TOMOGRAPHY OF ADAPTIVE MULTI-AGENT NETWORKS UNDER LIMITED OBSERVATION

Vincenzo Matta

DIEM, University of Salerno
via Giovanni Paolo II, I-84084, Fisciano, Italy
e-mail: vmatta@unisa.it

Ali H. Sayed *

EPFL, School of Engineering
CH-1015 Lausanne, Switzerland
e-mail: ali.sayed@epfl.ch

ABSTRACT

This work studies the problem of inferring from streaming data whether an agent is directly influenced by another agent over an adaptive network of interacting agents. Agent i influences agent j if they are connected, and if agent j uses the information from agent i to update its inference. The solution of this inference task is challenging for at least two reasons. First, only the output of the learning algorithm is available to the external observer and not the raw data. Second, only observations from a fraction of the network agents is available, with the total number of agents itself being also unknown. This work establishes, under reasonable conditions, that consistent tomography is possible, namely, that it is possible to reconstruct the interaction profile of the observable portion of the network, with negligible error as the network size increases. We characterize the decaying behavior of the error with the network size, and provide a set of numerical experiments to illustrate the results.

Index Terms— Diffusion networks, network tomography, combination policy, Erdös-Rényi model.

1. INTRODUCTION

One fundamental challenge of network science is the inverse modeling problem. In this problem, the network structure (topology) is unknown and one is interested in inferring relationships between network agents based on data arising from their activities. Inverse network modeling is usually challenging because: i) the inference of inter-agent relations must be based on *indirect* observations, with direct access to the data at the agents being impossible or impractical; and ii) the access to observations is *limited to a subset of the network agents*. The process of discovering inter-agent interactions from indirect/partial measurements is broadly referred to as *network tomography*. This work addresses network tomography for the class of *adaptive* networks, which can be succinctly described as an ensemble of dispersed agents exchanging information through diffusion mechanisms, so as to deliver simultaneous *adaptation* and *learning* from streaming data [1–6].

The problem of retrieving a graph topology from indirect measurements taken at some accessible network locations falls under the umbrella of signal processing over graphs [1, 2, 7–9]. Several works have considered similar problems, albeit with different specific goals. For space limitations, we provide here only a compact list of works that are most related to the present article. In [10], the locality properties of Wiener filters are exploited to provide exact reconstruction for specific network types. In [11], *directed* information graphs are considered in order to exploit the effect of *causal* dynamics. Particularly relevant for our work is the case where the relationship between agents obeys an autoregressive diffusion model, see, e.g., [12–14], and [15, 16], where additional sparsity constraints are exploited in conjunction with the spectral graph properties. The case of unobserved nodes is considered in [17], where sufficient conditions for the identification of a given link are offered, in terms of some structural constraints on the network topology. Estimation of covariance matrices with missing data is addressed in [18].

This work complements these previous efforts by answering the following questions. We consider a diffusion network solving some inference task (such as a distributed detection problem). The network size is unknown, and the outputs of the diffusion algorithm are available from only a *limited* subset of agents. The goal is establishing whether an individual agent is influencing another individual agent. For the class of Erdös-Rényi *random* graphs running diffusion strategies with *symmetric* combination matrices, we discover that the group of interacting agent pairs and the group of non-interacting agent pairs split into two well-defined clusters, for any (nonzero) fraction of observable agents. These clusters emerge as clearly separate for sufficiently large network sizes, so that the interaction relationships within the observable network can be recovered with negligible error. In this way, we establish that *consistent tomography from partially observed networks is achievable*.

Notation. Boldface letters denote random variables, normal font their realizations. Capital letters refer to matrices, small letters to vectors or scalars. The (i, j)-th entry of a matrix Z is denoted by z_{ij}, or by $[Z]_{ij}$. The sub-matrix of Z corresponding to the rows indexed by $\mathcal{S}_1$ and the columns indexed by $\mathcal{S}_2$ is denoted by $Z_{\mathcal{S}_1\mathcal{S}_2}$ (or $[Z]_{\mathcal{S}_1\mathcal{S}_2}$). The notation $Z_{\mathcal{S}}$ (or $[Z]_{\mathcal{S}}$) is used when $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$.

