LEARNING BOLLOBÁS-RIORDAN GRAPHS UNDER PARTIAL OBSERVABILITY

Michele Cirillo[⋆] *Vincenzo Matta*[⋆] *Ali H. Sayed*[†]

[⋆]Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, Italy
[†]School of Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
E-mails: {micirillo,vmatta}@unisa.it - ali.sayed@epfl.ch

ABSTRACT

This work examines the problem of learning the topology of a network (*graph learning*) from the signals produced at a subset of the network nodes (*partial observability*). This challenging problem was recently tackled assuming that the topology is drawn according to an Erdős-Rényi model, for which it was shown that graph learning under partial observability is achievable, exploiting in particular homogeneity across nodes and independence across edges. However, several real-world networks do not match the optimistic assumptions of homogeneity/independence, for example, high heterogeneity is often observed between very connected nodes (hubs) and scarcely connected peripheral nodes. Random graphs with *preferential attachment* were conceived to overcome these issues. In this work, we discover that, over first-order vector autoregressive systems with a stable Laplacian combination matrix, graph learning is achievable under partial observability, when the network topology is drawn according to a popular preferential attachment model known as the Bollobás-Riordan model.

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

1. INTRODUCTION AND MOTIVATION

In this work we consider a network with N nodes where at time instant $i \geqslant 1$, each node k collects a random input signal $\boldsymbol{x}_{k,i}$ and forms the output signal $\boldsymbol{y}_{k,i}$ according to the following diffusion model, a.k.a. first-order Vector AutoRegressive (VAR) model [1]:

$$\boldsymbol{y}_{k,i} = \sum_{\ell=1}^{N} a_{k\ell} \boldsymbol{y}_{\ell,i-1} + \boldsymbol{x}_{k,i}. \tag{1}$$

The random variables $\{\boldsymbol{x}_{k,i}\}$ are independent and identically distributed (i.i.d.) w.r.t. to both $k \in \{1, 2, \ldots, N\}$ and $i \geqslant 1$, with zero mean and finite variance. A critical feature of (1) is that the *combination weights* $a_{k\ell}$ reflect the network topology: $a_{k\ell}$ is strictly positive if k and ℓ are connected, and is zero otherwise. Thus, it is legitimate to ask whether the topology can be retrieved from the dynamical evolution of the nodes' signals in (1). The resulting graph learning problem (also referred to as topology inference or network tomography) has been extensively studied in the last decades, with fruitful application to several domains: in social learning, to discover how individual network agents contribute to opinion formation [2]; in neuroscience, to unveil the relationships between functional and structural connectivity [3]; in biology, to reveal how fish in a school coordinate their movements to escape a predator [4, 5].

Unfortunately, in many practical situations it is impossible to gather signals from the overall ensemble of nodes. This limitation leads to the *partial observability* setting, where one can observe the signals in (1) emitted only by nodes belonging to a subset $\mathcal{S} \subset \{1, 2, \ldots, N\}$. Formally, we assume that the cardinality of $\mathcal{S}$ can scale with the network size in such a way that the asymptotic fraction of probed nodes stays constant, namely,

$$\frac{|\mathcal{S}|}{N} \overset{N\to\infty}{\longrightarrow} \xi \in (0, 1). \tag{2}$$

The partial observability assumption adds significant complexity to the topology estimation task, since the collected state measurements are influenced by the latent unobserved nodes which act as noise sources. This work focuses on this challenging paradigm, namely, we address *graph learning under partial observability*, where the goal is to estimate the interconnections among nodes in $\mathcal{S}$, by gathering only signals from $\mathcal{S}$.

