LEARNING DYNAMIC GRAPHS UNDER PARTIAL OBSERVABILITY

(Invited paper)

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

[⋆] DIEM, University of Salerno, Fisciano (SA), Italy
[†] EPFL, School of Engineering, Lausanne, Switzerland

ABSTRACT

This work examines the problem of learning a network graph from signals emitted by the network nodes, according to a diffusion model ruled by a Laplacian combination policy. The challenging regime of *partial observability* is considered, where signals are collected from a limited subset of nodes, and we wish to estimate the subgraph of connections between these probed nodes. For the static setting where the network graph is fixed during the estimation process, we examine the sample complexity (number of time samples necessary to achieve consistent learning as the network size grows) of Erdős-Rényi and Bollobás-Riordan graphs. This complexity is almost quadratic for the former and almost linear for the latter class of graphs. We then examine the *dynamic graph* setting where the graph of latent nodes grows over time, while the probed subset remains fixed. We show that in this case the sample complexity *can be reduced*, implying the unexpected conclusion that dynamic graphs might help topology inference under partial observability.

Index Terms— Topology inference, partial observability, sample complexity, dynamic graphs, preferential attachment.

1. INTRODUCTION AND RELATED WORK

Network graphs are useful to describe a significant number of complex systems and phenomena arising in real-world applications, including social learning systems [1, 2], collective cognition in animal groups [3], or "brain networks" [4]. One critical problem in network science is the estimation of the graph topology (i.e., the connections between nodes) from signals collected from the network nodes. This graph learning problem has been addressed in several works before. We refer the reader to [5] for a broad literature survey. Most works consider the *full observability* setting, where signals from all network nodes are available, and the *static* setting, where both the node signals and the graph structure do *not* evolve over time. We consider here the challenging situation where: i) signals from only a limited subset of nodes are collected (*partial observability*); ii) signals at different nodes evolve over time, with a combination matrix ruling the information exchange between neighboring nodes; and iii) the latent (or unobserved) network graph is allowed to grow over time.

For the case of static graphs that do not grow over time, there exist useful works dealing with high-dimensional graphical models with latent variables [6, 7]. However, these works do not consider the time dynamics of the signals emitted by the nodes. For the case where time dynamics is taken into account, there are results under *full observability* [8–12], results on partial observability for graph topologies obeying specific structural conditions [13–17], and results focusing on asymptotic performance guarantees over large and/or random graphs [18–22]. Recent works address the case where the graph is allowed to grow over time [23, 24]. These works do not account for partial observability. Moreover, their focus is to learn the growing graphs and tracking the topology evolution dictated by the new incoming nodes. In comparison, in our work we probe only a partial subset, which is fixed, while the unobserved part of the network grows over time. The evolution of the probed signals depends on the entire network (i.e., on both the probed and latent nodes), and the goal is to estimate the topology linking the probed nodes. Our main results concern the *sample complexity* (i.e., the number of samples necessary to achieve consistent learning as the network size scales to infinity) of a graph learning strategy based on the Granger estimator, which is a popular estimator in the context of causal inference [19–22]. We first characterize the sample complexity associated with two useful models of *random* graphs: Erdős-Rényi (ER) and Bollobás-Riordan (BR) graphs [25–28]. This characterization is provided for the case where the network size is fixed. Then, we present some experiments regarding the *dynamic graph* setting, where the network size is allowed to increase over time. The results are compared with the sample complexity arising in the static case, with some novel and interesting trends emerging.

Notation. We use boldface font to denote random variables. 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 $\mathcal{P} \subset \{1, 2, \ldots, N\}$ is denoted by $A_{\mathcal{P}}$, or alternatively by $[A]_{\mathcal{P}}$. The symbols $\xrightarrow{\text{p}}$ and $\xrightarrow{\text{a.s.}}$ denote convergence in probability and almost-sure convergence, respectively. In both cases the parameter that goes to infinity is the network size (i.e., the number of nodes in the graph).

2. RANDOM GRAPH MODELS

An undirected graph is defined as a set of nodes (or vertices) and a set of edges connecting some pairs of nodes. No self-loops are permitted. A *random* graph with n nodes will be denoted by $\mathcal{G}_n$ (bold font emphasizes randomness). One important graph descriptor is the *degree* of node k, which counts the number of nodes connected to k, that is, the neighbors of k.

