ell\ell}^{(2i)} + \sum_{\substack{h=1 \\ h \neq \ell}}^N a_{kh}^{(2)} (\beta_i a_{h\ell} + \gamma_i) \\
&\leq \alpha_i a_{k\ell}^{(2)} + \beta_i \left(a_{kk}^{(2)} a_{k\ell} + \sum_{\substack{h=1 \\ h \neq k, \ell}}^N a_{kh}^{(2)} a_{h\ell} \right) + \gamma_i \sum_{\substack{h=1 \\ h \neq \ell}}^N a_{kh}^{(2)},
\end{aligned} \tag{139}$$

where the first inequality follows by applying the induction hypothesis to term $a_{h\ell}^{(2i)}$, while in the second inequality we used (127). Let us now bound the individual terms that multiply the quantities α_i , β_i , and γ_i in (139).

- From (138) we have:

$$a_{k\ell}^{(2)} \leq 2\rho a_{k\ell} + M. \tag{140}$$

- From (133) we have:

$$a_{kk}^{(2)} \leq \rho^2, \quad \sum_{\substack{h=1 \\ h \neq \ell}}^N a_{kh}^{(2)} \leq \rho^2. \tag{141}$$

- We can write:

$$\begin{aligned}
\sum_{\substack{h=1 \\ h \neq k, \ell}}^N a_{kh}^{(2)} a_{h\ell} &\leq \sum_{\substack{h=1 \\ h \neq k, \ell}}^N (2\rho a_{kh} + M) a_{h\ell} \\
&\leq 2\rho M + \rho M = 3\rho M,
\end{aligned} \tag{142}$$

where in the first inequality we applied (138), while in the second inequality we applied (125) and (126).

Using (140), (141), and (142) in (139), we get:

$$\begin{aligned}
a_{k\ell}^{(2i+2)} &\leq \alpha_i (2\rho a_{k\ell} + M) + \beta_i (\rho^2 a_{k\ell} + 3\rho M) + \gamma_i \rho^2 \\
&= \beta_{i+1} a_{k\ell} + \gamma_{i+1},
\end{aligned} \tag{143}$$

where the equality comes from (130) and (131), and the proof is complete. ■

Lemma 8 (Bounds on a Useful Matrix Power Series): Let A be an $N \times N$ scaled left-stochastic matrix as in (125). Let

$$C \triangleq [A^2]_{\mathcal{P}'}, \quad H \triangleq (I_{\mathcal{P}'} - C)^{-1} = \sum_{i=0}^{\infty} C^i, \tag{144}$$

where we recall that $\mathcal{P}' \triangleq \{1, 2, \dots, N\} \setminus \mathcal{P}$. Let further

$$\bar{\alpha} \triangleq 1 + \frac{\rho^2}{(1 - \rho^2)^2}, \quad \bar{\beta} \triangleq 2\rho \frac{\bar{\alpha}}{1 - \rho^2}, \quad \bar{\gamma} \triangleq \frac{\bar{\alpha} + 3\rho\bar{\beta}}{1 - \rho^2}. \tag{145}$$

Then, for $i = 1, 2, \dots$, we have that:

- The main diagonal entries of matrix H satisfy the inequalities:

$$0 < h_{kk} \leq \bar{\alpha}. \tag{146}$$

- The off-diagonal entries of matrix H satisfy the inequalities:

$$0 \leq h_{k\ell} \leq \bar{\beta} a_{k\ell} + M\bar{\gamma}. \tag{147}$$

Proof: The fact that $h_{k\ell} \geq 0$ for any k and ℓ is an immediate consequence of the definition of matrix H in (144), since the entries of matrix powers C^i are nonnegative for any i . So in the next we will focus on the upper bounds in (146) and (147).

As it can be trivially verified by induction, we first note that for any $k, \ell \in \mathcal{P}'$ and any $i = 1, 2, \dots$:

$$c_{k\ell}^{(i)} \leq a_{k\ell}^{(2i)}, \tag{148}$$

implying that the upper bounds provided in Lemma 7 are also valid for the matrix powers C^i . Therefore, by the definition of H in (144), we have:

$$h_{kk} \leq 1 + \sum_{i=1}^{\infty} \alpha_i, \quad h_{k\ell} \leq \sum_{i=1}^{\infty} \beta_i a_{k\ell} + \sum_{i=1}^{\infty} \gamma_i, \tag{149}$$

where the sequences α_i , β_i and γ_i are defined in (128), (130) and (131), respectively. According to (149), to establish the upper bounds in (146) and (147) it suffices to show that:

$$1 + \sum_{i=1}^{\infty} \alpha_i = \bar{\alpha}, \quad \sum_{i=1}^{\infty} \beta_i = \bar{\beta}, \quad \sum_{i=1}^{\infty} \gamma_i = M\bar{\gamma}. \tag{150}$$

To this aim, we note that the sequence α_i in (128) matches (212) in Lemma 12 with the choices:

$$f_1 = a = d = \rho^2, \quad b = c = 0. \tag{151}$$

Therefore, we can apply (213) to obtain:

$$\alpha_i = \rho^{2(i-1)} (\rho^2 + \rho^2(i-1)) = i \rho^{2i}, \tag{152}$$

which, recalling the series $\sum_{i=1}^{\infty} i a^i = \frac{a}{(1-a)^2}$ (for $|a| < 1$), allows us to write:

$$1 + \sum_{i=1}^{\infty} \alpha_i = 1 + \frac{\rho^2}{(1 - \rho^2)^2} = \bar{\alpha}. \tag{153}$$

Thus, we proved (146).

