Divide-and-Conquer Tomography for Large-Scale Networks

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

***Abstract*—This work considers the problem of reconstructing the topology of a network of interacting agents via observations of the state-evolution of the agents. A divide-and-conquer strategy is developed where the network is reconstructed via repeated applications of local reconstructions. In each step, observations from only a subset of the nodes are collected, and the information is used to infer their local connectivity. A fundamental result is to establish that local tomography is possible with high probability. Once established, this step is then used as a building block for reconstructing the larger network.**

***Index Terms*—Network inference, local inference, big-data, large-scale networks, network tomography.**

I. Introduction

In networked dynamical systems [1]–[4] the state of the agents comprising the network evolves over time and is affected by peer-to-peer interactions among the connected nodes. In general, information about the network graph is unavailable and one may be able to observe measurements from only a subset of the network nodes over time. It is the goal of *network tomography* to extract information about the network connectivity from *observables*. In this work we propose an algorithm to perform network tomography for large-scale graphs. A divide-and-conquer strategy is developed where the network is reconstructed via repeated applications of *local tomography* steps to different patches of the network, denoted by $S_1, \ldots, S_P$. In each step, observations from only a subset of the nodes (or a patch) are collected, and the information is used to infer their local connectivity, as illustrated in Fig. 1. A fundamental result is to establish that local tomography is possible with high probability. Once established, this step is then used as a building block for reconstructing the larger network graph.

Some key factors motivate the divide-and-conquer approach developed in this work. Thus note that in large-scale networked systems, such as brain neuron networks or large Internet-of-Things networks, one can only observe and/or process limited portions of the system *per experiment* in order to make inference or extract information about the system. There are two fundamental reasons that cause this limitation due to limited probing abilities and limited processing-time abilities:

- *Probing-limit.* The acquisition of data and storage capacities are often far smaller than the scale of the network.

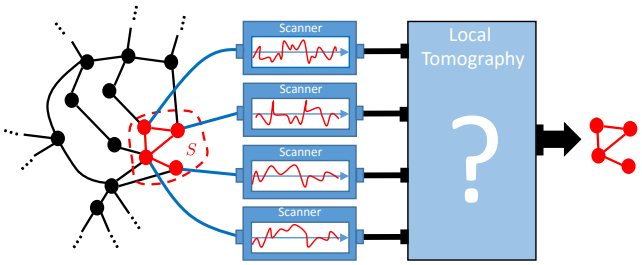

Fig. 1. Illustration of the *local tomography problem*. The goal is to devise a mechanism to reconstruct the underlying subnetwork topology by appropriately processing the state-evolution of the observable nodes.

- *Processing-limit.* The complexity of the data-mining further constrains the size of the data that can be processed.

For instance, one may probe the activity of a subset of neurons – as it is unfeasible to track the activity of all the brain neuron network – in order to reconstruct its underlying profile of interactions (a.k.a. connectome). This requires to partially observe the system at each experiment and extract information about its underlying subnetwork of interactions. One could hopefully sequentially integrate the inference results of the various local tomographic experiments as an attempt to deduce global information about the large-scale networked system, in particular, its topology (which is object of inference).

Related work. Most works focus on full-observability where the state of *all* agents can be tracked and processed. In the context of linear (or linearized) dynamical systems, particular attention is paid to (autoregressive) diffusion models [5]–[9]. The majority of the literature addresses tomography by exploring commonalities between covariance constructs on the data and the underlying graph of interactions; and by further resorting to optimization-based methods that reinforce some (application-dependent) structural constraints such as, e.g., sparsity, stability and symmetry. For instance, references [6], [7] explore the spectral commonalities between the underlying combination matrix and the covariance of the observed samples, which reduces the set of candidate topologies, and the inverse problem is then addressed via an optimization method that reinforces sparsity. A Wiener filter based approach is explored in [10] yielding an *approximation* of the true topology, although certain non-interacting pairs end up being classified as interacting. Nonlinear dynamics are often dealt with via linear variational characterizations of the dynamics [11]–[13] (which yield linear models resembling the dynamics in equation (1)) or by appropriately increasing