We focus on two popular models of *random* graphs: Erdős-Rényi [25] and Bollobás-Riordan [27, 28] random graphs. The former model builds the graph through a sequence of independent Bernoulli experiments: for each pair of nodes, an edge is drawn with probability p, and independently across all pairs of nodes. The resulting graph is *homogeneous*, meaning that, on average, all nodes are treated equally.

In contrast, Bollobás-Riordan random graphs are highly *inhomogeneous*, with central nodes acting as hubs with many neighbors,

This work was supported in part by grant 205121-184999 from the Swiss National Science Foundation (SNSF).

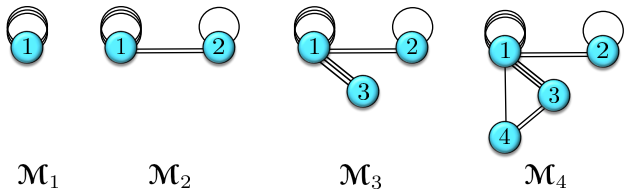

Fig. 1. One example of iterative construction of a Bollobás-Riordan multigraph with parameter $\eta = 3$.

and with peripheral nodes having fewer connections; a dichotomy often arising over real-world networks [29]. Bollobás-Riordan graphs belong to the family of *preferential attachment* graphs [29], and they are built in the following iterative manner. Starting from some initial graph, at each iteration a new node is added, along with a fixed number of edges, say η, linking this node to the current graph. The probability that the new node is connected to an existing node is proportional to the degree of the latter, which explains the terminology "preferential attachment". Note that this procedure follows "*the rich get richer*" philosophy: it promotes connections in favor of nodes that experienced a large amount of connections in previous steps of the iterative construction. This explains why BR graphs will end up being inhomogeneous. Note also that over BR graphs, there is strong dependence in the edge formation process, as opposed to the edge independence characterizing ER graphs. Let us now illustrate more closely the BR construction.

First, this construction focuses on *multigraphs*, namely, graphs where multiple self-loops and multiple edges are permitted. We denote by $\mathfrak{M}_n$ a random multigraph with n nodes. The BR model with parameter $\eta \in \mathbb{N}$ generates iteratively a random sequence of multigraphs as follows — see Fig. 1. The initial multigraph $\mathfrak{M}_1$ is a deterministic multigraph with one node and η self-loops. The multigraph $\mathfrak{M}_n$ is constructed starting from the previous multigraph $\mathfrak{M}_{n-1}$ by adding the fresh node n and η new connections (edges or self-loops). The η new connections are added sequentially. At each step $s = 1, 2, \ldots, \eta$, we build an intermediate multigraph $\mathfrak{M}_{n,s}$ by connecting n to a node randomly chosen from the set $\{1, 2, \ldots, n\}$. We denote by $\boldsymbol{v}_{n,s}$ the particular node that becomes connected to n through the edge introduced at step s, and by $\boldsymbol{d}_{n,s}^{(k)}$ the degree of node k in the *intermediate* multigraph $\mathfrak{M}_{n,s}$. We adopt the convention that the degree of a node in a multigraph is the number of edges connected to the node plus *twice* the number of its self-loops [30]. The description of the multigraph construction is completed by assigning the probability that a particular node $k \in \{1, 2, \ldots, n\}$ is picked:

$$\mathbb{P}\left[\boldsymbol{v}_{n,s} = k | \mathfrak{M}_{n,s-1}\right] = \begin{cases} \dfrac{\boldsymbol{d}_{n,s-1}^{(k)}}{1 + \sum_{\ell=1}^{n} \boldsymbol{d}_{n,s-1}^{(\ell)}}, & k \neq n, \\ \\ \dfrac{1 + \boldsymbol{d}_{n,s-1}^{(n)}}{1 + \sum_{\ell=1}^{n} \boldsymbol{d}_{n,s-1}^{(\ell)}}, & k = n. \end{cases} \tag{1}$$

Rule (1) implements the preferential attachment concept, since the new node n is more likely to be connected to nodes featuring higher degrees in $\mathfrak{M}_{n;s-1}$. The final Bollobás-Riordan graph $\mathcal{G}_n$ is obtained from the multigraph $\mathfrak{M}_n$ by simply removing all self-loops and by replacing multiple edges with a single edge [27, 28].

