

Consistent Tomography under Partial Observations over Adaptive Networks

Vincenzo Matta and Ali H. Sayed

Abstract—This work studies the problem of inferring whether an agent is directly influenced by another agent over a network. Agent i influences agent j if they are connected (according to the network topology), and if agent j uses the data from agent i to update its online learning algorithm. The solution of this inference task is challenging for two main reasons. First, only the output of the learning algorithm is available to the external observer that must perform the inference based on these indirect measurements. Second, only output measurements from a fraction of the network agents is available, with the total number of agents itself being also unknown. The main focus of this article is ascertaining under these demanding conditions whether *consistent tomography is possible*, namely, whether it is possible to reconstruct the interaction profile of the observable portion of the network, with negligible error as the network size increases. We establish a critical achievability result, namely, that for symmetric combination policies and for any given fraction of observable agents, the interacting and non-interacting agent pairs split into two separate clusters as the network size increases. This remarkable property then enables the application of clustering algorithms to identify the interacting agents influencing the observations. We provide a set of numerical experiments that verify the results for finite network sizes and time horizons. The numerical experiments show that the results hold for asymmetric combination policies as well, which is particularly relevant in the context of *causation*.

Index Terms—Adaptive networks, network tomography, causation, combination policy, Erdős-Renyi model.

I. INTRODUCTION

A significant number of complex real-world systems are modeled well by *networks*. At an abstract level, a network is an ensemble of interconnected agents. The interactions among neighboring agents enable the flow of information across the graph, and give rise to interesting and complex patterns of coordinated behavior.

One problem of immense value in network science is the inverse modeling problem. In this problem, the network structure (topology) is assumed to be largely unknown and one is interested in inferring relationships between agents based on some dataset arising from the agents' operations. Formulations of this type are of great interest in several application domains, such as communications, computer science, cyber-security, control, physics, biology, economics, and social sciences.

Inverse problems over networks are challenging because, in the vast majority of applications, direct access to the data ex-

changed between agents is often unavailable, and the inference of inter-agent relations must be based on *indirect* observations. Another source of difficulty in such inverse problems is that the access to observations is generally *limited to a subset of the network agents*. The process of discovering inter-agent interactions from indirect measurements is broadly referred to as *network tomography*. Some useful applications of network tomography include, among other possibilities: tracing the routes of clandestine information flows across communications networks [1]–[6]; revealing agent interactions over social networks, where disclosing commonalities within groups of agents might be useful for commercial as well as security purposes [7], [8]; inferential problems related to group testing and identification of defective items [9]; brain networks, where interactions among neurons might be of paramount importance to the understanding of particular diseases [10], [11]; anomaly detection in communications networks, where one tries to reveal the activities of intermediate nodes through destination-only measurements [12].

One useful type of networks is the class of adaptive networks [13], [14]. These networks consist of spatially dispersed agents that continually exchange information through diffusion mechanisms, and which are specifically designed to enable simultaneous *adaptation* and *learning* from streaming data (such as tracking targets moving in formation from streaming spatial observations; modeling the prey-predator behavior of animal groups on the move; allocating frequency resources over cognitive communication networks) [13], [14]. Using a powerful form of agent cooperation, adaptive diffusion networks are able to solve rather sophisticated inferential tasks, including: estimation tasks [14], detection tasks [15], [16], optimization and online learning tasks [17], [18]. This work is focused on *tomography over partially observed diffusion networks*.

A. Overview of Related Work

From a merely theoretical perspective, the aforementioned problems fall under the umbrella of signal processing over graphs [13], [14], [19]–[21]. More specifically, in this work we deal with the objective of retrieving a graph topology (a connection topology, or an “effective” topology corresponding to the exchange of information) from a set of indirect measurements taken at some accessible network locations. There are several works addressing a similar construction, albeit with different specific goals, and with different models. With no pretence of exhaustiveness, we start by giving a brief summary of the works that we believe are most related to the present

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

A. H. Sayed is with the École Polytechnique Fédérale de Lausanne EPFL, School of Engineering, CH-1015 Lausanne, Switzerland (e-mail: ali.sayed@epfl.ch). His work was supported in part by NSF grants CCF-1524250 and ECCS-1407712.

article. Later on, in Sec. II-C, we will refine the analysis of the connections of our model and results to other existing results once some necessary technical details have been introduced.

In [22], the problem of topology recovery for a general class of linear stochastic dynamical systems is considered. A solution based on Wiener filtering is proposed, which provides exact reconstruction for self-kin networks and where reconstruction of the smallest self-kin network embracing the true network topology is guaranteed. In [23], the authors introduce the concept of directed information graphs, which is used to capture the dependencies in networks of interacting processes linked by causal dynamics. The setting is further enriched in [24], where a new metric to learn causal relationships by means of *functional dependencies*, in a possibly nonlinear dynamical network, is proposed. Closely related to the network model considered in our work is the problem of estimating a graph when the relationships are encoded into a Vector AutoRegressive (VAR) or diffusion model — see, e.g., [25], where several methods to address causal inference in such a context are reviewed. Causal graph processes are also exploited in [26], where a computationally tractable algorithm for graph structure recovery is proposed, along with a detailed convergence analysis. Some recent works exploit spectral graph properties, in conjunction with sparsity constraints. In [27], [28], the problem of inferring the graph topology from observations of a random signal diffusing over the network is addressed. It is shown that the space of feasible matrices is a convex polytope, and two inferential strategies are proposed to select a point in this space as the topology estimate. In [29], the authors exploit convex optimization and sparsity in order to reconstruct an unknown graph from observable indirect relationships generated by diffusive signals defined on the graph nodes. The main idea is identifying a graph shift operator given only its eigenvectors, with the corresponding spectral templates being obtained from the sample covariance of independent graph signals diffused over the network.

The aforementioned works assume that data from *all* network nodes are available or observed. A distinctive feature of our work is that we consider the challenging scenario where *only partial observation* is permitted. Some results on partial observation settings appear in [30], [31], for the case of polytrees. In particular, reference [30] studies the linear stochastic model of [22] under partial observability, whereas reference [31] proposes methods to learning *directed information* polytrees where samples are available from only a subset of processes. The case of VAR models with unobserved nodes and with more general topologies is addressed in [32], [33], where detailed technical conditions for exact or partial reconstruction are provided.

B. Main Result

This work complements the previous efforts: it establishes an important identifiability condition and clarifies the asymptotic behavior of the recovery error as a function of the network size. An identification procedure is also developed to carry out the tomography calculations using a standard clustering technique. We summarize the main contributions

of this work as follows. We consider a network of agents running a diffusion strategy to solve some inference task of interest (such as a distributed detection problem). The overall network size is unknown, and the outputs of the diffusion algorithm are available from only a *limited* fraction of agents. The goal is determining the relationships among these agents, namely, establishing whether the datum of an individual agent is influencing another individual agent. We show that, under some regularity conditions, the interaction relationships existing within the observable portion of the network can be recovered, with negligible error for sufficiently large network sizes.

More specifically, we consider the class of Erdős-Rényi random graphs, where the probability of two agents being connected follows a Bernoulli distribution. For these graphs we discover that, for any fraction of observable agents, the group of interacting agent pairs and the group of non-interacting agent pairs split into two well-defined clusters. These clusters emerge as clearly separate for sufficiently large network sizes. This result is established for diffusion strategies employing *symmetric* combination matrices, a feature that enables analytical tractability of the problem. However, from a “physical” viewpoint, separability of the clusters does not appear to be limited to the symmetry assumption. Accordingly, we examine also the relevant case of asymmetric matrices, which is of interest especially in the context of *causal* networks. With reference to a typical form of asymmetric combination matrices, numerical simulations show that separability of the clusters can be preserved.

One distinctive feature of our work is that we establish *theoretical achievability* results for *consistent* tomography from *partially observed* networks. In answering these particular questions, in addition to what has been answered before, our work complements well existing results and recent progresses in the field of network tomography.

Notation. We use boldface letters to denote random variables, and normal font letters for their realizations. Capital letters refer to matrices, small letters to both vectors and scalars. Sometimes we violate the latter convention, for instance, we denote the total number of network agents by N .

For $i, j = 1, 2, \dots, N$, the (i, j) -th entry of an $N \times N$ matrix Z will be denoted by z_{ij} , or alternatively by $[Z]_{ij}$. Moreover, the sub-matrix that lies in the rows of Z indexed by the set $\mathcal{S}_1 \subseteq \{1, 2, \dots, N\}$ and in the columns indexed by the set $\mathcal{S}_2 \subseteq \{1, 2, \dots, N\}$, will be denoted by $Z_{\mathcal{S}_1 \mathcal{S}_2}$, or alternatively by $[Z]_{\mathcal{S}_1 \mathcal{S}_2}$. If $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$, the sub-matrix $Z_{\mathcal{S}_1 \mathcal{S}_2}$ will be abbreviated as $Z_{\mathcal{S}}$, or as $[Z]_{\mathcal{S}}$.

The symbols \mathbb{P} , \mathbb{E} , and \mathbb{V} are used to denote the probability, expectation, and variance operators, respectively. The notation $\xrightarrow{\mathbb{P}}$ denotes convergence in probability as $N \rightarrow \infty$.

II. THE PROBLEM

A. The Adaptive Diffusion Network

A network of N agents collects streaming data from the environment. The datum collected by the i -th agent at time n is denoted by $\mathbf{x}_i(n)$, and the global sequence of data is assumed to be formed by spatially (i.e., w.r.t. index i) and

temporally (i.e., w.r.t. index n) independent and identically distributed random variables. Without loss of generality, we assume that the variables have zero mean and unit variance.

In order to track drifts in the phenomenon that the network is monitoring, the agents implement a distributed adaptive strategy, where each agent relies on sharing information with local neighbors. In this work we focus on the Combine-Then-Adapt (CTA) diffusion mechanism, whose useful properties in terms of estimation and online learning performance have been already studied in some revealing detail in [13], [14], [17], [18]. The CTA algorithm can be described as follows.

First, during the *combination* step, agent i mixes the data received from its neighbors by using a sequence of convex (i.e., nonnegative and adding up to 1) combination weights $w_{i\ell}$, for $\ell = 1, 2, \dots, N$, giving rise to the following intermediate variable:

$$\mathbf{v}_i(n-1) = \sum_{\ell=1}^N w_{i\ell} \mathbf{y}_\ell(n-1), \quad (1)$$

where a weight $w_{i\ell}$ is equal to 0 if agent ℓ does not influence agent i . Then, during the *adaptation* step, agent i updates its output variable by comparison with the incoming streaming data $\mathbf{x}_i(n)$, employing a (typically small) step-size $\mu \in (0, 1)$:

$$\mathbf{y}_i(n) = \mathbf{v}_i(n-1) + \mu[\mathbf{x}_i(n) - \mathbf{v}_i(n-1)]. \quad (2)$$

Equations (1) and (2) can be merged into a single step as:

$$\mathbf{y}_i(n) = (1-\mu) \sum_{\ell=1}^N w_{i\ell} \mathbf{y}_\ell(n-1) + \mu \mathbf{x}_i(n). \quad (3)$$

For later use, it is convenient to introduce the *scaled* combination matrix A , whose entries are defined as:

$$a_{ij} \triangleq (1-\mu)w_{ij}, \quad (4)$$

which, in the sequel, will be simply referred to as the combination matrix. We remark that, since we use a sequence of *convex* combination weights, the (nonnegative) matrix $A/(1-\mu)$ is a *right-stochastic* matrix, namely,

$$\sum_{\ell=1}^N a_{i\ell} = 1 - \mu, \quad \text{for all } i = 1, 2, \dots, N. \quad (5)$$

If we now stack the observations across the network at time n into the $N \times 1$ vector \mathbf{x}_n , and the state variables at time n into the $N \times 1$ vector \mathbf{y}_n , Eq. (3) can be compactly rewritten as:

$$\mathbf{y}_n = A \mathbf{y}_{n-1} + \mu \mathbf{x}_n, \quad (6)$$

which, by iteration, allows us to express the diffusion output vector as a function of the streaming data:

$$\mathbf{y}_n = \mu \sum_{m=1}^n A^{n-m} \mathbf{x}_m, \quad n \geq 1 \quad (7)$$

starting from state $\mathbf{y}_0 = 0$, i.e., neglecting the *transient* term.

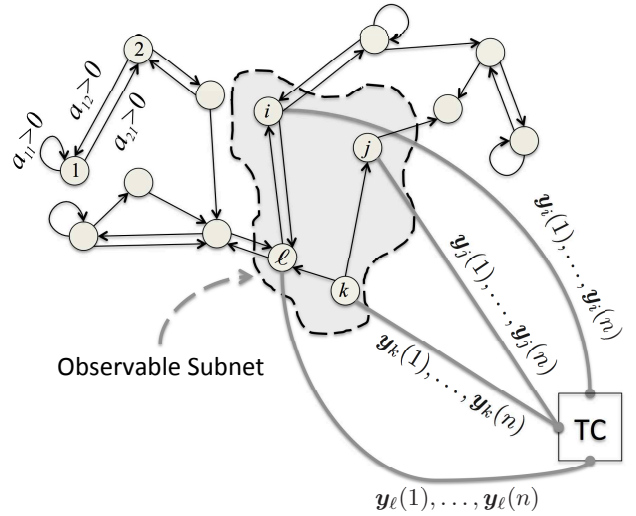


Fig. 1. Conceptual sketch of the tomography problem addressed in this article.

B. Network Tomography

An illustration of the setting considered in this work is given in Fig. 1. An entity external to the network (hereafter named Tomography Center, TC) is interested in reconstructing the interaction profile of the network, namely, it is interested in ascertaining which agent is influencing which other agent.

The TC is assumed to have access to a *subset* of the network agents, and is able to collect the streams of outputs exchanged by such agents during their communication activity. Letting $\Omega \subset \{1, 2, \dots, N\}$ be the observable subnet, the data available to the TC at time n are $\{\mathbf{y}_i(1), \mathbf{y}_i(2), \dots, \mathbf{y}_i(n)\}_{i \in \Omega}$. We shall focus on the asymptotic regime of large networks ($N \rightarrow \infty$), for the meaningful case where the fraction of observed agents does not vanish. Letting $K = |\Omega|$, such regime is defined in terms of the following condition:

$$\frac{K}{N} \rightarrow \xi \quad (8)$$

where $\xi \in (0, 1)$ takes on the meaning of the asymptotic fraction of observed agents.

In our setting, the TC does not know the overall number of agents in the network. Accordingly, the main goal of the TC is producing an estimate of the interaction profile for the *observed* agents. This problem is challenging for the following reasons. Let us ignore for a moment the fact that the network is partially observed, and assume that the TC is able to collect all output sequences from all agents at all times. Using such a dataset, there exist several well-established strategies to make inference about the influence that one agent has on another agent. The most intuitive is an estimate of the correlation between the outputs of two agents, which, however, is problematic for *directed* flows of information (where agent i can influence agent j but not vice versa). Such asymmetry would be reflected into the (i, j) -th and (j, i) -th entries of the combination matrix, yielding $a_{ij} = 0$ while $a_{ji} > 0$. The case of asymmetric influence is well studied within the framework

of *causal* inference, or *causation*. Many solutions exist for disclosing causal relationships [23]–[26], [31], [32].

There is, however, a challenging problem that is peculiar to the network setting of this work due to the streaming nature of the data. In general, when the TC starts collecting data, the network would have been in operation since some time already. Therefore, the output signals at the agents would have benefited from sufficient exchanges of information. While this exchange of information is beneficial for solving inference tasks, it nevertheless can become detrimental for reconstructing the network tomography. This is because, over a strongly connected network and after a “transient” phase, all agents would have become “correlated.”

In order to overcome this difficulty, we exploit knowledge of the diffusion mechanism. To this aim, let us introduce the correlation matrix of the diffusion output vector:

$$R_0(n) \triangleq \mathbb{E}[\mathbf{y}_n \mathbf{y}_n^T], \quad (9)$$

which, from (7), admits the following closed-form representation:

$$R_0(n) = \mu^2 \sum_{i=0}^{n-1} A^i (A^i)^T \xrightarrow{n \rightarrow \infty} R_0 = \mu^2 \sum_{i=0}^{\infty} A^i (A^i)^T, \quad (10)$$

where the latter series is guaranteed to converge since all eigenvalues of A are strictly inside the unit disc. The limiting correlation matrix, R_0 , can be interpreted as the (unique) solution to the discrete-time Lyapunov equation [13]:

$$R_0 - AR_0A^T = \mu^2 I_N, \quad (11)$$

where I_N denotes the $N \times N$ identity matrix. We note that $R_0(n)$ is positive definite for each n , and so is R_0 , due to the stability of A [34]. We also introduce the one-lag correlation matrix, which, in view of (6), takes on the form:

$$R_1(n) \triangleq \mathbb{E}[\mathbf{y}_n \mathbf{y}_{n-1}^T] = AR_0(n-1) \xrightarrow{n \rightarrow \infty} R_1 = AR_0. \quad (12)$$

Therefore, we obtain the following relationship:

$$A = R_1(n)(R_0(n-1))^{-1}, \quad n \geq 2, \quad (13)$$

and, at steady-state,

$$A = R_1 R_0^{-1} \quad (14)$$

In principle, since there exist many ways to estimate R_0 and R_1 consistently as $n \rightarrow \infty$, expression (14) reveals one possible strategy to estimate A from the output of the diffusion process. However, the approach described so far suffers from a critical problem: the network is only *partially* observed and, hence, not all entries in the matrices R_0 and R_1 can be estimated. This in turn means that the evaluation of (14) is in general problematic.

In order to estimate the combination (sub-)matrix corresponding to the observable agents, i.e.,

$$A^{(\text{obs})} \triangleq A_\Omega, \quad (15)$$

one might be tempted to simply replace the matrices involved in (14) by their *observable* counterparts, $R_0^{(\text{obs})} \triangleq [R_0]_\Omega$, and $R_1^{(\text{obs})} \triangleq [R_1]_\Omega$, obtaining the rough estimate:¹

$$\hat{A}^{(\text{obs})} = R_1^{(\text{obs})} (R_0^{(\text{obs})})^{-1} \quad (16)$$

In order to avoid confusion, we remark that $R_0^{(\text{obs})}$ is *not* obtained from R_0 by replacing with zeros the entries located in the unobserved set. In fact, $R_0^{(\text{obs})} = [R_0]_\Omega$ is the sub-matrix (of reduced size, i.e., $K \times K$) obtained by extracting from R_0 the entries corresponding to the observable set Ω . The same consideration holds for the other sub-matrices involved in our treatment.

Needless to say, calculation on the right-hand side of (16) does not lead to the true $A^{(\text{obs})}$, except in some special cases. For this reason, we are denoting the result of (16) by using the hat notation. If we could recover $A^{(\text{obs})}$ exactly, then we could deduce the desired influence relations. However, given that we only have the estimate $\hat{A}^{(\text{obs})}$, it is not clear at all whether the mutual influence relationships existing between the observed nodes can be consistently retrieved from $\hat{A}^{(\text{obs})}$. Answering this nontrivial question in the affirmative is one of the main contributions of this work.

We focus in this article on the case of symmetric combinations matrices. The symmetry assumption is made because, if A is symmetric, the autocorrelation matrix of the diffusion output takes on the following convenient form — see (10):

$$R_0(n) = \mu^2 \sum_{i=0}^{n-1} A^{2i} \Rightarrow R_0 = \mu^2 (I_N - A^2)^{-1}, \quad (17)$$

which is exploited in the proof of Theorem 1. The extension of the results to the asymmetric case requires adjustments in the arguments used in the proof of Theorem 1. For example, this can be pursued by appealing instead to the Kronecker representation of the solution of discrete-time Lyapunov equations [35]. We note in passing that, since the matrix $A/(1-\mu)$ is always right-stochastic, the symmetry assumption makes $A/(1-\mu)$ *doubly*-stochastic.

Before ending this section, we would like to comment on the two assumptions that we have made on the combination matrix, namely, convexity of the combination weights — see (5) — and symmetry. These assumptions are commonly encountered in the context of distributed data processing.