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

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

the dimension of the observable space [14], [15]. Regarding partial observability, references [16], [17] provide sufficient conditions on the network structure that allows reconstruction. Such conditions depend strongly on the network structure and may not be suitable to a large-scale network setting. In contrast, our result matches well the large-scale paradigm. In fact, by modeling the large network through a *random* graph, we are able to obtain simple *average* conditions (e.g., in terms of the probability that two nodes are connected) that guarantee topology reconstruction with high probability as the network size scales to infinity.

Preliminary notation.

- In what follows, we use boldface letters to denote random variables, and normal font letters for their realizations.
- Given an $N \times N$ matrix Z, the submatrix that lies in the rows and columns of Z indexed by the set $S \subseteq \{1, 2, \ldots, N\}$, will be denoted by $[Z]_S$.

II. Network and Data Model

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

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

where $\boldsymbol{y}_n = [\boldsymbol{y}_1(n), \boldsymbol{y}_2(n), \ldots, \boldsymbol{y}_N(n)]$ is a column vector collecting the states of all nodes at time n, $A = [a_{i\ell}]$ is a *combination matrix*, and $\boldsymbol{x}_n = [\boldsymbol{x}_1(n), \boldsymbol{x}_2(n), \ldots, \boldsymbol{x}_N(n)]$ is a column vector modeling a random input (e.g., streaming data or noise). We assume that $\{\boldsymbol{x}_i(n)\}$ are independent and identically distributed (i.i.d.) both spatially (i.e., w.r.t. to i) and temporally (i.e., w.r.t. to n), with zero-mean and unit variance. We observe from (1) that, if $a_{i\ell} = 0$, then agent i does not use the information arriving from agent ℓ to update its own state. Therefore, the *support-graph* of A reflects the underlying topology.

The stochastic dynamical system (1) arises naturally in the context of *adaptive diffusion networks* [18], [19]. It also arises in economics, e.g., [5], and is also used as a variational characterization of nonlinear networked dynamical systems, e.g., [11]. In other words, tomography analysis over such family of stochastic dynamical systems is useful for a broad class of networked systems.

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

$$R_1(n) = AR_0(n-1) \overset{n\to\infty}{\longrightarrow} R_1 = AR_0, \tag{2}$$

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

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

$$\boxed{\hat{A}_S = [R_1]_S \left([R_0]_S\right)^{-1}} \tag{3}$$

It is clear from basic linear algebra that $\hat{A}_S \neq A_S$. However, a key result established in [18] is that $\widehat{A}_S$ provides a consistent estimator for A_S, with high probability, under certain reasonable conditions on the combination matrix A. The result in Theorem 1 builds on reference [18] and is established in [20].

III. Key result and algorithm: Patch-and-Catch (PAC)

Let G_S be the adjacency matrix corresponding to the connecting observed agents S, with arbitrary topology. Let $\boldsymbol{G}$ be the global adjacency matrix for the network of N nodes. We assume that the connections of $\boldsymbol{G}$ are drawn as follows.

$$\begin{cases} \mathbb{P}\left(\boldsymbol{g}_{ij} = 1\right) = p_N, & \text{if } i \notin S \text{ or } j \notin S \\ \boldsymbol{g}_{ij} = [G_S]_{ij}\,, & \text{if } i, j \in S \end{cases} \tag{4}$$

where the edges $\boldsymbol{g}_{ij}$ are drawn with probability p_N and independently over the pairs ij when $i \notin S$ or $j \notin S$. In other words, G_S is a subnetwork of $\boldsymbol{G}$ with arbitrary (deterministic) topology and the complementary network is drawn as an Erdös-Rényi. We assume that $\boldsymbol{G}$ (and hence G_S) is unknown. The goal is to estimate G_S via observing the state evolution of the observable agents in S.