1.1. Related Work

There exist several useful works addressing the graph learning problem. We now list some references more closely related to our work, and refer the reader to [6] for a broader literature survey. Conditions for graph learning under partial observability are available for specific types of graphs, e.g., for polytrees [7, 8]. More general graph structures are considered in [9, 10], 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 [11, 12], in the context of high-dimensional graphical models with latent variables. In [11], provable learning guarantees are offered under an appropriate local separation criterion, whereas in [12] graph learning is shown to be achievable under the so-called *sparsity and low-rank* condition, namely, when the probed subnetwork is sparsely connected, whereas 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 VAR model considered here. For dynamical graph models, relevant results under full observability were presented in [13–19], whereas partial observability was recently addressed in [20–25]. Particularly relevant to our work is the setting considered in [23–25], where the VAR model (1) runs on top of an Erdős-Rényi random graph [26, 27]. Under reasonable technical assumptions, it is shown that graph learning is achievable under different regimes of connectivity. However, despite their popularity, Erdős-Rényi graphs are often not a faithful representation of real-world networks. For this reason, other random graphs have been proposed, which are able to capture several useful features emerging over real-world networks. One notable model is the *preferential attachment* model, which can be traced back to the work of Barabási

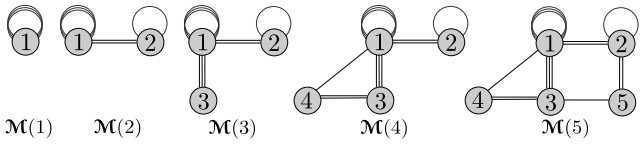

Fig. 1. One example of iterative construction of a BR multigraph with parameter $\eta = 3$.

and Albert [28]. For example, these types of graphs are able to embody the appearance of *hubs* with many connections (as opposed to peripheral, scarcely connected nodes), while Erdős-Rényi graphs are *homogeneous* by their own nature [26, 27]. As explained in Sec. 2, 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). Unfortunately, owing to this dependence, extending results valid for Erdős-Rényi graphs to preferential attachments graphs is a highly nontrivial task. The main contribution of the present work is to solve this challenging task, which allows us to provide a set of sufficient conditions for *consistent graph learning over preferential attachment graphs*.

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. If A is a matrix, we denote its (k, ℓ)-entry by $a_{k\ell}$. Sets and graphs are denoted by upper-case calligraphic letters. For an $N \times N$ matrix A, the submatrix spanning the rows and columns indexed by a set $\mathcal{S} \subset \{1, 2, \ldots, N\}$ is denoted by $A_{\mathcal{S}}$, or alternatively by $[A]_{\mathcal{S}}$. For a graph $\mathcal{G}$, the corresponding capital letter G is used to denote its adjacency matrix.

2. BOLLOBÁS-RIORDAN MODEL

Preferential attachment graphs are usually obtained as the result of an iterative procedure where, starting from an initial graph with a certain predetermined structure, at each subsequent iteration one node is added, along with some edges connecting this node to the rest of the graph constructed until the current iteration. The term preferential attachment stems from the fact that, at each iteration, the probability that the new node is connected to an existing node is proportional to the degree of the latter. This way, 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. Following the work of Barabási and Albert [28], several preferential attachment models have been proposed. One of the most popular variants is the Bollobás-Riordan (BR) random graph, which is the model examined in this work. The BR model is able to capture many features of real-world networks, through an elegant mathematical formulation that allows obtaining clean analytical results as regards the main quantitative descriptors of the graph (e.g., node degrees, minimum and maximum degrees, centrality measures).

Actually, the Bollobás-Riordan graph is a *multi*graph, which means that multiple self-loops and multiple edges are permitted. The multigraph structure was chosen by Bollobás and Riordan since it is instrumental to prove a number of theoretical results [29, 30]. The final goal of their model is 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, while the simple graph inherits the fundamental properties of the multigraph. Accordingly, once a BR multigraph is constructed, the corresponding simple graph that will be used to draw the weights governing the VAR model in (1), is obtained by merely uprooting all self-loops and multiple edges from the BR multigraph.