Let us move on to prove (147). By substituting (152) in (130), we get:

$$\beta_{i+1} = \rho^2 \beta_i + 2\rho^{2i+1} i, \tag{154}$$

and therefore we see that the sequence β_i matches (212) in Lemma 12 with the choices:

$$f_1 = c = 2\rho, \quad a = \rho^2, \quad b = d = 0. \tag{155}$$

In view of (213), we conclude that:

$$\beta_i = \rho^{2(i-1)} \left(2\rho + 2\rho \frac{i(i-1)}{2} \right) = \rho^{2i-1} (i^2 - i + 2), \tag{156}$$

which implies that the series $\sum_{i=1}^{\infty} \beta_i$ converges. Since we showed that also the series $\sum_{i=1}^{\infty} \alpha_i$ is convergent, by summing over index i in (130) we can write:

$$\sum_{i=1}^{\infty} \beta_{i+1} = 2\rho \sum_{i=1}^{\infty} \alpha_i + \rho^2 \sum_{i=1}^{\infty} \beta_i, \tag{157}$$

or

$$\sum_{i=1}^{\infty} \beta_i - \beta_1 = 2\rho \sum_{i=1}^{\infty} \alpha_i + \rho^2 \sum_{i=1}^{\infty} \beta_i, \tag{158}$$

which, using $\beta_1 = 2\rho$ and (153), yields:

$$\sum_{i=1}^{\infty} \beta_i = 2\rho \frac{1 + \sum_{i=1}^{\infty} \alpha_i}{1 - \rho^2} = 2\rho \frac{\bar{\alpha}}{1 - \rho^2} = \bar{\beta}. \quad (159)$$

It remains to examine the behavior of the summation in (149) involving the sequence γ_i in (131). Substituting (152) and (156) in (131) we get:

$$\begin{aligned} \gamma_{i+1} &= \rho^2 \gamma_i + M\rho^{2i} i + 3M\rho^{2i} (i^2 - i + 2) \\ &= \rho^2 \gamma_i + M\rho^{2i} (3i^2 - 2i + 6), \end{aligned} \quad (160)$$

which shows that the sequence γ_i matches (212) in Lemma 12 with the choices:

$$f_1 = M, \quad a = \rho^2, \quad b = 3M, \quad c = -2M, \quad d = 6M. \quad (161)$$

We conclude that the series $\sum_{i=1}^{\infty} \gamma_i$ is convergent. Thus, by summing over i in (131), we can write:

$$\sum_{i=1}^{\infty} \gamma_i - \gamma_1 = M \sum_{i=1}^{\infty} \alpha_i + 3M \sum_{i=1}^{\infty} \beta_i + \rho^2 \sum_{i=1}^{\infty} \gamma_i, \quad (162)$$

which, using $\gamma_1 = M$ along with (153) and (159), yields:

$$\sum_{i=1}^{\infty} \gamma_i = M \frac{\bar{\alpha} + 3\rho\bar{\beta}}{1 - \rho^2} = M \bar{\gamma}, \quad (163)$$

and the proof is complete. \blacksquare

We are now ready to apply Lemmas 7 and 8 to obtain a bound on the error of the limiting Granger estimator. By definition, the limiting Granger estimator $\hat{A}_{\mathcal{P}}(N)$ is a *deterministic* function of the combination matrix $\mathbf{A}(N)$. In fact, for a realization A of $\mathbf{A}(N)$, the limiting Granger estimator is:

$$\hat{A}_{\mathcal{P}} \triangleq [R_1]_{\mathcal{P}} [R_0]_{\mathcal{P}}^{-1}, \quad (164)$$

with:

$$R_0 \triangleq \sum_{i=0}^{\infty} A^i [A^i]^\top, \quad R_1 \triangleq AR_0. \quad (165)$$

When A is symmetric, we have (recall that I denotes the $N \times N$ identity matrix):

$$R_0 = (I - A^2)^{-1}, \quad (166)$$

and the limiting Granger estimator admits the following expression [29, App. A, Eq. (66)]:

$$\hat{A}_{\mathcal{P}} = A_{\mathcal{P}} + A_{\mathcal{P}\mathcal{P}'} (I_{\mathcal{P}'} - [A^2]_{\mathcal{P}'})^{-1} [A^2]_{\mathcal{P}'\mathcal{P}}, \quad (167)$$

which is critical to prove the following lemma.

Lemma 9 (Bound on the Error of the Limiting Granger Estimator): Let A be an $N \times N$ scaled left-stochastic matrix as in (125). If A is symmetric, then we have:

$$\|\hat{A}_{\mathcal{P}} - A_{\mathcal{P}}\|_{\max\text{-off}} \leq \kappa M \quad \forall \mathcal{P} \subseteq \{1, 2, \dots, N\}, \quad (168)$$

where κ is a positive constant, and M is defined in (126).

Proof: Since A is symmetric, the limiting Granger estimator admits the representation in (167), which, with the notation introduced in (144), becomes:

$$\hat{A}_{\mathcal{P}} - A_{\mathcal{P}} = A_{\mathcal{P}\mathcal{P}'} H[A^2]_{\mathcal{P}'\mathcal{P}}, \quad (169)$$

or, in terms of the individual (k, ℓ) -entry:

$$\begin{aligned} e_{k\ell} &\triangleq [A_{\mathcal{P}\mathcal{P}'} H[A^2]_{\mathcal{P}'\mathcal{P}}]_{k\ell} = \sum_{j, m \in \mathcal{P}'} a_{kj} h_{jm} a_{m\ell}^{(2)} \\ &= \sum_{j \in \mathcal{P}'} a_{kj} h_{jj} a_{j\ell}^{(2)} + \sum_{\substack{j, m \in \mathcal{P}' \\ j \neq m}} a_{kj} h_{jm} a_{m\ell}^{(2)}. \end{aligned} \quad (170)$$

We note that $e_{k\ell} \geq 0$ since all involved matrices are nonnegative — see (146) and (147) for what concerns H . Therefore, it suffices to prove that:

$$e_{k\ell} \leq \kappa M, \quad (171)$$

for some positive constant κ . To this aim, let us consider two indices $k, \ell \in \mathcal{P}$. Calling upon Lemma 8, we can apply (146) and (147) in (170), yielding:

$$e_{k\ell} \leq \bar{\alpha} \sum_{j \in \mathcal{P}'} a_{kj} a_{j\ell}^{(2)} + \bar{\beta} \sum_{\substack{j, m \in \mathcal{P}' \\ j \neq m}} a_{kj} a_{jm} a_{m\ell}^{(2)} + M \bar{\gamma} \sum_{\substack{j, m \in \mathcal{P}' \\ j \neq m}} a_{kj} a_{m\ell}^{(2)}. \quad (172)$$