2. THE NETWORK TOMOGRAPHY PROBLEM

A network of N agents gathers streaming data from the environment. The datum of the i-th agent at time n is denoted by $\boldsymbol{x}_i(n)$, and the data are spatially and temporally independent and identically distributed random variables, with zero mean and unit variance. 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 strategy, whose properties in terms of estimation and online learning performance have been already studied in detail — see, e.g. [1, 2, 5, 6]. The CTA algorithm can be described as follows. The output variable of the i-th agent at time n is denoted by $\boldsymbol{y}_i(n)$ During the *combination* step, agent i mixes the output variables received from its neighbors by using a sequence of convex (i.e., nonnegative and adding up to 1) combination weights $w_{i\ell}$, giving rise to the intermediate variable:

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

*This work was supported in part by NSF grants CCF-1524250 and ECCS- 1407712.

Then, during the *adaptation* step, agent i updates its output variable by comparing with the incoming streaming data $\boldsymbol{x}_i(n)$ and using some small step-size $\mu \in (0,1)$:

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

These two steps can be compactly represented by a single equation:

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

It is convenient to introduce the *scaled* combination matrix A, with entries $a_{ij} \triangleq (1-\mu)w_{ij}$. Since we use a sequence of convex combination weights, the matrix $A/(1-\mu)$ is a *right-stochastic* matrix. Moreover, in this work we focus on the case of symmetric combinations matrices, and, hence, $A/(1-\mu)$ is *doubly*-stochastic. If we now stack the data into the $N\times 1$ vector $\boldsymbol{x}_n$, and the output variables into the $N\times 1$ vector $\boldsymbol{y}_n$, Eq. (3) becomes:

$$\boldsymbol{y}_n = A\, \boldsymbol{y}_{n-1} + \mu\, \boldsymbol{x}_n \Rightarrow \boldsymbol{y}_n = \mu \sum_{m=1}^{n} A^{n-m}\boldsymbol{x}_m, \quad n \geq 1, \quad (4)$$

where, in the rightmost equality, we assume $\boldsymbol{y}_0 = 0$. That is, we neglect the *transient* term.

2.1. Network Tomography

A Tomography Center (TC) is interested in establishing which agent is influencing which other agent. The TC collects the streams of outputs exchanged by a *subset* of the network agents. Letting $\Omega \subset \{1,2,\dots,N\}$ be the observable subnet, the data available to the TC at time n are $\{\boldsymbol{y}_i(1), \boldsymbol{y}_i(2), \dots, \boldsymbol{y}_i(n)\}_{i\in\Omega}$. We consider the regime of large networks ($N \to \infty$) and the case where the fraction of observed agents does not vanish. In other words, if we let $K = |\Omega|$, then we assume that $(K/N) \to \xi$, where $\xi \in (0,1)$ is the asymptotic fraction of observed agents. In our setting, the overall network size is unknown, and the goal of the TC is to produce an estimate of the interaction profile for the *observed* agents.

Let us ignore for the moment the partial-observability limitation. Assuming that the TC is able to collect all output sequences from all agents at all times, there exist in principle several well-established solutions to infer whether an agent influences another agent. When choosing one particular solution, however, it is necessary to take into account the following peculiar aspect related to the streaming nature of the data. In general, when the TC starts working, the network would have been in operation since some time already. As a result, the agents' outputs would have benefited from substantial exchange of information. While this can be beneficial for the solution of the inference problem by the agents, it can however become detrimental for retrieving the network topology. This is because (over a strongly connected network and after a transient stage), *all* agents would become mutually correlated.