A random multigraph of size n will be denoted by $\mathfrak{M}(n)$. Its *adjacency matrix*, denoted by $\boldsymbol{M}(n)$, is the symmetric $n \times n$ matrix whose off-diagonal (k, ℓ)-entry $\boldsymbol{m}_{k\ell}(n)$ counts the number of edges between nodes k and ℓ, and whose diagonal entry $\boldsymbol{m}_{kk}(n)$ counts the number of self-loops of node k. We adopt the standard convention that, in a multigraph, the *degree* $\boldsymbol{d}_k(n)$ of node k is the number of edges connected to k, with self-loops counted twice[1] [31]:

$$\boldsymbol{d}_k(n) = \sum_{\ell=1}^{n} \boldsymbol{m}_{k\ell}(n) + 2\,\boldsymbol{m}_{kk}(n). \tag{3}$$

A Bollobás-Riordan graph with parameter $\eta \in \mathbb{N}$ consists of a random sequence of multigraphs $\{\mathfrak{M}(n)\}_{n\geqslant 1}$, which are iteratively constructed as follows — see Fig. 1 for a graphical illustration. The initial multigraph $\mathfrak{M}(1)$ is a deterministic multigraph with one node and η self-loops. Multigraph $\mathfrak{M}(n)$ is constructed starting from $\mathfrak{M}(n-1)$ adding a new node n and η new edges. Specifically, η steps are performed, and at each step an edge is introduced to connect node n to a node randomly chosen from the set $\{1, 2, \ldots, n\}$. The intermediate multigraph obtained at step $s = 1, 2, \ldots, \eta$, is denoted by $\mathfrak{M}(n; s)$. Accordingly, since after η steps we obtain the updated multigraph $\mathfrak{M}(n)$, we have the identity $\mathfrak{M}(n; \eta) = \mathfrak{M}(n)$. Likewise, we adopt the convention that $\mathfrak{M}(n; 0) = \mathfrak{M}(n-1)$. The particular node that becomes connected to n through the edge introduced at step s is denoted by $\boldsymbol{v}(n; s)$. At each step s, the degree of a node $k \neq n$ in the intermediate multigraph $\mathfrak{M}(n; s)$ increases by 1 if the node picked at step s is equal to k, namely,

$$\boldsymbol{d}_k(n; s) = \boldsymbol{d}_k(n; s-1) + \mathbb{I}(\boldsymbol{v}(n; s) = k), \tag{4}$$

where $\mathbb{I}(\cdot)$ is the indicator function, which is equal to 1 if its argument is true, and is zero otherwise. The degree of the new node n increases by 1 if the node picked at step s is equal to $k < n$, while increases by 2 if node n itself is picked, since each self-loop is counted twice in the degree, namely,[2]

$$\boldsymbol{d}_n(n; s) = \boldsymbol{d}_n(n; s-1) + 1 + \mathbb{I}(\boldsymbol{v}(n; s) = n). \tag{5}$$

The description of the BR graph construction is now completed by assigning the probability that a particular node is picked:

$$\mathbb{P}\left[\boldsymbol{v}(n; s) = k | \mathfrak{M}(n; s-1)\right] = \frac{\boldsymbol{d}_k(n; s-1)}{\sum_{\ell=1}^{n} \boldsymbol{d}_\ell(n; s-1)}. \tag{6}$$

The probability distribution in (6) reflects the preferential attachment paradigm, since we see that nodes with higher degrees in $\mathfrak{M}(n, s-1)$ 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.

Finally, to obtain the graph of the network that governs the diffusion model in (1), we first generate a multigraph $\mathfrak{M}(N)$ through

[1] With this convention, the (half-)sum of all degrees is the total number of edges and self-loops in the network. This property is particularly useful in controlling the behavior of the denominator in the probability update rule (6).

[2] Actually, the update rule in (5) is undetermined when $s = 1$, since the degree $\boldsymbol{d}_n(n; 0) = \boldsymbol{d}_n(n-1)$ is in principle undefined because node n is not present in $\mathfrak{M}(n-1)$. In order to guarantee symmetry in the multigraph construction, in [30] the customary choice $\boldsymbol{d}_n(n; 0) = 1$ is adopted.

the BR construction, and then obtain the corresponding simple graph $\mathcal{G}(N)$ by uprooting all repeated edges and self-loops in $\mathfrak{M}(N)$.

