Tomography of Large Adaptive Networks under the Dense Latent Regime

(Invited Paper)

Vincenzo Matta†, Augusto Santos⋆, and Ali H. Sayed⋆

***Abstract*—This work examines the problem of graph learning over a diffusion network when measurements can only be gathered from a limited fraction of agents (*latent* regime). Under this setting, most works in the literature rely on a degree of sparsity to provide guarantees of consistent graph recovery. This work moves away from this condition and shows that, even under *dense* connectivity, the Granger estimator ensures an identifiability gap that enables the discrimination between connected and disconnected nodes within the observable subnetwork.**

***Index Terms*—Graph learning, dense networks, Granger estimator, diffusion network, identifiability gap.**

I. Introduction

Learning the graph structure of a network from measurements at the agents is a problem of paramount importance. Applications include revealing relationships between functional connectivity (the signals) and structural connectivity (the topology) in the brain; understanding the role of topology for the diffusion of information over social networks; and modeling the exchange of information over biological networks.

The network tomography problem is challenging for at least three reasons: i) direct access to the data exchanged between nodes is often impractical, and the inference about node interconnections must be based on some *indirect* observations; ii) observations from only a *limited subset of the network* are usually available (*local* tomography); iii) and the gathered signals are influenced not only by the observed nodes, but also by the latent (i.e., unobserved) nodes because information diffuses across the network. Figure 1 illustrates the proposed local tomography paradigm. Data collected from a subnetwork are used to estimate a partial combination matrix. The entries of this matrix are then subjected to a clustering procedure to decide on whether agents are connected or not.

There have been several works in the literature that deal with the network tomography problem. We comment briefly on their features and on the contribution of our work in relation to these earlier efforts. Reference [1] provides a more detailed commentary on the pertinent literature. We note that the majority of existing works focus on linear and autoregressive diffusion models [2]–[4]. When observations are available from all nodes, causal graph processes are used in [4], while optimization methods based on structural constraints such sparsity are proposed in [3], [5]. Under the partial-observation formulation, there are results holding for particular graphs, e.g., for polytrees [6], as well as results holding under particular conditions (on the network structure and/or on the statistical model) that are usually impractical in large-scale network settings [2], [7]. In comparison, there are works specifically tailored to the large-scale network setting, where identifiability relies on average *macroscopic* indicators (e.g., node connection probability). In the latter category there are works dealing with graphical models [8], as well as works dealing with diffusion networks [1], [9]–[11].

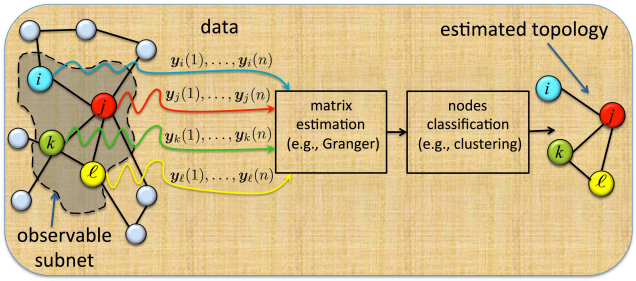

Fig. 1. Illustration of the local tomography problem.

However, the majority of these earlier contributions rely on a sparsity condition in the network structure to ensure consistent tomography. In this work, we move away from this condition and show that consistent tomography is still possible over densely connected networks with performance guarantees. More specifically, we will establish identifiability of the observable subnetwork for *dense* Erdös-Rényi random graphs and for the class of regular diffusion matrices described by Assumption 1 further ahead. The main observation is that a *fortunate coupling between the Erdös-Rényi model and the regular diffusion matrices renders the problem of topology inference a local problem*. The analysis will reveal that sparsity is not necessarily the key-enabler for local tomography. Instead, the main enabling feature will be seen to be the error concentration induced by the aforementioned coupling.

Notation. Boldface letters denote random variables. Sets are represented with calligraphy fonts, set cardinalities with the corresponding normal fonts, and set complements with the superscript $'$. Given a matrix Z, the submatrix that lies in the rows and columns indexed by $\mathcal{S}$ is denoted by $Z_{\mathcal{S}}$, and the entries of $Z_{\mathcal{S}}$ are indexed by indices (i,j) spanning $\mathcal{S}$. The (i,j)-th entry of the matrix power Z^k is denoted by $z_{ij}^{(k)}$.

† V. Matta is with DIEM, University of Salerno, via Giovanni Paolo II, I-84084, Fisciano (SA), Italy (email: vmatta@unisa.it).

⋆ A. Santos (email: augusto.santos@epfl.ch) and A. H. Sayed (email: ali.sayed@epfl.ch) are with École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. The work of A. H. Sayed was also supported in part by NSF grant CCF-1524250.