In order to overcome this difficulty, one can exploit knowledge of the diffusion mechanism. Let us introduce the correlation matrix, $R_0(n) \triangleq \mathbb{E}[\boldsymbol{y}_n\boldsymbol{y}_n^T]$, and the one-lag correlation matrix, $R_1(n) \triangleq \mathbb{E}[\boldsymbol{y}_n\boldsymbol{y}_{n-1}^T]$, of the diffusion output vector, which, using (4), can be written as:

$$\begin{aligned} R_0(n) &= \mu^2 \sum_{i=0}^{n-1} A^{2i} \stackrel{n\to\infty}{\longrightarrow} R_0 = \mu^2(I_N - A^2)^{-1}, \\ R_1(n) &= AR_0(n-1) \stackrel{n\to\infty}{\longrightarrow} R_1 = AR_0, \end{aligned} \quad (5)$$

where I_N is the $N\times N$ identity matrix. From (5) we obtain the following relationships, for $n \geq 2$:

$$A = R_1(n)(R_0(n-1))^{-1} \Rightarrow A = R_1 R_0^{-1}, \quad (6)$$

and since there are several ways to estimate R_0 and R_1 consistently as $n \to \infty$, expression (6) reveals that estimating A from the diffusion output is in principle feasible.

Unfortunately, in our setting the approach described so far suffers from a problem: not all entries in the matrices R_0 and R_1 are available since the network is only *partially* observed. In order to estimate the (sub-)matrix, $A^{(\mathrm{obs})} \triangleq A_\Omega$, corresponding to the observable subnet, it is tempting to replace R_0 and R_1 by their *observable* counterparts, $R_0^{(\mathrm{obs})} \triangleq [R_0]_\Omega$, and $R_1^{(\mathrm{obs})} \triangleq [R_1]_\Omega$, yielding:

$$\boxed{\hat{A}^{(\mathrm{obs})} = R_1^{(\mathrm{obs})}(R_0^{(\mathrm{obs})})^{-1}} \quad (7)$$

Needless to say, the calculation on the right-hand side of (7) does not lead to the true $A^{(\mathrm{obs})}$. Therefore, it is not clear at all whether the mutual influence relationships existing between the observed nodes can be consistently retrieved from $\hat{A}^{(\mathrm{obs})}$. Answering this nontrivial question in the affirmative is the main contribution of this work.

3. ERROR DUE TO PARTIAL OBSERVATIONS

We are interested in establishing whether the estimated values $\hat{a}_{ij}^{(\mathrm{obs})}$ (for $i \neq j$) allow us to identify the condition $a_{ij}^{(\mathrm{obs})} > 0$ or $a_{ij}^{(\mathrm{obs})} = 0$, which would reveal whether agents i and j influence each other. To this end, we start by introducing, with reference to (7), an error matrix E:

$$\hat{a}_{ij}^{(\mathrm{obs})} = a_{ij}^{(\mathrm{obs})} + e_{ij}, \quad (i,j = 1,2,\dots,K), \quad (8)$$

whose behavior is characterized in the next theorem, stated without proof for space limitations. Details can be found in [21].

Theorem 1 (Concentration of errors) *For a symmetric combination matrix, the entries of the error matrix defined in (8) are nonnegative, and satisfy for all $i = 1,2,\dots,K$:*

$$\boxed{\sum_{j=1}^{K} e_{ij} \leq 1-\mu} \quad (9)$$

□

Theorem 1 provides useful information about the concentration of the entries in the error matrix. Consider now a small threshold $\epsilon > 0$, and the fraction of off-diagonal (because we are interested in inter-agent interactions) entries exceeding such threshold, namely,

$$\frac{1}{K(K-1)}\sum_{i=1}^{K}\sum_{j\neq i}\mathbb{I}\{e_{ij} > \epsilon\}, \quad (10)$$

with $\mathbb{I}\{\cdot\}$ being the indicator function. Using Theorem 1, it can be shown that such fraction is upper bounded by the quantity $\frac{1-\mu}{\epsilon(K-1)}$, which vanishes because in our setting $K \to \infty$ as $N \to \infty$. As a result, most entries of the error matrix will be small for large networks, implying, for $i \neq j$:

$$\hat{a}_{ij}^{(\mathrm{obs})} = \begin{cases} a_{ij}^{(\mathrm{obs})} + \text{small quantity}, & \text{if } a_{ij}^{(\mathrm{obs})} > 0, \\ \text{small quantity}, & \text{if } a_{ij}^{(\mathrm{obs})} = 0. \end{cases} \quad (11)$$