The first summation in (172) can be upper bounded as follows:

$$\begin{aligned} \sum_{j \in \mathcal{P}'} a_{kj} a_{j\ell}^{(2)} &\leq \sum_{j \in \mathcal{P}'} a_{kj} (2\rho a_{j\ell} + M) \\ &= 2\rho \sum_{j \in \mathcal{P}'} a_{kj} a_{j\ell} + M \sum_{j \in \mathcal{P}'} a_{kj} \\ &\leq 2\rho M + M = 3M, \end{aligned} \quad (173)$$

where in the first inequality we used (138), while in the second inequality we used the following bounds:

$$\sum_{j \in \mathcal{P}'} a_{kj} a_{j\ell} \leq \sum_{\substack{j=1 \\ j \neq k, \ell}}^N a_{kj} a_{j\ell} \leq M, \quad \sum_{j \in \mathcal{P}'} a_{kj} \leq \rho. \quad (174)$$

Here, we remark that the inequality on the left exploits the fact that $k, \ell \in \mathcal{P}$ and $j \in \mathcal{P}'$, so we are allowed to extend the sum across $j \in \mathcal{P}'$ to a sum across $j \in \{1, 2, \dots, N\} \setminus \{k, \ell\}$.

Next we focus on the second summation in (172), which can be upper bounded as follows:

$$\begin{aligned} \sum_{\substack{j, m \in \mathcal{P}' \\ j \neq m}} a_{kj} a_{jm} a_{m\ell}^{(2)} &\leq \sum_{\substack{j, m \in \mathcal{P}' \\ j \neq m}} a_{kj} a_{jm} (2\rho a_{m\ell} + M) \\ &= 2\rho \sum_{\substack{j, m \in \mathcal{P}' \\ j \neq m}} a_{kj} a_{jm} a_{m\ell} + M \sum_{\substack{j, m \in \mathcal{P}' \\ j \neq m}} a_{kj} a_{jm}, \\ &\leq 2\rho^2 M + \rho^2 M = 3\rho^2 M, \end{aligned} \quad (175)$$

where in the first inequality we used (138), while in the second inequality we used (125) and (126) to get:

$$\sum_{\substack{j, m \in \mathcal{P}' \\ j \neq m}} a_{kj} a_{jm} a_{m\ell} = \sum_{j \in \mathcal{P}'} a_{kj} \sum_{\substack{m \in \mathcal{P}' \\ j \neq m}} a_{jm} a_{m\ell} \leq \sum_{j \in \mathcal{P}'} a_{kj} M \leq \rho M, \quad (176)$$

and:

$$\sum_{\substack{j, m \in \mathcal{P}' \\ j \neq m}} a_{kj} a_{jm} = \sum_{j \in \mathcal{P}'} a_{kj} \sum_{\substack{m \in \mathcal{P}' \\ j \neq m}} a_{jm} \leq \sum_{j \in \mathcal{P}'} a_{kj} \rho \leq \rho^2. \quad (177)$$

Finally, the third summation in (172) can be manipulated as follows:

$$\begin{aligned} \sum_{\substack{j,m \in \mathcal{P}' \\ j \neq m}} a_{kj} a_{m\ell}^{(2)} &= \sum_{\substack{j,m \in \mathcal{P}' \\ j \neq m}} a_{kj} \sum_{h=1}^N a_{mh} a_{h\ell} \\ &= \sum_{j \in \mathcal{P}'} a_{kj} \sum_{h=1}^N a_{h\ell} \sum_{\substack{m \in \mathcal{P}' \\ m \neq j}} a_{mh} \leq \rho^3, \end{aligned} \quad (178)$$

where in the last step we applied repeatedly (125), further noticing that in view of the symmetry of A we can write:

$$\sum_{\substack{m \in \mathcal{P}' \\ m \neq j}} a_{mh} = \sum_{\substack{m \in \mathcal{P}' \\ m \neq j}} a_{hm} \leq \rho. \quad (179)$$

Using (173), (175), and (178) in (172), we get:

$$e_{k\ell} \leq (3\rho\bar{\alpha} + 3\rho^2\bar{\beta} + \rho^3\bar{\gamma}) M \triangleq \kappa M, \quad (180)$$

which proves the claim. \blacksquare

APPENDIX F PROOF OF THEOREM 2

Since Lemma 9 holds for any symmetric matrix A fulfilling (125) and any subset $\mathcal{P} \subseteq \{1, 2, \dots, N\}$, in view of (168) we can write:

$$\|\widehat{\mathbf{A}}_{\mathcal{S}}(N) - \mathbf{A}_{\mathcal{S}}(N)\|_{\max\text{-off}} \leq \kappa \mathfrak{M}(N), \quad (181)$$

where $\mathfrak{M}(N)$ is the random variable defined in (111). Applying Theorem 4 to (181), we conclude that:

$$\sqrt{N} \|\widehat{\mathbf{A}}_{\mathcal{S}}(N) - \mathbf{A}_{\mathcal{S}}(N)\|_{\max\text{-off}} \xrightarrow{p} 0, \quad (182)$$

for any sequence of probed subsets \mathcal{S}_N . Now, by applying the triangle inequality we can write:

$$\begin{aligned} &\|\sqrt{N} \widehat{\mathbf{A}}_{\mathcal{S}}(N) - \gamma \mathbf{G}_{\mathcal{S}}(N)\|_{\max\text{-off}} \\ &\leq \sqrt{N} \|\widehat{\mathbf{A}}_{\mathcal{S}}(N) - \mathbf{A}_{\mathcal{S}}(N)\|_{\max\text{-off}} \\ &\quad + \|\sqrt{N} \mathbf{A}_{\mathcal{S}}(N) - \gamma \mathbf{G}_{\mathcal{S}}(N)\|_{\max\text{-off}}. \end{aligned} \quad (183)$$