II. Background

A. Useful Facts about Random Graphs

A *random* graph of N nodes, where edges are drawn independently and with identical probability p, is an Erdös-Rényi graph and is denoted by $\boldsymbol{G} \sim \mathscr{G}(N,p)$. In our treatment, to rule out trivial/pathological cases, we assume that $0 < p < 1$. In the following, we use the same symbol $\boldsymbol{G}$ to denote also the adjacency matrix of the graph, with $\boldsymbol{g}_{ij} = 1$ if nodes i and j are connected and $\boldsymbol{g}_{ij} = 0$ otherwise. We assume also that each node "stays connected to itself," namely, each node has (deterministically) a self-loop (i.e., $\boldsymbol{g}_{ii} = 1$).

One classic way to characterize random graphs is to examine their behavior as N gets large (*graph evolution*), in order to capture average behavior patterns that emerge with high probability over large networks. In the most general case, the connection probability is allowed to scale with N, and can be also vanishing with N.[1] In this work, we restrict attention to the regime of evolution where the connection probability stays constant when N gets large (*dense* regime).

One useful graph descriptor is the *degree* of a node, which is defined as the number of neighbors of that node (including the node itself). The degree of node i will be denoted by $\boldsymbol{d}_i$. It is readily verified that the random variable $\boldsymbol{d}_i - 1$ is distributed as a binomial random variable over $N-1$ trials (the $N-1$ edges possibly stemming from node i) with success probability p. We have subtracted 1 from $\boldsymbol{d}_i$ because, according to our definition, $\boldsymbol{d}_i$ counts node i itself. In our setting, it would be particularly useful to focus on the minimum and maximum degrees, which will be denoted by $\boldsymbol{d}_{\min}$ and $\boldsymbol{d}_{\max}$, respectively. In particular, the following scaling laws (arising from standard application of the Chernoff bounding technique) will play a fundamental role in the forthcoming analysis:

$$\boxed{\frac{\boldsymbol{d}_{\min}}{Np} \xrightarrow{\text{p}} 1, \qquad \frac{\boldsymbol{d}_{\max}}{Np} \xrightarrow{\text{p}} 1} \tag{1}$$

where $\xrightarrow{\text{p}}$ denotes convergence in probability as $N \to \infty$.

B. Network and Data Model

We assume that the state of each agent evolves over time as a result of the interaction with its neighbors. Formally, the state $\boldsymbol{y}_i(n)$ of each agent $i \in \{1,2,\ldots,N\}$, at time n, is assumed to obey the following stochastic dynamical system:

$$\boxed{\boldsymbol{y}_i(n) = \sum_{\ell=1}^{N} a_{i\ell}\, \boldsymbol{y}_\ell(n-1) + \boldsymbol{x}_i(n) \Leftrightarrow \boldsymbol{y}_n = A\, \boldsymbol{y}_{n-1} + \boldsymbol{x}_n} \tag{2}$$

where $\boldsymbol{y}_n = [\boldsymbol{y}_1(n), \boldsymbol{y}_2(n), \ldots, \boldsymbol{y}_N(n)]$ is a column vector collecting the states of all nodes at time n, $A = [a_{i\ell}]$ is a *combination matrix*, and $\boldsymbol{x}_n = [\boldsymbol{x}_1(n), \boldsymbol{x}_2(n), \ldots, \boldsymbol{x}_N(n)]$ is a column vector modeling a random input at the agents (e.g., streaming data or noise). We assume that $\{\boldsymbol{x}_i(n)\}$ are independent and identically distributed both spatially (i.e., w.r.t. to i) and temporally (i.e., w.r.t. to n), with zero-mean and unit variance. We observe from (2) that, if $a_{i\ell} = 0$, then agent i does not use the information arriving from agent ℓ to update its own state. Therefore, the *support-graph* of A reflects the underlying topology. The stochastic dynamical system (2) arises naturally in the context of *adaptive diffusion networks* [1], [13]. It also arises in economics and is used as a variational characterization of nonlinear dynamical systems.

We observe that the dynamical system in (2) implies the following relationship among the correlation matrix $R_0(n) \triangleq \mathbb{E}[\boldsymbol{y}_n \boldsymbol{y}_n^T]$, the one-lag correlation matrix $R_1(n) \triangleq \mathbb{E}[\boldsymbol{y}_n \boldsymbol{y}_{n-1}^T]$, and the combination matrix A:

$$R_1(n) = AR_0(n-1) \overset{n\to\infty}{\longrightarrow} R_1 = AR_0, \tag{3}$$

where R_0 and R_1 are the limiting correlation matrices (assuming A stable). Therefore, since there exist many ways to estimate R_0 and R_1 consistently as $n \to \infty$, the relationship $A = R_1 R_0^{-1}$ reveals one possible strategy to estimate A from the output of the diffusion process, $\boldsymbol{y}_n$. Such a strategy is sometimes referred to as Granger estimator [2].