This useful splitting suggests that the nonzero entries of $A^{(\mathrm{obs})}$ will make the estimated entries $\hat{a}_{ij}^{(\mathrm{obs})}$ stand out above the error floor as N increases, which in turn suggests that the network graph can be retrieved by comparing the estimated value, $\hat{a}_{ij}^{(\mathrm{obs})}$, against some threshold. However, the behavior of the error matrix alone is not sufficient to conclude that this procedure is effective. This is because, for typical combination matrices, the nonzero entries $a_{ij}^{(\text{obs})}$ vanish with N as well, implying that the estimated entries, $\hat{a}_{ij}^{(\text{obs})}$, would vanish even when agents i and j are interacting. For this reason, it is necessary to assess how fast the error signals e_{ij} decay to zero in relation to the desired entries $a_{ij}^{(\text{obs})}$. The forthcoming sections address such analysis, with reference to some popular random models used to describe the network structure formation.

4. INTERACTING AGENTS ON RANDOM GRAPHS

The interaction profile will be described through a symmetric matrix G, with $g_{ij} = 1$ if agents i and j interact, and $g_{ij} = 0$ otherwise. We assume that an agent always uses its own output variable in the combination step, implying that $g_{ii} = 1$. The combination matrix will be constructed through the following two-step procedure. First, an interaction matrix G is generated according to a *random graph* model [19, 20]. Then, $A = \gamma(G)$ is determined by a *combination policy*, $\gamma(G)$, which assigns the values a_{ij} corresponding to the nonzero entries of G. In particular, $\gamma(G)$ always assigns *positive* weights at the locations corresponding to *nonzero* entries of G, and $[\gamma(G)]_{ij} = [\gamma(G)]_{ji}$.

The random graph model we consider is the classic Erdös-Rényi model [19, 20]. This model, denoted by $\mathscr{G}(N, p_N)$, is an undirected graph where the presence of the $N(N-1)/2$ edges is determined by a sequence of $N(N-1)/2$ independent Bernoulli random variables with success probability p_N. A well-known result that holds true for the Erdös-Rényi graph is that the scaling law $p_N = \frac{1}{N}(\ln N + c_N)$, with $c_N \to \infty$, ensures that the graph is connected, with probability tending to 1 as N diverges [20]. In the following, we focus on the regime of connected Erdös-Rényi graph with *vanishing* p_N, namely, on the regime where $c_N \to \infty$ and $p_N \to 0$. This regime will be denoted by the symbol $\mathscr{G}^\star(N, p_N)$.

Let us now examine on the asymptotic behavior of the (off-diagonal) nonzero entries in the combination matrix. We start by considering a popular combination policy, the Laplacian rule, which is defined, for $0 < \lambda \le 1$, as [1]:

$$a_{ij} = \begin{cases} g_{ij}\,(1-\mu)\,\lambda/d_{\text{max}}, & \text{for } i \neq j, \\ (1-\mu) - \sum_{\ell \neq i} a_{i\ell} & \text{for } i = j, \end{cases} \tag{12}$$

where d_{max} is the *maximal* degree of the graph, with the degree of agent i (denoted by d_i), being the number of its neighbors including i itself. Now, for an Erdös-Rényi graph, the random variable $\boldsymbol{d_i} - 1$ is a binomial random variable with parameters $N-1$ and p_N. Therefore, we see that the degrees of the nodes scale as Np_N. Since $\boldsymbol{d}_{\text{max}}$ is the maximum of N degrees, it is expected to grow faster than Np_N. Interestingly, the next lemma shows that it cannot grow *much* faster.

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

$$\boxed{\mathbb{P}[\boldsymbol{d}_{\text{max}} \ge Np_N e \,|\, \boldsymbol{g}_{ij} = 1] \le \left(e + \frac{2e^2}{N}\right) e^{-c_N} \overset{N\to\infty}{\longrightarrow} 0} \tag{13}$$

where e is Euler's number. ☐

Applying Lemma 1 to (12) we can write, for all $i \neq j$:

$$\begin{aligned} &\mathbb{P}[Np_N \boldsymbol{a}_{ij} > (1-\mu)\lambda/e \,|\, \boldsymbol{g}_{ij} = 1] \\ &= \quad \mathbb{P}[\boldsymbol{d}_{\text{max}} < Np_N e \,|\, \boldsymbol{g}_{ij} = 1] \overset{N\to\infty}{\longrightarrow} 1. \end{aligned} \tag{14}$$