The first term on the RHS of (183) vanishes in probability in view of (182). Moreover, in view of (32) we have that:

$$\mathbf{A}_{\mathcal{S}}(N) - \text{diag}(\mathbf{A}_{\mathcal{S}}(N)) = \rho\lambda \frac{\mathbf{G}_{\mathcal{S}}(N)}{1 + \mu_{\mathcal{G}}(N)}, \quad (184)$$

where $\text{diag}(\cdot)$ is a diagonal matrix having on the main diagonal the entries of its matrix argument. Using (184), we see that the second term on the RHS of (183) is upper bounded by:

$$\left| \frac{\rho\lambda\sqrt{N}}{1 + \mu_{\mathcal{G}}(N)} - \gamma \right| \xrightarrow{p} 0, \quad (185)$$

where the convergence follows from (64) and (47) since μ is strictly positive. We conclude that:

$$\|\sqrt{N} \widehat{\mathbf{A}}_{\mathcal{S}}(N) - \gamma \mathbf{G}_{\mathcal{S}}(N)\|_{\max\text{-off}} \xrightarrow{p} 0, \quad (186)$$

which concludes the proof. \blacksquare

APPENDIX G SAMPLE COMPLEXITY ANALYSIS

The following lemmas characterize the rate of convergence of the sample covariance estimators. Preliminarily, it is convenient to introduce the following auxiliary function:

$$f_T(x) \triangleq |\mathcal{S}_N|^2 e^{-T/2} + |\mathcal{S}_N|^2 e^{-[\sqrt{T}x - \sqrt{2}]^2}, \quad (187)$$

and the error matrices, for $j \in \{0, 1\}$:

$$\mathbf{E}_j \triangleq \left[\widehat{\mathbf{R}}_j(T, N) - [\mathbf{R}_j(N)] \right]_{\mathcal{S}_N}. \quad (188)$$

Lemma 10 (Sample Covariance Errors): Let us consider the dynamical system (2), with Laplacian combination matrix as in (32), and with network graph $\mathcal{G}(N)$ being a simple graph obtained from a Bollobás-Riordan multigraph $\mathcal{M}(N)$ with step parameter η . Then, under Assumption 1, there exists a constant C such that:

$$\mathbb{P}[\|\mathbf{E}_0\|_{\max} > \epsilon] \leq 3 f_T(\epsilon C) \quad \text{for } T > \frac{2}{(\epsilon C)^2}, \quad (189)$$

$$\mathbb{P}[\|\mathbf{E}_1\|_{\max} > \epsilon] \leq 4 f_{T-1}(\epsilon C) \quad \text{for } T > 1 + \frac{2}{(\epsilon C)^2}. \quad (190)$$

Proof: In this proof we will often consider conditional probabilities given $\mathbf{A}(N) = A$. This is tantamount to assuming that the dynamical system (3) is run with a *deterministic* matrix A . In such a scenario, matrix $\mathbf{R}_0(N)$ becomes deterministic, since according (24) it is a deterministic function of $\mathbf{A}(N)$. Accordingly, we will conveniently denote by the normal-font symbol R_0 the realization of $\mathbf{R}_0(N)$ corresponding to A . In contrast, the quantity $\mathbf{R}_0(T, N)$ in (30) will remain random, since by definition it depends also on the source of randomness given by $\{\mathbf{x}_i(N)\}_{i=1}^T$ and $\mathbf{y}_0(N)$.

The proof of (189) and (190) is a slight variation of the bounding technique used in [25, Lemma 1]. In particular, let us define for $i, j = 0, 1, \dots$, with $j \leq i$, the conditional cross-covariance between $\mathbf{y}_i(N)$ and $\mathbf{y}_j(N)$, namely,

$$\Sigma_{i,j} \triangleq \mathbb{E}[\mathbf{y}_i(N) \mathbf{y}_j^\top(N) | \mathbf{A}(N) = A] = A^{i-j} \Sigma_{j,j} = A^{i-j} R_0, \quad (191)$$

where the intermediate equality is a classical result on VAR models [39], while the last equality comes from the stationarity enforced by Assumption 1. Starting from (191), in [25, Lemma 1] the following bound is used:

$$\|\Sigma_{i,j}\|_{\max} \leq \|\Sigma_{i,j}\|_2 \leq \|A^{i-j}\|_2 \|R_0\|_2. \quad (192)$$

In our case we can exploit additional constraints on A to replace (192) by:

$$\|\Sigma_{i,j}\|_{\max} \leq \|A^{i-j}\|_{\infty} \|R_0\|_{\max} = \rho^{i-j} \max_{k=1,2,\dots,N} [R_0]_{kk}, \quad (193)$$

where the inequality comes from the fact that, for any two matrices M_1, M_2 of compatible dimensions, we have $\|M_1 M_2\|_{\max} \leq \|M_1\|_{\infty} \|M_2\|_{\max}$. The equality in (193) follows by using (132) and by applying Cauchy-Schwarz inequality to obtain $|[R_0]_{k\ell}| \leq \sqrt{[R_0]_{kk} [R_0]_{\ell\ell}}$.

Using (193) in place of (192), and leaving other arguments in the proof of [25, Lemma 1] unaltered, we get, for any T such that¹¹:

$$T > \frac{2}{(\epsilon \varphi(R_0))^2}, \quad (194)$$

the following bound:

$$\mathbb{P}[\|\mathbf{E}_0\|_{k\ell} > \epsilon | \mathbf{A}(N) = A] \leq 3 \left(e^{-T/2} + e^{-[\sqrt{T}\epsilon\varphi(R_0) - \sqrt{2}]^2} \right), \quad (195)$$

where:

$$\varphi(R_0) \triangleq \frac{1 - \rho}{16\sqrt{2}} \frac{\min_{k=1,2,\dots,N} [R_0]_{kk}}{\max_{k=1,2,\dots,N} [R_0]_{kk}^2}. \quad (196)$$

From (50), (127), and (153) we have the inequalities:

$$\min_{k=1,2,\dots,N} [R_0]_{kk} \geq 1, \quad \max_{k=1,2,\dots,N} [R_0]_{kk} \leq \bar{\alpha}, \quad (197)$$

which can be used to bound the quantity $\varphi(R_0)$ in (196) as:

$$\varphi(R_0) \geq \frac{1}{\bar{\alpha}^2} \frac{1 - \rho}{16\sqrt{2}} \triangleq C. \quad (198)$$