In particular, the choice of *convex* combination weights is standard and stems from the fact that the agents in the network are typically interested in converging to some average statistic of the locally processed information [36], [37]; the convexity assumption helps avoid the introduction of bias. Several useful forms of distributed algorithms have been developed under such convexity conditions to address different types of problems, including: distributed detection [38] and estimation [39],

¹Actually, the matrices computed by the TC will generally be renumbered versions of $[R_0]_\Omega$ and $[R_1]_\Omega$, since the labeling used by the TC need not correspond to the ordering in $[R_0]_\Omega$ and $[R_1]_\Omega$. This is because the TC does not know the original labeling of the agents in the network, nor the network size. However, our focus is on evaluating the interaction between *pairs* of agents, irrespective of their labels, and, hence, we can safely keep the notation used in (16).

distributed signal processing [40], and distributed optimization [41], [42]. For these types of distributed inference and optimization problems, *adaptive diffusion* strategies refer to a class of algorithms that endow networks with the ability to respond to streaming data in real-time and to track drifts in the statistical properties of the data. As such, they are particularly useful for inference problems that require online implementations [13], [14] and have been studied extensively by many authors since their inception over a decade ago.

For what concerns the symmetry of the combination matrix, we remark that this condition is not necessary for the operation of the diffusion strategy. There are actually useful instances where asymmetric combination policies are directly applicable (e.g., when the uniform averaging rule is used) and diffusion strategies have been studied extensively under such conditions as well [13], [14]. However, for the analysis approach adopted in this paper, the symmetry assumption simplifies the derivation. Extension of the results to the non-symmetric scenario is left for future work.

C. Connections with Existing Results

— *Existing approaches to tomography with full observations.* A comprehensive account of the problem of topology reconstruction from node measurements, *under the full observations regime*, is provided in [22]. The model considered in [22] is a *general* linear model, for example, it can be a VAR model of arbitrary order, or a non-causal linear model. To address this general case, in [22] an approach based on Wiener filtering is proposed to infer the topology.

However, for the diffusion model considered in the present work, the solution (still in the case of full observations) is straightforward and given by (14). One useful interpretation of this solution is that the combination weights $\{a_{ij}\}_{j=1}^N$ obtained through (14) are the coefficients of the *best one-step linear predictor* (also referred to, in the context of causal analysis, as the *Granger estimator or operator*) that minimizes the expected squared error in estimating $y_i(n)$ given the past samples $\{y_j(n-1)\}_{j=1}^N$ — see, e.g., [35]. Such a well-known result is mentioned also in [22] with reference to topology estimation over strictly causal systems that include our diffusion model as a special case. Likewise, the case where (14) is applied with correlation matrices *estimated empirically from the available samples* corresponds to the best one-step linear predictor in a least-squares sense (where the *expected* squared error is replaced by the *empirical* squared error computed over the samples collected across time).

— *Tomography with partial observations, achievability.* As mentioned in Sec. I-A, sufficient conditions for retrieving the topology under partial observations, with reference to dynamical systems pertinent to our model, are offered in [32], [33]. It is useful to contrast these conditions with the approach pursued in the present work.

The identifiability conditions presented in [32], [33] can be broadly identified as conditions operating at a “microscopic” level, namely, they are related to some detailed knowledge of the topology and/or of the statistical model (e.g., type of noise, joint distribution of the observable data) [32], [33]. Such

a detailed knowledge makes those methodologies clearly not practical for realistic large-scale network settings (which are the main focus of this work).

In order to overcome this difficulty, in the present work we propose an approach that is natively tailored to the large-scale setting. We take a statistical perspective where the conditions on the network connection topology are described at a *macroscopic* level through some average/summary indicators, such as, e.g., the connection probability between any two nodes. Under these conditions, we focus on establishing an achievability result that holds (in a *statistical* sense) as the size of the network scales to infinity.

— *Tomography with partial observations, estimators.* As we have remarked, with full observations the solution in (14) arises from the solution of a one-step linear prediction problem, and is a classic solution adopted in the VAR context [22], [32]. In our context with partial observations, we propose to pursue the same kind of approach, but for the fact that the best one-step linear prediction is enforced *only with respect to the observable nodes*. As a matter of fact, the solution in (16) provides the combination weights yielding the best one-step linear prediction that minimizes the expected squared error in estimating the *observable* sample $y_i(n)$ (for $i \in \Omega$) given the past *observable* samples $\{y_j(n-1)\}_{j \in \Omega}$. This solution can be still interpreted as a Granger estimator, and is in fact widely adopted in causal inference from time series when one ignores and/or neglects the existence of the unobserved components. However, there is no claim of optimality for this estimator, and even no guarantee that it could provide faithful tomography. Establishing that it actually can is one nontrivial result of the present work.

For what concerns achievability, the use of $\hat{A}^{(\text{obs})}$ is instrumental to establish the achievability result: we show that such estimator induces a *threshold effect* by splitting the interacting and non-interacting pairs into two well-separated regions. At a more practical level, and as a consequence of the ascertained threshold effect, we will show in the sequel that the proposed estimator, followed by a thresholding operator (where the threshold can be conveniently determined through standard non-parametric clustering algorithms), provides also a *viable* recipe to accomplish the tomography task.

III. ERROR CAUSED BY PARTIAL OBSERVATIONS

Using the estimate (16), we can write:

$$\hat{A}^{(\text{obs})} = A^{(\text{obs})} + E, \quad (18)$$

where E denotes the *error* matrix. In terms of the individual entries, we can write, for $i, j = 1, 2, \dots, K$:²

$$\hat{a}_{ij}^{(\text{obs})} = a_{ij}^{(\text{obs})} + e_{ij}. \quad (19)$$

In this work we are interested in establishing whether the estimated values $\hat{a}_{ij}^{(\text{obs})}$ allow us to identify the condition $a_{ij}^{(\text{obs})} > 0$ or $a_{ij}^{(\text{obs})} = 0$, which would reveal whether the agents i and j influence each other. In order to enable this

²A pair (i, j) refers to agents ω_i and ω_j , where $\omega_1 < \omega_2 < \dots < \omega_K$ span the *observable* subnet Ω .

determination, we start by characterizing the behavior of the error terms e_{ij} in (19).

Theorem 1 (Concentration of errors): For a symmetric combination matrix, the entries of the error matrix defined in (18) are nonnegative, and satisfy for all $i = 1, 2, \dots, K$:

$$\boxed{\sum_{j=1}^K e_{ij} \leq 1 - \mu} \quad (20)$$

Proof: See Appendix A. ■

We have the following corollary, which is particularly tailored to our application.

Corollary 1 (Number of errors exceeding a threshold): For a symmetric combination matrix, we have, for all $i = 1, 2, \dots, K$, and for any $\epsilon > 0$:

$$\boxed{\sum_{j=1}^K \mathbb{I}\{e_{ij} > \epsilon\} \leq \frac{1 - \mu}{\epsilon}} \quad (21)$$

where $\mathbb{I}\{\cdot\}$ is the indicator function (which is equal to one when its argument is true and zero otherwise).

Proof: Suppose that (21) is false. Then, we would have (recall that $e_{ij} \geq 0$ for all $i, j = 1, 2, \dots, K$):

$$\begin{aligned} \sum_{j=1}^K e_{ij} &= \sum_{j:e_{ij} > \epsilon} e_{ij} + \sum_{j:e_{ij} \leq \epsilon} e_{ij} \geq \sum_{j:e_{ij} > \epsilon} e_{ij} \\ &> \epsilon \sum_{j=1}^K \mathbb{I}\{e_{ij} > \epsilon\} > \frac{1 - \mu}{\epsilon} \times \epsilon = 1 - \mu, \end{aligned} \quad (22)$$

which contradicts (20). ■

Theorem 1 and its corollary provide useful information about the concentration of the entries in the error matrix. In particular, Eq. (20) reveals that the row-sum of the entries in the error matrix cannot exceed $1 - \mu$, whereas Eq. (21) places an upper bound on the number of entries that exceed any positive threshold.

Consider now a small threshold ϵ , and the fraction of off-diagonal (because we are interested in inter-agent interactions) entries that exceed ϵ :

$$\frac{1}{K(K-1)} \sum_{i=1}^K \sum_{j \neq i} \mathbb{I}\{e_{ij} > \epsilon\}. \quad (23)$$

In the regime dictated by (8), where $K \rightarrow \infty$ as $N \rightarrow \infty$, we see from (21) that such fraction vanishes, namely, that most entries of the error matrix will be small in the asymptotic regime of large networks. Therefore, in view of definition (19), we can loosely say that, for large networks and for $i \neq j$, the following ‘‘typical’’ behavior will emerge:

$$\hat{a}_{ij}^{(\text{obs})} = \begin{cases} a_{ij}^{(\text{obs})} + \text{small quantity}, & \text{if } a_{ij}^{(\text{obs})} > 0, \\ \text{small quantity}, & \text{if } a_{ij}^{(\text{obs})} = 0. \end{cases} \quad (24)$$

This useful dichotomy suggests that the nonzero entries of $A^{(\text{obs})}$ will make the estimated entries $\hat{a}_{ij}^{(\text{obs})}$ stand out above the error floor as N increases. As a result, we should be able

to decide whether $a_{ij}^{(\text{obs})} = 0$ or $a_{ij}^{(\text{obs})} > 0$ by comparing the estimated value, $\hat{a}_{ij}^{(\text{obs})}$, against some threshold.

However, and unfortunately, the behavior of the error matrix alone is not sufficient to conclude that this inferential procedure is feasible, for one crucial reason. For most typical combination matrices, the nonzero entries $a_{ij}^{(\text{obs})}$ decrease with N as well, so that the nonzero entries appearing in the first line of (24) vanish for large network sizes, along with the errors. This means that the estimated entries, $\hat{a}_{ij}^{(\text{obs})}$, would vanish even when agents i and j are interacting (i.e., even when $a_{ij}^{(\text{obs})} > 0$). For this reason, a closer examination of the behavior of the error quantities is necessary before we can conclude that this inference procedure is viable. Specifically, it is necessary to assess how fast the error signals e_{ij} decay to zero in relation to the desired entries $a_{ij}^{(\text{obs})}$. In order to carry out this analysis, we need to make some assumptions about the network structure. We will select some typical and popular random models, as explained next.

IV. BEHAVIOR OF INTERACTING AGENTS ($a_{ij} > 0$)

The interaction profile can be conveniently described in terms of a symmetric *interaction matrix* G , defined by the following conditions:

$$g_{ij} = \begin{cases} 1, & \text{if } a_{ij} > 0, \\ 0, & \text{if } a_{ij} = 0. \end{cases} \quad (25)$$

We shall set $g_{ii} = 1$ for all $i = 1, 2, \dots, N$, which corresponds to the assumption that an agent does always use its own state variable in the combination step.

In our analysis, the combination matrix, A , will be constructed through the following two-step procedure. First, an interaction matrix G is generated according to a *random graph model* [43], [44]. Then, A is determined by a *combination policy*, $\gamma(G)$, which sets the values of the combination weights corresponding to the nonzero entries of G . Formally:

$$\boxed{A = \gamma(G)} \quad (26)$$

with the prescription that $\sum_{\ell=1}^N a_{i\ell} = 1 - \mu$ for all $i = 1, 2, \dots, N$. Note that $\gamma(G)$ must always assign *positive* weights at the locations corresponding to *nonzero* entries of G , otherwise the interaction matrix related to A would be different from G . Moreover, since in this article we focus on symmetric combination matrices, we must have that $[\gamma(G)]_{ij} = [\gamma(G)]_{ji}$ (a condition that would not be directly implied by the symmetry of G).

A. Random Model for the Interaction Profile

As we have stated, in this work we examine the useful case where the interaction profile of the network is generated according to a *random graph model* [43], [44]. In particular, we consider the classic Erdős-Rényi model. This model, which we denote by $\mathcal{G}(N, p_N)$, is an undirected (i.e., symmetric) graph, where the presence/absence of the $N(N-1)/2$ edges is determined by a sequence of $N(N-1)/2$ independent Bernoulli random variables with success (i.e., interaction) probability p_N . Accordingly, the variables g_{ij} ,

for $i = 1, 2, \dots, N$ and $j > i$, are independent Bernoulli random variables with $\mathbb{P}[g_{ij} = 1] = p_N$, and the matrix G is symmetric.

The Erdős-Rényi construction is one of the most popular statistical models for examining random graphs. Among the positive attributes of such model, we might mention: *i*) uniformity across nodes, which makes this model a good candidate to represent random interactions in a maximum uncertainty scenario where little is known about the underlying model; *ii*) analytical tractability, which stems from independence and identical distributions across edges, and which makes this model practical in order to gain insights and obtain closed-form results. At the same time, we must mention that the Erdős-Rényi model is known to have limitations in capturing some effects that characterize certain networks. For example, the power-law behavior (of the nodes' degree) that has been observed in some network applications, can be captured through more sophisticated models such as the preferential attachment model [45].

One meaningful regime to examine the properties of the Erdős-Rényi model is the regime where the probability p_N decreases as N increases [44]. Examining the asymptotic regime where p_N is kept constant while N increases would generally produce networks that are too much connected with respect to what happens in typical applications. For instance, in the regime of constant p_N , the network diameter (i.e., the maximum of the shortest distance between any pair of nodes) is equal to 2 with high probability [46]. On the other hand, p_N could not vanish in an arbitrary fast way, otherwise the number of neighbors of an agent would be too small, and the network would become very scarcely connected. A well-known result that holds true for the Erdős-Rényi graph is that the following scaling law:

$$p_N = \frac{\ln N + c_N}{N} \quad (27)$$

with $c_N \rightarrow \infty$ (in an arbitrarily slow fashion, i.e., even with $p_N \rightarrow 0$), ensures that the graph remains connected with probability tending to 1 as N diverges [44]. In the following, we shall focus on the regime of connected Erdős-Rényi graph with vanishing p_N , namely, on the regime where (27) is satisfied with $c_N \rightarrow \infty$ and $p_N \rightarrow 0$. Such a regime will be denoted by the symbol $\mathcal{G}^*(N, p_N)$.

B. Statistical Properties of Fundamental Graph Descriptors

Some useful descriptors of the network graph can be conveniently represented in terms of the matrix G in (25). In particular, the (interaction) neighborhood of agent i (which includes i itself) is defined as:

$$\mathcal{N}_i = \{\ell \in \{1, 2, \dots, N\} : g_{i\ell} = 1\}, \quad (28)$$

while the *degree* of agent i is:

$$d_i = |\mathcal{N}_i| = \sum_{\ell=1}^N g_{i\ell} = 1 + \sum_{\ell \neq i} g_{i\ell} = 1 + \sum_{\ell \neq i} g_{\ell i}. \quad (29)$$

We define the network maximal degree as:

$$d_{\max} \triangleq \max_{i=1,2,\dots,N} d_i. \quad (30)$$

Let us now highlight some useful statistical properties of the degree and maximal degree variables for the Erdős-Rényi model.

We observe from (29) that, for an Erdős-Rényi graph, the random variable $d_i - 1$ (we recall that boldface notation is used to emphasize randomness) is a binomial random variable with parameters $N - 1$ and p_N , which shall be denoted by $\beta(N - 1, p_N)$. Note also that d_i and d_j , for $i \neq j$, are *not* independent, because of the implied graph symmetry. Now, for a binomial random variable $\beta(N, p_N)$, we have:

$$\mathbb{E} \left[\frac{\beta(N, p_N)}{N p_N} \right] = 1, \quad \mathbb{V} \left[\frac{\beta(N, p_N)}{N p_N} \right] = \frac{N p_N (1 - p_N)}{(N p_N)^2}. \quad (31)$$

Under the $\mathcal{G}^*(N, p_N)$ model, we see from (27) that the product $N p_N$ diverges as $N \rightarrow \infty$, and, hence, the variance in (31) vanishes as $N \rightarrow \infty$, implying in particular the following convergence in probability [47]:

$$\boxed{\frac{\beta(N, p_N)}{N p_N} \xrightarrow{p} 1} \quad (32)$$

which reveals that the degrees of the nodes scale as $N p_N$.

It is also of interest to characterize the asymptotic behavior of d_{\max} , which, by being the maximum of N degrees, is expected to grow faster than $N p_N$. However, and interestingly, the following lemma shows that it cannot grow *much* faster.

Lemma 1 (Behavior of maximal degree): Under the $\mathcal{G}^*(N, p_N)$ model we have, for all $i, j = 1, 2, \dots, N$, with $i \neq j$:

$$\boxed{\mathbb{P}[d_{\max} \geq N p_N e \mid g_{ij} = 1] \leq \left(e + \frac{2e^2}{N} \right) e^{-c_N} \xrightarrow{N \rightarrow \infty} 0} \quad (33)$$

where e is Euler's number.

Proof: See Appendix B. ■

According to (33), the maximal degree exceeds the level $N p_N e$ with negligible probability, i.e., it cannot grow substantially faster than $N p_N$. We note that the probability in (33) is computed conditionally on the event that two agents interact. This choice is made because, in the following, we need to know the behavior of the maximal degree in relation to *interacting* agent pairs.

C. Stable Combination Policies

In order to characterize the asymptotic behavior of the (off-diagonal) nonzero entries in the combination matrix, we start by examining a popular combination policy, known as the Laplacian rule, which is given by [13]:

$$a_{ij} = \begin{cases} g_{ij} (1 - \mu) \lambda / d_{\max}, & \text{for } i \neq j, \\ (1 - \mu) - \sum_{\ell \neq i} a_{i\ell} & \text{for } i = j. \end{cases} \quad (34)$$

for some λ , with $0 < \lambda \leq 1$. Therefore, from Lemma 1 we can write, for all $i \neq j$:

$$\begin{aligned} & \mathbb{P}[N p_N a_{ij} > (1 - \mu) \lambda / e \mid g_{ij} = 1] \\ &= \mathbb{P}[d_{\max} < N p_N e \mid g_{ij} = 1] \\ &\geq 1 - \left(e + \frac{2e^2}{N} \right) e^{-c_N} \xrightarrow{N \rightarrow \infty} 1. \end{aligned} \quad (35)$$

Equation (35) has the following important implication: for large enough N , any nonzero entry of the Laplacian combination matrix, scaled by the factor Np_N , stays “almost always” above a certain threshold (namely, the value $(1 - \mu)\lambda/e$). Therefore, multiplying a_{ij} by the scaling factor Np_N would keep the nonzero entries stable (in the sense that they will not vanish) as N diverges. The same scaling factor is relevant for other combination policies. Therefore, it makes sense to introduce the following general class of combination policies.

Combination-policy class \mathcal{C}_τ . A combination policy belongs to class \mathcal{C}_τ if there exists $\tau > 0$ such that, for all $i, j = 1, 2, \dots, N$, with $i \neq j$:

$$\mathbb{P}[Np_N \mathbf{a}_{ij} > \tau \mid \mathbf{g}_{ij} = 1] \geq 1 - \epsilon_N \quad (36)$$

where ϵ_N goes to zero as $N \rightarrow \infty$, and where the probability is evaluated under the $\mathcal{G}^*(N, p_N)$ model. \square

Let us now examine the physical meaning of (36) in connection to network tomography applications. For a policy belonging to class \mathcal{C}_τ , we can rephrase (24) as:

$$Np_N \hat{a}_{ij}^{(\text{obs})} = \begin{cases} \underbrace{Np_N a_{ij}^{(\text{obs})}}_{\text{not vanishing}} + \underbrace{Np_N e_{ij}}_{\text{small quantity?}}, & \text{if } a_{ij}^{(\text{obs})} > 0, \\ \underbrace{Np_N e_{ij}}_{\text{small quantity?}}, & \text{if } a_{ij}^{(\text{obs})} = 0, \end{cases} \quad (37)$$

where the qualification of being “not vanishing” is a consequence of (36). In light of (37), if we will be able to show that $Np_N e_{ij}$ is *still* a small quantity, then $\hat{a}_{ij}^{(\text{obs})}$ would be effectively useful for tomography purposes, because the nonzero entry $Np_N a_{ij}^{(\text{obs})}$ would stand out from the error floor as N gets large. Actually, this heuristic argument will be made rigorous in the proof of Theorem 2, which is presented in the next section.

Before ending this section, we would like to introduce another useful combination policy that belongs to class \mathcal{C}_τ , namely, the Metropolis rule, which is given by [13]:

$$a_{ij} = \begin{cases} g_{ij} (1 - \mu) / \max(d_i, d_j), & \text{for } i \neq j, \\ (1 - \mu) - \sum_{\ell \neq i} a_{i\ell}, & \text{for } i = j. \end{cases} \quad (38)$$

Since $\max(d_i, d_j) \leq d_{\max}$, for the Metropolis rule we have the following implication, for all $i \neq j$:

$$\begin{aligned} \{d_{\max} < Np_N e, \mathbf{g}_{ij} = 1\} \\ \Rightarrow \{\max(d_i, d_j) < Np_N e, \mathbf{g}_{ij} = 1\} \\ = \{Np_N \mathbf{a}_{ij} > (1 - \mu)/e, \mathbf{g}_{ij} = 1\}, \end{aligned} \quad (39)$$

where the latter equality comes from (38). Now, since for two events \mathcal{E}_1 and \mathcal{E}_2 , the condition $\mathcal{E}_1 \Rightarrow \mathcal{E}_2$ implies that $\mathbb{P}[\mathcal{E}_2] \geq \mathbb{P}[\mathcal{E}_1]$, from (39) we can write:

$$\begin{aligned} \mathbb{P}[Np_N \mathbf{a}_{ij} > (1 - \mu)/e \mid \mathbf{g}_{ij} = 1] \\ \geq \mathbb{P}[d_{\max} < Np_N e \mid \mathbf{g}_{ij} = 1] \\ \geq 1 - \left(e + \frac{2e^2}{N}\right) e^{-cN} \xrightarrow{N \rightarrow \infty} 1, \end{aligned} \quad (40)$$

where the latter inequality follows by Lemma 1. Equation (40) reveals that the Metropolis rule satisfies (36) with the choice $\tau = (1 - \mu)/e$.