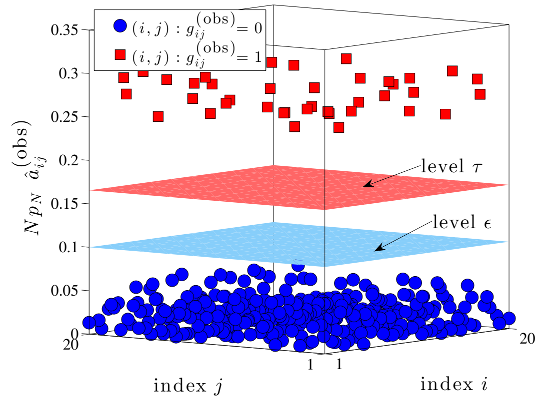

Fig. 1. A pictorial illustration of Theorem 2: the scaled estimated entries $\hat{a}_{ij}^{(\text{obs})}$ cluster in two groups depending on whether agents i and j are interacting or not.

Equation (14) reveals that, for large N, the nonzero entries of the Laplacian combination matrix, magnified by the factor Np_N, stay "almost always" above a certain threshold, and, hence, the magnified nonzero entries will not vanish as N diverges. The same scaling factor is in fact relevant for other combination policies, leading to the formulation of the following general class of combination policies.

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

$$\boxed{\mathbb{P}[Np_N \boldsymbol{a}_{ij} > \tau \,|\, \boldsymbol{g}_{ij} = 1] \ge 1 - \epsilon_N} \tag{15}$$

where ϵ_N goes to zero as $N \to \infty$, and where the probability is evaluated under the $\mathscr{G}^\star(N, p_N)$ model. ☐

Let us now examine the physical meaning of (15) in connection to network tomography applications. For a policy belonging to class $\mathscr{C}_\tau$, we can rephrase (11) as:

$$Np_N \hat{a}_{ij}^{(\text{obs})} = \begin{cases} \underbrace{Np_N a_{ij}^{(\text{obs})}}_{\text{not vanishing}} + \underbrace{Np_N e_{ij}}_{\text{small quantity?}}, & \text{if } a_{ij}^{(\text{obs})} > 0, \\ \underbrace{Np_N e_{ij}}_{\text{small quantity?}}, & \text{if } a_{ij}^{(\text{obs})} = 0, \end{cases} \tag{16}$$

where the "not vanishing" behavior follows by (15). According to (16), if we can prove that $Np_N e_{ij}$ is *still* a small quantity, then $\hat{a}_{ij}^{(\text{obs})}$ would be useful for tomography purposes, because the magnified nonzero entry $Np_N a_{ij}^{(\text{obs})}$ would stand out from the error floor as N gets large. Actually, using Theorem 1 it is possible to show that this desirable property holds true, enabling therefore consistent tomography, as will be formally stated in Theorem 2 further ahead.

5. CONSISTENT TOMOGRAPHY

Let us introduce the number of non-interacting ($\boldsymbol{N}_0$) and the number of interacting ($\boldsymbol{N}_1$) agent pairs over the observed set:

$$\boldsymbol{N}_0 \triangleq \sum_{i=1}^{K} \sum_{j \neq i} (1 - \boldsymbol{g}_{ij}^{(\text{obs})}), \qquad \boldsymbol{N}_1 \triangleq \sum_{i=1}^{K} \sum_{j \neq i} \boldsymbol{g}_{ij}^{(\text{obs})}, \tag{17}$$

where $\boldsymbol{g}_{ij}^{(\text{obs})} = \mathbb{I}\{\boldsymbol{a}_{ij}^{(\text{obs})} > 0\}$. Next we introduce the number of entries in $Np_N \hat{A}^{(\text{obs})}$ that stay below some positive level α, for non-