Under the considered large-scale network setting, only a subset $\mathcal{S}$ of the network is accessible. For this case, the combination matrix pertaining to the observed subset $\mathcal{S}$ of nodes is $A_{\mathcal{S}} = \left[R_1 R_0^{-1}\right]_{\mathcal{S}}$, which cannot be computed in the framework of local tomography as only the submatrices associated with the observable agents, $[R_0]_{\mathcal{S}}$ and $[R_1]_{\mathcal{S}}$, are available (or can be estimated). One could certainly consider an approximation for $A_{\mathcal{S}}$ as:

$$\boxed{\widehat{A}_{\mathcal{S}} = [R_1]_{\mathcal{S}} \left([R_0]_{\mathcal{S}}\right)^{-1}} \tag{4}$$

The estimator $\widehat{A}_{\mathcal{S}}$ corresponds to applying the Granger estimator at the observable nodes $\mathcal{S}$ (i.e., ignoring the latent part). It is clear that $\widehat{A}_{\mathcal{S}} \neq A_{\mathcal{S}}$. However, it was established in [1], [9] that $\widehat{A}_{\mathcal{S}}$ contains sufficient information to retrieve the support graph of $A_{\mathcal{S}}$, with high probability, under certain sparsity constraints on the network structure and stability conditions on the combination matrix A. This work brings a new contribution: it establishes structural consistency of the Granger estimator in (4) under the more demanding regime of dense connectivity.

C. The Random Graph and the Combination Rule

Let $\boldsymbol{G} \sim \mathscr{G}(N,p)$ be the Erdös-Rényi random graph linking the N agents. We can assign (nonnegative) weights to the edges of $\boldsymbol{G}$ and denote the resulting matrix of weights as $\boldsymbol{A}$ (where the randomness of $\boldsymbol{A}$ follows from the randomness of $\boldsymbol{G}$). Some popular choices are the Laplacian and the Metropolis rules, defined as follows.

Laplacian rule. For the Laplacian rule, the combination-matrix entries are:

$$\boldsymbol{a}_{ij} = \begin{cases} \rho\, \lambda\, \dfrac{\boldsymbol{g}_{ij}}{\boldsymbol{d}_{\max}}, & \text{for } i \neq j \\ \rho - \sum_{\ell \neq i} a_{i\ell}, & \text{for } i = j \end{cases} \tag{5}$$

with $0 < \rho < 1$ and $0 < \lambda \leq 1$.

[1] A graph can be connected in the sparse regime. For example, it is well known that, if $p_N = \frac{\log N + c_N}{N}$, with $c_N \to \infty$, then the graph is connected with high probability even if c_N is such that $p_N \to 0$ [12].

Metropolis rule. For the Metropolis rule, we have instead:

$$\boldsymbol{a}_{ij} = \begin{cases} \rho \dfrac{\boldsymbol{g}_{ij}}{\max\{\boldsymbol{d}_i, \boldsymbol{d}_j\}}, & \text{for } i \neq j \\ \rho - \sum_{\ell \neq i} a_{i\ell}, & \text{for } i = j \end{cases} \tag{6}$$

These rules arise naturally in many applications, for instance, they are one fundamental ingredient of *adaptive* networks [13]–[15].

III. Main Result

In the following treatment, the overall network of agents is denoted by $\mathcal{N} = \{1, 2, \ldots, N\}$. We start by introducing a useful class of combination matrices.

Assumption 1 (Regular diffusion matrices). *We assume that the combination matrix $\boldsymbol{A}$ is symmetric with entries satisfying:*

$$\sum_{\ell=1}^{N} \boldsymbol{a}_{i\ell} = \rho, \qquad 0 < \rho < 1, \tag{7}$$

and $\forall i \neq j$:

$$\boxed{\frac{\kappa}{\boldsymbol{d}_{\max}} \boldsymbol{g}_{ij} \leq \boldsymbol{a}_{ij} \leq \frac{\kappa}{\boldsymbol{d}_{\min}} \boldsymbol{g}_{ij}} \tag{8}$$

for some $0 < \kappa \leq 1$. □

The relevance of this class of matrices stems from the fact that the most common combination matrices encountered in the literature automatically satisfy Assumption 1. For instance, it is readily seen that the Laplacian combination matrix in (5) is a regular diffusion matrix with parameters ρ and $\kappa = \rho\lambda$, whereas the Metropolis combination matrix in (6) is a regular diffusion matrix with parameters ρ and $\kappa = \rho$.

