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 examples to illustrate the results.

Index Terms— Diffusion networks, network tomography, combination policy, Erdős-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 interactions. 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 Erdős-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 $S_1$ and the columns indexed by $S_2$ is denoted by $Z_{S_1,S_2}$ (or $[Z]_{S_1,S_2}$). The notation $Z_{S_0}$ (or $[Z]_{S_0}$) is used when $S_1 = S_2 = S$.

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 $x_i(n)$, 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 $y_i(n)$, 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 $w_{i,:}$, giving rise to the intermediate variable:

$$w_i(n-1) = \sum_{\ell=1}^{N} w_{i,\ell} y_i(n-1).$$

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

$$y_i(n) = v_i(n-1) + \mu[x_i(n) - v_i(n-1)].$$

(2)

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

$$y_i(n) = (1 - \mu) \sum_{\ell=1}^{N} w_{i\ell} y_i(n-1) + \mu x_i(n).$$

(3)

It is convenient to introduce the scaled combination matrix $A$, with entries $a_{ij} \overset{\Delta}{=} (1 - \mu)w_{ij}$. Since we use a sequence of convex combinations 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 $N \times N$ matrices, and, hence, $N$ all stack the data into the $N \times N$ matrices, and, hence, $N$ solutions to infer whether an agent influences another agent. When agents at all times, there exist in principle several well-established accounts the following peculiar aspect related to the streaming nature of the diffusion mechanism. Let us introduce the correlation matrix, written as:

$$A_{(1-\mu)} = \hat{R}_1 - \hat{R}_0^{-1}.$$  

(7)

Needless to say, the calculation on the right-hand side of (7) does not lead to the true $A_{(1-\mu)}$. Therefore, it is not clear at all whether the mutual influence relationships existing between the observed nodes can be consistently retrieved from $\hat{A}_{(1-\mu)}$. 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 $\hat{a}_{ij}$ (for $i \neq j$) allow us to identify the condition $\hat{a}_{ij} > 0$ or $\hat{a}_{ij} = 0$, 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 $E^i$:

$$e_{ij}^i = a_{ij} - \hat{a}_{ij}, \quad (i, j = 1, 2, \ldots, K),$$

(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 $i = 1, 2, \ldots, K$:

$$\sum_{j=1}^{K} e_{ij}^i \leq 1 - \mu$$

(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,

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

(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 $\frac{1 - \mu}{\epsilon(N - 1)}$, which vanishes because in our setting $K \rightarrow \infty$ as $N \rightarrow \infty$. As a result, most entries of the error matrix will be small for large networks, implying, for $i \neq j$:

$$\hat{a}_{ij} = \begin{cases} a_{ij} + \text{small quantity}, & \text{if } a_{ij} > 0, \\ \text{small quantity}, & \text{if } a_{ij} = 0. \end{cases}$$

(11)

This useful splitting suggests that the nonzero entries of $A_{(1-\mu)}$ will make the estimated entries $\hat{a}_{ij}$ 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, $\hat{a}_{ij}$, 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 $a_{ij}^{(obs)}$ vanish with $N$ as well, implying that the estimated entries $\hat{a}_{ij}^{(obs)}$ 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 $a_{ij}^{(obs)}$. 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 $g_{ii} = 1$. This model, denoted by $G_{graph}$ where the presence of the $p_{N}$ nonzero entries of $G$ is determined, for $0 < \lambda \leq 1$, as [1]:

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

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 Erdős-Rényi graph, the random variable $d_i - 1$ is a binomial random variable with parameters $N - 1$ and $p_N$. Therefore, we see that the degrees of the nodes scale as $N p_N$. Since $d_{max}$ is the maximum of $N$ degrees, it is expected to grow faster than $N p_N$. Interestingly, the next lemma shows that it cannot grow much faster.