Since the function $e^{-(\sqrt{T}x - \sqrt{2})^2}$ is decreasing for any $x > \sqrt{2}/T$, we conclude from (195) and (198) that, under the condition on T in (189), we have:

$$\mathbb{P}[\|\mathbf{E}_0\|_{k\ell} > \epsilon | \mathbf{A}(N) = A] \leq 3 \left(e^{-T/2} + e^{-[\sqrt{T}\epsilon C - \sqrt{2}]^2} \right), \quad (199)$$

which is a bound independent of the current realization A . Therefore, by applying the law of total probability in (199) we get:

$$\mathbb{P}[\|\mathbf{E}_0\|_{k\ell} > \epsilon] \leq 3 \left(e^{-T/2} + e^{-[\sqrt{T}\epsilon C - \sqrt{2}]^2} \right). \quad (200)$$

Now, using the union bound over the set of probed nodes \mathcal{S}_N we can write:

$$\begin{aligned} \mathbb{P}[\|\mathbf{E}_0\|_{\max} > \epsilon] &\leq \sum_{k,\ell \in [1,N]} \mathbb{P}[\|\mathbf{E}_0\|_{k\ell} > \epsilon] \\ &\leq 3|\mathcal{S}_N|^2 \left(e^{-T/2} + e^{-[\sqrt{T}\epsilon C - \sqrt{2}]^2} \right), \end{aligned} \quad (201)$$

and the claim in (189) follows from the definition of f_T in (187). In order to obtain (190), we must apply the same steps shown above to the proof of [25, Lemma 2]. ■

Lemma 11 (Scaling Law Useful for Sample Complexity): Assume the same conditions used in Lemma 10, and consider the following scaling law for the number of samples:

$$T_N = \omega_N N \log N, \quad (202)$$

for some positive sequence ω_N diverging in an arbitrarily slow fashion as $N \rightarrow \infty$. Then, for any sequence \mathcal{S}_N satisfying (8), and for $j \in \{0, 1\}$, we have that:

$$\sqrt{N} \|\mathbf{R}_j(T_N, N)\|_{\mathcal{S}} - \|\mathbf{R}_j(N)\|_{\mathcal{S}} \xrightarrow{P} 0, \quad (203)$$

¹¹Condition (194) is explicitly stated in [25, Lemma 3], and basically requires that the function $e^{-(\sqrt{T}x - \sqrt{2})^2}$ appearing in (195) is evaluated in the region where it is decreasing, i.e., for $x > \sqrt{2}/T$.

Proof: We need to show that, for any $\delta > 0$:

$$\lim_{N \rightarrow \infty} \mathbb{P} \left[\sqrt{N} \|\mathbf{R}_j(T_N, N)\|_{\mathcal{S}} - \|\mathbf{R}_j(N)\|_{\mathcal{S}} > \delta \right] = 0. \quad (204)$$

We will prove the claim with reference to the case $j = 0$, with the proof being identical for $j = 1$. Let us consider Lemma 10 with the choice $\epsilon = \delta/\sqrt{N}$. Since Eq. (202) implies that:

$$T_N (\epsilon C)^2 = T_N \frac{(\delta C)^2}{N} \xrightarrow{N \rightarrow \infty} \infty, \quad (205)$$

we see that condition on T in (189) is met for N sufficiently large. We conclude that to prove (204) it suffices to show that, for any $\delta > 0$:

$$\lim_{N \rightarrow \infty} f_{T_N} \left(\frac{\delta C}{\sqrt{N}} \right) = 0. \quad (206)$$

Now, the first term on the RHS of (187) converges to zero since T_N in (202) tends to $+\infty$ faster than $\log |\mathcal{S}_N|$. On the other hand, the second term on the RHS of (187) can be written as:

$$\begin{aligned} &\exp \left\{ - \left(\sqrt{T_N} \frac{\delta C}{\sqrt{N}} - \sqrt{2} \right)^2 + \log |\mathcal{S}_N|^2 \right\} \\ &= \exp \left\{ - \left[\left(\sqrt{\delta^2 C^2 \frac{\omega_N \log N}{\log |\mathcal{S}_N|}} - \frac{1}{\sqrt{\log |\mathcal{S}_N|}} \right)^2 - 1 \right] \log |\mathcal{S}_N|^2 \right\}, \end{aligned} \quad (207)$$

and vanishes as $N \rightarrow \infty$ in view of (8) and the fact that $\omega_N \rightarrow \infty$ by assumption. ■

Proof of Theorem 3: Calling upon Lemma 11 in [33], we have the following bound:

$$\begin{aligned} &\sqrt{N} \|\widehat{\mathbf{A}}_{\mathcal{S}}(T, N) - \widehat{\mathbf{A}}_{\mathcal{S}}(N)\|_{\max} \\ &\leq 2 \|\mathbf{R}_0\|_{\mathcal{S}}^{-1} \|\mathbf{R}_0\|_{\mathcal{S}} \sqrt{N} (\|\mathbf{R}_0(T, N)\|_{\mathcal{S}} - \|\mathbf{R}_0(N)\|_{\mathcal{S}})_{\max} \\ &\quad + \|\mathbf{R}_1(T, N)\|_{\mathcal{S}} - \|\mathbf{R}_1(N)\|_{\mathcal{S}}. \end{aligned} \quad (208)$$

Moreover, in [33, Eq. (321)], it was shown that:

$$\|\mathbf{R}_0\|_{\mathcal{S}}^{-1} \|\mathbf{R}_0\|_{\mathcal{S}} \leq 1 + \rho^2. \quad (209)$$

If we now use (209) in (208), from Lemma 11 we conclude that:

$$\sqrt{N} \|\widehat{\mathbf{A}}_{\mathcal{S}}(T_N, N) - \widehat{\mathbf{A}}_{\mathcal{S}}(N)\|_{\max} \xrightarrow{P} 0. \quad (210)$$

Since we know from Theorem 2 that the *limiting* Granger estimator fulfills (11), by application of the triangle inequality we have that:

$$\|\sqrt{N} \widehat{\mathbf{A}}_{\mathcal{S}}(T_N, N) - \gamma \mathbf{G}_{\mathcal{S}}(N)\|_{\max\text{-off}} \xrightarrow{P} 0, \quad (211)$$

which in turn implies (see footnote 4) that (62) holds true with high probability as $N \rightarrow \infty$ with the sample scaling law in (202). This means that the clustering algorithm graphclu is able to reconstruct correctly the subgraph of probed nodes, provided that the latter is neither fully connected nor fully disconnected. However, the probability that $\mathcal{G}_{\mathcal{S}}(N)$ is fully connected or fully disconnected vanishes $N \rightarrow \infty$ in view of Lemma 6, and the proof is complete. ■