In order to ascertain whether or not it is possible to discriminate interacting (i.e., connected) agents from non-interacting agents via inspection of an estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}}$, we introduce the concept of margin and identifiability gap.

Definition 1 (Margins). *Let $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ be a certain estimated combination matrix, corresponding to the subnetwork $\mathcal{S}$. The lower and upper margins corresponding to the* disconnected *pairs are defined as, respectively*[2]*:*

$$\underline{\boldsymbol{\delta}}_N \triangleq \min_{\substack{i,j \in \mathcal{S}: \boldsymbol{a}_{ij} = 0 \\ i \neq j}} \widehat{\boldsymbol{a}}_{ij}, \quad \overline{\boldsymbol{\delta}}_N \triangleq \max_{\substack{i,j \in \mathcal{S}: \boldsymbol{a}_{ij} = 0 \\ i \neq j}} \widehat{\boldsymbol{a}}_{ij}. \tag{9}$$

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

$$\underline{\boldsymbol{\Delta}}_N \triangleq \min_{\substack{i,j \in \mathcal{S}: \boldsymbol{a}_{ij} > 0 \\ i \neq j}} \widehat{\boldsymbol{a}}_{ij}, \quad \overline{\boldsymbol{\Delta}}_N \triangleq \max_{\substack{i,j \in \mathcal{S}: \boldsymbol{a}_{ij} > 0 \\ i \neq j}} \widehat{\boldsymbol{a}}_{ij}. \tag{10}$$

□

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

[2]The definitions in (9) and (10) are void if the nodes in $\mathcal{S}$ are all connected or all disconnected, respectively. Since these events are irrelevant as $N \to \infty$, for these singular cases we can formally assign arbitrary values to the margins.

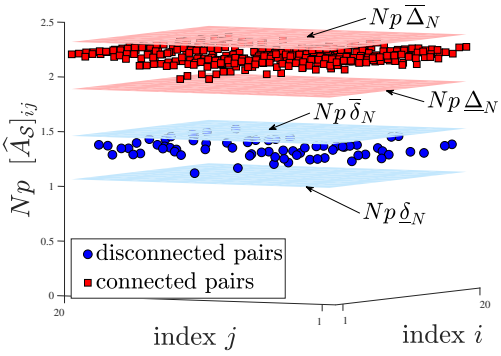

Fig. 2. Emergence of the identifiability gap. The example refers to a Granger estimator and to a scaling sequence $s_N = Np$. The scaled entries of the estimated matrix corresponding to connected pairs are sandwiched between the (red) margins, $Np\,\underline{\Delta}_N$ and $Np\,\overline{\Delta}_N$. Likewise, the scaled entries of the estimated matrix corresponding to non-connected pairs are sandwiched between the (blue) margins, $Np\,\underline{\delta}_N$ and $Np\,\overline{\delta}_N$.

Definition 2 (Identifiability Gap). *Let $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ be an estimated combination matrix. If there exists a sequence s_N, a real value η, and a strictly positive value Γ, such that:*

$$\boxed{\begin{array}{ll} s_N\,\underline{\boldsymbol{\delta}}_N \xrightarrow{\text{p}} \eta, & s_N\,\underline{\boldsymbol{\Delta}}_N \xrightarrow{\text{p}} \eta + \Gamma \\ s_N\,\overline{\boldsymbol{\delta}}_N \xrightarrow{\text{p}} \eta, & s_N\,\overline{\boldsymbol{\Delta}}_N \xrightarrow{\text{p}} \eta + \Gamma \end{array}} \tag{11}$$

we say that the estimated matrix $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ possesses an identifiability gap equal to Γ, and with scaling sequence s_N. □

The relationships in (11) imply the following three fundamental properties: separability, clustering, and bias.

Separability. The condition $s_N\,\overline{\boldsymbol{\delta}}_N \xrightarrow{\text{p}} \eta$ means that the *maximum* entry of $s_N\widehat{\boldsymbol{A}}_{\mathcal{S}}$ taken over the *disconnected* pairs converges to η. Likewise, condition $s_N\,\underline{\boldsymbol{\Delta}}_N \xrightarrow{\text{p}} \eta + \Gamma$ means that the *minimum* entry of $s_N\widehat{\boldsymbol{A}}_{\mathcal{S}}$ taken over the *connected* pairs converges to $\eta + \Gamma > \eta$. Joining these two relationships, we discover that the connected node pairs stand clearly separated from the disconnected node pairs, and the amount of separation is quantified by the gap, Γ.

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

Bias. For the *true* combination matrix, the entries corresponding to disconnected pairs are zero. In contrast, Eq. (11) reveals that the scaled entries for disconnected pairs tend to cluster around η, which results therefore in a *bias*. However, and remarkably, *this bias does not constitute a problem for consistent classification of connected/non-connected nodes*, because the bias does not affect in any manner the aforementioned separability and clustering properties.