V. CONSISTENT TOMOGRAPHY

In order to establish whether the estimated matrix, $\hat{A}^{(\text{obs})}$, can be used to infer the interaction pattern contained in $A^{(\text{obs})}$, we still need to provide a statistical characterization for the entries of $\hat{A}^{(\text{obs})}$. We pursue this goal by characterizing the asymptotic behavior of two conditional distributions: the empirical distribution given that two agents do not interact, and the empirical distribution given that two agents interact. In particular, owing to the Np_N scaling factor that we have obtained in the previous section, we shall focus on the *scaled* matrix $Np_N \hat{A}^{(\text{obs})}$.

Let us preliminarily introduce the number of non-interacting (N_0) and the number of interacting (N_1) agent pairs over the observed set, defined respectively as:

$$N_0 \triangleq \sum_{i=1}^K \sum_{j \neq i} (1 - \mathbf{g}_{ij}^{(\text{obs})}), \quad N_1 \triangleq \sum_{i=1}^K \sum_{j \neq i} \mathbf{g}_{ij}^{(\text{obs})}, \quad (41)$$

where $\mathbf{g}_{ij}^{(\text{obs})} = \mathbb{I}\{\mathbf{a}_{ij}^{(\text{obs})} > 0\}$ is the (i, j) -th entry of the interaction sub-matrix corresponding to the observable agents. Next we introduce the number of entries in $Np_N \hat{A}^{(\text{obs})}$ that stay below some positive level α , for the case of non-interacting and interacting agent pairs, respectively:

$$N_0(\alpha) \triangleq \sum_{i=1}^K \sum_{j \neq i} \mathbb{I}\{Np_N \hat{a}_{ij}^{(\text{obs})} \leq \alpha, \mathbf{g}_{ij}^{(\text{obs})} = 0\}, \quad (42)$$

$$N_1(\alpha) \triangleq \sum_{i=1}^K \sum_{j \neq i} \mathbb{I}\{Np_N \hat{a}_{ij}^{(\text{obs})} \leq \alpha, \mathbf{g}_{ij}^{(\text{obs})} = 1\}. \quad (43)$$

Using (41), (42) and (43), we can compute the conditional empirical distributions given that the agents are not interacting and given that they are interacting, defined respectively as:

$$\mathcal{F}_0(\alpha) = \frac{N_0(\alpha)}{N_0}, \quad \mathcal{F}_1(\alpha) = \frac{N_1(\alpha)}{N_1}, \quad (44)$$

where $\mathcal{F}_0(\alpha)$ (resp., $\mathcal{F}_1(\alpha)$) is conventionally set to $1/2$ when $N_0 = 0$ (resp., $N_1 = 0$). It is also convenient to introduce the complementary distribution $\bar{\mathcal{F}}_1(\alpha) = 1 - \mathcal{F}_1(\alpha)$. Accordingly, the quantity $\mathcal{F}_0(\alpha)$ represents the fraction of entries in $Np_N \hat{A}^{(\text{obs})}$ that correspond to *non-interacting* agent pairs and that stay *below* α . In contrast, the quantity $\mathcal{F}_1(\alpha)$ represents the fraction of entries in $Np_N \hat{A}^{(\text{obs})}$ that correspond to *interacting* agent-pairs and that stay *above* α .

The next theorem establishes achievability of consistent tomography through the asymptotic characterization of the aforementioned empirical distributions.

Theorem 2 (Achievability of consistent tomography): Let the network interaction profile obey a $\mathcal{G}^*(N, p_N)$ model, and let the combination policy belong to class \mathcal{C}_τ . Then, for a given $\xi > 0$ in (8), we have in the limit as the network size increases ($N \rightarrow \infty$):

$$\mathcal{F}_0(\epsilon) \xrightarrow{P} 1 \quad \forall \epsilon > 0, \quad \bar{\mathcal{F}}_1(\tau) \xrightarrow{P} 1 \quad (45)$$

Proof: See Appendix C. \blacksquare

Theorem 2 reveals that, from the knowledge of the estimated combination matrix, $\hat{A}^{(\text{obs})}$, it is possible to separate the zero/nonzero entries of the true combination matrix, $A^{(\text{obs})}$. In

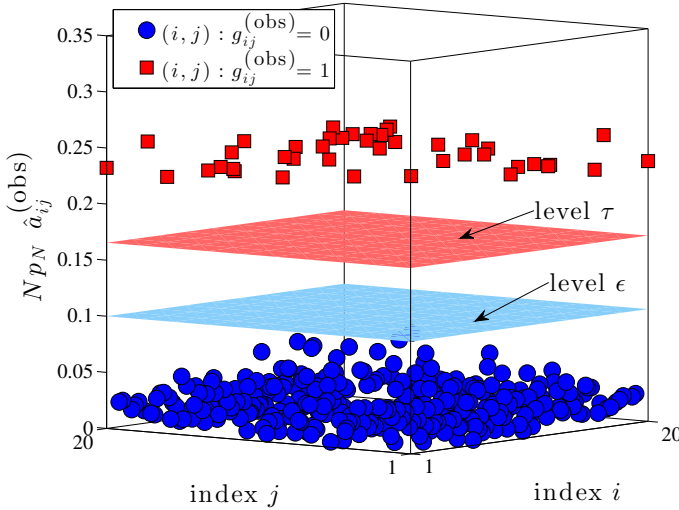


Fig. 2. A pictorial illustration of Theorem 2, which establishes that the scaled estimated entries $\hat{a}_{ij}^{(obs)}$ cluster in two groups depending on whether agents i and j are interacting or not.

fact, we see from Theorem 2 that the following dichotomous behavior is observed, asymptotically with N : *i*) when agents i and j are *not* interacting, most of the scaled estimated matrix entries, $Np_N \hat{a}_{ij}^{(obs)}$, stay below *any* level ϵ (and, hence, also below an arbitrarily small value $\epsilon < \tau$); *ii*) when agents i and j are interacting, most of the scaled estimated matrix entries stay above a positive value τ . Therefore, a separation between the two classes of non-interacting and interacting agent pairs arises, which translates into the emergence of two separate clusters, one corresponding to the region $Np_N \hat{a}_{ij}^{(obs)} \leq \epsilon$, and the other one corresponding to the region $Np_N \hat{a}_{ij}^{(obs)} > \tau$. This situation is illustrated in Fig. 2.

Theorem 2 establishes the separability of the two classes based upon knowledge of $\hat{A}^{(obs)}$, whose computation requires knowledge of the *exact* correlation sub-matrices, $[R_0]_\Omega$ and $[R_1]_\Omega$ — see (16). Since, in principle, such matrices can be estimated with arbitrarily large precision as the steady-state is approached (i.e., as the number of collected diffusion output samples increases) the result in Theorem 2 is an *achievability* result. In addition, for many of the methods available to estimate correlation matrices, the behavior of the estimation error is known, at least for n sufficiently large. A useful extension of the present treatment would be examining the interplay between the two sources of error, namely, the error caused by partial observations, and the error caused by estimation of the correlation sub-matrices.

When some prior knowledge about the value of τ and the typical number of neighbors, Np_N , is available, Theorem 2 provides a constructive recipe to perform the reconstruction of the interaction profile across the observed network. In fact, the separation between the two classes of interacting and non-interacting agent pairs can be performed by comparing each entry $\hat{a}_{ij}^{(obs)}$ to an intermediate threshold lying between 0 and $\tau/(Np_N)$.

On the other hand, in many contexts it is unrealistic to assume precise knowledge of the parameters τ and Np_N .

When such knowledge is not available, the aforementioned reconstruction strategy is not applicable, but the achievability result in Theorem 2 still has a fundamental implication in that it guarantees the existence of the clustering structure. The existence of two separate thresholds, $\epsilon < \tau$, and the related (asymptotic) separation of the scaled estimated entries in two separate clusters ($Np_N \hat{a}_{ij}^{(obs)} \leq \epsilon$ and $Np_N \hat{a}_{ij}^{(obs)} > \tau$), opens up the possibility of employing universal and non-parametric pattern recognition strategies to perform cluster separation. In particular, in our numerical experiments, we shall verify the validity of this argument by employing the k -means clustering algorithm.

VI. HIGHER ORDER ASYMPTOTIC ANALYSIS

As illustrated in the previous section, and as we can infer from Fig. 2, it is undesirable to have $Np_N \hat{a}_{ij}^{(obs)} > \epsilon$ when $g_{ij} = 0$. For the sake of brevity, let us refer to the agent pair for which such event occurs as being identified as “mistakenly-interacting”. Examining (41), (42) and (44), we deduce that the number of mistakenly-interacting pairs is given by:

$$N_0 - N_0(\epsilon) = N_0[1 - \mathcal{F}_0(\epsilon)] \approx K^2 \underbrace{[1 - \mathcal{F}_0(\epsilon)]}_{\rightarrow 0, \text{ see (45)}}, \quad (46)$$

where we used the fact that, under the $\mathcal{G}^*(N, p_N)$ model, N_0 scales as $K(K-1)(1-p_N)$, and p_N vanishes as $N \rightarrow \infty$.

On the other hand, for the number of *truly-interacting* agent pairs we have that:

$$N_1 \approx K^2 p_N. \quad (47)$$

The fact that p_N vanishes as N increases causes the following issue. According to (46)–(47) and also to (85)–(86), without any information about the speed of convergence of $1 - \mathcal{F}_0(\epsilon)$ (in comparison to p_N), we cannot exclude that the number of *mistakenly-interacting* pairs is larger than the number of *truly-interacting* pairs. In order to ward off the presence of such unpleasant feature, we should prove that the quantity $1 - \mathcal{F}_0(\epsilon)$ goes to zero faster than p_N , formally:

$$\frac{1 - \mathcal{F}_0(\epsilon)}{p_N} \xrightarrow{p} 0, \quad (48)$$

a condition that will be abbreviated as $1 - \mathcal{F}_0(\epsilon) = o_p(p_N)$. Verification of such desired requirement is addressed in the forthcoming Proposition 1. In order to prove Proposition 1, we require that the combination policy possesses two additional regularity properties, which are again automatically satisfied by the Laplacian and Metropolis rules.

Property P1. There exists $\kappa > 0$ such that, for all $i, j = 1, 2, \dots, N$, with $i \neq j$:

$$\mathbf{a}_{ij} \leq \frac{\kappa}{d_i} \mathbf{g}_{ij} \quad (49)$$

with the inequality being trivially an equality for $\mathbf{g}_{ij} = 0$. \square

It is useful to make the following comparison between (36) and (49). Equation (36) means that the nonzero entries in the combination matrix, scaled by Np_N , do not collapse to zero, i.e., they are stable *from below*. On the other hand, in view of (32), the upper bound in (49) implies that the

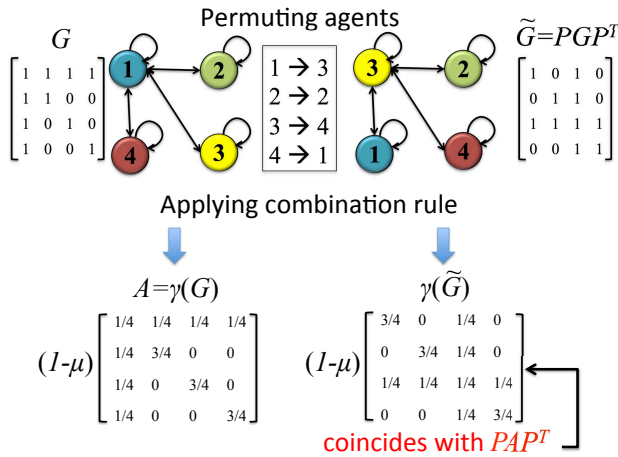


Fig. 3. Illustration of property P2: permuting the agents, leads to a permuted version of the combination policy.

nonzero entries in the combination matrix, scaled by Np_N , are asymptotically stable *from above*. This is because d_i approaches Np_N asymptotically as $N \rightarrow \infty$ so that (49) translates into the asymptotic condition that $Np_N a_{ij} \leq \kappa$.

Moreover, it is straightforward to verify that (49) holds for the Laplacian rule, with $\kappa = (1 - \mu)\lambda$, as well as for the Metropolis rule, with $\kappa = 1 - \mu$.

Property P2. Consider a certain permutation of the agents, i.e., a renumbering of the rows and columns of the interaction matrix G , which leads to the interaction matrix \tilde{G} . Such operation can be represented by means of a permutation matrix P (see Appendix E), as $\tilde{G} = PGP^T$. Property P2 holds if:

$$\gamma(PGP^T) = P\gamma(G)P^T, \quad (50)$$

namely, if applying the combination policy $\gamma(\cdot)$ to the *renumbered* interaction matrix is equivalent to renumbering (through the same P) the initial combination matrix $A = \gamma(G)$. \square

The meaning of property P2 is illustrated in Fig. 3. In the leftmost panel, we represent a network graph, along with the corresponding combination matrix obtained by applying the Metropolis rule. In the rightmost panel, the agents are exchanged as detailed in the figure, using the permutation matrix:

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (51)$$

Then, the Metropolis rule is applied to the new (i.e., renumbered) interaction matrix. It is seen that the resulting combination matrix corresponds to renumbering the original combination matrix using the permutation (51).

Property P2 is particularly relevant for the following reasons. First, under the Erdős-Renyi model, the statistical distribution of the interaction matrix is invariant to agents' permutations. Owing to (50), such invariance is automatically inherited by the combination matrix. Moreover, property P2 is satisfied by typical combination policies, such as the Laplacian rule and the Metropolis rule. It is also useful to provide a counterexample that shows why property P2 is not always

verified. Consider a network with three agents, and with the following interaction matrix:

$$G = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad (52)$$

The combination policy of our counterexample works as follows. First, a Metropolis rule is applied. Second, the resulting self-combination weight assigned to agent 1 is slightly increased by adding a small extra-weight. The extra-weight assigned to agent 1 is then subtracted, in equal parts, from the other nonzero weights of agent 1, in order to meet the requirement $\sum_{\ell \neq 1} a_{1\ell} = 1 - \mu$. Finally, the other entries of the combination matrix are adjusted so as to make $A/(1 - \mu)$ symmetric and right-stochastic. The final result is (with $\epsilon \ll 1$):

$$A = (1 - \mu) \begin{bmatrix} 2/3 + \epsilon & 1/3 - \epsilon & 0 \\ 1/3 - \epsilon & 1/3 + \epsilon & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix} \quad (53)$$

Assume now that agents 1 and 2 are exchanged, which would yield the following interaction matrix:

$$G = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad (54)$$

Applying the combination policy described before, we would end up with the following combination matrix:

$$A = (1 - \mu) \begin{bmatrix} 1/3 + \epsilon & 1/3 - \epsilon/2 & 1/3 - \epsilon/2 \\ 1/3 - \epsilon/2 & 2/3 + \epsilon/2 & 0 \\ 1/3 - \epsilon/2 & 0 & 2/3 + \epsilon/2 \end{bmatrix} \quad (55)$$

We see that: *i*) the interaction matrix in (54) is a renumbered version of the interaction matrix in (52), which corresponds to exchanging agents 1 and 2, while *ii*) the combination matrix in (55) is *not obtained by applying the same renumbering* to the combination matrix in (53). Therefore, property P2 is violated. The presented counterexample shows that, while in practice constructing a combination policy that violates property P2 might look rather artificial, from a purely theoretical viewpoint it must be stated that property P2 is not always verified.

A combination policy satisfying the additional properties P1 and P2 describes the following class of combination policies.

Combination-policy class \mathcal{C}_τ^l . A combination policy belongs to class \mathcal{C}_τ^l if it belongs to class \mathcal{C}_τ and possesses properties P1 and P2. \square

Proposition 1, which is stated below, will be proved by resorting to an approximation that will be referred to as the *independence approximation*. More specifically, the proof of the proposition will rely on characterizing the variance of the entries in the error matrix (see Appendix D). Unfortunately, the exact evaluation of the error variance is generally a formidable task. In order to gain insight into the asymptotic behavior of the variance, in the proof of Proposition 1 we make a simplified evaluation by treating the entries of the various matrices involved in the calculations as *independent* random variables. The accuracy of the results arising from this approximation is tested by means of numerical experiments

— see Fig. 9 in Appendix D. A more rigorous analysis would require estimating the order of the error introduced by the independence approximation.

Proposition 1 (Rate of convergence of $1 - \mathcal{F}_0(\epsilon)$): Let the network interaction profile obey a $\mathcal{G}^*(N, p_N)$ model, and let the combination policy belong to class \mathcal{C}'_τ . Then, the result in (45) holds true because $\mathcal{C}'_\tau \subset \mathcal{C}_\tau$. In addition, for a given $\xi > 0$ in (8), and for all $\epsilon > 0$, we have:

$$1 - \mathcal{F}_0(\epsilon) \approx o_p(p_N) \quad (56)$$

where the approximation in (56) arises from the independence approximation used in the proof.

Proof: See Appendix D. ■

VII. ILLUSTRATIVE EXAMPLES

We now examine the practical significance of the asymptotic results derived in the previous sections, with reference to three combination matrices that are rather popular in the literature of adaptive networks. The first two strategies lead to symmetric combination matrices, which therefore match the hypotheses of our theorems. The third strategy corresponds to an asymmetric combination matrix. Even if the asymmetric case is not contemplated by our theorems, it is relevant in practice and, as we shall see from the forthcoming experiments, consistent tomography works also for such a relevant case. The presentation of the examples is organized through the following steps.

S1) We consider first the case that the projections of the correlation matrices, R_0 and R_1 , onto the observable part of the network, are available without error. For this case, we compute the inversion of the observable part, which leads to the matrix $\hat{A}^{(\text{obs})}$, through:

$$\hat{A}^{(\text{obs})} = R_1^{(\text{obs})} (R_0^{(\text{obs})})^{-1} \quad (57)$$

S2) We use the off-diagonal entries of $\hat{A}^{(\text{obs})}$, and apply a k -means clustering algorithm in order to split these entries into two clusters. The cluster with smallest arithmetic average is labeled as “non-interacting”, while the other cluster is labeled as “interacting”. We remark that such *classification strategy is implemented in a universal, fully non-parametric way.*

S3) Then, we enlarge the setting to the case that the projections of the correlation matrices are estimated from the diffusion outputs. In the simulations, the observations fed into the diffusion algorithm, $\{\mathbf{x}_i(n)\}_{i,n}$, follow a standard normal distribution. As an estimator for $R_0^{(\text{obs})} = [R_0]_\Omega$, we use the empirical correlation, namely:

$$\hat{R}_0^{(\text{obs})} = \frac{1}{n} Y Y^T, \quad (58)$$

where, for $i = 1, 2, \dots, K$, the i -th row of the $K \times n$ matrix Y is given by:

$$\mathbf{y}_{\omega_i}(1), \mathbf{y}_{\omega_i}(2), \dots, \mathbf{y}_{\omega_i}(n), \quad (59)$$

with the indices $\omega_1 < \omega_2 < \dots < \omega_K$ spanning the observable set Ω . The estimate $\hat{R}_1^{(\text{obs})}$ is computed similarly.

We remark that in this work we do not focus on optimizing the performance of the correlation matrix estimators, since our focus is on ascertaining the fundamental limits of tomography. There are already considerable works in the literature on perfecting correlation estimation from ensemble data. One challenge that arises in the network case is the interplay between the matrix size and the number of samples used for the estimation.

Now, using (57) with the true correlation matrices replaced by their estimated counterparts, we get the following estimate:

$$\hat{A}^{(\text{obs})} = \hat{R}_1^{(\text{obs})} (\hat{R}_0^{(\text{obs})})^{-1} \quad (60)$$

S4) We run the k -means clustering algorithm over the entries of $\hat{A}^{(\text{obs})}$ in (60).