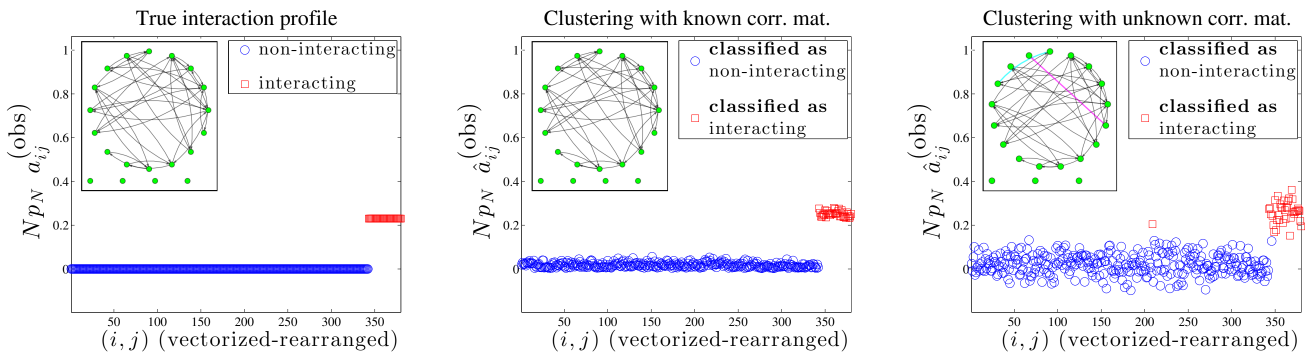

Fig. 2. Network tomography for the case of a Laplacian combination rule in (12), with parameter $\lambda = 0.5$. The network size is $N = 100$, where only $K = 20$ agents are observable. The interaction probability is $p_N = 2\,(\ln N)/N \approx 0.092$, and the step-size is $\mu = 0.1$.

interacting and interacting agent pairs, respectively:

$$\boldsymbol{N}_0(\alpha) \triangleq \sum_{i=1}^{K}\sum_{j\neq i}\mathbb{I}\{Np_N\hat{\boldsymbol{a}}_{ij}^{(\mathrm{obs})} \leq \alpha,\, \boldsymbol{g}_{ij}^{(\mathrm{obs})} = 0\}, \quad (18)$$

$$\boldsymbol{N}_1(\alpha) \triangleq \sum_{i=1}^{K}\sum_{j\neq i}\mathbb{I}\{Np_N\hat{\boldsymbol{a}}_{ij}^{(\mathrm{obs})} \leq \alpha,\, \boldsymbol{g}_{ij}^{(\mathrm{obs})} = 1\}. \quad (19)$$

Finally, we introduce the conditional empirical distributions given that the agents are not interacting and given that they are interacting, namely, $\boldsymbol{\mathcal{F}}_0(\alpha) = \boldsymbol{N}_0(\alpha)/\boldsymbol{N}_0$, and $\boldsymbol{\mathcal{F}}_1(\alpha) = \boldsymbol{N}_1(\alpha)/\boldsymbol{N}_1$, where $\boldsymbol{\mathcal{F}}_0(\alpha)$ (resp., $\boldsymbol{\mathcal{F}}_1(\alpha)$) is conventionally set to $1/2$ when $\boldsymbol{N}_0 = 0$ (resp., $\boldsymbol{N}_1 = 0$). The next theorem establishes achievability of consistent tomography through the asymptotic characterization of the aforementioned empirical distributions (Details on the proof can be found in [21]).

Theorem 2 (Achievability of consistent tomography) *Let the network interaction profile obey a $\mathscr{G}^\star(N, p_N)$ model, and let the combination policy belong to class $\mathscr{C}_\tau$. Then, for any asymptotic fraction of observable agents, $\xi > 0$, we have:*

$$\boxed{\boldsymbol{\mathcal{F}}_0(\epsilon) \xrightarrow{\mathrm{p}} 1 \;\; \forall \epsilon > 0, \qquad \boldsymbol{\mathcal{F}}_1(\tau) \xrightarrow{\mathrm{p}} 0} \quad (20)$$

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

Theorem 2 reveals the following useful dichotomy: *i*) when agents i and j are *not* interacting, most of the (magnified) estimated matrix entries stay below *an arbitrarily small* level ϵ; *ii*) when agents i and j are interacting, most of the (magnified) estimated matrix entries stay above a positive value τ. Therefore, two separate clusters emerge, one corresponding to the region $Np_N\hat{a}_{ij}^{(\mathrm{obs})} \leq \epsilon$, and the other one corresponding to the region $Np_N\hat{a}_{ij}^{(\mathrm{obs})} > \tau$. This situation is illustrated in Fig. 1. Theorem 2 reveals also that, in the limit of large network sizes, the consistent tomography property does *not* depend on the fraction of observable nodes, ξ.