In order to evaluate the accuracy of an estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ (not necessarily the Granger estimator), we introduce the error matrix $\boldsymbol{E} \triangleq \widehat{\boldsymbol{A}}_{\mathcal{S}} - \boldsymbol{A}$. In the next lemma, we provide a sufficient condition for identifiability by showing that a strictly positive gap exists when the (scaled) *maximum* and *minimum* errors are asymptotically concentrated around a fixed value.

Lemma 1 (Sufficient condition for an identifiability gap). *Let $\boldsymbol{G} \sim \mathscr{G}(N,p)$, and let $\boldsymbol{A}$ be a regular diffusion matrix with parameters ρ and κ. Let $\mathcal{S}$ be the set of observable nodes, of cardinality S, and let $S/N \to \xi > 0$ as $N \to \infty$. Consider then an estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}} = \boldsymbol{A}_{\mathcal{S}} + \boldsymbol{E}$, and assume that, for all $i,j \in \mathcal{S}$, with $i \neq j$:*

$$\underline{\boldsymbol{Z}}_N \leq \boldsymbol{e}_{ij} \leq \overline{\boldsymbol{Z}}_N, \tag{12}$$

where the quantities $\underline{\boldsymbol{Z}}_N$ and $\overline{\boldsymbol{Z}}_N$ do not depend on (i,j), and fulfill the following convergences:

$$Np\underline{\boldsymbol{Z}}_N \xrightarrow{\text{p}} \eta, \qquad Np\overline{\boldsymbol{Z}}_N \xrightarrow{\text{p}} \eta. \tag{13}$$

Then, the estimator $\widehat{\boldsymbol{A}}_{\mathcal{S}}$ possesses an identifiability gap equal to κ, with scaling sequence $s_N = Np$.

Proof: From (12) we know that:

$$\boldsymbol{a}_{ij} + \underline{\boldsymbol{Z}}_N \leq \boldsymbol{a}_{ij} + \boldsymbol{e}_{ij} \leq \boldsymbol{a}_{ij} + \overline{\boldsymbol{Z}}_N. \tag{14}$$

Substituting into (10) gives (when at least one pair in $\mathcal{S}$ is connected):

$$Np\,(\boldsymbol{m}_N + \underline{\boldsymbol{Z}}_N) \leq Np\,\underline{\boldsymbol{\Delta}}_N \leq Np\,\overline{\boldsymbol{\Delta}}_N \leq Np\,(\boldsymbol{M}_N + \overline{\boldsymbol{Z}}_N), \tag{15}$$

where we introduced

$$\boldsymbol{m}_N \triangleq \min_{\substack{i,j\in\mathcal{S}:\boldsymbol{a}_{ij}>0\\ i\neq j}} \boldsymbol{a}_{ij}, \qquad \boldsymbol{M}_N \triangleq \max_{\substack{i,j\in\mathcal{S}:\boldsymbol{a}_{ij}>0\\ i\neq j}} \boldsymbol{a}_{ij}. \tag{16}$$

Using now (8) along with (1), we easily see that both $Np\,\boldsymbol{m}_N$ and $Np\,\boldsymbol{M}_N$ converge in probability to κ. Using this result in conjunction with (13), and further observing that the event that all pairs in $\mathcal{S}$ are disconnected has asymptotically zero probability, we conclude that both $Np\,\underline{\boldsymbol{\Delta}}_N$ and $Np\,\overline{\boldsymbol{\Delta}}_N$ converge in probability to $\kappa+\eta$. Reasoning along the same lines, it is possible to show that the remaining convergences listed in (11) hold true, with $\Gamma = \kappa$. ∎

The next theorem ascertains that there exists a (positive) identifiability gap for the Granger estimator in (4), under the regime of dense connectivity. In [1] it is shown that such error matrix admits a convenient representation. Specifically, one has first to introduce the matrices $\boldsymbol{C} \triangleq [\boldsymbol{A}^2]_{\mathcal{S}'}$ (where $\mathcal{S}'$ is the complement of set $\mathcal{S}$) and $\boldsymbol{H} = (I_{N-K} - \boldsymbol{C})^{-1}$. Then, the entries of the error matrix are [1]:

$$\boldsymbol{e}_{ij} = \sum_{\ell,m\in\mathcal{S}'} \boldsymbol{a}_{i\ell}\boldsymbol{h}_{\ell m}\boldsymbol{a}^{(2)}_{mj}, \quad i,j \in \mathcal{S}. \tag{17}$$

