

Consistent Tomography over Diffusion Networks under the Low-Observability Regime

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

Abstract—This work considers a diffusion network responding to streaming data, and studies the problem of identifying the topology of a subnetwork of *observable* agents by tracking their output measurements. Topology inference from indirect and/or incomplete datasets (*network tomography*) is in general an ill-posed problem. Under an appropriate Erdős-Rényi random graph modeling the unobserved part, the problem of network tomography is well-posed in the thermodynamic limit: when the number of network agents grows to infinity, *any arbitrary* subnetwork topology associated with the observed agents can be recovered with high probability.

I. INTRODUCTION

This work addresses the problem of inferring the subnetwork topology associated with a subset S of observable interacting agents or nodes in a network by tracking their states or output measurements over time. We assume that the state of each agent evolves over time as a result of the interaction with its neighbors. Formally, we assume that the state $\mathbf{y}_i(n)$ of each node $i \in \{1, 2, \dots, N\}$, at time n , evolves according to the following stochastic dynamical system (a.k.a. first-order vector autoregressive model):

$$\mathbf{y}_i(n) = \sum_{\ell=1}^N a_{i\ell} \mathbf{y}_\ell(n-1) + \mathbf{x}_i(n) \Leftrightarrow \mathbf{y}_n = A \mathbf{y}_{n-1} + \mathbf{x}_n \quad (1)$$

where $\mathbf{y}_n = [\mathbf{y}_1(n), \mathbf{y}_2(n), \dots, \mathbf{y}_N(n)]$ is a column vector collecting the nodes' states at time n , $A = [a_{i\ell}]$ is a *combination matrix*, and $\mathbf{x}_n = [\mathbf{x}_1(n), \mathbf{x}_2(n), \dots, \mathbf{x}_N(n)]$ is a column vector modeling a random input (e.g., streaming data or noise). We assume that $\{\mathbf{x}_i(n)\}$ are zero-mean and unit-variance random variables, independent and identically distributed (i.i.d.) both spatially (i.e., w.r.t. to i) and temporally (i.e., w.r.t. to n). For example, the dynamical system in (1) emerges naturally in the context of *adaptive diffusion networks* [1]. We observe from (1) that, if $a_{i\ell} = 0$, then agent i does not use the information coming from agent ℓ to update its own state. Therefore, the *support-graph* of A entails the underlying topology of connections among the agents.

The problem of topology retrieval addressed in this work is referred to as *network tomography* because only *indirect and partial* observations about the network topology are available. More specifically, only the output measurements from a *subset*

* A. Santos (email: augusto.santos@epfl.ch) and A. H. Sayed (email: ali.sayed@epfl.ch) are with STI at EPFL, CH-1015 Lausanne, Switzerland. The work of A. H. Sayed was also supported in part by NSF grants CCF-1524250 and ECCS-1407712.

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

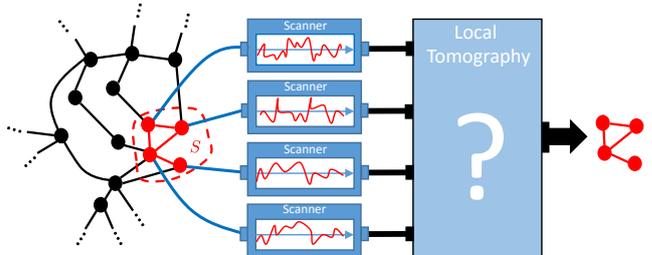


Fig. 1. Illustration of the problem of topology inference under partial observations. The goal is to estimate the underlying topology via processing the state-evolution of the *observable nodes* over time.

of nodes are accessible, and no information is available as regards the unobserved nodes neither their number — see Fig. 1 for a sketch of the problem. Under these challenging conditions, the ultimate goal in this paper is to infer the topology associated with the subset S of *observable* agents.