When prior knowledge about τ and Np_N is available, Theorem 2 provides a direct recipe to reconstruct the interaction profile. In the lack of such knowledge, the existence itself of the clustering structure opens up the possibility of employing non-parametric pattern recognition strategies to perform cluster separation We give an example of this possibility in the next section.

6. ILLUSTRATIVE EXAMPLES

We now examine the practical significance of the asymptotic results derived in the previous sections. In the simulations, we use the Laplacian combination rule, while the observations $\boldsymbol{x}_i(n)$ fed into the diffusion algorithm follow a standard normal distribution. The interaction profile is retrieved by using the off-diagonal entries of $\hat{A}^{(\mathrm{obs})}$, and applying a k-means clustering algorithm in order to split these entries into two clusters. We consider two cases: (a) the case in which the exact correlation matrices are known, and (b) the case in which they must be estimated from the diffusion outputs. In the latter case, as an estimator for $R_0^{(\mathrm{obs})}$ and $R_1^{(\mathrm{obs})}$ we use the empirical correlations, and we compute (7) by replacing R_0 and R_1 with their estimates.

In Fig. 2, leftmost panel, we display the off-diagonal entries of the (magnified) *true* combination matrix corresponding to the *observable* subnet. 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 $A^{(\mathrm{obs})}$ will be then applied to the matrices displayed in the remaining panels. Interacting agent pairs are marked by a red square, whereas non-interacting pairs with a blue circle. 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 display the estimated matrix, $Np_N\hat{A}^{(\mathrm{obs})}$, computed under perfect knowledge of $R_0^{(\mathrm{obs})}$ and $R_1^{(\mathrm{obs})}$, and we show the classification performed by the k-means algorithm. In the rightmost panel the same type of analysis is reported, but the estimated matrix is computed using the *empirical* correlation matrices. In the latter two panels, matrix entries are marked in a way that depends on the results of the clustering: blue-circles if agents i and j are *classified as* non-interacting, whereas red-square markers if they are *classified as* interacting. Examining the middle panel, we see that the experiments confirm the theoretical analysis, since the entries of the matrix $\hat{A}^{(\mathrm{obs})}$ are well-separable. Examining the rightmost panel in comparison with the middle panel, we see that the estimated clusters are more "noisy", which produces a few classification errors. Such a behavior makes sense since the procedure applied in the rightmost panel *must* be affected by the error in estimating $R_0^{(\mathrm{obs})}$ and $R_1^{(\mathrm{obs})}$.

In the inset plots we display the network graphs (of the observable subnet) corresponding to the strategy addressed in the pertinent panel. Such graphs 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 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.

7. REFERENCES

[1] 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.

[2] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] C. Quinn, N. Kiyavash, and T. Coleman, "Directed information graphs," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6887–6909, Dec. 2015.

[12] 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.

[13] J. Mei and J. Moura, "Signal processing on graphs: causal modeling of *un*structured data," *IEEE Trans. Signal Process., to appear*, DOI: 10.1109/TSP.2016.2634543, 2016.

[14] 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.

[15] 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.

[16] 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.

[17] D. Materassi and M. V. Salapaka, "Identification of network components in presence of unobserved nodes," in *Proc. of the 54th Annual Conference on Decision and Control (CDC)*, Osaka, Japan, Dec. 2015, pp. 1563–1568.

[18] M. Zamanighomi, Z. Wang, and G. B. Giannakis, "Estimating high-dimensional covariance matrices with misses for Kronecker product expansion models," in *Proc. of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 2667–2671.

[19] P. Erdös and N. Rényi, "On random graphs, I.," *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.

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

[21] V. Matta and A. H. Sayed, "Consistent tomography under partial observations over adaptive networks," *submitted for publication*. Available as arXiv:1707.06444v1 [cs.MA], Jul. 2017.