A. Laplacian Combination Rule

Under the Laplacian combination rule, the off-diagonal combination weights are zero when the agents are not interacting, and are otherwise equal to a constant across the network. Therefore, we see that the weights reflect perfectly the structure of the underlying network graph. In fact, several important properties of the graph are encoded in the properties of the Laplacian matrix [21].

In Fig. 4, leftmost panel, we display the off-diagonal entries of the (scaled) *true* combination matrix, $N p_N A^{(\text{obs})}$, corresponding to the *observable* network portion. For clarity of presentation, the matrix has been vectorized by means of column-major ordering, and the (vectorized) (i, j) pairs have been rearranged in such a way that the zero entries appear before the nonzero entries. The same ordering used for $A^{(\text{obs})}$ will be then applied to the matrices displayed in the remaining panels, i.e., the horizontal axis is homogeneous across the different panels, so as to allow a fair comparison. If agents i and j do not interact, the pertinent matrix entry is marked with a blue circle, whereas if they do interact, a red square is used. The observed step-function behavior comes from the fact that, for the Laplacian combination rule, the nonzero weights are constant across the network.

In the middle panel we focus on steps S1) and S2): we display the scaled estimated matrix, $N p_N \hat{A}^{(\text{obs})}$, computed under perfect knowledge of $R_0^{(\text{obs})}$ and $R_1^{(\text{obs})}$, and we display the classification performed by the k -means algorithm. In the rightmost panel we focus instead on steps S3) and S4): we display the scaled estimated matrix, $N p_N \hat{A}^{(\text{obs})}$, computed with the *empirical estimates* $\hat{R}_0^{(\text{obs})}$ and $\hat{R}_1^{(\text{obs})}$, and we display the classification performed by the k -means algorithm. In the latter two panels, matrix entries are marked with different colors and symbols, depending on the results of the k -means clustering algorithm: blue-circle markers if agents i and j have been *classified as* non-interacting, and red-square markers if agents i and j have been *classified as* interacting.

The network considered in Fig. 4 consists of $N = 100$ agents, where only $K = 20$ agents are observable ($\xi = 0.2$). According to the connection properties of the Erdős-Rényi model, see (27), the interaction probability is chosen as $p_N = 2 (\ln N) / N \approx 0.092$. The parameter of the Laplacian

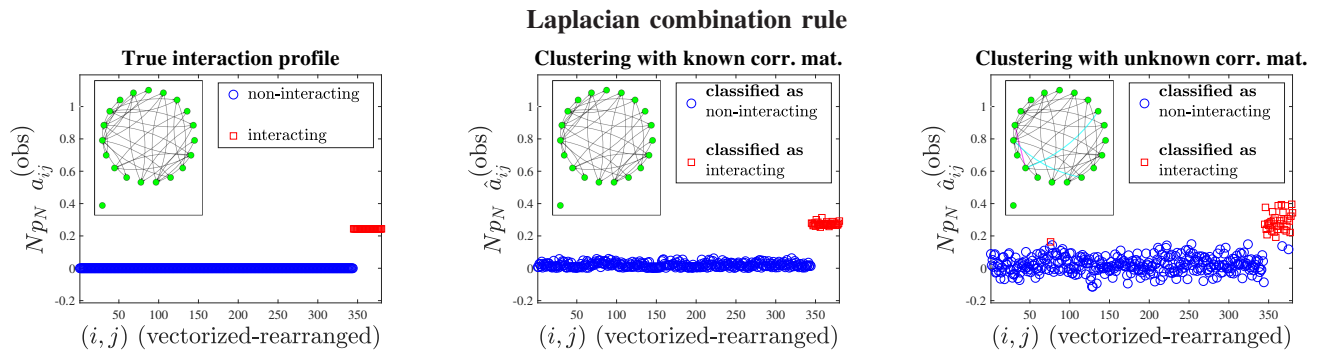


Fig. 4. Network tomography for the case of a Laplacian combination rule with parameter $\lambda = 0.5$ — see Sec. VII-A. The network consists of $N = 100$ agents, where only $K = 20$ agents are observable ($\xi = 0.2$). The interaction probability is $p_N = 2(\ln N)/N \approx 0.092$, and the value of the step-size is $\mu = 0.1$. **Leftmost panel:** the *true* combination matrix, $A^{(\text{obs})}$, is vectorized with column-major ordering, and the (vectorized) (i, j) pairs are rearranged in such a way that the zero entries come before the nonzero entries (the same ordering is applied to the estimated matrix, $\hat{A}^{(\text{obs})}$, in the other two panels). The different markers highlight the *true* interaction profile of the *observable* network portion. **Middle panel:** matrix $\hat{A}^{(\text{obs})}$, computed under perfect knowledge of the steady-state correlation matrices of the *observable* diffusion output, see (57). The different markers highlight the interaction profile as *reconstructed* by the k -means algorithm. **Rightmost panel:** same as middle panel, with matrix $\hat{A}^{(\text{obs})}$ computed using the correlation matrices *estimated empirically* with $n = 2 \times 10^4$ samples, see (58), (59), and (60). **Inset plots:** interaction profiles represented through the corresponding network graphs. In the inset plots of the middle and of the rightmost panels, erroneously detected edges (the edge is not present but it is “seen” by the tomography algorithm) are marked in magenta, while missed edges (the edge is present but the tomography algorithm misses it) are marked in cyan.

combination matrix is $\lambda = 0.5$. As regards the diffusion algorithm, we choose a typical value for the step-size, i.e., $\mu = 0.1$.

Let us begin with examining the output of steps S1) and S2), middle panel. As we can see, the experiments confirm *in toto* what is expected from the theoretical analysis: the entries of the matrix $\hat{A}^{(\text{obs})}$ appear to be well-separable, since: *i*) the unavoidable error induced by *partial* observation of the network, is relatively small, implying that zero entries of $\hat{A}^{(\text{obs})}$ are concentrated around zero; *ii*) the nonzero entries of the *true* combination matrix (leftmost panel) are bounded from below, and, since the error e_{ij} is positive, this fact keeps the nonzero entries of $\hat{A}^{(\text{obs})}$ (middle panel) sufficiently far away from the origin.

Next we move on to steps S3) and S4), namely, we move on to examining the behavior in the presence of imperfect knowledge of the correlation matrices. In particular, we use $n = 2 \times 10^4$ samples to perform estimation of the correlation matrix from the diffusion output. This situation is examined in the rightmost panel of Fig. 4. By comparison with the middle panel, we see that the estimated clusters are more “noisy”, which makes sense since the procedure *must* be affected by the error in estimating $R_0^{(\text{obs})}$ and $R_1^{(\text{obs})}$. This notwithstanding, tomography is still effective, meaning that the number of observations collected to estimating the correlation matrices is sufficiently high. Few classification errors are committed: the red-square markers appearing among the blue-circle markers, and the blue-circle marker appearing among the red-square markers (approximately at position 350), represent mistakenly classified agent pairs.

The results arising from the above example are collected, perhaps in a more revealing form, in the inset plots of Fig. 4. The displayed network graphs (corresponding only to the observable subnet) are drawn with the following rules. An edge drawn from j to i means that agent i is influenced (leftmost panel) or is *estimated to be* influenced (middle and rightmost panels) by agent j . When an edge is erroneously detected (i.e.,

the edge is in fact not present but the tomography algorithm “sees” it), it is marked in magenta. Likewise, when an edge is not detected (i.e., the edge is present but the tomography algorithm misses it), it is marked in cyan. The impact of imperfect knowledge of the correlation matrices can be noticed in the inset plot, where we see that some errors are committed in reconstructing the profile of the observed subnet.

B. Metropolis Combination Rule

In Fig. 5, the analysis presented in the previous section is applied to a different combination matrix, namely, the Metropolis combination matrix. This example is useful because, in a Metropolis rule, the nonzero weights are no longer constant, and exhibit a certain dynamics related to the different neighborhood sizes corresponding to the different agents. It is important to examine the impact of such dynamics on the performance of the tomography algorithm. From the leftmost panel of Fig. 5, we see that the nonzero entries of the *true* combination matrix exhibit a certain variability. This notwithstanding, since the Metropolis rule belongs to class \mathcal{C}_τ , with $\tau = (1 - \mu)/e$, the scaled *nonzero* entries are still lower bounded in probability. The separability of the clusters is preserved, and considerations similar to those drawn in the previous section as regards the Laplacian rule apply.

Still, there is an important difference between the two kinds of combination policies. By inspecting carefully the rightmost panel in Fig. 5 and the related inset, we see that, differently from what happened in the rightmost panel of Fig. 4, the network profile is reconstructed perfectly. This suggests that estimation of the correlation matrices is easier for Metropolis rules than for Laplacian rules. The latter effect could be probably ascribed to the improved convergence properties of Metropolis combination matrices (with respect to Laplacian combination matrices), which should allow a more accurate estimation of the correlation matrices for a given number of samples.

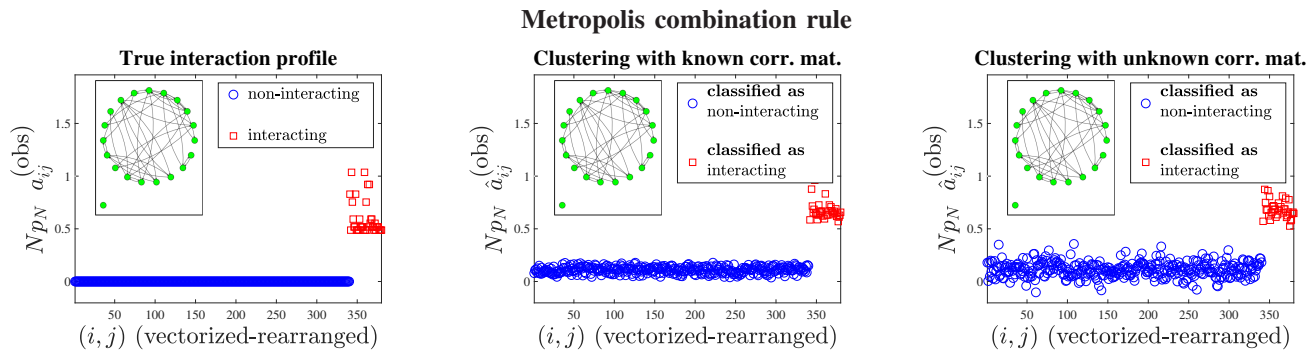


Fig. 5. Network tomography for the case of a Metropolis combination rule addressed in Sec. VII-B. The network consists of $N = 100$ agents, where only $K = 20$ agents are observable ($\xi = 0.2$). The interaction probability is $p_N = 2(\ln N)/N \approx 0.092$, and the value of the step-size is $\mu = 0.1$. **Leftmost panel:** the *true* combination matrix, $A^{(\text{obs})}$, is vectorized with column-major ordering, and the (vectorized) (i, j) pairs are rearranged in such a way that the zero entries come before the nonzero entries (the same ordering is applied to the estimated matrix, $\hat{A}^{(\text{obs})}$, in the other two panels). The different markers highlight the *true* interaction profile of the *observable* network portion. **Middle panel:** matrix $\hat{A}^{(\text{obs})}$, computed under perfect knowledge of the steady-state correlation matrices of the *observable* diffusion output, see (57). The different markers highlight the interaction profile as *reconstructed* by the k -means algorithm. **Rightmost panel:** same as middle panel, with matrix $\hat{A}^{(\text{obs})}$ computed using the correlation matrices *estimated empirically* with $n = 2 \times 10^4$ samples, see (58), (59), and (60). **Inset plots:** interaction profiles represented through the corresponding network graphs. In the inset plots of the middle and of the rightmost panels, erroneously detected edges (the edge is not present but it is “seen” by the tomography algorithm) are marked in magenta, while missed edges (the edge is present but the tomography algorithm misses it) are marked in cyan.

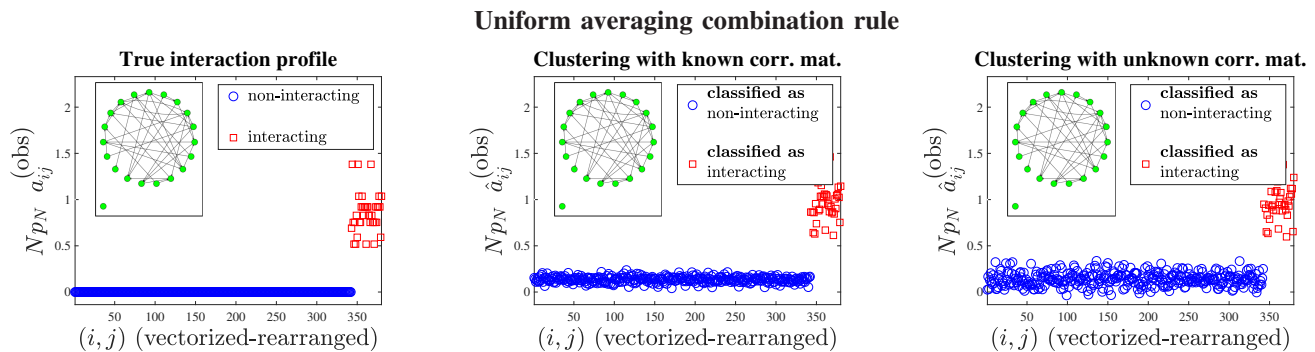


Fig. 6. Network tomography for the case of a uniform averaging combination rule addressed in Sec. VII-C. The network consists of $N = 100$ agents, where only $K = 20$ agents are observable ($\xi = 0.2$). The interaction probability is $p_N = 2(\ln N)/N \approx 0.092$, and the value of the step-size is $\mu = 0.1$. **Leftmost panel:** the *true* combination matrix, $A^{(\text{obs})}$, is vectorized with column-major ordering, and the (vectorized) (i, j) pairs are rearranged in such a way that the zero entries come before the nonzero entries (the same ordering is applied to the estimated matrix, $\hat{A}^{(\text{obs})}$, in the other two panels). The different markers highlight the *true* interaction profile of the *observable* network portion. **Middle panel:** matrix $\hat{A}^{(\text{obs})}$, computed under perfect knowledge of the steady-state correlation matrices of the *observable* diffusion output, see (57). The different markers highlight the interaction profile as *reconstructed* by the k -means algorithm. **Rightmost panel:** same as middle panel, with matrix $\hat{A}^{(\text{obs})}$ computed using the correlation matrices *estimated empirically* with $n = 2 \times 10^4$ samples, see (58), (59), and (60). **Inset plots:** interaction profiles represented through the corresponding network graphs. In the inset plots of the middle and of the rightmost panels, erroneously detected edges (the edge is not present but it is “seen” by the tomography algorithm) are marked in magenta, while missed edges (the edge is present but the tomography algorithm misses it) are marked in cyan.

C. Uniform Averaging Rule

One of the most important goals of causation is ascertaining whether one agent influences another agent. Such influence is not necessarily symmetrical. Although we have obtained an analytical characterization of the error matrix for the case of symmetric combination matrices, it is useful to examine whether the conclusions drawn for the symmetric case find some correspondence in the asymmetric case. To this aim, the network interaction profile is now generated according to a *binomial random graph* model, $\mathcal{B}(N, p_N)$, where the variables g_{ij} , for $i = 1, 2, \dots, N$ and $j \neq i$, are independent Bernoulli random variables with $\mathbb{P}[g_{ij} = 1] = p_N$ — see [46]. Such a procedure leads (in general) to an asymmetric matrix G , i.e., to a *directed* graph. As regards the nonzero entries of the combination matrix, we consider the well-known uniform averaging rule. According to such rule, if agent i is influenced by agent j (i.e., if $g_{ij} = 1$), agent i sets $w_{ij} = 1/\sum_{\ell=1}^N g_{i\ell}$, namely,

uniform weights are assigned to all agents that influence directly agent i . We note in passing that, in general, the uniform averaging rule gives rise to an asymmetric combination matrix even if the underlying graph is symmetric. The results of our experiments are reported in Fig. 6, with reference to the case of perfect and imperfect knowledge of the correlation matrices, respectively. Remarkably, the tomography algorithm exhibits proper operation even in the asymmetric case. From the inset plots, we can appreciate that, differently from the previous examples, it is now possible that the influence of an agent over another agent is in general unidirectional.

D. Role of the Step-Size

The step-size, μ , is a fundamental parameter ruling the performance of the adaptive diffusion algorithm implemented by the network agents. As examined in detail in recent works [17], [18], the following adaptation/learning trade-off

The effect of different system parameters

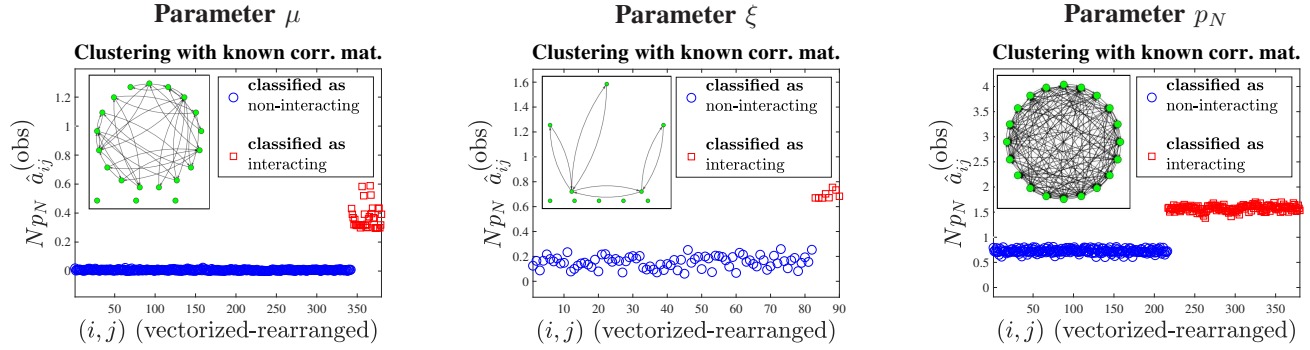


Fig. 7. Network tomography for the case of a Metropolis combination rule addressed in Sec. VII-B, and for a network with $N = 100$ agents. **Leftmost panel:** the case of “high” values for the step-size addressed in Sec. VII-D. The relevant parameters are: $\mu = 0.5$, $\xi = 0.2$, and $p_N = 2(\ln N)/N \approx 0.092$. **Middle panel:** the case of “small” fractions of observed agents addressed in Sec. VII-E. The relevant parameters are: $\xi = 0.1$, $p_N \approx 0.092$, and $\mu = 0.1$. **Rightmost panel:** the case of “high” values for the interaction probability addressed in Sec. VII-F. The relevant parameters are: $p_N \approx 0.46$, $\xi = 0.2$, and $\mu = 0.1$. In all panels, the matrix $\hat{A}^{(\text{obs})}$ (displayed with the same ordering criterion used in the previous figures) is computed under perfect knowledge of the steady-state correlation matrices of the *observable* diffusion output, see (57). The different markers highlight the interaction profile as *reconstructed* by the k -means algorithm. **Inset plots:** interaction profiles represented through the corresponding network graphs. In the inset plots of the lowermost panels, erroneously detected edges (the edge is not present but it is “seen” by the tomography algorithm) are marked in magenta, while missed edges (the edge is present but the tomography algorithm misses it) are marked in cyan.