Once a particular graph $\mathcal{G}(N)$ is obtained, a combination policy must be chosen to determine the combination weights appearing in (1). We focus on the *Laplacian* policy, which is defined as follows, for $k \neq \ell$, and for some parameters $0 < \rho < 1$ and $0 < \lambda \leq 1$ [32]:

$$\boldsymbol{a}_{k\ell}(N) = \frac{\rho\lambda}{\boldsymbol{d}_{\max}(N)}\, \boldsymbol{g}_{k\ell}(N), \qquad \boldsymbol{a}_{kk}(N) = \rho - \sum_{\substack{\ell=1 \\ \ell\neq k}}^{N} \boldsymbol{a}_{k\ell}(N), \tag{7}$$

where $\boldsymbol{g}_{k\ell}(N) \in \{0,1\}$ is the (k,ℓ)-entry of the adjacency matrix $\boldsymbol{G}(N)$ of the *simple* graph $\mathcal{G}(N)$, and $\boldsymbol{d}_{\max}(N)$ is the maximum degree of $\mathcal{G}(N)$. We remark that we added explicit dependence of the combination matrix, $\boldsymbol{A}(N) = [\boldsymbol{a}_{k\ell}(N)]$, on the network size. The Laplacian policy is a popular choice (e.g., in distributed optimization and graph signal processing) which, despite its simplicity, encodes full information on the network graph, and generates a *doubly-stochastic* combination matrix. This property is often desirable for distributed strategies like (1) since (over a strictly connected network) it leads to a uniform Perron eigenvector, implying that all nodes are able to gain equal importance through diffusion, irrespective of their different individual levels of connectivity [32].

3. MAIN RESULT

Whenever the network graph $\mathcal{G}(N)$ is the support graph of the combination matrix $\boldsymbol{A}(N)$ (as in (7)), a natural approach to graph learning is to first estimate the combination matrix, and then retrieve the connected pairs through some thresholding/clustering strategy applied to the matrix entries [6]. Accordingly, we start by seeing how the combination matrix can be estimated.

By evaluating the steady-state ($i \to \infty$) covariance matrix $\boldsymbol{R}_0(N)$ and the one-lag covariance matrix $\boldsymbol{R}_1(N)$ of the signals $\{\boldsymbol{y}_{k,i}\}_{k=1}^{N}$ in (1), it is readily verified that the submatrix of $\boldsymbol{A}(N)$ corresponding to the probed subset $\mathcal{S}$ is given by:

$$\boldsymbol{A}_{\mathcal{S}}(N) = \left[\boldsymbol{R}_1(N)(\boldsymbol{R}_0(N))^{-1}\right]_{\mathcal{S}}. \tag{8}$$

Equation (8) reveals that the combination matrix can be computed from the covariance matrices. This is useful because covariance matrices can be estimated within any precision when the number of samples i grows.[3] However, in our partial observability setting we can only probe the subgraph of nodes in $\mathcal{S}$ and, hence, only the submatrices $[\boldsymbol{R}_0(N)]_{\mathcal{S}}$ and $[\boldsymbol{R}_1(N)]_{\mathcal{S}}$ can be available. Therefore, the matrix $\boldsymbol{R}_1(N)(\boldsymbol{R}_0(N))^{-1}$ in (8) cannot be computed, due to the lack of information about the latent nodes. For this reason, the following surrogate of (8) is considered:

$$\widehat{\boldsymbol{A}}_{\mathcal{S}}(N) \triangleq [\boldsymbol{R}_1(N)]_{\mathcal{S}}\, ([\boldsymbol{R}_0(N)]_{\mathcal{S}})^{-1}. \tag{9}$$

The estimator in (9) clearly differs from the true matrix in (8). However, in the next theorem we show that the error introduced by the presence of latent nodes can be sufficiently small to allow faithful reconstruction of the adjacency matrix of the graph of probed nodes.