APPENDIX H
AUXILIARY TECHNICAL RESULTS

Lemma 12: Let f_i be the sequence recursively defined as:

$$f_{i+1} = af_i + a^i (bi^2 + ci + d), \quad i = 1, 2, \dots \quad (212)$$

with $0 < a < 1$ and $b, c, d \in \mathbb{R}$. Then we have that:

$$f_i = a^{i-1} \left(f_1 + b \frac{i(i-1)(2i-1)}{6} + c \frac{i(i-1)}{2} + d(i-1) \right). \quad (213)$$

Proof: Unfolding the recursion in (212), we conclude that, for all $i > 1$:

$$f_i = a^{i-1} \left(f_1 + \sum_{j=1}^{i-1} (bj^2 + cj + d) \right). \quad (214)$$

Thus, to obtain (213) we use the well-known results:

$$\sum_{j=1}^{i-1} j = \frac{i(i-1)}{2}, \quad \sum_{j=1}^{i-1} j^2 = \frac{i(i-1)(2i-1)}{6}. \quad (215)$$

The following lemma is an adaptation of Theorem 2.1 in [59], useful for the proofs of Lemma 2 and Theorem 4.

Lemma 13: Let us consider a family of random sequences $\{\mathbf{u}_\theta(n)\}_{n \geq 1}$ spanned by the parameter $\theta \in \Theta$ and defined on the same probability space. Assume that the following conditions are met for all $\theta \in \Theta$:

$$\mathbf{u}_\theta(1) = 0, \quad 0 \leq \mathbf{u}_\theta(n) \leq b \quad \text{for all } n > 1, \quad (216)$$

for a positive constant b . Let us further define the first two conditional moment sequences $\{\boldsymbol{\nu}_\theta(n)\}_{n \geq 1}$ and $\{\boldsymbol{\chi}_\theta(n)\}_{n \geq 1}$ w.r.t. a given filtration $\{\mathcal{F}(n)\}_{n \geq 1}$ of the underlying space:

$$\boldsymbol{\nu}_\theta(1) \triangleq 0, \quad \boldsymbol{\nu}_\theta(n) \triangleq \mathbb{E}[\mathbf{u}_\theta(n) | \mathcal{F}(n-1)], \quad (217)$$

$$\boldsymbol{\chi}_\theta(1) \triangleq 0, \quad \boldsymbol{\chi}_\theta(n) \triangleq \mathbb{E}[\mathbf{u}_\theta^2(n) | \mathcal{F}(n-1)], \quad (218)$$

and finally consider:

$$\mathbf{U}_\theta(N) \triangleq \sum_{n=1}^N \mathbf{u}_\theta(n), \quad \mathbf{C}_\theta(N) \triangleq \sum_{n=1}^N \boldsymbol{\nu}_\theta(n), \quad (219)$$

$$\mathbf{Q}_\theta(N) \triangleq \sum_{n=1}^N \boldsymbol{\chi}_\theta(n), \quad \bar{\mathbf{U}}_\theta(N) \triangleq \mathbf{U}_\theta(N) - \mathbf{C}_\theta(N). \quad (220)$$

Then, for any subset $\mathcal{T} \subseteq \Theta$ and any $u > 0$ we have:

$$\mathbb{P} \left[\max_{\theta \in \mathcal{T}} \mathbf{U}_\theta(N) > u \right] \leq |\mathcal{T}| e^{-\frac{3}{16} \tau u} + \mathbb{P} \left[\max_{\theta \in \mathcal{T}} \mathbf{C}_\theta(N) > \frac{u}{2} \right]. \quad (221)$$

Proof: For any two events \mathcal{E}_1 and \mathcal{E}_2 , it is true that ($\bar{\mathcal{E}}_2$ is the complement of event \mathcal{E}_2):

$$\mathbb{P}[\mathcal{E}_1] = \mathbb{P}[\mathcal{E}_1, \mathcal{E}_2] + \mathbb{P}[\mathcal{E}_1, \bar{\mathcal{E}}_2] \leq \mathbb{P}[\mathcal{E}_1, \mathcal{E}_2] + \mathbb{P}[\bar{\mathcal{E}}_2], \quad (222)$$

so that we can write:

$$\mathbb{P} \left[\max_{\theta \in \mathcal{T}} \mathbf{U}_\theta(N) > u \right] \leq \mathbb{P} \left[\max_{\theta \in \mathcal{T}} \mathbf{U}_\theta(N) > u, \max_{\theta \in \mathcal{T}} \mathbf{C}_\theta(N) \leq \frac{u}{2} \right] + \mathbb{P} \left[\max_{\theta \in \mathcal{T}} \mathbf{C}_\theta(N) > \frac{u}{2} \right]. \quad (223)$$

Let us focus on the first term on the RHS of (223). We have the following relations:

$$\begin{aligned} & \left\{ \bigcup_{\theta \in \mathcal{T}} \{ \mathbf{U}_\theta(N) > u \} \right\} \cap \left\{ \bigcap_{\theta' \in \mathcal{T}} \{ \mathbf{C}_{\theta'}(N) \leq u/2 \} \right\} \\ & \stackrel{(a)}{=} \left\{ \bigcup_{\theta \in \mathcal{T}} \{ \bar{\mathbf{U}}_\theta(N) > u - \mathbf{C}_\theta(N) \} \right\} \cap \left\{ \bigcap_{\theta' \in \mathcal{T}} \{ \mathbf{C}_{\theta'}(N) \leq u/2 \} \right\} \\ & \stackrel{(b)}{\subseteq} \bigcup_{\theta \in \mathcal{T}} \{ \bar{\mathbf{U}}_\theta(N) > u - \mathbf{C}_\theta(N) \} \cap \{ \mathbf{C}_\theta(N) \leq u/2 \} \\ & \stackrel{(c)}{\subseteq} \bigcup_{\theta \in \mathcal{T}} \{ \bar{\mathbf{U}}_\theta(N) > u - u/2 \} \cap \{ \mathbf{C}_\theta(N) \leq u/2 \} \\ & \stackrel{(d)}{\subseteq} \bigcup_{\theta \in \mathcal{T}} \{ \bar{\mathbf{U}}_\theta(N) > u - u/2 \} \cap \{ \mathbf{Q}_\theta(N) \leq bu/2 \}, \quad (224) \end{aligned}$$