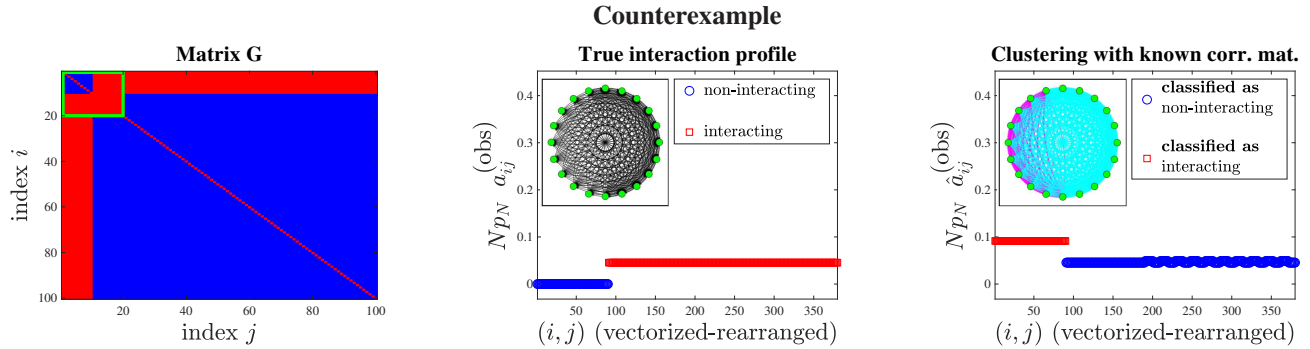


Fig. 8. Counterexample where the tomography algorithm proposed in this work breaks. The adopted combination policy is a Laplacian combination rule with parameter $\lambda = 0.5$ — see Sec. VII-A — and the value of the step-size is $\mu = 0.1$. **Leftmost panel:** Interaction profile (i.e., adjacency matrix) of the network, with a total number of $N = 100$ agents, and $K = 20$ agents in the observable set (square region highlighted by the green contour). The interacting (i, j) pairs are displayed in red, whereas the non-interacting pairs are displayed in blue. **Middle panel:** the *true* combination matrix, $A^{(\text{obs})}$, is vectorized with column-major ordering, and the (vectorized) (i, j) pairs are rearranged in such a way that the zero entries come before the nonzero entries (the same ordering is applied to the estimated matrix, $\hat{A}^{(\text{obs})}$, in the rightmost panel). The different markers highlight the *true* interaction profile of the *observable* network portion. **Rightmost panel:** matrix $\hat{A}^{(\text{obs})}$, computed under perfect knowledge of the steady-state correlation matrices of the *observable* diffusion output, see (57). The different markers highlight the interaction profile as *reconstructed* by the k -means algorithm. **Inset plots:** interaction profiles represented through the corresponding network graphs. In the inset plots of the middle and of the rightmost panels, erroneously detected edges (the edge is not present but it is “seen” by the tomography algorithm) are marked in magenta, while missed edges (the edge is present but the tomography algorithm misses it) are marked in cyan.

arises: small values of μ enhance learning precision, while relatively high values of μ enhance adaptation.

We now examine the impact of the step-size on the tomography algorithm. For the sake of concreteness, we focus on one of the combination rules examined in the previous sections, namely, the Metropolis rule, and consider the noiseless case, namely, steps S1) and S2). In the previous simulations concerning the Metropolis rule (see Fig. 5), we have used $\mu = 0.1$. In the new simulation, reported in Fig. 7 (leftmost panel), we run the diffusion algorithm with $\mu = 0.5$, which turns out to be a relatively high value for the step-size in practical applications.

By joint inspection of Fig. 5 (middle panel) and Fig. 7 (leftmost panel), we can appreciate the presence of different effects. The first effect relates to the amplitude of the nonzero entries. We see that increasing μ corresponds to diminishing

the amplitude of the nonzero entries in the combination matrix. This effect is rather obvious, since our definition of A embodies a scaling factor $1 - \mu$, see (4). The reduction of the distance between zero and nonzero entries could have a negative impact on tomography performance, especially when the correlation matrices must be estimated from the diffusion output, and the additional error associated to such estimation reduces further the separation between the two classes.

The second effect relates to the amplitude of the entries in the error matrix. We see that increasing μ corresponds to diminishing the amplitude of the errors, which could be beneficial for tomography. This effect can be grasped by Theorem 1, since we see that the (columnwise) sum of errors is upper bounded by $1 - \mu$. However, Theorem 1 provides only an upper bound, and, hence, the exact behavior of the error as a function of μ is not known.

As a result, it is not easy to anticipate which of the aforementioned two effects will prevail, and elucidating the trade-offs related to the effect of the step-size on the tomography performance, in connection to the adaptation/learning trade-off, is an interesting issue that deserves further investigation.

E. Degree of Observability

It is not unreasonable to presume that the error induced by *partial* observation of the network depends inversely on the observable fraction of agents, ξ . Accordingly, it is expected that the error grows when the size of the monitored portion of the network decreases, i.e., as ξ decreases. This effect is confirmed by our numerical experiments. A sample of such experiments is reported in Fig. 7 (middle panel). In particular, by comparison with Fig. 5 (middle panel), it is observed that the averages of the errors shift upward for both the zero and the nonzero entries, and that the error's amplitude is generally increased. However, even in the presence of higher errors, separability between the two clusters seems to be preserved.

F. "High" Interaction Probability

The theoretical analysis conducted in this work focused on a classic asymptotic regime for random graphs, namely, the regime where the interaction probability goes to zero as the network size increases. We were able to conclude that, as $N \rightarrow \infty$, the (scaled) error, $Np_N e_{ij}$, is driven to zero in probability.

On the other hand, in practice one obviously operates with fixed (finite) values of N and nonzero values of p_N , and, hence, the error will be never exactly equal to zero. In particular, for a fixed value of N , a higher value of p_N (which, in the asymptotic setting described by (27), would correspond to select a sequence c_N diverging faster), is expected to imply higher values of $Np_N e_{ij}$. Such features are observed in Fig. 7, rightmost panel. Comparison with the middle panel of Fig. 5 (where the interaction probability was smaller) reveals that the error in Fig. 7 is, on average, higher.

However, there is an important aspect that emerges here. We see clearly that the average increase of the error does not impair the capability of the system to discriminate between interacting and non-interacting agent pairs. Clustering is still effective, since the two clusters are merely shifted upward. This suggests the following important observation. In this case, what matters is the *variance* of the error, which should be sufficiently small in order to guarantee separability of the clusters.

G. All that Glitters is Not Gold

In order to gain a deeper understanding of the proposed tomography algorithm, as well as to clarify the connections with the adopted assumptions, we now present a counterexample where the algorithm breaks. In particular, we construct a network of $N = 100$ agents, whose interaction profile (i.e., adjacency matrix) is shown in the leftmost panel of Fig. 8. In this panel, interacting pairs are displayed in red, non-interacting pairs are displayed in blue, and the observable set ($K = 20$ agents) is contained in the green square. We observe

the following peculiar behavior. The agents corresponding to the first half of the observable set (rows from 1 to 10) do not interact with each other, because their mutual combination weights (columns from 1 to 10) are all zero (except for the self-weights on the main diagonal). In contrast, these agents are connected to *all* the remaining agents, as witnessed by the fact that, over the rows from $i = 1$ to $i = 10$, the columns from $j = 11$ to $j = 100$ are all displayed in red.

The behavior of the remaining observable agents (rows from 11 to 20) is completely different, since these agents are connected *only* to, and to *all* the agents within the observable set. In summary, we could say that the non-interacting pairs in the observable set (first ten agents) are severely affected by the unobserved agents, while the interacting pairs in the observable set (agents from 11 to 20) are not directly affected by the unobserved agents.

The main consequence of the aforementioned behavior is clearly visible in the rightmost panel of Fig. 8, which shows the output of the tomography algorithm that works under the favorable case where the correlation matrices are known. By comparison with the middle panel (which shows the true interaction profile), we see that in the rightmost panel the entries of the combination matrix, over the non-interacting pairs, are raised up by an error floor that make them higher than the entries corresponding to the interacting pairs. This effect clearly impairs the tomography algorithm, which is no longer able to discriminate correctly the two classes. This example is useful because it shows at least two useful things: *i*) it explains one physical reason why the algorithm could break, namely, the case that the non-interacting pairs are treated very differently from the interacting pairs, and in such a way that their magnitude is increased so as to make it impossible to identify properly the clusters; *ii*) the ad-hoc way in which we construct the counterexample to violate our assumptions suggests also that, provided a certain degree of isotropy (which is granted, for instance, in the Erdős-Rényi model), in practice the algorithm is expected to work properly.

VIII. CONCLUDING REMARKS AND OPEN ISSUES

We considered the problem of tomography over partially observed diffusion networks, where the goal is to infer whether an agent is directly influenced by another agent. Under the challenging scenario where only the output of the diffusion algorithm is observable (i.e., there is no direct information about who's talking to whom) *and* only a fraction of the agents is accessible, we prove that consistent tomography is possible: as the network size increases, the interacting and non-interacting agent pairs split in two separate clusters, for any given fraction of accessible agents. The theoretical results are proved under some technical assumptions: *i*) the network graph is an Erdős-Rényi graph; *ii*) the combination matrices belong to a certain symmetric class; *iii*) perfect knowledge is assumed for the correlation matrices of the *observable* diffusion output (steady-state regime). As another contribution, we show by means of numerical experiments that the results hold for asymmetric matrices as well (which is useful in the context of *causal* inference), and that the theoretical correlation matrices can be

replaced by their empirical counterparts, estimated from the diffusion outputs observed for a sufficiently long time.

There are many interesting questions and open issues that deserve future investigation. First, it would be useful to generalize the theoretical results to broader settings, including: the case of asymmetric combination matrices; graph generative models different from the Erdős-Rényi model; different classes of combination matrices.

We see from the numerical analysis that the separation between the agents' clusters is very marked, suggesting that vanishing of the error (which, in this work, has been established in terms of the *fraction* of erroneously classified agents) might hold under more restrictive criteria, e.g., a maximum error criterion, where we require that the *largest* entry in the error matrix vanishes. It would be also of interest to see whether the characterization of the result in Proposition 1 can be obtained without resorting to the independence approximation. Moreover, we notice that the additive structure of the entries in the error matrix encourages to pursue an asymptotic normality characterization.

From a more practical perspective, an interesting problem is optimizing the algorithms for tomography (exploiting, when possible, further structural properties, such as sparsity), which requires careful managing of the interplay between network size and number of samples collected for the inferential task.

APPENDIX A

Proof of Theorem 1: In view of (12) and (17), the steady-state zero-lag and one-lag correlation matrices can be written as:

$$R_0 = \mu^2 Z^{-1}, \quad R_1 = \mu^2 A Z^{-1}, \quad (61)$$

where we let $Z = I_N - A^2$. Applying the rules for multiplication between partitioned matrices [48, p. 17], the observable part of the one-lag correlation matrix is given by:

$$[R_1]_{\Omega} = \mu^2 A_{\Omega} [Z^{-1}]_{\Omega} + \mu^2 A_{\Omega\Omega'} [Z^{-1}]_{\Omega'\Omega}, \quad (62)$$

where $\Omega' = \{1, 2, \dots, N\} \setminus \Omega$ denotes the complement of Ω . Therefore, using (61) and (62), the matrix $\hat{A}^{(\text{obs})}$ in (16) can be written as:

$$\begin{aligned} \hat{A}^{(\text{obs})} &= (A_{\Omega} [Z^{-1}]_{\Omega} + A_{\Omega\Omega'} [Z^{-1}]_{\Omega'\Omega}) ([Z^{-1}]_{\Omega})^{-1} \\ &= A_{\Omega} + A_{\Omega\Omega'} [Z^{-1}]_{\Omega'\Omega} ([Z^{-1}]_{\Omega})^{-1}. \end{aligned} \quad (63)$$

Now, from the block matrix inversion formula we have that [13, p. 541]:

$$[Z^{-1}]_{\Omega'\Omega} = -(Z_{\Omega'})^{-1} Z_{\Omega'\Omega} [Z^{-1}]_{\Omega}. \quad (64)$$

Substituting (64) into (63), and making explicit the definition of Z , we get:

$$\begin{aligned} \hat{A}^{(\text{obs})} &= A_{\Omega} - A_{\Omega\Omega'} (Z_{\Omega'})^{-1} Z_{\Omega'\Omega} \\ &= A_{\Omega} - A_{\Omega\Omega'} \left(\underbrace{[I_N - A^2]_{\Omega'}}_{I_{N-K} - [A^2]_{\Omega'}} \right)^{-1} \underbrace{[I_N - A^2]_{\Omega'\Omega}}_{-[A^2]_{\Omega'\Omega}}, \end{aligned} \quad (65)$$

which, in view of (18), and noting that $A_{\Omega} = A^{(\text{obs})}$, gives the following expression for the error matrix E :

$$E = A_{\Omega\Omega'} (I_{N-K} - [A^2]_{\Omega'})^{-1} [A^2]_{\Omega'\Omega}. \quad (66)$$

By introducing the following matrices:

$$B \triangleq A^2, \quad H \triangleq (I_{N-K} - B_{\Omega'})^{-1}, \quad (67)$$

the error matrix can be represented as:

$$E = A_{\Omega\Omega'} H B_{\Omega'\Omega}. \quad (68)$$

Accordingly, the entries of the error matrix, for $i, j = 1, 2, \dots, K$, take on the form:

$$e_{ij} = \sum_{\ell, m=1}^{N-K} a_{\omega_i \omega'_\ell} h_{\ell m} b_{\omega'_m \omega_j}, \quad (69)$$

where indices $\omega_1 < \omega_2 < \dots < \omega_K$ span the observable set Ω , while indices $\omega'_1 < \omega'_2 < \dots < \omega'_{N-K}$ span the complement set Ω' . We remark that the error matrix E is a nonnegative matrix (i.e., $e_{ij} \geq 0$ for all $i, j = 1, 2, \dots, K$), because all terms in the summation (69) are nonnegative. In fact, $A_{\Omega\Omega'}$ and $B = A^2$ are nonnegative matrices since A is nonnegative. With regard to H , from (67) it can be expressed as:

$$H = (I_{N-K} - B_{\Omega'})^{-1} = \sum_{k=0}^{\infty} (B_{\Omega'})^k, \quad (70)$$

and, hence, H is nonnegative because so is $B_{\Omega'}$.

Let us now consider the (columnwise) sum of the errors, which, in view of (69), can be written as:

$$\sum_{j=1}^K e_{ij} = \sum_{\ell, m=1}^{N-K} a_{\omega_i \omega'_\ell} h_{\ell m} \sum_{j=1}^K b_{\omega'_m \omega_j}. \quad (71)$$

It is convenient to introduce the following auxiliary matrix:

$$F \triangleq H B_{\Omega'\Omega}, \quad (72)$$

whose entries, for $\ell = 1, 2, \dots, N - K$, and $j = 1, 2, \dots, K$, are accordingly:

$$f_{\ell j} = \sum_{m=1}^{N-K} h_{\ell m} b_{\omega'_m \omega_j}. \quad (73)$$

With such definition, the matrix entry e_{ij} in (69) becomes:

$$e_{ij} = \sum_{\ell=1}^{N-K} a_{\omega_i \omega'_\ell} f_{\ell j}. \quad (74)$$

We start by proving that:

$$\boxed{\sum_{j=1}^K f_{\ell j} \leq 1} \quad (75)$$

Recalling the definition of the matrix B in (67), we have that $\sum_{j=1}^N b_{mj} = (1 - \mu)^2$. Therefore, we can write:

$$\begin{aligned} \sum_{j=1}^K b_{\omega'_m \omega_j} &= (1 - \mu)^2 - \sum_{j=1}^{N-K} b_{\omega'_m \omega'_j} \\ &= 1 - \mu(2 - \mu) - \sum_{j=1}^{N-K} b_{\omega'_m \omega'_j} \\ &= -\mu(2 - \mu) + \sum_{j=1}^{N-K} \underbrace{(\delta_{mj} - \{B_{\Omega'}\}_{mj})}_{=\{H^{-1}\}_{mj}}, \end{aligned} \quad (76)$$

and, hence, from (73) we get:

$$\begin{aligned} \sum_{j=1}^K f_{\ell j} &= -\mu(2-\mu) \sum_{m=1}^{N-K} h_{\ell m} + \sum_{j=1}^{N-K} \underbrace{\sum_{m=1}^{N-K} h_{\ell m} \{H^{-1}\}_{mj}}_{\delta_{\ell j}} \\ &= -\mu(2-\mu) \sum_{m=1}^{N-K} h_{\ell m} + 1 \leq 1, \end{aligned} \quad (77)$$

where the inequality holds since H is a nonnegative matrix, and since $\mu(2-\mu) > 0$ because $0 < \mu < 1$. Using now (74), we can write:

$$\sum_{j=1}^K e_{ij} = \sum_{\ell=1}^{N-K} a_{\omega_i \omega'_\ell} \sum_{j=1}^{N-K} f_{\ell j} \leq \sum_{\ell=1}^{N-K} a_{\omega_i \omega'_\ell} \leq 1 - \mu, \quad (78)$$

where i) the first inequality comes from (75) (since A is a nonnegative matrix), and ii) the second inequality holds since $\sum_{\ell=1}^N a_{i\ell} = 1 - \mu$. ■

APPENDIX B

Proof of Lemma 1: We observe that for any nonnegative real number d and $i \neq j$, it holds that:

$$\begin{aligned} \mathbb{P}[\mathbf{d}_{\max} \geq d \mid \mathbf{g}_{ij} = 1] &\leq \sum_{\ell=1}^N \mathbb{P}[\mathbf{d}_\ell \geq d \mid \mathbf{g}_{ij} = 1] \\ &= \sum_{\ell \neq i, j} \mathbb{P}[\mathbf{d}_\ell \geq d \mid \mathbf{g}_{ij} = 1] + \\ &\quad \mathbb{P}[\mathbf{d}_i \geq d \mid \mathbf{g}_{ij} = 1] + \mathbb{P}[\mathbf{d}_j \geq d \mid \mathbf{g}_{ij} = 1], \end{aligned} \quad (79)$$

where the inequality is an application of the union bound. Now, for $\ell \neq i, j$, we have that \mathbf{d}_ℓ is independent of \mathbf{g}_{ij} , and that $\mathbf{d}_\ell = 1 + \beta(N-1, p_N)$, implying that $\mathbb{P}[\mathbf{d}_\ell \geq d \mid \mathbf{g}_{ij} = 1] = \mathbb{P}[1 + \beta(N-1, p_N) \geq d]$. On the other hand, for $\ell = i$ or $\ell = j$, we have that, conditioned on the event $\mathbf{g}_{ij} = 1$, the degree \mathbf{d}_ℓ is equal to 2 (because $\mathbf{g}_{ii} = 1$ and $\mathbf{g}_{ij} = 1$) plus the number of neighbors arising from the remaining $N-2$ agents, which yields $\mathbf{d}_\ell = 2 + \beta(N-2, p_N)$. Therefore, it is legitimate to write the following chain of inequalities, for $\eta \in \mathbb{R}$ and $t > 0$:

$$\begin{aligned} \mathbb{P}[\mathbf{d}_{\max} \geq \eta N p_N \mid \mathbf{g}_{ij} = 1] &\leq (N-2) \mathbb{P}[1 + \beta(N-1, p_N) \geq \eta N p_N] + \\ &\quad 2 \mathbb{P}[2 + \beta(N-2, p_N) \geq \eta N p_N] \\ &\leq N e^{-\eta N p_N t} \mathbb{E}[e^{(1+\beta(N-1, p_N))t}] + \\ &\quad 2 e^{-\eta N p_N t} \mathbb{E}[e^{(2+\beta(N-2, p_N))t}] \\ &= N e^t e^{-\eta N p_N t} [1 + p_N (e^t - 1)]^{N-1} + \\ &\quad 2 e^{2t} e^{-\eta N p_N t} [1 + p_N (e^t - 1)]^{N-2} \\ &\leq (N e^t + 2 e^{2t}) e^{-N p_N (\eta t + 1 - e^t)}, \end{aligned} \quad (80)$$

where: i) the first inequality follows from (79); ii) the second inequality is a classic Chernoff bound³; iii) the equality follows by evaluating explicitly the moment generating function,

³For a random variable \mathbf{x} , and a positive t , the Chernoff bound is given by $\mathbb{P}[\mathbf{x} \geq \eta] \leq \mathbb{E}[e^{t\mathbf{x}}] / e^{\eta t}$.

$\mathbb{E}[e^{\beta(N, p)t}]$, of the binomial random variable⁴; iv) the latter inequality follows from $(1+x)^N \leq e^{Nx}$ (which holds true for $x > 0$), having further replaced the exponents $N-1$ and $N-2$ with the (higher) exponent N . Now, maximizing the exponent $\eta t - e^t$ (w.r.t. t) yields the solution $e^t = \eta$, which in turn implies:

$$\mathbb{P}[\mathbf{d}_{\max} \geq \eta N p_N \mid \mathbf{g}_{ij} = 1] \leq (\eta N + 2\eta^2) e^{-N p_N (\eta \ln \eta + 1 - \eta)}. \quad (81)$$

Using now the particular value $\eta = e$ yields:

$$\begin{aligned} \mathbb{P}[\mathbf{d}_{\max} \geq e N p_N \mid \mathbf{g}_{ij} = 1] &\leq (e N + 2e^2) e^{-N p_N} \\ &\leq \left(e + \frac{2e^2}{N}\right) e^{-c_N}, \end{aligned} \quad (82)$$

where, in the latter inequality, we used the fact that $p_N = (\ln N + c_N)/N$. ■

APPENDIX C

Proof of Theorem 2: The empirical distributions $\mathcal{F}_0(\alpha)$ and $\mathcal{F}_1(\alpha)$ in (44) can be written as:

$$\mathcal{F}_0(\alpha) = \frac{N_0(\alpha)}{N_0 + \mathbb{I}\{N_0 = 0\}} + \frac{1}{2} \mathbb{I}\{N_0 = 0\}, \quad (83)$$

$$\mathcal{F}_1(\alpha) = \frac{N_1(\alpha)}{N_1 + \mathbb{I}\{N_1 = 0\}} + \frac{1}{2} \mathbb{I}\{N_1 = 0\}, \quad (84)$$

where we have simply used a formal representation to state that $\mathcal{F}_0(\alpha)$ (resp., $\mathcal{F}_1(\alpha)$) are set to 1/2 when $N_0 = 0$ (resp., $N_1 = 0$).