[3] We focus on an *achievability* result, i.e., on establishing if graph learning is asymptotically possible as the network grows, and when an arbitrarily large number of samples is available. We do not address the *sample complexity* aspect, i.e., how the number of samples must grow with the network size.

Theorem 1 *There exists a positive random variable*[4] $\boldsymbol{\Gamma}$ *such that the estimator* $\widehat{\mathbf{A}}_{\mathcal{S}}(N)$ *defined in* (9) *satisfies the following properties with high probability as* $N \to \infty$. *Let* $\epsilon > 0$. *For* $k, \ell \in \mathcal{S}$, *if* k *and* ℓ *are connected we have:*

$$(1-\epsilon)\boldsymbol{\Gamma} < \sqrt{N}\,\widehat{\boldsymbol{a}}_{k\ell}(N) < (1+\epsilon)\boldsymbol{\Gamma}, \tag{10}$$

whereas if k *and* ℓ *are unconnected we have:*

$$-\epsilon\,\boldsymbol{\Gamma} < \sqrt{N}\,\widehat{\boldsymbol{a}}_{k\ell}(N) < \epsilon\,\boldsymbol{\Gamma}. \tag{11}$$

Sketch of proof. A known property of BR graphs is that a positive random variable $\boldsymbol{\delta}$ exists such that $\boldsymbol{d}_{\max}(N)$ scales as [33]:

$$\frac{\boldsymbol{d}_{\max}(N)}{\sqrt{N}} \xrightarrow{\text{a.s.}} \boldsymbol{\delta}, \tag{12}$$

where $\xrightarrow{\text{a.s.}}$ denotes almost-sure convergence. Exploiting (7), we conclude that the nonzero entries of the *true* combination matrix fulfill:

$$\sqrt{N}\boldsymbol{a}_{k\ell}(N) = \sqrt{N}\frac{\rho\lambda}{\boldsymbol{d}_{\max}(N)} \xrightarrow{\text{a.s.}} \boldsymbol{\Gamma} \triangleq \frac{\rho\lambda}{\boldsymbol{\delta}}. \tag{13}$$

The next step of the proof consists in showing that the estimated matrix $\widehat{\boldsymbol{A}}_{\mathcal{S}}(N)$ is close to the true matrix with high probability, namely,

$$\sqrt{N}\|\widehat{\boldsymbol{A}}_{\mathcal{S}}(N) - \boldsymbol{A}_{\mathcal{S}}(N)\|_{\text{max-off}} \xrightarrow{\text{p}} 0, \tag{14}$$

where $\|\cdot\|_{\text{max-off}}$ computes the maximum absolute value across the off-diagonal entries of its matrix argument, and $\xrightarrow{\text{p}}$ denotes convergence in probability. The proof of (14) is rather sophisticated and cannot be reported here for space limitations. We deem it useful to illustrate shortly the main idea behind the proof. First, from the symmetry of $\boldsymbol{A}(N)$ we have $\boldsymbol{R}_0(N) = (I - \boldsymbol{A}^2(N))^{-1}$, where I is the $N \times N$ identity matrix. Exploiting the explicit form of $\boldsymbol{R}_0(N)$, it is possible to construct a series of recursive bounds for the error matrix $\widehat{\boldsymbol{A}}_{\mathcal{S}}(N) - \boldsymbol{A}_{\mathcal{S}}(N)$, see the error series provided in [25, Eqs. (157)–(159)]. Then, it can be shown that, to get (14), it is necessary to prove that the following random variable converges to zero in probability:

$$\max_{\substack{k,k'\in\{1,2,\ldots,N\} \\ k\neq k'}} \sum_{\substack{\ell=1 \\ \ell\neq k,k'}}^{N} \boldsymbol{a}_{k\ell}\boldsymbol{a}_{\ell k'}. \tag{15}$$