To each graph G, we can assign (positive) weights to the edges of G and denote the resulting matrix of weights by A. Some useful choices are the Laplacian and the Metropolis rules, defined as follows. Let $\rho \in (0, 1)$ and $\lambda \in (0, 1]$:

• **Laplacian rule.**

$$a_{ij} = \begin{cases} \rho\lambda g_{ij}/d_{\max}, & \text{for } i \neq j \\ \rho - \sum_{k\neq i} a_{ik}, & \text{for } i = j \end{cases}, \tag{5}$$

• **Metropolis rule.**

$$a_{ij} = \begin{cases} \rho g_{ij}/\max\{d_i, d_j\}\,, & \text{for } i \neq j \\ \rho - \sum_{k\neq i} a_{ik}, & \text{for } i = j \end{cases}, \tag{6}$$

where d_i is the degree of agent i and $d_{\max}$ is the maximum degree in the network. These rules arise naturally in the context of adaptive diffusion networks [19].

Theorem 1 (Exact local tomography). *Let $\boldsymbol{G}$ be a graph generated as described above and let $\boldsymbol{A}$ be any nonnegative symmetric $N \times N$ matrix with support graph $\boldsymbol{G}$ (as constructed*

above) and obeying certain stability conditions[1]*. Furthermore, assume that the connection probabilities are of the form* $p_N = (1/N)(\log N + c_N)$*, with* c_N *being any divergent sequence such that* $[\log(\log N + c_N)]^2/\log N \to 0$*. Then, there exists* $\tau > 0$ *so that*

$$\boxed{\lim_{N\to\infty} \mathbb{P}[\Gamma_\tau(Np_N\hat{A}_S) = G_S] = 1} \tag{7}$$

where $\Gamma_\tau(B) = I_{\{b_{ij}>\tau\}}$ *corresponds to the adjacency matrix obtained from the matrix* B *by* τ*-thresholding its entries and* $\hat{A}_S$ *is the truncated estimator given by* (3). □

Theorem 1 is a key (yet highly non-trivial) result; its proof is omitted – it builds on [18] and is established in [20]. The theorem essentially asserts the possibility of performing local tomography over large-scale diffusion networks. This is because it establishes the existence of a threshold τ such that the entries of the naïve estimator $\hat{A}_S$ provide correct local tomography with high-probability. In particular, let $\eta \triangleq \tau/(Np_N)$. Then, the topology of the observable network can be recovered for sufficiently large N as follows: If $\hat{a}_{ij} \leq \eta$, then classify ij as non-interacting, otherwise if $\hat{a}_{ij} > \eta$, then classify ij as interacting. We still need to select an appropriate threshold τ in order to correctly classify interacting and non-interacting pairs from the truncated estimator $\hat{A}_S$. We observe that prior information on the dynamical system (1) can help in the choice. In particular, if we know the average degree Np_N of the network and the nature of the combination matrix, e.g., Laplacian or Metropolis, then we can estimate a threshold. For the Laplacian or Metropolis rules, we could set

$$\eta_L = \rho/\left(eNp_N\right), \qquad \eta_M = \rho\lambda/\left(eNp_N\right), \tag{8}$$

respectively, as these are thresholds that meet the limit in (7). We remark that Theorem 1 is an asymptotic (in N) result, and that even in the presence of prior information there can be a range of possible thresholds yielding correct classification *in the limit*. It is not clear at this stage how to optimize this threshold in order to grant high performance in the *finite-network* scenario. A higher threshold reduces the likelihood of classifying non-interacting pairs as interacting, whereas a lower threshold increases that likelihood.

When prior information is not available, we can follow an alternative non-parametric approach by applying a k-means clustering algorithm (using $k = 2$) to the off-diagonal entries of $\hat{A}_S$. The clustering algorithm will end up splitting the off-diagonal entries into two clusters (connected and non-connected). The cluster with higher arithmetic mean is classified as the one corresponding to connected nodes.