2.1. Maximum Degrees

A graph descriptor that plays an important role in our treatment is the *maximum degree*. The maximum degree over a graph $\mathcal{G}_n$ is:

$$\boldsymbol{\mu}_n \triangleq \max_{k=1,2,\ldots n} \boldsymbol{d}_n^{(k)} \qquad \text{[maximum degree]}. \tag{2}$$

In particular, it is of interest to characterize how the maximum degree scales asymptotically as $n \to \infty$. This scaling law can be precisely characterized for both ER graphs [5, 21, 30] and BR graphs [22, 31]. We have that:

$$\frac{\boldsymbol{\mu}_n}{np} \xrightarrow{\text{P}} 1 \quad \text{[ER graphs]}, \qquad \frac{\boldsymbol{\mu}_n}{\sqrt{n}} \xrightarrow{\text{a.s.}} \boldsymbol{\mu} \quad \text{[BR graphs]}, \tag{3}$$

where we recall that p is the connection probability of the ER graph, and where $\boldsymbol{\mu}$ is a certain *positive* random variable [28].

3. STATIC GRAPHS

In this section we illustrate the static setting where the network size is fixed beforehand. Denoting this fixed size by N, and given an underlying graph $\mathcal{G}_N$, the network nodes emit signals that evolve according to a diffusion model, a.k.a. first-order vector autoregressive (VAR) model [32]. More precisely, each node $k = 1, 2, \ldots, N$, at time $t = 1, 2, \ldots$, is driven by a stochastic input source $\boldsymbol{x}_{k,t}$ and produces the output signal $\boldsymbol{y}_{k,t}$ according to the linear model:

$$\boldsymbol{y}_{k,t} = \sum_{\ell=1}^{N} \boldsymbol{a}_{k\ell}\, \boldsymbol{y}_{\ell,t-1} + \boldsymbol{x}_{k,t} \quad \Leftrightarrow \quad \boldsymbol{y}_t = \boldsymbol{A}\, \boldsymbol{y}_{t-1} + \boldsymbol{x}_t, \tag{4}$$

where the $N \times 1$ vectors $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ collect the entries $\boldsymbol{x}_{k,t}$ and $\boldsymbol{y}_{k,t}$, and where $\boldsymbol{A} = [\boldsymbol{a}_{k\ell}]$ is a *Schur stable matrix* collecting the nonnegative combination weights $\boldsymbol{a}_{k\ell}$. We stick to a popular choice in distributed optimization and learning, namely, the *Laplacian* combination policy (so named since it is related to the graph Laplacian [33, 34]), which is defined as follows. For some positive parameters $\rho < 1$ and $\lambda \leq 1$, and for $k \neq \ell$:

$$\begin{cases} \boldsymbol{a}_{k\ell} = 0, & (k, \ell) \text{ unconnected over } \mathcal{G}_N, \\ \boldsymbol{a}_{k\ell} = \dfrac{\rho\lambda}{1 + \boldsymbol{\mu}_N}, & (k, \ell) \text{ connected over } \mathcal{G}_N, \end{cases} \tag{5}$$

with $\boldsymbol{a}_{kk} = \rho - \sum_{\ell \neq k} \boldsymbol{a}_{k\ell}$. We see that the combination weights in (5) reflect the network topology, with null weights corresponding to unconnected node pairs, and positive weights to connected pairs. Note also that the positive weights are inversely proportional to (1 plus) the maximum degree of the graph.

Once a graph realization is fixed, the combination matrix becomes deterministic, and the system in (4) evolves according to the randomness of the initial state $\boldsymbol{y}_0$ and the input signals $\boldsymbol{x}_{k,t}$. The latter random variables have zero mean and unit variance, are independent and identically distributed (i.i.d.) both spatially (index k) and temporally (index t), and they are also independent of the graph. The initial vector $\boldsymbol{y}_0$ has finite second moment, is independent of signals $\boldsymbol{x}_{k,t}$, but it can depend on $\mathcal{G}_N$.