In [25], this task is accomplished with reference to ER graphs, exploiting the fact that the edges over these graphs are drawn as independent Bernoulli random variables [26, 27]. Over BR graphs, this property is lost, and in particular the node degrees are dependent and non-identically distributed. In order to overcome this difficulty, we appeal to the martingale structure of the degrees and to a maximal martingale inequality [34], which allow us to prove (15), and then (14). Using now (13), (14) and the triangle inequality we get:

$$\|\sqrt{N}\widehat{\boldsymbol{A}}_{\mathcal{S}}(N) - \boldsymbol{\Gamma}\,\boldsymbol{G}_{\mathcal{S}}(N)\|_{\text{max-off}} \xrightarrow{\text{p}} 0. \tag{16}$$

Since $\boldsymbol{\Gamma}$ is positive, Eq. (16) implies, for all $\epsilon > 0$:

$$\lim_{N\to\infty} \mathbb{P}\left(\|\sqrt{N}\widehat{\boldsymbol{A}}_{\mathcal{S}}(N) - \boldsymbol{\Gamma}\,\boldsymbol{G}_{\mathcal{S}}(N)\|_{\text{max-off}} > \epsilon\boldsymbol{\Gamma}\right) = 0, \tag{17}$$

which is equivalent to the claim of the theorem. □

[4] The particular realization of $\boldsymbol{\Gamma}$ arises from the a.s. limit of the maximum degree in (12). As a result, $\boldsymbol{\Gamma}$ does *not* vary across different entries $\boldsymbol{a}_{k\ell}(N)$, whereas it varies across different graph realizations.

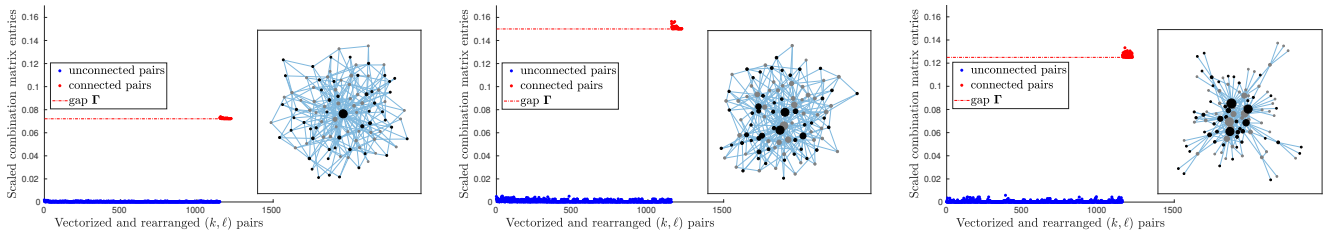

Fig. 2. Illustration of Theorem 1 for three networks of $N = 100$ nodes. The probed subset $\mathcal{S}$ has cardinality $\xi N = 50$, and its nodes (displayed with black circles, while latent nodes are gray) are randomly picked without replacement from $\{1, 2, \ldots, N\}$. The radius of each circle is proportional to the degree of the corresponding node. We show the entries of the estimated matrix $\widehat{\boldsymbol{A}}_{\mathcal{S}}(N)$ (scaled by $\sqrt{N}$, vectorized and rearranged so that the entries corresponding to unconnected nodes come first). The broken line displays the gap $\boldsymbol{\Gamma}$. *Left and middle*: Two realizations of a BR graph. *Right*: Graph of 100 nodes extracted from the real-world network of routers named "*tech-internet-as*", taken from the data repository *networkrepository.com* [35].

The practical significance of Theorem 1 resides in the fact that Eqs. (10) and (11) entail the possibility of *clustering* the entries of the estimated matrix so as to classify connected vs. unconnected pairs. In particular, as suggested in [25], it is possible to devise a variant of the k-means algorithm (with $k = 2$) that: i) finds the stationary points of the optimization problem solved by the k-means algorithm, namely, the configurations fulfilling simultaneously the nearest-neighbor and the centroid conditions; and ii) selects, among the stationary configurations, that with the highest distance between the centroids. By virtue of (10) and (11), for a sufficiently small ϵ this clustering algorithm ends up with correct classification with high probability. We remark that a similar property was shown for ER graphs in [6]. However, over ER graphs the gap separating connected and unconnected entries is deterministic, whereas a fundamental difference arising over BR graphs is the randomness of the limiting gap $\boldsymbol{\Gamma}$. We can show that, despite this additional uncertainty, consistent graph learning can be achieved.