In this work, we empirically estimate the truncated correlation matrices $[R_0]_S$ and $[R_1]_S$ from the observed data as

$$\begin{aligned} \overline{[R_1]}_S &= \sum_{n=0}^{n_{\max}-1} [y_{n+1}]_S \, [y_n]_S^\top \\ \overline{[R_0]}_S &= \sum_{n=0}^{n_{\max}} [y_n]_S \, [y_n]_S^\top \end{aligned} . \tag{9}$$

[1]We omit the details, but the conditions can be found in [18] and are satisfied by the Laplacian and the Metropolis rules.

We remark that it is possible to optimize such estimates by exploiting prior information on the structural properties of the system, and such an optimization could boost the performance of the algorithm. In equation (9), we have not used boldface notation to emphasize that the observed y_n refers to a particular realization.

Algorithm. We now describe our divide-and-conquer tomography algorithm. We let M be the probing capacity (so that no more than M agents can be observed per experiment). Let S be the subset of nodes in which the subnetwork G_S is to be reconstructed. The set S is partitioned into the subsets (or patches) $S_1, \ldots, S_P$ with cardinality $|S_i| \leq M/2$, for $i = 1, 2, \ldots, P$. Each local reconstruction experiment will correspond to probing two patches, resulting in a total of $P(P-1)/2$ experiments. Note that only probing and reconstructing each patch S_i is not sufficient to reconstruct G_S as all connections between nodes across distinct patches will be missed.

Algorithm 1 Patch-and-Catch (PAC) Strategy for Network Tomography

Input: Ensemble of patches $\{S_1, \ldots, S_P\}$ and observables $\left\{[y_n]_{S_i}\right\}$ over the patches $i = 1, \ldots, P$ and for time $n = 0, 1, \ldots, n_{max}$

Output: Estimated adjacency matrix $\overline{A}_S$ associated with the subnetwork G_S of observable nodes

while $i \leq P$ **do**
  while $j < i$ **do**
    $\overline{[R_1]}_{S_i\cup S_j} = \sum_{n=0}^{n_{\max}-1} [y_{n+1}]_{S_i\cup S_j} \, [y_n]_{S_i\cup S_j}^\top$
    $\overline{[R_0]}_{S_i\cup S_j} = \sum_{n=0}^{n_{\max}} [y_n]_{S_i\cup S_j} \, [y_n]_{S_i\cup S_j}^\top$
    $\overline{A}_{S_i\cup S_j} = \overline{[R_1]}_{S_i\cup S_j} \left(\overline{[R_0]}_{S_i\cup S_j}\right)^{-1}$
    $\overline{A}_{i,j} = k\text{-means}\left(\overline{A}_{S_i\cup S_j}\right)$
    $j = j + 1$
  end while
  $j = 1$
  $i = i + 1$
end while

IV. Numerical Experiments

In this section, we illustrate the practical performance of the PAC algorithm. We start with the single-patch case in order to illustrate the asymptotic result of Theorem 1 while providing a preview of the multi-patch case.

Single-patch. Figure 2 displays the (empirically-estimated) topology-recovery probability, with reference to an overall network with number of nodes N ranging from 10 to 200, and for the case of Laplacian and Metropolis combination rules with $\rho = 0.8$ and $\lambda = 1$. The observable network is made up of $K = 10$ nodes. The probability of correct recovery gets close to 1 as N increases for all the considered scenarios: parametric *versus* k-means thresholding, and empirically estimated truncated correlation matrices (as in equation (9))