The tomography task amounts to design a scheme to process the observables — i.e., the state evolution of the observable nodes — and then estimate the topology. To this aim, it is useful to observe that the dynamical system in (1) implies the following relationship among the correlation matrix $R_0(n) \triangleq \mathbb{E}[\mathbf{y}_n \mathbf{y}_n^T]$, the one-lag correlation matrix $R_1(n) \triangleq \mathbb{E}[\mathbf{y}_n \mathbf{y}_{n-1}^T]$, and the combination matrix A :

$$R_1(n) = A R_0(n-1) \xrightarrow{n \rightarrow \infty} R_1 = A R_0, \quad (2)$$

where R_0 and R_1 are the limiting correlation matrices (we assume A stable). Therefore, since there exist many ways to estimate R_0 and R_1 consistently as $n \rightarrow \infty$, the relationship $A = R_1 R_0^{-1}$ reveals one possible strategy to estimate A (and hence its support) from the output of the diffusion process, \mathbf{y}_n . This is a scheme to consider when the state evolution of *all* nodes over time can be observed (*full* observation).

Let us now consider the fundamental restriction, which is at the core of this work, that only a subset S of the network is accessible (*partial* observation). For this case, the combination matrix pertaining to the observed subset of nodes is $A_S = [R_1 R_0^{-1}]_S$, which cannot be computed because only the submatrices associated with the observable agents, $[R_0]_S$ and $[R_1]_S$, are available. One could certainly mimic the relationship $A_S = [R_1 R_0^{-1}]_S$ using the truncated estimator:

$$\hat{A}_S = [R_1]_S ([R_0]_S)^{-1} \quad (3)$$

It is nonetheless clear from basic linear algebra that $\hat{A}_S \neq A_S$. This notwithstanding, it has been recently shown [2] that the

support of the observable network can be in fact recovered (*consistent tomography*) through the truncated estimator in (3), under certain conditions that will be discussed later. In the present work, this idea is pursued by addressing the consistent tomography problem under a novel setting, whose distinguishing features are now contrasted to their relatives in [2].

— *Arbitrary network topology*. In this work we assume that the graph of the accessible subnetwork has an *arbitrary* topology, modeled through a *deterministic* graph. The unobserved network graph is instead drawn from an Erdős-Rényi *random* graph. Such construction matches the classic “signal + noise” inferential paradigm where the object of the inference (i.e., the graph of the observable subnet), is modeled as an arbitrary deterministic signal, whereas the undesired component (i.e., the unobserved subnet), is modeled as a *noisy* component. To get insightful results, one must choose some model for such random component, and, without any particular prior information, we opt for a uniform connection model, where the presence/absence of each edge is determined through a sequence of i.i.d. Bernoulli experiments (the Erdős-Rényi graph). In contrast, in [2] it is assumed that the *overall* network (observed + unobserved) is drawn from an Erdős-Rényi random graph. Such a construction poses limitations on the topology object of the inference, which cannot be selected in an arbitrary fashion. Moreover, the network construction used in [2] enforces a *vanishing fraction of connected nodes* within the observable set, which is another limitation.

— *Low-observability regime*. In [2] it is assumed that the cardinality of the observed set, $|S|$, scales linearly with N , so as to ensure that the ratio $|S|/N$ converges to some positive fraction as $N \rightarrow \infty$. In contrast, we assume here that $|S|$ is fixed. Thus, we want to retrieve the topology of a fixed subnet embedded in a network that becomes infinitely dominant as $N \rightarrow \infty$. The resulting regime is accordingly referred to as a *low-observability* regime. We remark that the case where the ratio $|S|/N$ goes to zero is not addressed (nor can be obtained from the results) in [2].

A. Related Work and Main Contributions

At a highly-stylized level, the existing approaches to topology inference can be categorized in terms of three major features:

- \mathcal{F}_1 : Class of dynamical systems describing how the state of the agents evolves over time, e.g., the model in (1).
- \mathcal{F}_2 : Observable state variables, e.g., the process \mathbf{y}_n in (1).
- \mathcal{F}_3 : Topology-retrieval methods, which should exploit the *entanglement* between the observables in \mathcal{F}_2 and the underlying topology implicit in \mathcal{F}_1 . Such methods are sensitive to the dynamics in \mathcal{F}_1 and to the observables in \mathcal{F}_2 .

For what concerns \mathcal{F}_1 , most works focus on linear systems. Nonlinear dynamics are often dealt with by linearizing via considering variational characterizations of the dynamics (under small-noise regimes) [3]–[5] or by appropriately increasing the dimension of the observable space [6], [7]. In the context of linear (or linearized) systems, particular attention is paid to autoregressive diffusion models [8]–[11].