where (a) follows from the definition of $\bar{\mathbf{U}}_\theta(N)$ in (220); (b) is obtained by retaining only the event corresponding to $\theta' = \theta$ in the intersection; (c) holds since in the intersection $-u/2 \leq -\mathbf{C}_\theta(N)$; and (d) follows by observing that, in view of (216), (217) and (218) we have $\boldsymbol{\chi}_\theta(n) \leq b\nu_\theta(n)$, which in turn implies, from the definitions in (219) and (220), that $\mathbf{Q}_\theta(N) \leq b\mathbf{C}_\theta(N)$. Using (224) in the first term on the RHS of (223), and further applying the union bound, we obtain:

$$\begin{aligned} & \mathbb{P} \left[\max_{\theta \in \mathcal{T}} \mathbf{U}_\theta(N) > u, \max_{\theta \in \mathcal{T}} \mathbf{C}_\theta(N) \leq \frac{u}{2} \right] \\ & \leq \sum_{\theta \in \mathcal{T}} \mathbb{P} \left[\bar{\mathbf{U}}_\theta(N) > \frac{u}{2}, \mathbf{Q}_\theta(N) \leq \frac{bu}{2} \right]. \quad (225) \end{aligned}$$

The sequence $\{\bar{\mathbf{U}}_\theta(N)\}_{N \geq 1}$ is a martingale by construction, since it is a sum of random variables (i.e., $\mathbf{u}_\theta(n)$) minus their conditional expectation (i.e., $\boldsymbol{\nu}_\theta(n)$). Moreover, from (216) we have the bound:

$$\bar{\mathbf{U}}_\theta(N+1) - \bar{\mathbf{U}}_\theta(N) \leq \mathbf{u}_\theta(N+1) \leq b. \quad (226)$$

Therefore, it can be readily checked that the scaled sequence $\{\bar{\mathbf{U}}_\theta(N)/b\}_{N \geq 1}$ meets the hypotheses of Theorem 2.1 in [59], and in particular the upper bound obtained by combining Eqs. (10), (11) and (15) in [59], which finally yields:

$$\mathbb{P} \left[\bar{\mathbf{U}}_\theta(N) > \frac{u}{2}, \mathbf{Q}_\theta(N) \leq \frac{bu}{2} \right] \leq e^{-\frac{3}{16} \tau u}, \quad (227)$$

and the proof is complete. \blacksquare

REFERENCES

- [1] M. Cirillo, V. Matta, and A. H. Sayed, "Learning Bollobás–Riordan graphs under partial observability," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Toronto, ON, Canada, Jun. 2021, pp. 5360–5364.
- [2] I. D. Couzin, "Collective cognition in animal groups," *Trends Cogn. Sci.*, vol. 13, no. 1, pp. 36–43, Jan. 2009.
- [3] B. L. Partridge, "The structure and function of fish schools," *Sci. Amer.*, vol. 246, no. 6, pp. 114–123, Jun. 1982.