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

$$P[d_{max} \geq N p_N e \mid g_{ij} = 1] \leq \left( 1 + \frac{2e^2}{N} \right) e^{-c_N N \rightarrow \infty} 0 \quad (13)$$

where $c_N$ is Euler’s number. \H

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

$$P[N p_N a_{ij} > (1 - \mu) / e \mid g_{ij} = 1] = P[d_{max} < N p_N e \mid g_{ij} = 1] \quad (14)$$

Equation (14) reveals that, for large $N$, the nonzero entries of the Laplacian combination matrix, magnified by the factor $N p_N$, 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 $\mathcal{C}_e$.** A combination policy belongs to class $\mathcal{C}_e$ if there exists $\tau > 0$ such that, for all $i, j = 1, 2, \ldots, N$, with $i \neq j$:

$$P[N p_N a_{ij} > \tau \mid g_{ij} = 1] \geq 1 - 1 / e_N \quad (15)$$

where $e_N$ goes to zero as $N \rightarrow \infty$, and where the probability is evaluated under the $G^*(N, p_N)$ model. \H

We let now examine the physical meaning of (15) in connection to network tomography applications. For a policy belonging to class $\mathcal{C}_e$, we can rephrase (11) as:

$$N p_N a_{ij}^{(obs)} + N p_N e_{ij}, \quad \text{if } a_{ij}^{(obs)} > 0,$$

$$N p_N e_{ij}, \quad \text{if } a_{ij}^{(obs)} = 0, \quad (16)$$

where the “not vanishing” behavior follows by (15). According to (16), if we can prove that $N p_N e_{ij}$ is still a small quantity, then $a_{ij}^{(obs)}$ would be useful for tomography purposes, because the magnified nonzero entry $N p_N a_{ij}^{(obs)}$ 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 ($\mathcal{N}_0$) and the number of interacting ($\mathcal{N}_i$) agent pairs over the observed set:

$$\mathcal{N}_0 \triangleq \sum_{i=1}^{K} \sum_{j \neq i} (1 - g_{ij}^{(obs)}), \quad \mathcal{N}_i \triangleq \sum_{i=1}^{K} \sum_{j \neq i} g_{ij}^{(obs)} \quad (17)$$

where $g_{ij}^{(obs)} = \mathbb{I}\{a_{ij}^{(obs)} > 0\}$. Next we introduce the number of entries in $N p_N \hat{A}_{ij}^{(obs)}$ that stay below some positive level $c$, for non-
we now examine the practical significance of the asymptotic results of observable agents, \( \xi > 0 \), where \( F \) is a standard normal distribution. The interaction profile is retrieved by using the diagonal entries of \( A^{(obs)} \), 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 \( R_0^{(obs)} \) and \( R_1^{(obs)} \) 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 \( A^{(obs)} \) 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, \( N_{PN} \hat{A}^{(obs)} \), computed under perfect knowledge of \( R_0^{(obs)} \) and \( R_1^{(obs)} \), 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.

Examing the middle panel, we see that the experiments confirm the theoretical analysis, since the entries of the matrix \( A^{(obs)} \) are well-separated. 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 \( R_0^{(obs)} \) and \( R_1^{(obs)} \).

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.

### Theorem 2 (Achievability of consistent tomography)

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 \( \epsilon \); ii) when agents \( i \) and \( j \) are interacting, most of the (magnified) estimated matrix entries stay above a positive value \( \tau \). Therefore, two separate clusters emerge, one corresponding to the region \( N_{PN} \hat{A}^{(obs)} \leq \epsilon \), and the other one corresponding to the region \( N_{PN} \hat{A}^{(obs)} > \tau \). 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, \( \xi \).

When prior knowledge about \( \tau \) and \( N_{PN} \) 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 \( \mathbf{x}_i(n) \) fed into the diffusion algorithm follow a standard normal distribution. The interaction profile is retrieved by using the off-diagonal entries of \( A^{(obs)} \), and applying a k-means clustering algorithm 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 \( R_0^{(obs)} \) and \( R_1^{(obs)} \) 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 \( A^{(obs)} \) 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, \( N_{PN} \hat{A}^{(obs)} \), computed under perfect knowledge of \( R_0^{(obs)} \) and \( R_1^{(obs)} \), 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 \( A^{(obs)} \) are well-separated. 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 \( R_0^{(obs)} \) and \( R_1^{(obs)} \).

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