3.1. Partial Observability

In several applications of practical interest, it is seldom possible to collect signals from all nodes, whereas it is more likely to have access only to a subset of nodes. Under this regime of *partial observability*, given an observation time interval $t = 1, 2, \ldots, T$, and a probed subset $\mathcal{P}$, the collection of available signals is:

$$\boldsymbol{Y}_{\mathcal{P}}(T, N) \triangleq \{\, \boldsymbol{y}_{k,t} : k \in \mathcal{P}, t = 1, 2, \ldots, T \,\}. \tag{6}$$

In this work we focus on the following *graph learning* problem under partial observability. We want to estimate the topology of the partial graph $\mathcal{G}_{\mathcal{P}}$ (that is, the subgraph of nodes belonging to $\mathcal{P}$) starting from the collection $\boldsymbol{Y}_{\mathcal{P}}(T,N)$. In the *static* setting, the graph has fixed dimension N from the beginning of the time interval. In order to assess the performance of a graph estimator, one standard approach is to consider the asymptotic regime of large network sizes. Under this regime, it is necessary to establish how the number of samples $T = T_N$ must scale with the network size to provide faithful graph learning. The scaling law followed by T_N characterizes the so-called *sample complexity* of the estimator. In other words, the value T_N quantifies how many samples are needed to achieve good performance for a fixed network size N.

3.2. Granger Estimator

We introduce the covariance matrix and the one-lag covariance matrix corresponding to model (4), which are, respectively:

$$\boldsymbol{R}_0(t) = \mathbb{E}\left[\boldsymbol{y}_t\boldsymbol{y}_t^\top \middle| \boldsymbol{A}\right], \qquad \boldsymbol{R}_1(t) = \mathbb{E}\left[\boldsymbol{y}_t\boldsymbol{y}_{t-1}^\top \middle| \boldsymbol{A}\right]. \quad (7)$$

The bold notation for the covariance matrices is used because of the randomness of the underlying combination matrix $\boldsymbol{A}$. Exploiting (4), it is straightforward to show that [32]:

$$\boldsymbol{R}_1(t) = \boldsymbol{A}\boldsymbol{R}_0(t-1) \Rightarrow \boldsymbol{A}_{\mathcal{P}} = [\boldsymbol{R}_1(t)\boldsymbol{R}_0^{-1}(t-1)]_{\mathcal{P}}. \quad (8)$$

Unfortunately, under partial observability we can compute the covariance matrices only relative to the *probed subset*, which prevents the computation of the inversion formula in (8), owing to unavailability of signals from the *latent* nodes. It is nevertheless meaningful to compute the following *Granger estimator or predictor*:

$$\widehat{\boldsymbol{A}}_{\mathcal{P}}(t) = [\boldsymbol{R}_1(t)]_{\mathcal{P}}([\boldsymbol{R}_0(t-1)]_{\mathcal{P}})^{-1}, \quad (9)$$

which provides the best linear prediction of the future samples from the past one-lag samples collected *over the probed subset*. In the static case, the covariance matrices converge as $t \to \infty$, such that we can introduce the estimator:

$$\widehat{\boldsymbol{A}}_{\mathcal{P}} = [\boldsymbol{R}_1]_{\mathcal{P}}([\boldsymbol{R}_0]_{\mathcal{P}})^{-1}, \quad (10)$$

where $\boldsymbol{R}_j = \lim_{t\to\infty} \boldsymbol{R}_j(t)$, for $j = 0, 1$. Actually, to estimate the topology of the subgraph on $\mathcal{P}$ from the data, we will employ the *sample* Granger estimator:

$$\widehat{\boldsymbol{A}}_{\mathcal{P}}(T,N) = \left[\widehat{\boldsymbol{R}}_1(T,N)\right]_{\mathcal{P}}\left(\left[\widehat{\boldsymbol{R}}_0(T,N)\right]_{\mathcal{P}}\right)^{-1}, \quad (11)$$

which replaces the true covariance matrices used in (10) with the sample covariance matrices that we denote by $\widehat{\boldsymbol{R}}_0(T,N)$ and $\widehat{\boldsymbol{R}}_1(T,N)$. It is also useful to consider the following *regularized* version of (11). For $k \in \mathcal{P}$, the k-th row of the matrix estimator $\widehat{\boldsymbol{A}}_{\mathcal{P}}(T,N)$ is a solution to the following optimization problem:

$$\min_x \left\| x\left[\widehat{\boldsymbol{R}}_0(T,N)\right]_{\mathcal{P}} - \left[\widehat{\boldsymbol{R}}_1(T,N)\right]_{k\mathcal{P}}\right\|_\infty, \text{subject to } \|x\|_1 \le 1, \quad (12)$$

where $x \in \mathbb{R}^{|\mathcal{P}|}$ is a row vector, and, for a matrix M, the notation $[M]_{k\mathcal{P}}$ denotes the k-th row of the submatrix $M_{\mathcal{P}}$. Note that, when the non-regularized Granger estimator (11) exists (i.e., when the sample covariance is invertible) and fulfills the bounded-norm constraint in (12), it coincides with the regularized estimator.