- [4] V. Matta, V. Bordinon, A. Santos, and A. H. Sayed, "Interplay between topology and social learning over weak graphs," *IEEE Open J. Signal Process.*, vol. 1, pp. 99–119, 2020.
- [5] 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, Jan. 2020.
- [6] V. Matta, A. Santos, and A. H. Sayed, "Graph learning under partial observability," *Proc. IEEE*, vol. 108, no. 11, pp. 2049–2066, Nov. 2020.
- [7] G. Mateos, S. Segarra, A. G. 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] A. Moneta, N. Chlass, D. Entner, and P. Hoyer, "Causal search in structural vector autoregressive models," in *Proc. Neural Inf. Process. Syst. (NIPS)*, Vancouver, BC, Canada, Dec. 2009, pp. 95–118.
- [13] B. Passetoup, 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.
- [14] S. Segarra, M. T. Schaub, and A. Jadbabaie, "Network inference from consensus dynamics," in *Proc. IEEE 56th Annu. Conf. Decis. Control (CDC)*, Dec. 2017, pp. 3212–3217.
- [15] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 467–483, Sep. 2017.
- [16] J. Mei and J. M. F. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2077–2092, Apr. 2017.
- [17] A. Natali, M. Coutino, E. Isufi, and G. Leus, "Online time-varying topology identification via prediction-correction algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Toronto, ON, Canada, Jun. 2021, pp. 5400–5404.
- [18] D. Materassi and M. V. Salapaka, "Network reconstruction of dynamical polytrees with unobserved nodes," in *Proc. IEEE 51st Annu. Conf. Decis. Control (CDC)*, Maui, HI, USA, Dec. 2012, pp. 4629–4634.
- [19] J. Etesami, N. Kiyavash, and T. Coleman, "Learning minimal latent directed information polytrees," *Neural Comput.*, vol. 28, no. 9, pp. 1723–1768, Aug. 2016.
- [20] P. Geiger, K. Zhang, M. Gong, D. Janzing, and B. Schölkopf, "Causal inference by identification of vector autoregressive processes with hidden components," in *Proc. 32nd Int. Conf. Mach. Learn.*, Lille, France, vol. 37, Jul. 2015, pp. 1917–1925.
- [21] D. Materassi and M. V. Salapaka, "Identification of network components in presence of unobserved nodes," in *Proc. IEEE 54th Annu. Conf. Decis. Control*, Osaka, Japan, Dec. 2015, pp. 1563–1568.
- [22] 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, no. 1, pp. 2293–2337, Jan. 2012.
- [23] 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.
- [24] J. Bento, M. Ibrahim, and A. Montanari, "Learning networks of stochastic differential equations," in *Proc. Neural Inf. Process. Syst. (NIPS)*, Vancouver, QC, Canada, Dec. 2010, pp. 172–180.
- [25] F. Han, H. Lu, and H. Liu, "A direct estimation of high dimensional stationary vector autoregressions," *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 3115–3150, Dec. 2015.
- [26] 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.
- [27] 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 (CDC)*, Las Vegas, NV, USA, Dec. 2016, pp. 2938–2944.
- [28] A. Jalali and S. Sanghavi, "Learning the dependence graph of time series with latent factors," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Scotland, U.K., Jun. 2012, pp. 619–626.
- [29] 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.
- [30] 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.
- [31] V. Matta, A. Santos, and A. H. Sayed, "Tomography of large adaptive networks under the dense latent regime: Invited paper," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Oct. 2018, pp. 2144–2148.
- [32] V. Matta, A. Santos, and A. H. Sayed, "Graph learning with partial observations: Role of degree concentration," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 1312–1316.
- [33] V. Matta, A. Santos, and A. H. Sayed, "Graph learning over partially observed diffusion networks: Role of degree concentration," *IEEE Open J. Signal Process.*, vol. 3, pp. 335–371, 2022.
- [34] P. Erdős and A. Rényi, "On random graphs I," *Publicationes Math. Debrecen*, vol. 6, no. 1, pp. 290–297, Jan. 1959.
- [35] B. Bollobás, *Random Graphs*. Cambridge, U.K.: Cambridge Univ. Press, 2001.
- [36] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.
- [37] B. Bollobás and O. Riordan, "Mathematical results on scale-free random graphs," in *Handbook of Graphs and Networks: From the Genome to the Internet*. Hoboken, NJ, USA: Wiley, 2003, pp. 1–34.
- [38] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, "The degree sequence of a scale-free random graph process," *Random Struct. Algorithms*, vol. 18, no. 3, pp. 279–290, 2001.
- [39] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. Berlin, Germany: Springer, 2005.
- [40] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. New York, NY, USA: Springer, 2001.
- [41] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Found. Trends Mach. Learn.*, vol. 7, nos. 4–5, pp. 311–801, 2014.
- [42] U. A. Khan, W. U. Bajwa, A. Nedić, M. G. Rabbat, and A. H. Sayed, "Optimization for data-driven learning and control," *Proc. IEEE*, vol. 108, no. 11, pp. 1863–1868, Nov. 2020.
- [43] A. H. Sayed, "Adaptive networks," *Proc. IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.
- [44] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, Sep. 2004.
- [45] M. H. DeGroot, "Reaching a consensus," *J. Amer. Statist. Assoc.*, vol. 69, no. 345, pp. 118–121, Mar. 1974.
- [46] 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.
- [47] 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.
- [48] R. Diestel, *Graph Theory*. Heidelberg, Germany: Springer-Verlag, 2005.
- [49] R. Van Der Hofstad, *Random Graphs and Complex Networks*. Berlin, Germany: Springer, 2016.
- [50] M. C. Grant and S. P. Boyd. *CVX: MATLAB Software for Disciplined Convex Programming*. [Online]. Available: <http://cvxr.com/cvx>
- [51] M. C. Grant and S. P. Boyd, *Graph Implementations for Non-smooth Convex Programs*. London, U.K.: Springer-Verlag, 2008, pp. 95–110.
- [52] V. Bordinon, V. Matta, and A. H. Sayed, "Adaptive social learning," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, pp. 6053–6081, Sep. 2021.
- [53] V. Shumovskaia, K. Ntemos, S. Vlaski, and A. H. Sayed, "Online graph learning from social interactions," in *Proc. 55th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Oct./Nov. 2021, pp. 1263–1267.
- [54] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, Austin, TX, USA, Jan. 2015, pp. 4292–4293.
- [55] A. D. Broido and A. Clauset, "Scale-free networks are rare," *Nature Commun.*, vol. 10, no. 1, pp. 1–10, Mar. 2019.

- [56] B. Bollobás and O. Riordan, "Robustness and vulnerability of scale-free random graphs," *Internet Math.*, vol. 1, no. 1, pp. 1–35, Jan. 2004.
- [57] O. Stolz, *Vorlesungen Über Allgemeine Arithmetik: Nach Den Neueren Ansichten*. Leipzig, Germany: Teubner, 1885.
- [58] E. Cesàro, "Sur la convergence des séries," *Nouvelles Annales de Mathématiques*, vol. 7, pp. 49–59, Feb. 1888.
- [59] X. Fan, I. Grama, and Q. Liu, "Hoeffding's inequality for supermartingales," *Stochastic Process. Appl.*, vol. 122, no. 10, pp. 3545–3559, 2012.

Michele Cirillo (Graduate Student Member, IEEE) received the B.S. degree in computer engineering from the University of Salerno, Italy, and the M.S. degree in computer engineering from the University of Rome La Sapienza, Italy. He is currently pursuing the Ph.D. degree in information engineering with the Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno. His research interests include distributed learning over graphs and security in communication networks.

Vincenzo Matta (Senior Member, IEEE) is currently a Full Professor in telecommunications at the Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, Italy. He is the author of more than 130 articles published on international journals and proceedings of international conferences. His research interests include adaptation and learning over networks, social learning, statistical inference on graphs, and security in communication networks. He serves as an Associate Editor for the IEEE OPEN JOURNAL OF SIGNAL PROCESSING. He served as an Associate Editor for the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS, the IEEE SIGNAL PROCESSING LETTERS, and the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS, and a Senior Area Editor for the IEEE SIGNAL PROCESSING LETTERS. He is a member of the Sensor Array and Multichannel Technical Committee of the Signal Processing Society (SPS), and served as an IEEE SPS Representative to the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS.

Ali H. Sayed (Fellow, IEEE) is the Dean of Engineering at the EPFL, Switzerland, where he also leads the Adaptive Systems Laboratory (<https://asl.ep.ch/>). He is a member of the U.S. National Academy of Engineering (NAE) and The World Academy of Sciences (TWAS). He served as the President of the IEEE Signal Processing Society in 2018 and 2019. His research areas cover 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 received the 2014 Papoulis Award from the European Association for Signal Processing, the 2005 Terman Award from the American Society for Engineering Education, and several best paper awards. He is also a fellow of EURASIP and the American Association for the Advancement of Science (AAAS).

Open Access funding provided by 'Università degli Studi di Salerno' within the CRUI CARE Agreement