For what concerns \mathcal{F}_2 , the most common choice is to track the output measurements associated to the network nodes.

For what concerns \mathcal{F}_3 , the majority of the literature considers methods aimed at identifying commonalities between correlation constructs and graph topologies. For instance, in [9], a nontrivial construct on the correlation matrices is exploited to infer the underlying combination matrix of a partially observed system. The inverse problem of recovering the topology via correlation structures is often addressed through optimization-based methods, by reinforcing some (application-dependent) structural constraints such as, e.g., sparsity, stability, symmetry. For instance, in [10], [11], since the combination matrix and the correlation matrix share the same eigenvectors, the set of candidate topologies is reduced by computing these eigenvectors, and the inverse problem is then tackled with optimization methods under sparsity constraints.

This work complements the previous efforts by the following contributions. First, we consider a dynamical system consisting of a diffusion network, under the relevant *low-observability* regime where the number of observable agents is fixed and the size of the network scales to infinity. Second, we provide a straightforward inference strategy based on a thresholded version of (3). Third, we show in Theorem 1 further ahead that, for nonnegative symmetric combination matrices obeying suitable stability conditions (to be detailed later), any *arbitrary* topology of the observed set can be faithfully retrieved as the network size scales to infinity under an appropriate Erdős-Rényi random graph setting.

II. NOTATION AND PRELIMINARY DEFINITIONS

- We use boldface letters to denote random variables, and normal font letters for their realizations.
- Given an $N \times N$ matrix Z , the submatrix that lies in the rows of Z indexed by the set $S \subseteq \{1, 2, \dots, N\}$ and in the columns indexed by the set $T \subseteq \{1, 2, \dots, N\}$, is denoted by Z_{ST} , or alternatively by $[Z]_{ST}$. When $S = T$, the submatrix Z_{ST} is simply Z_S . In the indexing of the submatrix we will retain the index set of the original matrix. For example, if $S = \{2, 3\}$ and $T = \{2, 4, 5\}$, the submatrix $M = Z_{ST}$ is:

$$M = \begin{pmatrix} z_{22} & z_{24} & z_{25} \\ z_{32} & z_{34} & z_{35} \end{pmatrix} = \begin{pmatrix} m_{22} & m_{24} & m_{25} \\ m_{32} & m_{34} & m_{35} \end{pmatrix}. \quad (4)$$

- $\mathcal{G}(V)$ is the set of undirected graphs defined on a set of nodes (vertex set) V . When N is the number of nodes, the shortcut $\mathcal{G}(N)$ implicitly implies that the vertex set is $V = \{1, 2, \dots, N\}$. Self-loops will be permitted. Given $G \in \mathcal{G}(N)$, and a set $S \subseteq \{1, 2, \dots, N\}$, the subgraph corresponding to S is denoted by $G_S \in \mathcal{G}(S)$. The undirected support graph of a nonnegative symmetric matrix A is denoted by $G(A)$. The (i, j) -th entry of its adjacency matrix is $\mathbb{I}_{\{a_{ij} > 0\}}$, where $\mathbb{I}_{\mathcal{E}}$ denotes the indicator function.
- $\delta_{i,j}(G)$ is the distance between nodes i and j on graph G , i.e., the length of the *shortest* path binding i and j . The r -th order neighborhood of node i (including i itself), is given by: $\mathcal{N}_i^{(r)}(G) = \{j \in V : \delta_{i,j}(G) \leq r\}$. When $r = 1$, we simply talk of “the neighborhood”, and the superscript ⁽¹⁾ is omitted.

• A graph G is an Erdős-Rényi random graph if each edge of G is drawn, independently from the other edges, with identical probability p_N , and is denoted by $G \sim \mathcal{G}(N, p_N)$. It is known that, if the connection probability obeys the scaling law:

$$p_N = \frac{\log N + c_N}{N}, \quad \text{with } c_N \xrightarrow{N \rightarrow \infty} \infty, \quad (5)$$

the graph is connected with high probability [12]. Notably, arbitrariness of c_N implies that Erdős-Rényi graphs are connected with high probability even in the nontrivial regime where $p_N \rightarrow 0$ as $N \rightarrow \infty$.