Theorem 1 (Identifiability Gap for the Granger Estimator). *Let $\boldsymbol{G} \sim \mathscr{G}(N,p)$, and let $\boldsymbol{A}$ be a regular diffusion matrix with parameters ρ and κ. Let $\mathcal{S}$ be the set of observable nodes, of cardinality S, and let $S/N \to \xi > 0$ as $N \to \infty$. Then, for the Granger estimator, we have that (11) holds true with the choices $s_N = Np$, $\Gamma = \kappa$, and:*

$$\boxed{\eta = \kappa^2 p\,\frac{(2\rho-\kappa)\,(1-\xi)}{1-(\rho^2-2\rho\kappa\xi+\kappa^2\xi)}} \tag{18}$$

Random variable	Limit (in probability)
$\mathfrak{M}_a \triangleq \max_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \boldsymbol{a}_{ij}$	0
$\mathfrak{M}_{a,\text{self}} \triangleq \max_{i\in\mathcal{N}} \boldsymbol{a}_{ii}$	$\rho-\kappa$
$\mathfrak{M}_{c,\text{self}} \triangleq \max_{\ell\in\mathcal{S}'} \boldsymbol{c}_{\ell\ell}$	$(\rho-\kappa)^2$
$\mathfrak{M}^{(\mathcal{S}')}_{a,\text{sum}} \triangleq \max_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \sum_{\substack{h\in\mathcal{S}'\\ h\neq\ell,m}} \boldsymbol{a}_{hm}$	$\kappa(1-\xi)$
$\tilde{\mathfrak{M}}^{(\mathcal{S}')}_{a,\text{sum}} \triangleq \max_{i} \sum_{\ell\in\mathcal{S}'} \boldsymbol{a}_{i\ell}$	$\kappa(1-\xi)$
$\tilde{\tilde{\mathfrak{M}}}^{(\mathcal{S}')}_{a,\text{sum}} \triangleq \max_{i} \sum_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \boldsymbol{a}_{i\ell}$	$\kappa(1-\xi)^2$
$\tilde{\mathfrak{M}}^{(\mathcal{S}')} \triangleq \max_{i\in\mathcal{S}} \sum_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell m}$	$\kappa^2(1-\xi)^2$
$\tilde{\tilde{\mathfrak{M}}}^{(\mathcal{S}')} \triangleq \max_{\substack{i,j\in\mathcal{S}\\ i\neq j}} \sum_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{mj}$	$\kappa^2(1-\xi)^2$
$\mathfrak{M}_{c,\text{sum}} \triangleq \max_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \sum_{\substack{h\in\mathcal{S}'\\ h\neq m}} \boldsymbol{c}_{\ell h}$	$\rho^2 - 2\rho\kappa\xi + \kappa^2\xi$
$\mathfrak{M} \triangleq \max_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \sum_{\substack{\ell\in\mathcal{N}\\ \ell\neq i,j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j}$	$Np\,\mathfrak{M} \xrightarrow{\text{p}} \kappa^2 p$
$\mathfrak{M}^{(\mathcal{S}')} \triangleq \max_{\substack{i,j\in\mathcal{N}\\ i\neq j}} \sum_{\substack{\ell\in\mathcal{S}'\\ \ell\neq i,j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j}$	$Np\,\mathfrak{M}^{(\mathcal{S}')} \xrightarrow{\text{p}} \kappa^2 p(1-\xi)$
$\mathfrak{M}^{(\mathcal{S}')}_{a^3,\text{sum}} \triangleq \max_{\substack{i,j\in\mathcal{S}\\ i\neq j}} \sum_{\substack{\ell,m\in\mathcal{S}'\\ \ell\neq m}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell m}\boldsymbol{a}_{mj}$	$Np\,\mathfrak{M}^{(\mathcal{S}')}_{a^3,\text{sum}} \xrightarrow{\text{p}} \kappa^3 p(1-\xi)^2$

TABLE I
RANDOM VARIABLES RELEVANT FOR THEOREM 1.

Proof: Due to space constraints, we limit ourselves to a sketch of the proof. The main idea is obtaining some uniform (i.e., independent of the particular (i,j)-pair) bounds on the error matrix entries. Uniformity is critical because we need to examine the asymptotic regime where the number of agent pairs gets infinitely large.