We start by proving that $\mathcal{F}_0(\epsilon) \xrightarrow{p} 1$ for all $\epsilon > 0$. From (83) we can write:

$$1 - \mathcal{F}_0(\epsilon) = \frac{N_0 - N_0(\epsilon)}{N_0 + \mathbb{I}\{N_0 = 0\}} + \frac{1}{2} \mathbb{I}\{N_0 = 0\}. \quad (85)$$

Using (42), and since $\hat{\mathbf{a}}_{ij}^{(\text{obs})} = \mathbf{a}_{ij}^{(\text{obs})} + \mathbf{e}_{ij}$, from (85) we can write:

$$\begin{aligned} &\frac{N_0 - N_0(\epsilon)}{N_0 + \mathbb{I}\{N_0 = 0\}} \\ &= \frac{\sum_{i=1}^K \sum_{j \neq i} \mathbb{I}\{N p_N \mathbf{e}_{ij} > \epsilon, \mathbf{g}_{ij}^{(\text{obs})} = 0\}}{N_0 + \mathbb{I}\{N_0 = 0\}} \\ &\leq \frac{\sum_{i=1}^K \sum_{j \neq i} \mathbb{I}\{N p_N \mathbf{e}_{ij} > \epsilon\}}{N_0 + \mathbb{I}\{N_0 = 0\}} \\ &\leq \frac{K N p_N}{N_0 + \mathbb{I}\{N_0 = 0\}} \frac{1 - \mu}{\epsilon} \\ &= \left(\frac{K^2}{N_0 + \mathbb{I}\{N_0 = 0\}}\right) \frac{N}{K} \frac{1 - \mu}{\epsilon} p_N, \end{aligned} \quad (86)$$

where the latter inequality follows by applying Corollary 1. Now, the asymptotic properties of N_0 can be deduced from the asymptotic properties of binomial random variables. Indeed, since for the Erdős-Rényi model, $\mathbf{g}_{ij} = \mathbf{g}_{ji}$, we have:

$$N_0 = 2 \sum_{i=1}^K \sum_{j < i} (1 - \mathbf{g}_{ij}^{(\text{obs})}). \quad (87)$$

⁴The moment generating function of a Bernoulli random variable with success probability p is given by $p e^t + (1-p)$. Since a binomial random variable $\beta(N, p)$ is the sum of N independent Bernoulli variables with success probability p , we have $\mathbb{E}[e^{\beta(N, p)t}] = [p e^t + (1-p)]^N$.

Moreover, since the variables $1 - \mathbf{g}_{ij}^{(\text{obs})}$ form, for $j < i$, a collection of $L = K(K-1)/2$ independent Bernoulli random variables with parameter $1 - p_N$, we have that \mathbf{N}_0 is distributed as $2\beta(L, 1 - p_N)$. Now, in view of (8), K (and, hence, L) diverges as N goes to infinity. Since by assumption p_N vanishes as N goes to infinity, the product $L(1 - p_N)$ diverges as well, and we can invoke (32) to conclude that:

$$\frac{\mathbf{N}_0}{L(1 - p_N)} \xrightarrow{p} 2 \Rightarrow \frac{\mathbf{N}_0}{K^2} \xrightarrow{p} 1. \quad (88)$$

Since $\mathbb{I}\{\mathbf{N}_0 = 0\} \in \{0, 1\}$, using (88) and (8) into (86), and recalling that p_N vanishes by assumption, we have in fact proved that the first term on the RHS in (85) goes to zero as $N \rightarrow \infty$. With regard to the second term on the RHS in (85), it converges to zero in mean (and, hence, in probability [47]), since we have:

$$\begin{aligned} \mathbb{E}\{\mathbb{I}\{\mathbf{N}_0 = 0\}\} &= \mathbb{P}\{\text{all observed agents are interacting}\} \\ &= (p_N)^L \xrightarrow{N \rightarrow \infty} 0. \end{aligned} \quad (89)$$

We have in fact proved that $\mathcal{F}_0(\epsilon) \xrightarrow{p} 1$ for all $\epsilon > 0$.

We now show that $\mathcal{F}_1(\tau) \xrightarrow{p} 1$. Using (43), from (84) we can write:

$$\begin{aligned} \mathcal{F}_1(\tau) &= \frac{\sum_{i=1}^K \sum_{j \neq i} \mathbb{I}\{N p_N \hat{\mathbf{a}}_{ij}^{(\text{obs})} \leq \tau, \mathbf{g}_{ij}^{(\text{obs})} = 1\}}{K(K-1)p_N} \times \\ &\quad \underbrace{\hspace{10em}}_{\mathbf{T}_1} \\ &\quad \underbrace{\frac{K(K-1)p_N}{\mathbf{N}_1 + \mathbb{I}\{\mathbf{N}_1 = 0\}} + \frac{1}{2} \mathbb{I}\{\mathbf{N}_1 = 0\}}_{\mathbf{T}_2}. \end{aligned} \quad (90)$$

With regard to the term \mathbf{T}_1 , we have:

$$\mathbb{E}\{\mathbf{T}_1\} = \frac{\sum_{i=1}^K \sum_{j \neq i} \mathbb{P}[N p_N \hat{\mathbf{a}}_{ij}^{(\text{obs})} \leq \tau | \mathbf{g}_{ij}^{(\text{obs})} = 1]}{K(K-1)}, \quad (91)$$

where we used the fact that $\mathbb{P}[\mathbf{g}_{ij}^{(\text{obs})} = 1] = p_N$. Moreover, since $\hat{\mathbf{a}}_{ij}^{(\text{obs})} = \mathbf{a}_{ij}^{(\text{obs})} + \mathbf{e}_{ij}$, and since $\mathbf{e}_{ij} \geq 0$, we have that:

$$\begin{aligned} \mathbb{P}[N p_N \hat{\mathbf{a}}_{ij}^{(\text{obs})} \leq \tau | \mathbf{g}_{ij}^{(\text{obs})} = 1] \\ \leq \mathbb{P}[N p_N \mathbf{a}_{ij}^{(\text{obs})} \leq \tau | \mathbf{g}_{ij}^{(\text{obs})} = 1]. \end{aligned} \quad (92)$$

Combining (91) and (92), in view of (36) we can write:

$$\mathbb{E}\{\mathbf{T}_1\} \leq \epsilon_N \xrightarrow{N \rightarrow \infty} 0. \quad (93)$$

Thus, we have shown that the term \mathbf{T}_1 appearing in (90) converges to zero in mean and, hence, in probability [47]. Moreover, the term \mathbf{T}_2 in (90) goes to one in probability in view of (32), because: *i*) for the $\mathcal{G}^*(N, p_N)$ model, \mathbf{N}_1 is distributed as $2\beta(L, p_N)$, where $L = K(K-1)/2$; *ii*) the product $L p_N$ diverges as N goes to infinity in view of (8), since by assumption the product $N p_N$ diverges; *iii*) the term $\mathbb{I}\{\mathbf{N}_1 = 0\}$ is either 0 or 1. Finally, as regards the latter term in (90), it is straightforward to see that it converges to zero in mean (and, hence, in probability [47]), since we have:

$$\begin{aligned} \mathbb{E}\{\mathbb{I}\{\mathbf{N}_1 = 0\}\} &= \mathbb{P}\{\text{all observed agents are not interacting}\} \\ &= (1 - p_N)^L \leq e^{-L p_N} \xrightarrow{N \rightarrow \infty} 0, \end{aligned} \quad (94)$$

where the inequality holds because $p_N \leq 1$. ■

APPENDIX D

The main result of this appendix is the proof of Proposition 1. We start by proving two auxiliary lemmas.

Lemma 2: Let $\beta(N, p)$ be a binomial random variable of parameters N and p . We have, for $m = 1, 2$:

$$\mathbb{E}\left[\frac{1}{(1 + \beta(N, p))^m}\right] \leq \frac{m}{(Np)^m}. \quad (95)$$

Proof: We prove the claim for the case $m = 2$, since the proof for the case $m = 1$ follows similar steps. We have that:

$$\begin{aligned} \mathbb{E}\left[\frac{1}{(1 + \beta(N, p_N))^2}\right] \\ &= \sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} \frac{1}{(k+1)^2} \\ &= \sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} \frac{1}{(k+1)(k+2)} \underbrace{\frac{k+2}{k+1}}_{\leq 2} \\ &\leq \sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} \frac{2}{(k+1)(k+2)}. \end{aligned} \quad (96)$$

Now, since we have

$$\binom{N+2}{k+2} = \frac{(N+2)!}{(k+2)!(N-k)!} = \binom{N}{k} \frac{(N+1)(N+2)}{(k+1)(k+2)}, \quad (97)$$

we can write:

$$\begin{aligned} \sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} \frac{2}{(k+1)(k+2)} \\ \leq \frac{2}{(N+1)(N+2)} \sum_{k=0}^N \binom{N+2}{k+2} p^k (1-p)^{N-k} \\ \leq \frac{2}{(Np)^2} \sum_{k=0}^N \binom{N+2}{k+2} p^{k+2} (1-p)^{(N+2)-(k+2)} \\ = \frac{2}{(Np)^2} \underbrace{\sum_{k=2}^{N+2} \binom{N+2}{k} p^k (1-p)^{(N+2)-k}}_{\mathbb{P}[\beta(N+2, p) \geq 2] < 1} < \frac{2}{(Np)^2}, \end{aligned} \quad (98)$$

and the proof is complete. ■

We recall that \mathbb{E} and \mathbb{V} denote the expectation and variance operators, respectively. In the following, the expectation, the variance and the mean-square value of a random variable z will be denoted respectively by:

$$\bar{z} \triangleq \mathbb{E}[z], \quad \sigma_z^2 \triangleq \mathbb{V}[z] = \mathbb{E}[(z - \bar{z})^2], \quad \overline{z^2} \triangleq \mathbb{E}[z^2] = \sigma_z^2 + \bar{z}^2. \quad (99)$$

Moreover, given a sequence of *identically distributed* random variables, z_1, z_2, \dots, z_L , we shall still write:

$$\bar{z} \triangleq \mathbb{E}[z_\ell], \quad \sigma_z^2 \triangleq \mathbb{V}[z_\ell] = \mathbb{E}[(z_\ell - \bar{z})^2], \quad \overline{z^2} \triangleq \mathbb{E}[z_\ell^2] = \sigma_z^2 + \bar{z}^2, \quad (100)$$

since, owing to identical distribution across ℓ , both the expectation and the variance do not depend upon ℓ .

Lemma 3: Let $[\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_L]$ and $[z_1, z_2, \dots, z_L]$ be mutually independent random vectors, with the variables z_ℓ , for

$\ell = 1, 2, \dots, L$, being mutually uncorrelated and identically distributed scalar random variables across ℓ . Let also:

$$\mathbf{s} \triangleq \sum_{\ell=1}^L \mathbf{u}_\ell \mathbf{z}_\ell. \quad (101)$$

Then:

$$\sigma_s^2 = \sigma_z^2 \sum_{\ell=1}^L \mathbb{E}[\mathbf{u}_\ell^2] + \bar{z}^2 \mathbb{V} \left[\sum_{\ell=1}^L \mathbf{u}_\ell \right] \quad (102)$$

Proof: We have:

$$\begin{aligned} \mathbf{s}^2 &= \sum_{\ell=1}^L \sum_{m=1}^L \mathbf{u}_\ell \mathbf{u}_m \mathbf{z}_\ell \mathbf{z}_m \\ &= \sum_{\ell=1}^L \mathbf{u}_\ell^2 \mathbf{z}_\ell^2 + \sum_{\ell=1}^L \sum_{m \neq \ell} \mathbf{u}_\ell \mathbf{u}_m \mathbf{z}_\ell \mathbf{z}_m. \end{aligned} \quad (103)$$

Since $[\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_L]$ and $[\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_L]$ are mutually independent random vectors, and since the variables \mathbf{z}_ℓ are mutually uncorrelated, by taking expectations we get:

$$\mathbb{E}[\mathbf{s}^2] = \sum_{\ell=1}^L \mathbb{E}[\mathbf{u}_\ell^2] \mathbb{E}[\mathbf{z}_\ell^2] + \sum_{\ell=1}^L \sum_{m \neq \ell} \mathbb{E}[\mathbf{u}_\ell \mathbf{u}_m] \mathbb{E}[\mathbf{z}_\ell] \mathbb{E}[\mathbf{z}_m]. \quad (104)$$

Moreover, since the variables \mathbf{z}_ℓ are identically distributed, Eq. (104) yields:

$$\mathbb{E}[\mathbf{s}^2] = (\sigma_z^2 + \bar{z}^2) \sum_{\ell=1}^L \mathbb{E}[\mathbf{u}_\ell^2] + \bar{z}^2 \sum_{\ell=1}^L \sum_{m \neq \ell} \mathbb{E}[\mathbf{u}_\ell \mathbf{u}_m]. \quad (105)$$

Now, observing that:

$$\left(\sum_{\ell=1}^L \mathbf{u}_\ell \right)^2 = \sum_{\ell=1}^L \mathbf{u}_\ell^2 + \sum_{\ell=1}^L \sum_{m \neq \ell} \mathbf{u}_\ell \mathbf{u}_m, \quad (106)$$

we can rearrange (105) as:

$$\mathbb{E}[\mathbf{s}^2] = \sigma_z^2 \sum_{\ell=1}^L \mathbb{E}[\mathbf{u}_\ell^2] + \bar{z}^2 \mathbb{E} \left[\left(\sum_{\ell=1}^L \mathbf{u}_\ell \right)^2 \right]. \quad (107)$$

On the other hand, from (101) we have:

$$\bar{s} = \bar{z} \mathbb{E} \left[\sum_{\ell=1}^L \mathbf{u}_\ell \right] \Rightarrow \bar{s}^2 = \bar{z}^2 \left(\mathbb{E} \left[\sum_{\ell=1}^L \mathbf{u}_\ell \right] \right)^2, \quad (108)$$

implying, in view of (107):

$$\begin{aligned} \sigma_s^2 &= \mathbb{E}[\mathbf{s}^2] - \bar{s}^2 \\ &= \sigma_z^2 \sum_{\ell=1}^L \mathbb{E}[\mathbf{u}_\ell^2] + \\ &\quad \bar{z}^2 \left\{ \mathbb{E} \left[\left(\sum_{\ell=1}^L \mathbf{u}_\ell \right)^2 \right] - \left(\mathbb{E} \left[\sum_{\ell=1}^L \mathbf{u}_\ell \right] \right)^2 \right\} \\ &= \sigma_z^2 \sum_{\ell=1}^L \mathbb{E}[\mathbf{u}_\ell^2] + \bar{z}^2 \mathbb{V} \left[\sum_{\ell=1}^L \mathbf{u}_\ell \right]. \end{aligned} \quad (109)$$

Proof of Proposition 1: We start by observing that, using (85), we can write:

$$\frac{1 - \mathcal{F}_0(\epsilon)}{p_N} = \frac{N_0 - N_0(\epsilon)}{N_0 + \mathbb{I}\{N_0 = 0\}} \frac{1}{p_N} + \frac{1}{2} \frac{\mathbb{I}\{N_0 = 0\}}{p_N}. \quad (110)$$

In view of (89), the second term appearing on the RHS in (110) converges to zero in mean, and, hence, in probability [47]. With regard to the first term appearing on the RHS in (110), from the first inequality in (86) we have:

$$\begin{aligned} &\frac{N_0 - N_0(\epsilon)}{N_0 + \mathbb{I}\{N_0 = 0\}} \frac{1}{p_N} \\ &\leq \frac{\sum_{i=1}^K \sum_{j \neq i} \mathbb{I}\{N p_N \mathbf{e}_{ij} > \epsilon\}}{N_0 + \mathbb{I}\{N_0 = 0\}} \frac{1}{p_N} \\ &= \underbrace{\frac{\sum_{i=1}^K \sum_{j \neq i} \mathbb{I}\{N p_N \mathbf{e}_{ij} > \epsilon\}}{K(K-1)p_N}}_{\mathcal{T}_3} \times \\ &\quad \underbrace{\frac{K(K-1)}{N_0 + \mathbb{I}\{N_0 = 0\}}}_{\mathcal{T}_4}. \end{aligned} \quad (111)$$

The fact that \mathcal{T}_4 converges to 1 in probability has been already ascertained in the proof of Theorem 2 — see the rightmost relationship in (88). With regard to \mathcal{T}_3 , by taking expectations we get:

$$\mathbb{E}[\mathcal{T}_3] = \frac{\sum_{i=1}^K \sum_{j \neq i} \mathbb{P}[N p_N \mathbf{e}_{ij} > \epsilon]}{K(K-1)p_N} = \frac{\mathbb{P}[N p_N \mathbf{e}_{ij} > \epsilon]}{p_N}, \quad (112)$$

where we used the fact that, in view of (152), the random variables \mathbf{e}_{ij} , for $i \neq j$, are identically distributed. This is because any renumbering leaves unaltered the statistical distribution of matrix \mathbf{E} . In addition, identical distribution across the $K-1$ agents distinct from agent i , implies in particular the following identity for expectations:

$$\mathbb{E}[\mathbf{e}_{ij}] = \frac{1}{K-1} \sum_{j \neq i} \mathbb{E}[\mathbf{e}_{ij}]. \quad (113)$$

Multiplying the expected error by the factor $N p_N$, we get:

$$\begin{aligned} N p_N \mathbb{E}[\mathbf{e}_{ij}] &= N p_N \frac{1}{K-1} \sum_{j \neq i} \mathbb{E}[\mathbf{e}_{ij}] \\ &= \frac{N p_N}{K-1} \mathbb{E} \left[\sum_{j \neq i} \mathbf{e}_{ij} \right] \\ &\leq (1 - \mu) \frac{N}{K-1} p_N \xrightarrow{N \rightarrow \infty} 0, \end{aligned} \quad (114)$$

where the inequality follows from Theorem 1, while the convergence follows from (8) and from the fact that p_N vanishes. Now, we can write, for $i, j = 1, 2, \dots, K$, and $i \neq j$:

$$\begin{aligned} &\mathbb{P}[N p_N \mathbf{e}_{ij} > \epsilon] \\ &= \mathbb{P}[(N p_N \mathbf{e}_{ij} - N p_N \mathbb{E}[\mathbf{e}_{ij}]) > (\epsilon - N p_N \mathbb{E}[\mathbf{e}_{ij}])] \\ &\leq \mathbb{P}[(N p_N)^2 (\mathbf{e}_{ij} - \mathbb{E}[\mathbf{e}_{ij}])^2 > (\epsilon - N p_N \mathbb{E}[\mathbf{e}_{ij}])^2] \\ &\leq \frac{(N p_N)^2 \sigma_e^2}{(\epsilon - N p_N \mathbb{E}[\mathbf{e}_{ij}])^2} \leq \frac{4}{\epsilon^2} (N p_N)^2 \sigma_e^2, \end{aligned} \quad (115)$$

where we have applied Chebyshev's inequality, and we have assumed that N is sufficiently large to have $Np_N\mathbb{E}[e_{ij}] \leq \epsilon/2$ (a condition guaranteed by (114)). Therefore, in view of (112) and (115), the claim of the proposition will be proved if we are able to show that

$$\boxed{\frac{(Np_N)^2\sigma_e^2}{p_N} \xrightarrow{N \rightarrow \infty} 0} \quad (116)$$

Accordingly, we now focus on the evaluation of σ_e^2 . To this end, we use for e_{ij} the representation given in (74), and apply Lemma 3 to e_{ij} , for $i \neq j$, with the choices $L = N - K$, and:

$$\begin{aligned} [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_L] &\rightarrow [\mathbf{a}_{\omega_i 1}, \mathbf{a}_{\omega_i 2}, \dots, \mathbf{a}_{\omega_i L}], \\ [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_L] &\rightarrow [\mathbf{f}_{1j}, \mathbf{f}_{2j}, \dots, \mathbf{f}_{Lj}], \\ \mathbf{s} &\rightarrow e_{ij}. \end{aligned} \quad (117)$$

We remark that all the matrices are *random*, as a consequence of the randomness implied by the underlying Erdős-Renyi graph. Preliminarily, it is necessary to verify that the choices in (117) meet the assumptions of Lemma 3. First, observe that, in view of (153), row-permutations leave unaltered the statistical distribution of \mathbf{F} . This property implies in particular that the random variables $\mathbf{f}_{1j}, \mathbf{f}_{2j}, \dots, \mathbf{f}_{Lj}$ are identically distributed. Moreover, under the independence approximation, such entries are approximated as mutually independent (and, hence, uncorrelated). Under the same approximation, the random vectors $[\mathbf{a}_{\omega_i 1}, \mathbf{a}_{\omega_i 2}, \dots, \mathbf{a}_{\omega_i L}]$ and $[\mathbf{f}_{1j}, \mathbf{f}_{2j}, \dots, \mathbf{f}_{Lj}]$ are approximated as mutually independent. Therefore, we conclude that the hypotheses of Lemma 3 are met, and we can write:

$$\mathbb{V}[e_{ij}] = \sigma_e^2 \approx \sigma_f^2 \sum_{\ell=1}^{N-K} \mathbb{E}[\mathbf{a}_{\omega_i \omega'_\ell}^2] + \bar{f}^2 \mathbb{V} \left[\sum_{\ell=1}^{N-K} \mathbf{a}_{\omega_i \omega'_\ell} \right], \quad (118)$$

The approximation symbol in (118) comes from the fact that the pertinent calculations are done under the independence approximation.