A. Useful Graph Operations

Graph embedding. Given a vertex set V , and a subset thereof, $S \subset V$, the embedding of a graph $G^{(1)} \in \mathcal{G}(S)$ into the larger graph $G^{(2)} \in \mathcal{G}(V)$ will be denoted by:

$$G = G^{(1)} \rightsquigarrow G^{(2)}, \quad G \in \mathcal{G}(V), \quad (6)$$

and results in a graph where: *i)* the connections among nodes within S are determined by the graph $G^{(1)}$; *ii)* all the remaining connections – linking nodes from S to S' and among nodes within S' are determined by the graph $G^{(2)}$.

Local disconnection. The next operation consists of disconnecting some nodes of the graph. With the notation:

$$G_{U_1 \leftrightarrow U_2} \in \mathcal{G}(V), \quad (7)$$

we describe the graph obtained from G after removing all the edges that connect nodes from U_1 to nodes in U_2 .

III. NETWORK GENERATIVE MODEL

Following [2], the combination matrix is obtained through the following two steps.

– *Topology construction.* The structure of the graph within the observable set S will be deterministic and arbitrary. The structure of the graph in the complement set, S' , and as regards the connections between S and S' , will be random, following i.i.d. drawing of the pertinent edges. Accordingly, let $G^{(\text{obs})} \in \mathcal{G}(S)$ be a deterministic arbitrary graph on the observable set S , and let $\mathbf{G}^{(\text{unobs})} \sim \mathcal{G}(N, p_N)$ be an Erdős-Rényi random graph on N nodes. The final network graph, \mathbf{G} , is defined as:

$$\mathbf{G} = G^{(\text{obs})} \rightsquigarrow \mathbf{G}^{(\text{unobs})} \quad (8)$$

Connections within the observable set S are thus described through the graph $G^{(\text{obs})}$, whereas connections within S' , as well as the connections from S' to S , are described through the graph $\mathbf{G}^{(\text{unobs})}$. Equation (8) highlights the “signal + noise” construction, with the boldface notation emphasizing the random (i.e., noisy) component that corresponds to the unobserved network portion, and with the normal fonts emphasizing the deterministic component that corresponds to the arbitrary topology of the observed network portion. The aforementioned construction will be referred to as a *partial* Erdős-Rényi graph, and the resulting class of graphs with input deterministic graph $G^{(\text{obs})} = G_S$ placed on set S , will be represented as $\mathcal{G}(N, p_N, G_S)$. It can be shown that, under

condition (5), also the *partial* Erdős-Rényi graph is connected with high probability for any choice of the subgraph G_S .

– *Combination-weights assignment.* A combination policy γ is a nonnegative matrix-valued function that assigns a positive weight $[\gamma(G)]_{ij} = [\gamma(G)]_{ji} > 0$ to each connected pair (i, j) of an undirected graph G , and that assigns zero otherwise. Thus, G is the support graph of the matrix $A \triangleq \gamma(G)$. Also, $\gamma(G)$ has positive diagonal since we assume that $a_{ii} > 0$ for all nodes i . We remark that the combination matrix resulting from such an assignment will be *nonnegative* and *symmetric*. Among other possibilities, these matrices play a fundamental role in the context of *adaptive* networks [1].

Next we introduce two technical conditions on A .

Property 1 (Stability): The maximum row-sum norm, $\max_i \sum_{\ell=1}^N |a_{i\ell}|$, is upper bounded by some $\rho < 1$.

For nonnegative symmetric matrices, Property 1 becomes:

$$\max_{i=1,2,\dots,N} \sum_{\ell=1}^N a_{i\ell} = \max_{i=1,2,\dots,N} \sum_{\ell=1}^N a_{\ell i} \leq \rho \quad (9)$$

From Property 1 we see that (most of) the combination weights $a_{i\ell}$ typically vanish as N gets large, since a finite mass of value at most ρ must be allocated across an ever-increasing number of neighbors (on an Erdős-Rényi graph, the number of neighbors grows as $Np_N \rightarrow \infty$). The next property identifies a useful class of combination policies, which has been introduced in [2], for which degeneracy to zero of the combination weights is prevented by proper scaling.