We start by noting that, for all $i,j \in \mathcal{N}$, with $i \neq j$:

$$\begin{aligned} \boldsymbol{a}^{(2)}_{ij} &= \sum_{\ell\in\mathcal{N}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j} = (\boldsymbol{a}_{ii}+\boldsymbol{a}_{jj})\boldsymbol{a}_{ij} + \sum_{\substack{\ell\in\mathcal{N}\\ \ell\neq i,j}} \boldsymbol{a}_{i\ell}\boldsymbol{a}_{\ell j} \\ &\leq 2\,\mathfrak{M}_{a,\text{self}}\,\boldsymbol{a}_{ij} + \mathfrak{M} \leq 2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}_a + \mathfrak{M}. \end{aligned} \tag{19}$$

Recalling now that $\boldsymbol{C} = \boldsymbol{A}^2$, and using the relationship $\boldsymbol{C}^{k+1} = \boldsymbol{C}\boldsymbol{C}^k$ on an entrywise basis, it is possible (even if not straightforward) to work out the following bounds:

$$\boldsymbol{c}^{(k)}_{\ell\ell} \leq \overline{\boldsymbol{\alpha}}_k, \qquad \boldsymbol{c}^{(k)}_{\ell m} \leq \overline{\boldsymbol{\beta}}_k\,\boldsymbol{a}_{\ell m} + \overline{\boldsymbol{\gamma}}_k \text{ for } \ell \neq m, \tag{20}$$

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

$$\overline{\boldsymbol{\alpha}}_{k+1} = \mathfrak{M}_{c,\text{self}}\,\overline{\boldsymbol{\alpha}}_k + (2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}_a + \mathfrak{M})\,\rho^{2k}, \quad (21)$$

$$\overline{\boldsymbol{\beta}}_{k+1} = 2\,\mathfrak{M}_{a,\text{self}}\,\overline{\boldsymbol{\alpha}}_k + \mathfrak{M}_{c,\text{self}}\,\overline{\boldsymbol{\beta}}_k, \quad (22)$$

$$\overline{\boldsymbol{\gamma}}_{k+1} = \mathfrak{M}\,\overline{\boldsymbol{\alpha}}_k + (2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}^{(\mathcal{S}')} + \mathfrak{M}\,\mathfrak{M}^{(\mathcal{S}')}_{a,\text{sum}})\overline{\boldsymbol{\beta}}_k + \mathfrak{M}_{c,\text{sum}}\,\overline{\boldsymbol{\gamma}}_k, \quad (23)$$

with the initialization choices $\overline{\boldsymbol{\alpha}}_1 = \mathfrak{M}_{c,\text{self}}$, $\overline{\boldsymbol{\beta}}_1 = 2\,\mathfrak{M}_{a,\text{self}}$, and $\overline{\boldsymbol{\gamma}}_1 = \mathfrak{M}$. The explicit definitions of the pertinent random variables (e.g., $\mathfrak{M}_{a,\text{self}}$) appearing in (21)–(23) are listed in Table I. These variables serve to construct upper bounds that do not depend on (i,j), for example, $\mathfrak{M}_{a,\text{self}}$ is the maximum diagonal entry of matrix $\boldsymbol{A}$. Exploiting the fact that $\boldsymbol{H}$ can be expressed through a series involving powers of $\boldsymbol{C}$, and using (21)–(23), it is further possible to bound its entries as:

$$\boldsymbol{h}_{\ell\ell} \leq 1 + \overline{\boldsymbol{\Sigma}}_\alpha, \quad \boldsymbol{h}_{\ell m} \leq \overline{\boldsymbol{\Sigma}}_\beta\,\boldsymbol{a}_{\ell m} + \overline{\boldsymbol{\Sigma}}_\gamma \text{ for } \ell \neq m, \quad (24)$$

where we have introduced the series:

$$\overline{\boldsymbol{\Sigma}}_\alpha \triangleq \sum_{k=1}^{\infty} \overline{\boldsymbol{\alpha}}_k, \quad \overline{\boldsymbol{\Sigma}}_\beta \triangleq \sum_{k=1}^{\infty} \overline{\boldsymbol{\beta}}_k, \quad \overline{\boldsymbol{\Sigma}}_\gamma \triangleq \sum_{k=1}^{\infty} \overline{\boldsymbol{\gamma}}_k. \quad (25)$$

Next, we use the definition of the error in (17) to get:

$$\begin{aligned} \boldsymbol{e}_{ij} &\leq (1 + \overline{\boldsymbol{\Sigma}}_\alpha)\left[2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}^{(\mathcal{S}')} + \mathfrak{M}\,\tilde{\mathfrak{M}}^{(\mathcal{S}')}_{a,\text{sum}}\right] \\ &+ \overline{\boldsymbol{\Sigma}}_\beta\left[2\,\mathfrak{M}_{a,\text{self}}\,\mathfrak{M}^{(\mathcal{S}')}_{a_3,\text{sum}} + \mathfrak{M}\,\tilde{\mathfrak{M}}^{(\mathcal{S}')}\right] \\ &+ \overline{\boldsymbol{\Sigma}}_\gamma\left[2\,\mathfrak{M}_{a,\text{self}}\,\tilde{\tilde{\mathfrak{M}}}^{(\mathcal{S}')} + \mathfrak{M}\,\tilde{\tilde{\mathfrak{M}}}^{(\mathcal{S}')}_{a,\text{sum}}\right] \triangleq \overline{\boldsymbol{Z}}_N, \end{aligned} \quad (26)$$