Now, in view of (151), any renumbering leaves unaltered the statistical distribution of \mathbf{A} . This property implies that the random variables \mathbf{a}_{ij} , for all $i, j = 1, 2, \dots, N$, with $i \neq j$, are identically distributed, and, in particular, that they share the common mean-square value $\mathbb{E}[\mathbf{a}_{ij}^2] = \bar{a}^2$. Since $\omega_i \neq \omega'_\ell$ (because $\Omega \cap \Omega' = \emptyset$), Eq. (118) can be rewritten as:

$$\sigma_e^2 \approx (N - K)\sigma_f^2 \bar{a}^2 + \bar{f}^2 \mathbb{V} \left[\sum_{\ell=1}^{N-K} \mathbf{a}_{\omega_i \omega'_\ell} \right]. \quad (119)$$

Likewise, we apply Lemma 3 to $\mathbf{f}_{\ell j}$, with the choices $L = N - K$, and:

$$\begin{aligned} [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_L] &\rightarrow [\mathbf{h}_{\ell 1}, \mathbf{h}_{\ell 2}, \dots, \mathbf{h}_{\ell L}], \\ [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_L] &\rightarrow [\mathbf{b}_{1\omega_j}, \mathbf{b}_{2\omega_j}, \dots, \mathbf{b}_{L\omega_j}], \\ \mathbf{s} &\rightarrow \mathbf{f}_{\ell j}. \end{aligned} \quad (120)$$

In view of (165), any renumbering leaves unaltered the statistical distribution of \mathbf{B} . This property implies in particular that the random variables $\mathbf{b}_{1\omega_j}, \mathbf{b}_{2\omega_j}, \dots, \mathbf{b}_{L\omega_j}$ are identically distributed. Moreover, under the independence approximation, such entries are approximated as mutually independent

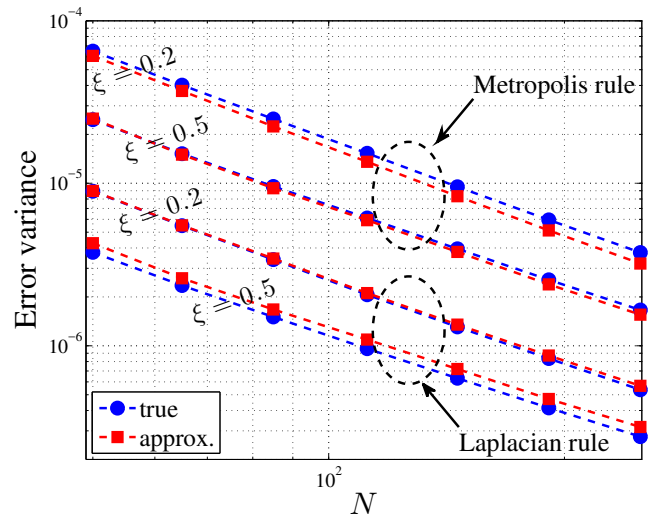


Fig. 9. Comparison between the true error variance, $\mathbb{V}[e_{ij}]$ (evaluated through 10^3 Monte Carlo runs), and the error variance arising from the independence approximation, see (122). The different curves are obtained by varying the type of combination matrix (Laplacian or Metropolis), and the fraction of observable agents ($\xi = 0.2$ or $\xi = 0.5$).

(and, hence, uncorrelated). Under the same approximation, the random vectors $[\mathbf{h}_{\ell 1}, \mathbf{h}_{\ell 2}, \dots, \mathbf{h}_{\ell L}]$ and $[\mathbf{b}_{1\omega_j}, \mathbf{b}_{2\omega_j}, \dots, \mathbf{b}_{L\omega_j}]$ are approximated as mutually independent. Therefore, we conclude that also in this case the hypotheses of Lemma 3 are met, which allows writing:

$$\sigma_f^2 \approx \sigma_b^2 \sum_{m=1}^{N-K} \mathbb{E}[\mathbf{h}_{\ell m}^2] + \bar{b}^2 \mathbb{V} \left[\sum_{m=1}^{N-K} \mathbf{h}_{\ell m} \right]. \quad (121)$$

Substituting (121) into (119) gives:

$$\begin{aligned} \sigma_e^2 &\approx (N - K)\bar{a}^2 \sigma_b^2 \sum_{m=1}^{N-K} \mathbb{E}[\mathbf{h}_{\ell m}^2] + \\ &\quad (N - K)\bar{a}^2 \bar{b}^2 \mathbb{V} \left[\sum_{m=1}^{N-K} \mathbf{h}_{\ell m} \right] + \\ &\quad \bar{f}^2 \mathbb{V} \left[\sum_{\ell=1}^{N-K} \mathbf{a}_{\omega_i \omega'_\ell} \right] \\ &\triangleq \tilde{\sigma}_e^2. \end{aligned} \quad (122)$$

A comparison between the true variance, σ_e^2 , and the approximate variance, $\tilde{\sigma}_e^2$, is given in Fig. 9. We see that, for the considered examples, the agreement between the two quantities is satisfying. Note that, in certain cases, the true curve and the approximate curve (pertaining to the same example) exhibit slightly different slopes, suggesting that the approximation does not always capture the precise asymptotic behavior.

We now compute an upper bound for the approximate variance in (122). First, we find an upper bound for f . In view of (154), column-permutations leave unaltered the statistical distribution of \mathbf{F} . This property implies in particular that the

random variables $\bar{f}_{\ell 1}, \bar{f}_{\ell 2}, \dots, \bar{f}_{\ell K}$ are identically distributed, yielding:

$$\bar{f} = \frac{1}{K} \sum_{j=1}^K \mathbb{E}[\bar{f}_{\ell j}] = \mathbb{E} \left[\frac{\sum_{j=1}^K \bar{f}_{\ell j}}{K} \right] \leq \frac{1}{K}, \quad (123)$$

where the inequality comes from (75).

Second, we observe that:

$$\begin{aligned} \sum_{m=1}^{N-K} \mathbf{h}_{\ell m} &\stackrel{(a)}{=} \sum_{k=0}^{\infty} \sum_{m=1}^{N-K} [(\mathbf{B}_{\Omega'}^k)]_{\ell m} \\ &\stackrel{(b)}{\leq} \sum_{k=0}^{\infty} \sum_{m=1}^{N-K} [(\mathbf{B}^k)_{\Omega'}]_{\ell m} \\ &\stackrel{(c)}{=} \sum_{k=0}^{\infty} \sum_{m=1}^{N-K} [\mathbf{B}^k]_{\omega'_\ell \omega'_m} \\ &\stackrel{(d)}{\leq} \sum_{k=0}^{\infty} \sum_{j=1}^N [\mathbf{B}^k]_{\omega'_\ell j} \\ &\stackrel{(e)}{=} \sum_{k=0}^{\infty} (1-\mu)^{2k} = \frac{1}{\mu(2-\mu)} \triangleq \sqrt{C}, \end{aligned} \quad (124)$$

where (a) follows from (70); (b) holds since for any two nonnegative matrices P and Q we have that $P_{\Omega'} Q_{\Omega'} \leq [PQ]_{\Omega'}$ (with entrywise inequality), and, hence, by iteration we get $(\mathbf{B}_{\Omega'})^k \leq [\mathbf{B}^k]_{\Omega'}$ because \mathbf{B} and its powers are nonnegative; (c) holds because indices $\omega'_1 < \omega'_2 < \dots < \omega'_{N-K}$ span the set Ω' ; (d) holds because \mathbf{B}^k is nonnegative, and, hence, adding more terms can only increase the value of the sum; and (e) holds because $\mathbf{B} = \mathbf{A}^2$, yielding $\sum_{j=1}^N \mathbf{b}_{ij} = (1-\mu)^2 \Rightarrow \sum_{j=1}^N [\mathbf{B}^k]_{ij} = (1-\mu)^{2k}$, for all $i = 1, 2, \dots, N$. Exploiting now the bounds (123) and (124), from (122) we get:

$$\begin{aligned} \tilde{\sigma}_e^2 &\leq (N-K) \bar{a}^2 \sigma_b^2 C + (N-K) \bar{a}^2 \bar{b}^2 C + \\ &\quad \frac{1}{K^2} \mathbb{V} \left[\sum_{\ell=1}^{N-K} \mathbf{a}_{\omega_i \omega'_\ell} \right], \end{aligned} \quad (125)$$

where we used the following known bounds:

$$\begin{aligned} \sum_{m=1}^{N-K} \mathbf{h}_{\ell m}^2 &\leq \left(\sum_{m=1}^{N-K} \mathbf{h}_{\ell m} \right)^2, \\ \mathbb{V} \left[\sum_{m=1}^{N-K} \mathbf{h}_{\ell m} \right] &\leq \mathbb{E} \left[\left(\sum_{m=1}^{N-K} \mathbf{h}_{\ell m} \right)^2 \right]. \end{aligned} \quad (126)$$

Since $\sigma_b^2 + \bar{b}^2 = \bar{b}^2$, inequality (125) can be also written as:

$$\tilde{\sigma}_e^2 \leq (N-K) \bar{a}^2 \bar{b}^2 C + \frac{1}{K^2} \mathbb{V} \left[\sum_{\ell=1}^{N-K} \mathbf{a}_{\omega_i \omega'_\ell} \right]. \quad (127)$$

We must now examine the asymptotic behavior of the terms \bar{a}^2 and \bar{b}^2 . We start with \bar{a}^2 . We can write:

$$\begin{aligned} \bar{a}^2 &= \frac{1}{N-1} \mathbb{E} \left[\sum_{\ell \neq i}^N \mathbf{a}_{i\ell}^2 \right] \leq \frac{\kappa^2}{(N-1)} \mathbb{E} \left[\frac{1}{\mathbf{d}_i^2} \sum_{\ell \neq i}^N \mathbf{g}_{i\ell} \right] \\ &\leq \frac{\kappa^2}{N-1} \mathbb{E} \left[\frac{1}{\mathbf{d}_i} \right] = \frac{\kappa^2}{N-1} \mathbb{E} \left[\frac{1}{1 + \beta(N-1, p_N)} \right], \end{aligned} \quad (128)$$

where the first equality holds because, as already observed, the random variables $\mathbf{a}_{i\ell}$, for $\ell \neq i$, are identically distributed in view of (151). The first inequality follows from (49), the second inequality follows because $\sum_{\ell \neq i} \mathbf{g}_{i\ell} = \mathbf{d}_i - 1$, and the last equality follows from the fact that $\mathbf{d}_i - 1$ is distributed as a binomial random variable of parameters $N-1$ and p_N . In view of Lemma 2, Eq. (128) implies the existence of an upper bound, $\epsilon_N^{(1)}$, such that:

$$\bar{a}^2 \leq \epsilon_N^{(1)} \sim \frac{1}{N^2 p_N}, \quad (129)$$

where the notation $a_N \sim b_N$ means that a_N is on the same order of b_N , namely, that the limit $\lim_{N \rightarrow \infty} a_N/b_N$ exists and is finite. It remains to examine the asymptotic behavior of the quantity \bar{b}^2 . Recalling that $\mathbf{B} = \mathbf{A}^2$ and that \mathbf{A} is symmetric, we have, for $i \neq j$:

$$\begin{aligned} \mathbf{b}_{ij} &= \sum_{\ell=1}^N \mathbf{a}_{i\ell} \mathbf{a}_{\ell j} = (\mathbf{a}_{ii} + \mathbf{a}_{jj}) \mathbf{a}_{ij} + \sum_{\ell \neq i, j} \mathbf{a}_{i\ell} \mathbf{a}_{\ell j} \\ &\leq 2 \mathbf{a}_{ij} + \kappa^2 \underbrace{\frac{\sum_{\ell \neq i, j} \mathbf{g}_{i\ell} \mathbf{g}_{\ell j}}{\mathbf{d}_i \mathbf{d}_j}}_{\triangleq \mathbf{z}_{ij}}, \end{aligned} \quad (130)$$

where, in the last step, we used (49). We can thus write:

$$\mathbf{b}_{ij}^2 \leq 4 \mathbf{a}_{ij}^2 + \kappa^4 \mathbf{z}_{ij}^2 + 4\kappa^2 \mathbf{a}_{ij} \mathbf{z}_{ij}. \quad (131)$$

Let us first focus on the term \mathbf{z}_{ij}^2 . We observe that:

$$\begin{aligned} \mathbf{z}_{ij}^2 &= \frac{1}{\mathbf{d}_i^2 \mathbf{d}_j^2} \left(\sum_{\ell \neq i, j} \mathbf{g}_{i\ell} \mathbf{g}_{\ell j} + \sum_{\substack{\ell \neq i, j \\ m \neq i, j \\ \ell \neq m}} \mathbf{g}_{i\ell} \mathbf{g}_{\ell j} \mathbf{g}_{im} \mathbf{g}_{jm} \right) \\ &\leq \underbrace{\sum_{\ell \neq i, j} \frac{\mathbf{g}_{i\ell} \mathbf{g}_{\ell j}}{(\mathbf{d}_i - \mathbf{g}_{i\ell})^2 (\mathbf{d}_j - \mathbf{g}_{\ell j})^2}}_{\mathbf{Z}_1} + \\ &\quad \underbrace{\sum_{\substack{\ell \neq i, j \\ m \neq i, j \\ \ell \neq m}} \frac{\mathbf{g}_{i\ell} \mathbf{g}_{\ell j} \mathbf{g}_{im} \mathbf{g}_{jm}}{(\mathbf{d}_i - \mathbf{g}_{i\ell} - \mathbf{g}_{im})^2 (\mathbf{d}_j - \mathbf{g}_{\ell j} - \mathbf{g}_{jm})^2}}_{\mathbf{Z}_2}, \end{aligned} \quad (132)$$

where the inequality holds because, in view of (29), we have $\mathbf{d}_i > \mathbf{d}_i - \mathbf{g}_{i\ell} > 0$ for $\ell \neq i$ and $\mathbf{d}_i > \mathbf{d}_i - \mathbf{g}_{i\ell} - \mathbf{g}_{im} > 0$

for $\ell \neq i$ and $m \neq i$. Using again (29), the terms \mathbf{Z}_1 and \mathbf{Z}_2 in (132) can be rewritten as, respectively:

$$\begin{aligned} \mathbf{Z}_1 &= \sum_{\substack{\ell \neq i, j \\ m \neq i, j \\ \ell \neq m}} \frac{\mathbf{g}_{i\ell} \mathbf{g}_{j\ell}}{(1 + \sum_{k \neq i, \ell} \mathbf{g}_{ik})^2 (1 + \sum_{k \neq j, \ell} \mathbf{g}_{jk})^2} \\ \mathbf{Z}_2 &= \sum_{\substack{\ell \neq i, j \\ m \neq i, j \\ \ell \neq m}} \frac{\mathbf{g}_{i\ell} \mathbf{g}_{j\ell} \mathbf{g}_{im} \mathbf{g}_{jm}}{(1 + \sum_{k \neq i, \ell, m} \mathbf{g}_{ik})^2 (1 + \sum_{k \neq j, \ell, m} \mathbf{g}_{jk})^2}. \end{aligned} \quad (133)$$

We are now interested in evaluating the expectation of \mathbf{Z}_1 and \mathbf{Z}_2 . To this aim, two observations are useful. First, we note that the random variables $\mathbf{g}_{i\ell}, \mathbf{g}_{j\ell}, \mathbf{g}_{im}, \mathbf{g}_{jm}$ are, for the indices considered in the pertinent summations, mutually independent Bernoulli variables. Second, the terms in the denominator are binomial random variables that are independent from the variables present in the numerator. This is because, for example, the quantity $\sum_{k \neq i, \ell, m} \mathbf{g}_{ik}$ appearing at the denominator of the second ratio in (133), does not contain any of the random variables $\mathbf{g}_{i\ell}, \mathbf{g}_{j\ell}, \mathbf{g}_{im}, \mathbf{g}_{jm}$ (nor their symmetric counterparts $\mathbf{g}_{\ell i}, \mathbf{g}_{\ell j}, \mathbf{g}_{mi}, \mathbf{g}_{mj}$) that appear at the numerator. Therefore, taking the expectation of \mathbf{Z}_1 in (132) yields:

$$\begin{aligned} \mathbb{E}[\mathbf{Z}_1] &= \sum_{\ell \neq i, j} \mathbb{E}[\mathbf{g}_{i\ell}] \mathbb{E}[\mathbf{g}_{j\ell}] \times \\ &\quad \mathbb{E} \left[\frac{1}{(1 + \sum_{k \neq i, \ell} \mathbf{g}_{ik})^2} \right] \mathbb{E} \left[\frac{1}{(1 + \sum_{k \neq j, \ell} \mathbf{g}_{jk})^2} \right]. \end{aligned} \quad (134)$$

Now, since $\mathbb{E}[\mathbf{g}_{i\ell}] = \mathbb{E}[\mathbf{g}_{j\ell}] = p_N$, since $\sum_{k \neq i, \ell} \mathbf{g}_{ik}$ is a binomial random variable $\beta(N-2, p_N)$, and since the summation in (134) contains no more than N terms, we have that:

$$\mathbb{E}[\mathbf{Z}_1] \leq N p_N^2 \left(\mathbb{E} \left[\frac{1}{(1 + \beta(N-2, p_N))^2} \right] \right)^2. \quad (135)$$

Likewise, taking the expectation of \mathbf{Z}_2 in (132) yields:

$$\begin{aligned} \mathbb{E}[\mathbf{Z}_2] &= \sum_{\substack{\ell \neq i, j \\ m \neq i, j \\ \ell \neq m}} \mathbb{E}[\mathbf{g}_{i\ell}] \mathbb{E}[\mathbf{g}_{j\ell}] \mathbb{E}[\mathbf{g}_{im}] \mathbb{E}[\mathbf{g}_{jm}] \times \\ &\quad \mathbb{E} \left[\frac{1}{(1 + \sum_{k \neq i, \ell, m} \mathbf{g}_{ik})^2} \right] \mathbb{E} \left[\frac{1}{(1 + \sum_{k \neq j, \ell, m} \mathbf{g}_{jk})^2} \right]. \end{aligned} \quad (136)$$

Since $\sum_{k \neq i, \ell, m} \mathbf{g}_{ik}$ is a binomial random variable $\beta(N-3, p_N)$, and since the summation in (136) contains no more than N^2 terms, we have that:

$$\mathbb{E}[\mathbf{Z}_2] \leq N^2 p_N^4 \left(\mathbb{E} \left[\frac{1}{(1 + \beta(N-3, p_N))^2} \right] \right)^2. \quad (137)$$

Finally, using (135) and (137) into (132), we have:

$$\begin{aligned} \mathbb{E}[\mathbf{z}_{ij}^2] &\leq N p_N^2 \left(\mathbb{E} \left[\frac{1}{(1 + \beta(N-2, p_N))^2} \right] \right)^2 + \\ &\quad N^2 p_N^4 \left(\mathbb{E} \left[\frac{1}{(1 + \beta(N-3, p_N))^2} \right] \right)^2. \end{aligned} \quad (138)$$