Property 2 (Non-degeneracy under (Np_N) -scaling): A combination policy, applied to a partial Erdős-Rényi graph $\mathbf{G} \sim \mathcal{G}(N, p_N, G_S)$, is non-degenerate under (Np_N) -scaling, if there exists $\tau > 0$ such that, for $i \neq j$:¹

$$\mathbb{P}[Np_N \mathbf{a}_{ij} > \tau | \mathbf{a}_{ij} > 0] \geq 1 - \epsilon_N \quad (10)$$

where ϵ_N goes to zero as $N \rightarrow \infty$.

The class of policies with both properties is denoted by $\mathcal{C}_{\rho, \tau}$.

Properties 1 and 2 are relevant because they are automatically satisfied by several useful combination policies of common use. For example, the well-known Metropolis and Laplacian rules obey both properties [1]. Moreover, Property 2 is useful for tomography purposes because it will let the interacting pairs stick out from the error floor when the truncated estimator in (3) is used for topology retrieval. In order to understand why, we introduce the error matrix:

$$\mathbf{E} \triangleq \hat{\mathbf{A}}_S - \mathbf{A}_S, \quad (11)$$

which allows representing the (i, j) -th entry (magnified by Np_N) of the truncated estimator in (3) as:

$$\begin{cases} \underbrace{Np_N \mathbf{a}_{ij} + Np_N e_{ij}}_{\text{not vanishing}}, & \text{if } i \text{ and } j \text{ are connected,} \\ Np_N e_{ij}, & \text{otherwise.} \end{cases} \quad (12)$$

¹For a connected pair (i, j) with $i, j \in S$, Eq. (10) becomes $\mathbb{P}[Np_N \mathbf{a}_{ij} > \tau] \geq 1 - \epsilon_N$, since conditioning on a deterministic event is immaterial.

Thanks to (10), the *magnified* combination matrix is bounded away from zero *at the interacting pairs*, which explains the qualification of being “not vanishing” in (12). According to (12), if we want that the nonzero entry $Np_N \hat{a}_{ij}$ sticks out from the error floor we should control the error term $Np_N e_{ij}$, and this is done in the next section.

IV. MAIN RESULT

Preliminarily, let $M = [m_{ij}]$, with $i, j \in S$, be a nonnegative symmetric matrix, and consider a thresholding operator that compares the off-diagonal entries of M against a threshold $\tau > 0$, and outputs a graph, $\Gamma_\tau(M) \in \mathcal{G}(S)$, whose adjacency matrix has (i, j) -th entry equal to $\mathbb{I}_{\{m_{ij} > \tau\}}$. The entries on the main diagonal are set to one.

Theorem 1 (Exact recovery of G_S): Let G_S be a deterministic graph (with arbitrary topology), and let $\mathbf{G} \sim \mathcal{G}(N, p_N, G_S)$ be a partial Erdős-Rényi random graph where:

$$p_N = \frac{\log N + c_N}{N}, \quad c_N \xrightarrow{N \rightarrow \infty} \infty, \quad \frac{\log(Np_N)}{\sqrt{\log N}} \xrightarrow{N \rightarrow \infty} 0 \quad (13)$$

Let also $\mathbf{A} = \gamma(\mathbf{G})$ be a combination matrix obtained from a combination policy $\gamma \in \mathcal{C}_{\rho, \tau}$, for some $0 < \rho < 1$, and some $\tau > 0$. Then, the graph obtained by applying the thresholding operator $\Gamma_\tau(\cdot)$ to the magnified truncated estimator, $Np_N \hat{\mathbf{A}}_S$, matches the true support graph, G_S , with high probability:

$$\lim_{N \rightarrow \infty} \mathbb{P}[\Gamma_\tau(Np_N \hat{\mathbf{A}}_S) = G_S] = 1 \quad (14)$$

□

We remark that condition (13) takes on the practical meaning of imposing a minimum degree of sparsity, since it places a limitation on the asymptotic growth of c_N , and automatically enforces a vanishing connection probability. As a result, all the proved results hold in the interesting regime where p_N vanishes and the graph is connected with high probability.