Another important difference between ER and BR graphs is related to the degree growth. Over ER graphs, the maximum degree scales roughly as $1/(Np)$, where p is the edge formation probability. Over BR graphs, the situation is markedly different. First, the minimum degree is upper bounded by η, because, e.g., node N (which is added last) has at most η edges. Second, the maximum degree grows as $\sqrt{N}$. Accordingly, BR graphs are simultaneously *inhomogeneous* (since they contain nodes whose degree blows up with N and nodes with bounded degree) and *sparse* (since even the very connected nodes feature a number of edges that is asymptotically negligible as compared to the N candidate neighbors of each node).

4. ILLUSTRATIVE EXAMPLES

In Fig. 2 we show the true and estimated combination matrix entries corresponding to three network topologies. The first two topologies (left and middle) are independent realizations of a BR graph with $N = 100$ and $\eta = 3$, whereas the third topology (right) is a graph of 100 nodes taken from a real-world network of routers [35]. In all three cases, a Laplacian combination policy with $\rho = 0.5$ and $\lambda = 0.75$ is applied to the pertinent topology, and used to drive the dynamical model (1). For what concerns the BR graphs, we observe the clear emergence of the clustering effect predicted by Theorem 1. Moreover, we see that the different realizations of the BR graph with the same parameters correspond to different values of the gap $\boldsymbol{\Gamma}$, which confirms that this gap is in fact random. Notably, the gap presence and the clustering effect are preserved in the real-world example, which exhibits some of the basic features of BR graphs (e.g., heterogeneous structure with few hubs and many peripheral nodes).

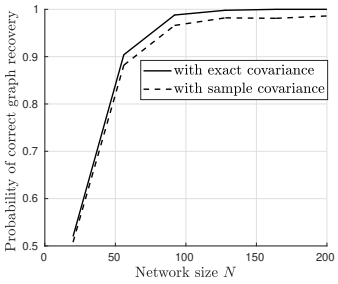

Fig. 3. Probability of correct graph recovery, computed over 10^3 Monte Carlo runs, for different values of the network size N. We consider: the *limiting* estimator obtained by using the true covariances (solid line); and the *empirical* estimator obtained by using the sample covariances evaluated over 10^5 samples (dashed line).

In Fig. 3 the graph learning problem is examined in a more quantitative fashion. Specifically, the dynamical evolution in (1) is simulated over a BR graph (with $\eta = 3$) of increasing size, and a subset of probed nodes with cardinality $\lfloor \xi N \rfloor$ is assumed, with $\xi = 0.15$. For each size, the probability of correct graph learning (using the estimator in (9) and the modified k-means algorithm described in the previous section) is evaluated over 10^3 Monte Carlo runs. In Fig. 3 we report the performance of both the limiting estimator obtained by using the true covariance matrices, and of the empirical estimator obtained by using the sample covariance matrices computed over 10^5 time samples. We see how the probability of correct graph learning approaches 1 as N grows, confirming the validity of our analysis.

5. CONCLUSION

We examined the problem of learning a network graph from the signals diffusing across the network according to (1). The distinguishing features of our work are: i) only part of the network is monitored (*partial observability*); and ii) the topology is modeled as a *preferential attachment* random graph. We establish that graph learning under partial observability is achievable (with high probability as the network grows) for the class of Bollobás-Riordan graphs when the combination weights in (1) follow a Laplacian combination policy.

6. REFERENCES

[1] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Berlin, Germany: Springer, 2005.

[2] V. Matta, V. Bordignon, 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.

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

[4] I. D. Couzin, "Collective cognition in animal groups," *Trends Cognit. Sci.*, vol. 13, no. 1, pp. 36–43, Jan. 2009.

[5] B. L. Partridge, "The structure and function of fish schools," *Sci. Amer.*, vol. 246, no. 6, pp. 114–123, Jan. 1982.