In view of Lemma 2, we conclude that:

$$\mathbb{E}[\mathbf{z}_{ij}^2] \leq \epsilon_N^{(2)} \sim \frac{1}{N^3 p_N^2} + \frac{1}{N^2}. \quad (139)$$

Let us move on to examining the term $\mathbf{a}_{ij} \mathbf{z}_{ij}$ in (131). Using (49) and the definition of \mathbf{z}_{ij} in (130), we can write:

$$\mathbf{a}_{ij} \mathbf{z}_{ij} \leq \frac{\kappa}{\mathbf{d}_i} \mathbf{g}_{ij} \mathbf{z}_{ij} = \kappa \mathbf{g}_{ij} \frac{\sum_{\ell \neq i, j} \mathbf{g}_{i\ell} \mathbf{g}_{j\ell}}{\mathbf{d}_i^2 \mathbf{d}_j}. \quad (140)$$

Working along the same lines as done to obtain (136) and (137), we can write:

$$\begin{aligned} \mathbb{E}[\mathbf{a}_{ij} \mathbf{z}_{ij}] &\leq \kappa \mathbb{E} \left[\mathbf{g}_{ij} \frac{\sum_{\ell \neq i, j} \mathbf{g}_{i\ell} \mathbf{g}_{j\ell}}{\mathbf{d}_i^2 \mathbf{d}_j} \right] \\ &\leq \kappa \mathbb{E} \left[\sum_{\ell \neq i, j} \frac{\mathbf{g}_{ij} \mathbf{g}_{i\ell} \mathbf{g}_{j\ell}}{(\mathbf{d}_i - \mathbf{g}_{i\ell} - \mathbf{g}_{ij})^2 (\mathbf{d}_j - \mathbf{g}_{j\ell} - \mathbf{g}_{ij})} \right] \\ &= \kappa \sum_{\ell \neq i, j} \mathbb{E}[\mathbf{g}_{ij}] \mathbb{E}[\mathbf{g}_{i\ell}] \mathbb{E}[\mathbf{g}_{j\ell}] \times \\ &\quad \mathbb{E} \left[\frac{1}{(1 + \sum_{k \neq i, j, \ell} \mathbf{g}_{ik})^2} \right] \mathbb{E} \left[\frac{1}{(1 + \sum_{k \neq i, j, \ell} \mathbf{g}_{jk})^2} \right] \\ &\leq \kappa N p_N^3 \times \\ &\quad \mathbb{E} \left[\frac{1}{(1 + \beta(N-3, p_N))^2} \right] \mathbb{E} \left[\frac{1}{1 + \beta(N-3, p_N)} \right], \end{aligned} \quad (141)$$

which, using again Lemma 2, yields:

$$\mathbb{E}[\mathbf{a}_{ij} \mathbf{z}_{ij}] \leq \epsilon_N^{(3)} \sim \frac{1}{N^2}. \quad (142)$$

We can now apply to (131) the upper bounds obtained in (129), (139) and (142), getting:

$$\begin{aligned} \mathbb{E}[\mathbf{b}_{ij}^2] &\leq 4 \mathbb{E}[\mathbf{a}_{ij}^2] + \kappa^4 \mathbb{E}[\mathbf{z}_{ij}^2] + 4\kappa^2 \mathbb{E}[\mathbf{a}_{ij} \mathbf{z}_{ij}] \\ &\leq 4 \epsilon_N^{(1)} + \kappa^4 \epsilon_N^{(2)} + 4\kappa^2 \epsilon_N^{(3)} \\ &\sim \frac{1}{N^2 p_N} + \frac{1}{N^3 p_N^2} + \frac{1}{N^2} \\ &= \frac{1}{N^2 p_N} \left(1 + \frac{1}{N p_N} + p_N \right) \sim \frac{1}{N^2 p_N}, \end{aligned} \quad (143)$$

where the last estimate follows because, under the $\mathcal{G}^*(N, p_N)$ model, $N p_N = \ln N + c_N \rightarrow \infty$, and $p_N \rightarrow 0$, as $N \rightarrow \infty$.

It remains to use the results obtained as regards the decaying rate of \bar{a}^2 and \bar{b}^2 into (127). To this end, it is useful to recast (127) in the following form:

$$\bar{\sigma}_e^2 \leq C(1 - K/N) N \bar{a}^2 \bar{b}^2 + \frac{1/N^2}{(K/N)^2} (1 - \mu)^2, \quad (144)$$

where we used the fact that $\sum_{\ell=1}^{N-K} \mathbf{a}_{\omega_i \omega'_\ell} \leq 1 - \mu$, and that the variance is upper bounded by the mean-square value. Using now (129) and (143) into (144), and noticing further that $K/N \rightarrow \xi$ in view of (8), we conclude that:

$$\bar{\sigma}_e^2 \leq \epsilon_N^{(4)} \sim \frac{1}{N^3 p_N^2} + \frac{1}{N^2}, \quad (145)$$

implying:

$$\frac{(Np_N)^2 \tilde{\sigma}_e^2}{p_N} \leq N^2 p_N \epsilon_N^{(4)} \sim \frac{1}{Np_N} + p_N \xrightarrow{N \rightarrow \infty} 0, \quad (146)$$

which, in view of (116), establishes the proposition. ■

APPENDIX E

We start by introducing the operations of matrix permutation and renumbering. A matrix P is a permutation matrix if exactly one entry in each row and column is equal to 1, and all other entries are 0 — see [48]. Left multiplication of a matrix Z by a permutation matrix P permutes the rows of Z , whereas right multiplication permutes the columns. For example, if:

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (147)$$

then $\tilde{Z} = PZ$ sends the first row of Z to the third row of \tilde{Z} , leaves the second row of Z in the second row of \tilde{Z} , sends the third row of Z to the fourth row of \tilde{Z} , and sends the fourth row of Z to the first row of \tilde{Z} .

Permutation matrices have the following property [48]:

$$P^T = P^{-1}. \quad (148)$$

Since P^T permutes columns in the same way that P permutes rows, the transformation $Z \rightarrow PZP^T$ permutes the rows and columns of Z . In our network setting, where the (i, j) -th matrix entry is associated to a property of agents i and j , the latter transformation corresponds to *renumbering* the agents.

A random matrix Z will be said to be statistically invariant to renumbering if, for any permutation matrix P :

$$PZP^T \stackrel{(d)}{=} Z, \quad (149)$$

where the symbol $\stackrel{(d)}{=}$ denotes equality in distribution. Likewise, Z is invariant to row (resp., to column) permutation, if we have, for any permutation matrix P :

$$PZ \stackrel{(d)}{=} Z \quad (\text{resp., } ZP^T \stackrel{(d)}{=} Z). \quad (150)$$

Lemma 4 (Renumbered combination matrix): Under the Erdős-Renyi model, if the combination policy possesses property P2, the $N \times N$ combination matrix A and the $K \times K$ error matrix E in (68) are statistically invariant to renumbering. Likewise, the $(N - K) \times K$ matrix F in (72) is invariant to row and to column permutation. Formally, letting \mathcal{P}_M be the ensemble of all $M \times M$ permutation matrices, we have that:

$$PAP^T \stackrel{(d)}{=} A, \quad \forall P \in \mathcal{P}_N, \quad (151)$$

$$PEP^T \stackrel{(d)}{=} E, \quad \forall P \in \mathcal{P}_K, \quad (152)$$

$$PF \stackrel{(d)}{=} F, \quad \forall P \in \mathcal{P}_{N-K}, \quad (153)$$

$$FP^T \stackrel{(d)}{=} F, \quad \forall P \in \mathcal{P}_K. \quad (154)$$

Proof: Under the Erdős-Renyi model, the variables g_{ij} , for $i = 1, 2, \dots, N$ and $j > i$, are independent Bernoulli random variables with $\mathbb{P}[g_{ij} = 1] = p_N$, and the matrix G is a symmetric matrix. Therefore, exchanging the agents does not

alter the statistical properties of G , namely, any renumbered version of G has the same probability of occurrence:

$$PGP^T \stackrel{(d)}{=} G \Rightarrow \gamma(PGP^T) \stackrel{(d)}{=} \gamma(G), \quad (155)$$

where the latter equality in distribution holds because $\gamma(\cdot)$ is a deterministic function. Moreover, by property P2 we have that:

$$\gamma(PGP^T) = P\gamma(G)P^T. \quad (156)$$

Since $A = \gamma(G)$, Eqs. (155) and (156) immediately imply (151).

Next we prove (152). Using (68) and (67), the matrix E can be formally written as a function of A as follows:

$$\begin{aligned} E &= A_{\Omega\Omega'} H B_{\Omega'\Omega} \\ &= A_{\Omega\Omega'} (I_{N-K} - [A^2]_{\Omega'})^{-1} [A^2]_{\Omega'\Omega} \\ &\triangleq \psi(A). \end{aligned} \quad (157)$$

Let now \tilde{A} and \tilde{E} denote arbitrarily renumbered versions of A and E , respectively. Since we know that A and \tilde{A} share the same distribution, and since $\psi(\cdot)$ is a deterministic function, also $\psi(A)$ and $\psi(\tilde{A})$ will have the same distribution. Therefore, in view of (157), claim (152) will be proved if we show that any renumbered error matrix, \tilde{E} , can be always written as $\tilde{E} = \psi(\tilde{A})$, for a certain renumbered combination matrix, \tilde{A} . To this end, let us introduce an $N \times N$ permutation matrix P that permutes rows belonging to the index set Ω only with rows belonging to Ω , and rows belonging to the complement set Ω' only with rows belonging to Ω' . Formally, this assumption implies that the principal sub-matrix P_Ω is a $K \times K$ permutation matrix, that $P_{\Omega'}$ is an $(N - K) \times (N - K)$ permutation matrix, and that $P_{\Omega\Omega'}$ and $P_{\Omega'\Omega}$ are matrices containing only zero entries. Therefore, since any renumbering of matrix E is accomplished by using a certain $K \times K$ permutation matrix, we can always write, without losing generality:

$$\tilde{E} = P_\Omega E P_\Omega^T = P_\Omega A_{\Omega\Omega'} P_{\Omega'}^T P_{\Omega'} H P_{\Omega'}^T P_{\Omega'} B_{\Omega'\Omega} P_\Omega^T, \quad (158)$$

where we used (148). From the rules for multiplication of partitioned matrices (and since $P_{\Omega\Omega'}$ and $P_{\Omega'\Omega}$ are matrices with all zeros), we have, for a generic $N \times N$ matrix Z :

$$[PZP^T]_\Omega = P_\Omega Z_\Omega P_\Omega^T, \quad (159)$$

$$[PZP^T]_{\Omega'} = P_{\Omega'} Z_{\Omega'} P_{\Omega'}^T, \quad (160)$$

$$[PZP^T]_{\Omega\Omega'} = P_\Omega Z_{\Omega\Omega'} P_{\Omega'}^T, \quad (161)$$

$$[PZP^T]_{\Omega'\Omega} = P_{\Omega'} Z_{\Omega'\Omega} P_\Omega^T, \quad (162)$$

Therefore, applying (161) and (162) to (158), we get:

$$\tilde{E} = P_\Omega E P_\Omega^T = [PAP^T]_{\Omega\Omega'} P_{\Omega'} H P_{\Omega'}^T [PBP^T]_{\Omega'\Omega}. \quad (163)$$

Moreover, we observe that:

$$\begin{aligned} P_{\Omega'} H P_{\Omega'}^T &= P_{\Omega'} (I_{N-K} - B_{\Omega'})^{-1} P_{\Omega'}^T \\ &= (P_{\Omega'} (I_{N-K} - B_{\Omega'}) P_{\Omega'}^T)^{-1} \\ &= (I_{N-K} - P_{\Omega'} B_{\Omega'} P_{\Omega'}^T)^{-1} \\ &= (I_{N-K} - [PBP^T]_{\Omega'})^{-1} \end{aligned} \quad (164)$$

where we used again (148). On the other hand, we have that:

$$PBP^T = PA^2P^T = (PAP^T)(PAP^T) = (PAP^T)^2. \quad (165)$$

Finally, letting $\tilde{A} = PAP^T$, and using (164) and (165) into (163), we get:

$$\tilde{E} = \tilde{A}_{\Omega'\Omega'}(I_{N-K} - [\tilde{A}^2]_{\Omega'})^{-1}[\tilde{A}^2]_{\Omega'\Omega} = \psi(\tilde{A}), \quad (166)$$

which completes the proof of (152).

Let us now focus on proving (153). Using (72) and (67), the matrix F can be formally written as a function of A as follows:

$$F = HB_{\Omega'\Omega} = (I_{N-K} - [A^2]_{\Omega'})^{-1}[A^2]_{\Omega'\Omega} \triangleq \varphi(A). \quad (167)$$

Let now \tilde{F} denote an arbitrarily row-permuted version of F . Claim (153) will be proved if we show that any row-permuted matrix \tilde{F} can be always written as $\tilde{F} = \varphi(\tilde{A})$, for a certain renumbered combination matrix, \tilde{A} . To this end, let us introduce an $N \times N$ permutation matrix P that permutes rows belonging to the complement set Ω' only with rows belonging to Ω' , while leaving unaltered the rows belonging to the index set Ω . Such assumption implies that the principal sub-matrix $P_{\Omega'}$ is an $(N - K) \times (N - K)$ permutation matrix, that $P_{\Omega} = I_K$, and that $P_{\Omega'\Omega'}$ and $P_{\Omega'\Omega}$ are matrices containing only zero entries. Thus, an arbitrary row-permutation of matrix F can be represented as:

$$\begin{aligned} \tilde{F} = P_{\Omega'}F &= P_{\Omega'}HP_{\Omega'}^TP_{\Omega'}B_{\Omega'\Omega}P_{\Omega'}^T \\ &= P_{\Omega'}HP_{\Omega'}^T[PBPT]_{\Omega'\Omega}, \end{aligned} \quad (168)$$

where we used (148), the equality $P_{\Omega} = I_K$, and (162). Letting now $\tilde{A} = PAP^T$, and using (164) and (165) into (168), we get:

$$\tilde{F} = (I_{N-K} - [\tilde{A}^2]_{\Omega'})^{-1}[\tilde{A}^2]_{\Omega'\Omega} = \varphi(\tilde{A}), \quad (169)$$

which completes the proof of (153). The proof of (154) is similar to the proof of (153), and is accordingly omitted. ■

REFERENCES

- [1] P. Venkatasubramanian, T. He, and L. Tong, "Anonymous networking amidst eavesdroppers," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, Jun. 2008.
- [2] T. He and L. Tong, "Distributed detection of information flows," *IEEE Trans. Inf. Forensics and Security*, vol. 3, no. 3, pp. 390–403, Sep. 2008.
- [3] T. He and L. Tong, "Detection of information flows," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4925–4945, Nov. 2008.
- [4] T. He, A. Agaskar, and L. Tong, "Distributed detection of multi-hop information flows with fusion capacity constraints," *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3373–3383, Jun. 2010.
- [5] J. Kim and L. Tong, "Unsupervised and nonparametric detection of information flows," *Signal Processing*, vol. 92, no. 11, pp. 2577–2593, Nov. 2012.
- [6] S. Marano, V. Matta, T. He, and L. Tong, "The embedding capacity of information flows under renewal traffic," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1724–1739, Mar. 2013.
- [7] S. Mahdizadehghadam, H. Wang, H. Krim and L. Dai, "Information diffusion of topic propagation in social media," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 4, pp. 569–581, Dec. 2016.
- [8] S. Marano, V. Matta, and P. Willett, "The importance of being Earnest: social sensing with unknown agent quality," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 3, pp. 306–320, Sep. 2016.
- [9] M. Cheraghchi, A. Karbasi, S. Mohajer, and V. Saligrama, "Graph-constrained group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 248–262, Jan. 2012.
- [10] H-J. Park and K. Friston, "Structural and functional brain networks: from connections to cognition," *Science*, vol. 342, no. 1658, Nov. 2013.
- [11] C. Quinn, T. Coleman, N. Kiyavash, and N. Hatsopoulos, "Estimating the directed information to infer causal relationships in ensemble neural spike train recordings," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 17–44, Feb. 2011.
- [12] M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: tracking network anomalies via sparsity and low rank," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 50–66, Feb. 2013.
- [13] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, NOW Publishers, Boston-Delft, 2014.
- [14] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.
- [15] V. Matta, P. Braca, S. Marano, and A. H. Sayed, "Diffusion-based adaptive distributed detection: steady-state performance in the slow adaptation regime," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4710–4732, Aug. 2016.
- [16] V. Matta, P. Braca, S. Marano, and A. H. Sayed, "Distributed detection over adaptive networks: refined asymptotics and the role of connectivity," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 4, pp. 442–460, Dec. 2016.
- [17] J. Chen and A. H. Sayed, "On the learning behavior of adaptive networks — Part I: Transient analysis," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3487–3517, Jun. 2015.
- [18] J. Chen and A. H. Sayed, "On the learning behavior of adaptive networks — Part II: Performance analysis," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3518–3548, Jun. 2015.
- [19] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [20] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3775–3789, Jul. 2016.
- [21] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4846–4861, Sep. 2016.
- [22] D. Materassi and M. V. Salapaka, "On the problem of reconstructing an unknown topology via locality properties of the Wiener filter," *IEEE Trans. Autom. Control*, vol. 57, no. 7, pp. 1765–1777, Jul. 2012.
- [23] C. Quinn, N. Kiyavash, and T. Coleman, "Directed information graphs," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6887–6909, Dec. 2015.
- [24] J. Etesami and N. Kiyavash, "Measuring causal relationships in dynamical systems through recovery of functional dependencies," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 4, pp. 650–659, Dec. 2017.
- [25] A. Moneta, N. Chlass, D. Entner, and P. Hoyer, "Causal search in structural vector autoregressive models," in *JMLR Workshop and Conference Proceedings: Causality in Time Series*, vol. 12, pp. 95–114, 2011.
- [26] J. Mei and J. Moura, "Signal processing on graphs: causal modeling of unstructured data," *IEEE Trans. Signal Process.*, to appear, DOI: 10.1109/TSP.2016.2634543, 2016.
- [27] 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.*, to appear, DOI: 10.1109/TSIPN.2017.2742940, 2017.
- [28] B. Pasdeloup, M. Rabbat, V. Gripon, D. Pastor, and G. Mercier, "Graph reconstruction from the observation of diffused signals," in *Proc. of the 53rd Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Sep. 29 - Oct. 2, 2015, pp. 1386–1390.
- [29] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology identification from imperfect spectral templates," in *Proc. of the Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Nov. 2016, pp. 1465–1469.
- [30] D. Materassi and M. V. Salapaka, "Network reconstruction of dynamical polytrees with unobserved nodes," in *Proc. of the IEEE 51th Annual Conference on Decision and Control*, Maui, Hawaii, USA, Dec. 10-13, 2012, pp. 4629–4634.
- [31] J. Etesami, N. Kiyavash, and T. Coleman, "Learning minimal latent directed information polytrees," *Neural Computation*, vol. 28, no. 9, pp. 1723–1768, Aug. 2016.
- [32] P. Geiger, K. Zhang, M. Gong, D. Janzing, and B. Schölkopf, "Causal inference by identification of vector autoregressive processes with hidden components," in *Proc. of the 32nd International Conference on Machine Learning*, Lille, France, Jul. 2015, vol. 37, pp. 1917–1925.

- [33] D. Materassi and M. V. Salapaka, "Identification of network components in presence of unobserved nodes," in *Proc. of the IEEE 54th Annual Conference on Decision and Control*, Osaka, Japan, Dec. 15-18, 2015, pp. 1563–1568.
- [34] T. Kailath, A. H. Sayed, and B. Hassibi *Linear Estimation*. Prentice Hall, 2000.
- [35] A. H. Sayed, *Adaptive Filters*. Wiley, NJ, 2008.
- [36] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, no. 1, pp. 65–78, Sep. 2004.
- [37] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.
- [38] D. Bajovic, D. Jakovetic, J. M. F. Moura, J. Xavier, and B. Sinopoli, "Large deviations performance of consensus+innovations distributed detection with non-Gaussian observations," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5987–6002, Nov. 2012.
- [39] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 674–690, Aug. 2011.
- [40] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," in *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.
- [41] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.
- [42] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multiagent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.
- [43] P. Erdős and N. Rényi, "On random graphs, I," *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.
- [44] B. Bollobás, *Random Graphs*. Cambridge University Press, 2001.
- [45] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.
- [46] A. Frieze and M. Karóński, *Introduction to Random Graphs*. Cambridge University Press, 2016.
- [47] H. Shao, *Mathematical Statistics*, 2nd ed., Springer, 2003.
- [48] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, New York, 1985.