Proof sketch of Theorem 1: It is shown in [2] that the entries of the error matrix defined in (11) are nonnegative, and, hence, we can write, for $i, j \in S$:

$$Np_N[\hat{\mathbf{A}}_S]_{ij} = Np_N a_{ij} + Np_N e_{ij} \geq Np_N a_{ij}. \quad (15)$$

Therefore, from Property 2 we conclude that, if i and j are interacting nodes, then $Np_N[\hat{\mathbf{A}}_S]_{ij}$ exceeds a positive threshold τ with high probability. If we further show that, for two *non-interacting* nodes i and j , and for any $\epsilon > 0$:

$$\mathbb{P}[Np_N e_{ij} > \epsilon] \xrightarrow{N \rightarrow \infty} 0 \quad (16)$$

then we can attain exact (with high probability) classification via inspection on the truncated estimator $\hat{\mathbf{A}}_S$: if $Np_N[\hat{\mathbf{A}}_S]_{ij} > \tau$, then classify (i, j) as an interacting pair, otherwise classify it as non-interacting. We now offer a succinct sketch of proof that is developed through the following two steps.

Step 1: Relating the error to “large” distances between nodes belonging to S' . It is shown in [2] that the entries of the error matrix in (11) can be represented as:

$$e_{ij} = \sum_{\ell, m \in S'} a_{i\ell} h_{\ell m} b_{mj}, \quad i, j \in S \quad (17)$$

where:

$$B \triangleq A^2, \quad H \triangleq (I_{S'} - B_{S'})^{-1} = \sum_{k=0}^{\infty} (B_{S'})^k. \quad (18)$$

From (17) we see that small values of the terms $h_{\ell m}$, for $\ell, m \in S'$, are desirable to get a small error. Examining the series in (18), one could haste to prospect a small $h_{\ell m}$ if nodes ℓ and m lay far apart in the support graph of $B_{S'}$, in light of the following known fact from matrix algebra [13]:

$$\delta_{\ell, m}(G(M)) = r \Leftrightarrow \text{the smallest } k \text{ with } [M^k]_{\ell m} > 0 \text{ is } r, \quad (19)$$

which holds for any nonnegative symmetric matrix M . However, in general the distance between ℓ and m will be contingent on the topology of subgraph G_S , which is unknown. To get rid of the dependencies on the particular graph G_S , it is therefore desirable to get some *universal* relationship between $h_{\ell m}$ and the distance between ℓ and m . To this aim, we observe that $B_{S'} = [A^2]_{S'} = A_{S'S} A_{SS'} + (A_{S'})^2$, which highlights that the matrix $B_{S'}$ (and, hence, the matrix H) does *not* depend on the matrix A_S . For this reason, we can remove the edges among the observable agents as far as computing bounds on H goes. This will imply that an appropriate distance between nodes in S' to bound $h_{\ell m}$ is $\delta_{\ell, m}(G_{S \leftrightarrow S})$, the distance between ℓ and m on the graph $G_{S \leftrightarrow S}$ where the edges among the observed agents in S have been removed. By exploiting these observations, the following theorem can be proved.

Theorem 2 (Bound on H): Given two distinct nodes $\ell, m \in S'$, we have that:

$$\delta_{\ell, m}(G_{S \leftrightarrow S}) = r \Rightarrow h_{\ell m} \leq \frac{\rho^r}{1 - \rho^2} \quad (20)$$

According to (20), it is convenient to split the analysis of the error by considering separately the case of “large” distances and the case of “small” distances. By simple application of the law of total probability in (16), we can write:

$$\mathbb{P}[Np_N e_{ij} > \epsilon] = \mathbb{P}[Np_N e_{ij} > \epsilon | \mathcal{D}_{\text{small}}] \mathbb{P}[\mathcal{D}_{\text{small}}] + \mathbb{P}[Np_N e_{ij} > \epsilon | \mathcal{D}'_{\text{small}}] \mathbb{P}[\mathcal{D}'_{\text{small}}], \quad (21)$$

where $\mathcal{D}_{\text{small}} \triangleq \bigcup_{\ell, m \in S'} \mathcal{D}_{\ell, m}$,

$$\mathcal{D}_{\ell, m} \triangleq \{\delta_{\ell, m}(G_{S \leftrightarrow S}) \leq r_N, \ell \in \mathcal{N}_i(\mathbf{G}), m \in \mathcal{N}_j^{(2)}(\mathbf{G})\}, \quad (22)$$

and where

$$r_N \triangleq \left\lceil \frac{1}{2} \frac{\log N}{\log(\log N + c_N)} \right\rceil \quad (23)$$