[6] V. Matta, A. Santos, and A. H. Sayed, "Graph learning under partial observability," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 2049–2066, Nov. 2020.

[7] D. Materassi and M. V. Salapaka, "Network reconstruction of dynamical polytrees with unobserved nodes," in *Proc. IEEE Conf. Decis. Control (CDC)*, Maui, HI, USA, Dec. 2012, pp. 4629–4634.

[8] J. Etesami, N. Kiyavash, and T. Coleman, "Learning minimal latent directed information polytrees," *Neural Comput.*, vol. 28, no. 9, pp. 1723–1768, Aug. 2016.

[9] P. Geiger, K. Zhang, B. Schölkopf, M. Gong, and D. Janzing, "Causal inference by identification of vector autoregressive processes with hidden components," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Lille, France, Jul. 2015, vol. 37, pp. 1917–1925.

[10] D. Materassi and M. V. Salapaka, "Identification of network components in presence of unobserved nodes," in *Proc. IEEE Conf. Decis. Control (CDC)*, Osaka, Japan, Dec. 2015, pp. 1563–1568.

[11] A. Anandkumar, V. Y. F. Tan, F. Huang, and A. S. Willsky, "High-dimensional Gaussian graphical model selection: Walk summability and local separation criterion," *J. Mach. Learn. Res.*, vol. 13, pp. 2293–2337, Jan. 2012.

[12] 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.

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

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

[15] 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.

[16] 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.

[17] J. Mei and J. M. F. Moura, "Signal processing on graphs: Causal modeling of *un*structured data," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2077–2092, Apr. 2017.

[18] G. B. Giannakis, Y. Shen, and G. V. Karanikolas, "Topology identification and learning over graphs: Accounting for nonlinearities and dynamics," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 787–807, May 2018.

[19] E. Isufi, A. Loukas, N. Perraudin, and G. Leus, "Forecasting time series with VARMA recursions on graphs," *IEEE Trans. Signal Process.*, vol. 67, no. 18, pp. 4870–4885, Sep. 2019.

[20] V. N. Ioannidis, Y. Shen, and G. B. Giannakis, "Semi-blind inference of topologies and dynamical processes over dynamic graphs," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2263–2274, May 2019.

[21] M. Coutino, E. Isufi, T. Maehara, and G. Leus, "State-space network topology identification from partial observations," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 6, pp. 211–225, 2020.

[22] A. Jalali and S. Sanghavi, "Learning the dependence graph of time series with latent factors," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Scotland, UK, Jun. 2012, pp. 619–626.

[23] 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.

[24] 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.

[25] V. Matta, A. Santos, and A. H. Sayed, "Graph learning over partially observed diffusion networks: Role of degree concentration," *submitted for publication*, available online as arXiv:1904.02963v2 [math.ST], Jun. 2020.

[26] P. Erdős and A. Rényi, "On random graphs I," *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959.

[27] B. Bollobás, *Random Graphs*, Cambridge, UK: Cambridge University Press, 2001.

[28] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.

[29] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, "The degree sequence of a scale-free random graph process," *Random Structures & Algorithms*, vol. 18, no. 3, pp. 279–290, 2001.

[30] B. Bollobás and O. Riordan, "Mathematical results on scale-free random graphs," *Handbook of Graphs and Networks: from the Genome to the Internet*, Hoboken, UK: Wiley, pp. 1–34, 2003.

[31] R. Diestel, *Graph Theory*, Heidelberg, Germany: Springer-Verlag, 2005.

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

[33] R. Van Der Hofstad, *Random Graphs and Complex Networks*, Berlin, Germany: Springer, 2016.

[34] X. Fan, I. Grama, and Q. Liu, "Hoeffding's inequality for supermartingales," *Stoch. Process. Their Appl.*, vol. 122, no. 10, pp. 3545–3559, 2012.

[35] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. AAAI Conference on Artificial Intelligence (AAAI)*, Austin, TX, USA, Jan. 2015, pp. 4292–4293.