Once we have a matrix estimator, the underlying graph on $\mathcal{P}$ can be estimated by means of a clustering algorithm that separates the matrix-estimator entries in two classes (connected/unconnected). Useful algorithms are the classical k-means algorithm [35] (run with $k = 2$) and one variation thereof proposed in [21] which selects, among the possible stationary points of the k-means algorithm, the one that maximizes the distance between clusters' centroids.

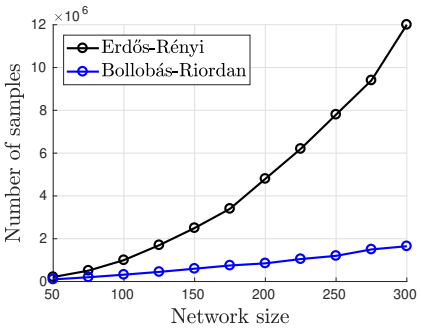

Fig. 2. Sample complexity of ER and BR graphs. We use the sample Granger estimator (11), followed by the modified k-means algorithm proposed in [21]. The curves represent the number of samples needed to attain 90% of the probability of correct graph learning achieved by the ideal estimator in (10) (which uses the exact covariance matrices, i.e., it would need an infinite number of samples). The ER graphs have connection probability $p = 0.5$. The BR graphs have parameter $\eta = 3$. The number of Monte Carlo runs is 10^3. For the Laplacian combination policy we use $\rho = 0.5$ and $\lambda = 0.75$.

It is possible to characterize the sample complexity of the Granger estimator when the underlying graph is Erdős-Rényi or Bollobás-Riordan, as stated in the next theorem, whose proof is omitted for space constraints. Details for the ER and BR models can be found in [21] and [22], respectively.

Theorem 1 (Sample Complexity for ER and BR Graphs). *Consider the diffusion model* (4) *when the input source is Gaussian and the initial state (conditioned on a realization of $\mathcal{G}_N$) is distributed according to the stationary Gaussian distribution of* (4)*. Let the cardinality of the probed subset scale with N such that $|\mathcal{P}|/N \to \xi > 0$. Consider the following sample-complexity laws:*

$$T_N \propto N^2 \log N \ \text{ [ER graphs]}, \quad T_N = \omega_N N \log N \ \text{ [BR graphs]}, \quad (13)$$

where ω_N is a positive sequence diverging in an arbitrary fashion. Then, the regularized Granger estimator in (12)*, followed by the modified k-means algorithm proposed in [21], learns the subgraph of probed nodes with probability converging to* 1 *as $N \to \infty$.* ■

In Fig. 2, we evaluate empirically the number of samples needed to attain a probability of correct learning equal to 90% of the probability achieved by the ideal (i.e., with an infinite number of samples) Granger estimator in (10). We see that the displayed scaling laws match well the predictions of Theorem 1: the sample complexity is almost quadratic for ER graphs, and almost linear for BR graphs.

The main reason behind this behavior is tightly coupled with the behavior of the maximum degree of the graph. According to the Laplacian matrix in (5), the growth of the maximum degree determines the way the nonzero entries of the matrix vanish as $N \to \infty$. The smaller these nonzero entries are, the higher the precision that is necessary for the estimators to distinguish the nonzero entries from the zero entries. As a result, a faster increase of the maximum degree corresponds to an increase of the number of samples. This argument is made rigorous in the proof of Theorem 1 to show that the sample