is a sequence of distances that diverges as $N \rightarrow \infty$ in view of condition (13). The event in (22) certifies that the distance

on graph $\mathbf{G}_{S \leftrightarrow S}$ between two distinct nodes, $\ell, m \in S'$, does not exceed a prescribed value r_N , and also certifies the membership of nodes ℓ and m to the pertinent neighborhoods defined on graph \mathbf{G} . Adding these membership is necessary because the nonzero terms in (17) correspond to nodes $\ell \in S'$ that are neighbors of $i \in S$ ($a_{i\ell} > 0$), and to nodes $m \in S'$ that are second-order neighbors of $j \in S$ ($b_{mj} > 0$).

Now, in view of Theorem 2, and owing to the definition of $\mathcal{D}_{\text{small}}$, it can be proved that large distances imply small errors, formally: $\mathbb{P}[N p_N e_{ij} > \epsilon | \mathcal{D}_{\text{small}}] = 0$ for sufficiently large N . As a result, the second term appearing on the RHS in (21) can be neglected, and it remains to show that small distances are rare, namely, that $\mathbb{P}[\mathcal{D}_{\text{small}}] \xrightarrow{N \rightarrow \infty} 0$.

Step 2: Managing “small” distances. Examining the relationship in (22), two sources of asymmetry emerge. First, the distance is computed with respect to graph $\mathbf{G}_{S \leftrightarrow S}$, while neighborhood memberships are defined in terms of the original graph, \mathbf{G} . Second, both $\mathbf{G}_{S \leftrightarrow S}$ and \mathbf{G} are non-homogeneous (because G_S is void in the former case, and is arbitrary in the latter). In order to manage (22) in a compelling way, it would be instead desirable to find some new graph $\tilde{\mathbf{G}}$ with the following properties: *i*) it is a homogeneous (i.e., classic) Erdős-Rényi graph; *ii*) it is coupled to the original graph in the sense that, if small distances are rare over a classic Erdős-Rényi graph, then small distances are rare also over the coupled partial Erdős-Rényi graph. This result is formally contained in the forthcoming theorem, stated without proof.

Theorem 3 (Homogenization and coupling): Let $\mathbf{G} \sim \mathcal{G}(N, p_N; G_S)$ be a partial Erdős-Rényi random graph, and let $\tilde{\mathbf{G}} \sim \mathcal{G}(N, \tilde{p}_N)$ be a pure Erdős-Rényi random graph where $\tilde{p}_N = |S|p_N$. Consider also the event $\tilde{\mathcal{D}}_{\text{small}} \triangleq \bigcup_{\ell, m \in S'} \tilde{\mathcal{D}}_{\ell, m}$,

$$\tilde{\mathcal{D}}_{\ell, m} \triangleq \{\delta_{\ell, m}(\tilde{\mathbf{G}}) \leq r_N, \ell \in \mathcal{N}_i(\tilde{\mathbf{G}}), m \in \mathcal{N}_j^{(2)}(\tilde{\mathbf{G}})\}. \quad (24)$$

Then, if $i, j \in S$ are non-interacting we have that:

$$\mathbb{P}[\mathcal{D}_{\text{small}}] \leq \mathbb{P}[\tilde{\mathcal{D}}_{\text{small}}] \quad (25)$$

It is finally possible to show that small distances are rare on a pure Erdős-Rényi graph, as stated in the following theorem.

Theorem 4 (Small distances): Let $\tilde{\mathbf{G}} \in \mathcal{G}(N, \tilde{p}_N)$ be a pure Erdős-Rényi random graph. Then, the sequence r_N in (23) ensures that, for sufficiently large N and for all $\epsilon > 0$:

$$\mathbb{P}[\tilde{\mathcal{D}}_{\text{small}}] \leq \tilde{\epsilon}_N \sim \tilde{p}_N (N \tilde{p}_N)^{r_N + 2} \quad (26)$$

and that the RHS of (26) vanishes as $N \rightarrow \infty$.