where we have exploited (24), and we have bounded the terms involving entries of $\boldsymbol{A}$ in terms of some random variables listed in Table I. It is now possible to prove that $Np\,\overline{\boldsymbol{Z}}_N \xrightarrow{\text{p}} \eta$, with η being defined in (18). To this end, one can apply to (26) all the convergences listed in Table I. In particular, applying these convergences to the recursion in (21)–(23) gives the limits of the series in (25). The above arguments can be repeated by replacing upper bounds with lower bounds, and maxima with minima, so as to obtain a uniform lower bound, $\underline{\boldsymbol{Z}}_N$, fulfilling $Np\,\underline{\boldsymbol{Z}}_N \xrightarrow{\text{p}} \eta$. Thus, we see that the error for the Granger estimator meets the hypotheses of Lemma 1, which completes the proof of the theorem. ■

IV. Illustrative Example

We now illustrate empirically the asymptotic result of Theorem 1. We generate a realization of an Erdös-Rényi random graph with $N = 100$ nodes and connection probability $p = 0.8$, and we assume that only $S = 20$ nodes are observed. The combination matrix is given by the Metropolis rule in (6), with $\rho = 0.9$. In Fig. 3, the blue and red markers depict the vectorized and rearranged entries of the true combination matrix $\boldsymbol{A}_{\mathcal{S}}$. The cyan and magenta markers depict the vectorized entries (rearranged with the same ordering used for $\boldsymbol{A}_{\mathcal{S}}$) of the Granger estimator, $\widehat{\boldsymbol{A}}_{\mathcal{S}}$, computed from the empirical sample correlations evaluated over 5×10^5 time samples. We see how the emergence and magnitude of the identifiability gap, Γ, and of the bias, η, match well the predictions of Theorem 1.

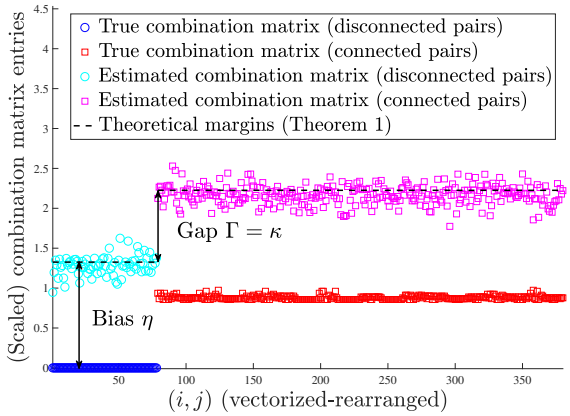

Fig. 3. Emergence of the identifiability gap for the Granger estimator, for the example described in Sec. IV.

References

[1] V. Matta and A. H. Sayed, "Consistent tomography under partial observations over adaptive networks," *IEEE Trans. Inf. Theory (early access)*, 2018, doi: 10.1109/TIT.2018.2839192.

[2] P. Geiger, K. Zhang, B. Schölkopf, M. Gong, and D. Janzing, "Causal inference by identification of vector autoregressive processes with hidden components," in *Proc. International Conference on Machine Learning*, vol. 37, Lille, France, July 2015, pp. 1917–1925.

[3] B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and M. G. Rabbat, "Characterization and inference of graph diffusion processes from observations of stationary signals," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 3, pp. 481–496, September 2018.

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

[5] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology identification from imperfect spectral templates," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, November 2016, pp. 1465–1469.

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

[7] D. Materassi and M. V. Salapaka, "Identification of network components in presence of unobserved nodes," in *Proc. IEEE Conference on Decision and Control*, Osaka, Japan, December 2015, pp. 1563–1568.

[8] A. Anandkumar and R. Valluvan, "Learning loopy graphical models with latent variables: Efficient methods and guarantees," *The Annals of Statistics*, vol. 41, no. 2, pp. 401–435, 2013.

[9] A. Santos, V. Matta, and A. H. Sayed, "Local tomography of large networks under the low-observability regime," May 2018, available as arXiv:1805.09081v1 [cs.MA].

[10] ——, "Divide-and-conquer tomography for large-scale networks," in *Proc. IEEE Data Science Workshop*, Lausanne, Switzerland, June 2018, pp. 1–5.

[11] ——, "Consistent tomography over diffusion networks under the low-observability regime," in *Proc. IEEE International Symposium on Information Theory*, Vail, CO, USA, June 2018, pp. 1–5.

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

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

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

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