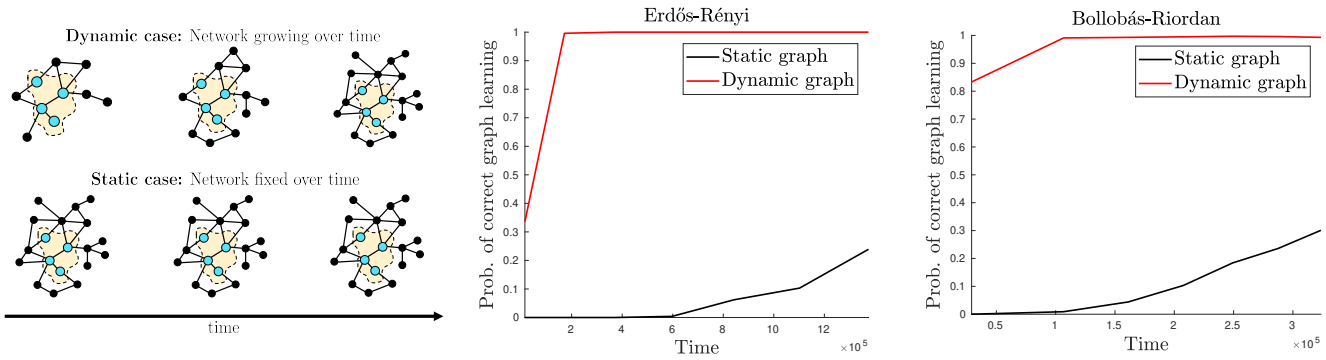

Fig. 3. *Left plot*. Dynamic versus static graphs. *Middle plot*. ER graphs with connection probability $p = 0.5$. In the dynamic case, the network size scales as $N_t \sim t^{4/5}$. *Right plot*. BR graphs with parameter $\eta = 3$. In the dynamic case, the network size scales as $N_t \sim t^{3/2}$. In both experiments: we use the sample Granger estimator (11), followed by the modified k-means algorithm proposed in [21]; the probed subset is $\{1, 2, \ldots, 10\}$; the parameters of the Laplacian combination matrix are $\lambda = 0.75$ and $\rho = 0.5$; and we use 10^3 Monte Carlo runs.

complexity goes (up to a $\log N$ factor arising from a union bound on the cardinality of $\mathcal{P}$) as the square of the maximum degree and, hence, quadratically in the ER case, and linearly in the BR case.

4. DYNAMIC GRAPHS

We now consider the dynamic setting where the latent graph grows as time t progresses. Specifically, we assume the network size increases according to some law N_t, and modify (4) to:

$$\boldsymbol{y}_t = \boldsymbol{A}_t \, \boldsymbol{y}_{t-1} + \boldsymbol{x}_t, \tag{14}$$

where the combination matrix $\boldsymbol{A}_t$ is obtained by using (5) over the *dynamic* graph $\mathcal{G}_{N_t}$. For example, when we consider a BR model, $\mathcal{G}_{N_t}$ can be constructed according to the preferential attachment rule described in Sec. 2, where a new node is added at each time instant when N_t increases by 1. We assume that the probed subset $\mathcal{P}$ is fixed, whereas the graph involving the latent nodes (including connections between latent and probed nodes) grows over time. The critical difference that distinguishes the dynamic setting from the static setting considered in Sec. 3 is illustrated in the left panel of Fig. 3, where probed nodes are displayed in cyan, while latent nodes in black. In the static setting (bottom diagram in the left panel), a fixed graph underlies the diffusion process for the *entire observation interval* during which topology inference is performed. In contrast, in the dynamic setting (top diagram) the latent graph grows over time.

We consider graph growths of the form $N_t = N_0 + \alpha \, t^{\beta}$, where N_0 is some initial graph size. Note that there is a connection between the graph growth N_t and the sample complexity. Indeed, saying that the network size at time t is $N_t \sim t^{\beta}$ means that the number of samples employed to estimate a graph of size N is $T_N \sim N^{1/\beta}$. Therefore, the following remarkable coupling between sample complexity and graph growth emerges in the dynamic setting. If a certain *minimum* sample complexity T_N is necessary to learn faithfully, this means that a *maximum* graph growth is permitted. In other words, *sample complexity places a limit on the maximum allowable velocity at which the dynamic graph can grow over time.* Building on the results available from Theorem 1, we would expect the following graph growth laws:

$$N_t \sim t^{1/2} \;\; \text{[ER graphs]}, \qquad N_t \sim t \;\; \text{[BR graphs]}. \tag{15}$$