Applying jointly Theorems 3 and 4, we conclude that the second term on the RHS of (21) vanishes as $N \rightarrow \infty$. ■

V. ILLUSTRATIVE EXAMPLE

In Fig. 2 we depict the error probability in the topology recovery for the setting detailed in the caption. Two algorithms are considered for topology recovery. The first algorithm uses directly Theorem 1, namely, it employs a thresholding operator with the threshold τ that characterizes the class $\mathcal{C}_{\rho, \tau}$ that the pertinent combination rule belongs to. The second algorithm

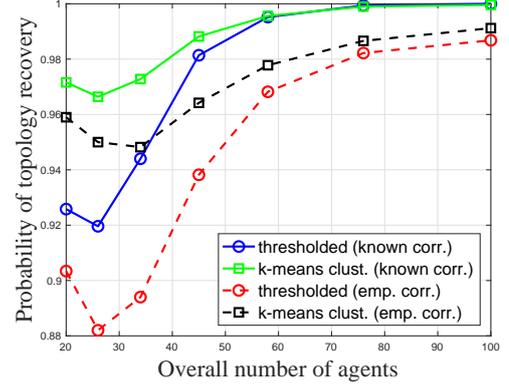


Fig. 2. Probability of topology recovery as a function of the overall network size, for an observable subnet with 10 nodes. The network evolution follows (1) with Gaussian data, $\rho = 0.9$, and Laplacian combination rule [1].

is a k -means algorithm (hence, *nonparametric*) that attempts splitting into two clusters the entries of the truncated estimator $\hat{\mathbf{A}}_S$, without any prior knowledge. We display also the curves corresponding to the relevant practical case where the correlation matrices $[R_0]_S$ and $[R_1]_S$ are not known beforehand, and are *estimated* from the streaming outputs collected from the observable set S . It is seen from Fig. 2 that the probability of correct recovery converges to one as the network size grows.

REFERENCES

- [1] A. H. Sayed, “Adaptation, Learning, and Optimization over Networks,” *Found. Trends Mach. Learn.*, vol. 7, no. 4-5, pp. 311–801, 2014.
- [2] V. Matta and A. H. Sayed, “Consistent Tomography under Partial Observations over Adaptive Networks,” July 2017. [Online]. Available: <http://arxiv.org/abs/1707.06444>
- [3] E. S. C. Ching and H. C. Tam, “Reconstructing links in directed networks from noisy dynamics,” *Phys. Rev. E*, vol. 95, p. 010301, 2017.
- [4] D. Napolitano and T. D. Sauer, “Reconstructing the topology of sparsely connected dynamical networks,” *Phys. Rev. E*, vol. 77, p. 026103, 2008.
- [5] J. Ren, W.-X. Wang, B. Li, and Y.-C. Lai, “Noise bridges dynamical correlation and topology in coupled oscillator networks,” *Phys. Rev. Lett.*, vol. 104, p. 058701, Feb 2010.
- [6] Y. Yang, T. Luo, Z. Li, X. Zhang, and P. S. Yu, “A robust method for inferring network structures,” in *Scientific Reports*, vol. 7, no. 5221, July 2017.
- [7] A. Mauroy and J. Goncalves, “Linear identification of nonlinear systems: A lifting technique based on the Koopman operator,” in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Las Vegas, USA, Dec 2016, pp. 6500–6505.
- [8] A. Moneta, N. Chlaß, D. Entner, and P. Hoyer, “Causal search in structural vector autoregressive models,” in *Proceedings of the 12th International Conference on Neural Information Processing Systems*, Vancouver, Canada, 2009, pp. 95–118.
- [9] P. Geiger, K. Zhang, B. Schölkopf, M. Gong, and D. Janzing, “Causal inference by identification of vector autoregressive processes with hidden components,” in *Proceedings of the 32nd International Conference on Machine Learning*, vol. 37, July 2015, pp. 1917–1925.
- [10] B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and M. Rabbat, “Characterization and inference of graph diffusion processes from observations of stationary signals,” *IEEE Transactions on Signal and Information Processing over Networks*, available as IEEE early access article, 2017.
- [11] S. Segarra, M. T. Schaub, and A. Jadbabaie, “Network inference from consensus dynamics,” 2017. [Online]. Available: <http://arxiv.org/abs/1708.05329>
- [12] P. Erdős and A. Rényi, “On Random Graphs I,” *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959.
- [13] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd edition. New York, NY, USA: Cambridge University Press, 2012.