In Fig. 3 we test instead the following alternative laws:

$$N_t \sim t^{4/5} \;\; \text{[ER graphs]}, \qquad N_t \sim t^{3/2} \;\; \text{[BR graphs]}. \tag{16}$$

The exponents $4/5$ and $3/2$ are based on numerical experiments, as there is currently no counterpart of Theorem 1 available for the dynamic case. Note that in (16) we increase the velocity at which the graph grows, for both the ER and the BR models. As was mentioned, increasing the velocity corresponds to reducing the number of samples, which, inverting the relations in (16), is $T_N \sim N^{5/4} < N^2$ for ER graphs and $T_N \sim N^{2/3} < N$ for BR graphs. Remarkably, the middle and right plots in Fig. 3 reveal that, despite the increased velocity (i.e., the reduced number of samples), the graph learning problem becomes feasible. In contrast, in the static case (where the graph has constant size $N = 200$, which is the size corresponding to the end of the observation window for the dynamic case), the performance is not good since we are violating the prescriptions of Theorem 1. It is worth reporting that we have also tested the *directed* counterparts of ER graphs (a.k.a. *binomial graphs*) and of BR graphs,[1] obtaining results similar to those shown in Fig. 3.

One conclusion arising from the obtained results is that, under partial observability, application of the Granger estimator – in our experiments we implemented the sample version in (11) – over dynamic graphs can deliver superior performance as compared to the static case. This is a remarkable and perhaps unexpected behavior. It is possible to provide an interpretation of this behavior based on Theorem 1. Even though this theorem characterizes only the static case, its proof reveals that the main factor determining the sample complexity is the magnitude of the nonzero entries in the combination matrix: the smaller they are, the higher the sample complexity will be. On the other hand, the nonzero entries are inversely proportional to the maximum degree of the graph, which increases with the network size, leading to an increase in sample complexity. Under a static model, the system works during the entire observation interval with the largest graph. In contrast, under the dynamic model the system works with growing graphs and, hence, on average the network size is smaller (i.e., more favorable) than the size considered in the static case. This is one reason why the dynamic case looks less demanding in terms of samples, ultimately implying that a faster growth is permitted for the sequence of dynamic graphs.

[1] In a binomial graph, directed edges corresponding to pairs (k, ℓ) and (ℓ, k) are drawn independently. Directed BR graphs can be constructed by considering preferential-attachment probabilities based on in-degrees and out-degrees to build directed edges.

5. REFERENCES

[1] V. Bordignon, V. Matta, and A. H. Sayed, "Interplay between topology and social learning over weak graphs," *IEEE Open J. Signal Process.*, vol. 1, pp. 99–119, 2020.

[2] V. Bordignon, V. Matta, and A. H. Sayed, "Adaptive social learning," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, pp. 6053–6081, Sep. 2021.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

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

[21] 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, Jul. 2022.

[22] M. Cirillo, V. Matta, and A. H. Sayed, "Estimating the topology of preferential attachment graphs under partial observability," *IEEE Trans. Inf. Theory*, 2022, doi: 10.1109/TIT.2022.3211078.

[23] B. Das, A. Hanjalic, and E. Isufi, "Task-aware connectivity learning for incoming nodes over growing graphs," *IEEE Trans. Signal Inf. Process. Netw.*, 2022, doi: 10.1109/TSIPN.2022.3206578.

[24] A. Natali, E. Isufi, M. Coutino, and G. Leus "Learning time-varying graphs from online data," *IEEE Open J. Signal Process.*, vol. 3, pp. 212–228, May 2022.

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

[26] B. Bollobás, *Random Graphs*, Cambridge University Press, 2001.

[27] 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.

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

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

[30] R. Diestel, *Graph Theory*, Springer-Verlag, 2005.

[31] R. Van Der Hofstad, *Random Graphs and Complex Networks*, Springer, 2016.

[32] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer, 2005.

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

[34] U. A. Khan, W. U. Bajwa, A. Nedić, M. G. Rabbat, and A. H. Sayed, *Editors*, "Optimization for data-driven learning and control," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 1863–1868, Nov. 2020

[35] A. H. Sayed, *Inference and Learning from Data*, Cambridge University Press, 2022.