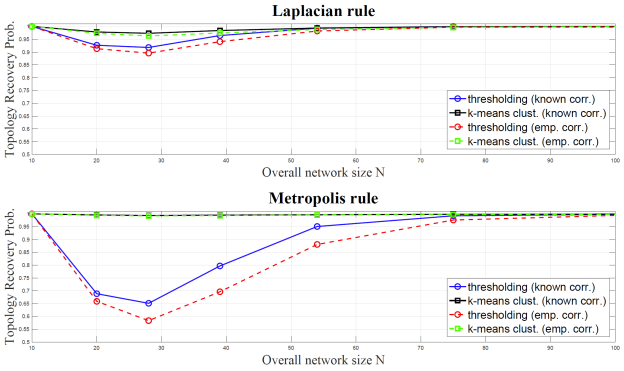

Fig. 2. Empirical topology-recovery probability *versus* size of overall network. *Thresholding* means that the entries of $\hat{A}_S$ were thresholded with η_L and η_M (refer to equation (8) in Section III). *Empirical correlation* means that the truncated correlations were estimated as in (9), whereas the curves with *known correlations* are also displayed as a superior limit in performance.

versus known truncated correlation matrices. We notice that the recovery probability curve is not monotonic. Indeed, when $N = K = 10$, all the network is observed, and in view of equation (2) (and the comments that follow this equation) the recovery probability must be equal to 1. On the other hand, Theorem 1 ensures that exact recovery should be also attained asymptotically (in N). As discussed, a break in performance in a *pre-limit region* is not surprising, especially for the non-optimal thresholding method based on the prior information as discussed in the previous section.

Multi-patch. We consider an overall network of $N = 300$ nodes. We apply the PAC algorithm in a sub-region S with K nodes and probing capacity of $M < K$. That is, at each tomographic experiment we can only probe and reconstruct a subnetwork of M nodes within S. The goal is to reconstruct the subnetwork G_S associated with the agents in S. The overall graph G is generated as an Erdös-Rényi radom graph with probability of connection $p_N = 5(\log N)/N$. The combination matrix A is obtained via the Metropolis rule. The k-means algorithm is adopted as the thresholding method. The truncated correlation matrices are empirically estimated from the data as in equation (9) with $n_{\max} = 10^5$ samples.

Regarding the patches, and as described in Section III, we assume that $S_1, \ldots, S_P$ form a partition of the subset S of interest with $|S_i| = M/2$. That is, $S = \bigcup_{i=1}^{P} S_i$ and $S_i \cap S_j = \emptyset$ for all $i \neq j$. Each local inference experiment corresponds to probing two patches, resulting in a total of $P(P-1)/2$ experiments.

We consider two cases. In the first, we assume $K = 20$ and in the second $K = 60$. In both cases we assume the probing limit $M = 10$. The smaller K in the first case allows to illustrate graphically the reconstruction of the subnetwork G_S by the PAC algorithm as depicted in Fig. 3. In the second case, we rather illustrate the evolution of the distance between the true graph G_S and the estimated graph as more experiments are performed, Fig. 4. We assume that initially the estimated graph has no edges and we define the (normalized) distance

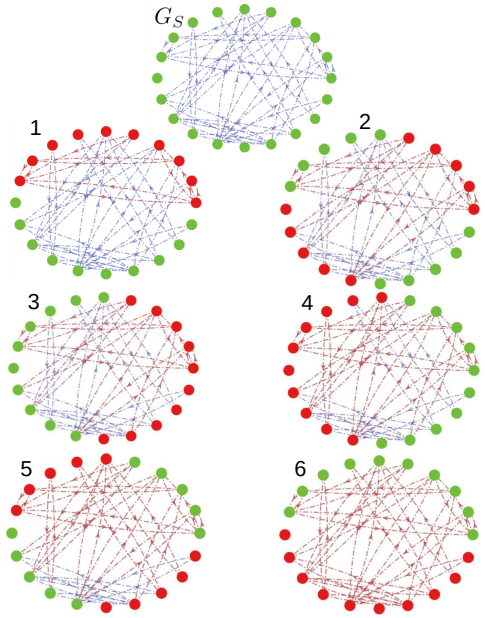

Fig. 3. Illustration of the graph reconstruction. We consider $K = 20$ nodes with probing limit $M = 10$. Each patch has $|S_i| = 5$ nodes. At each experiment two patches are probed. The red nodes represent the nodes being probed at each experiment and the red edges represent the inferred edges up to the current experiment. All pairs were correctly classified.

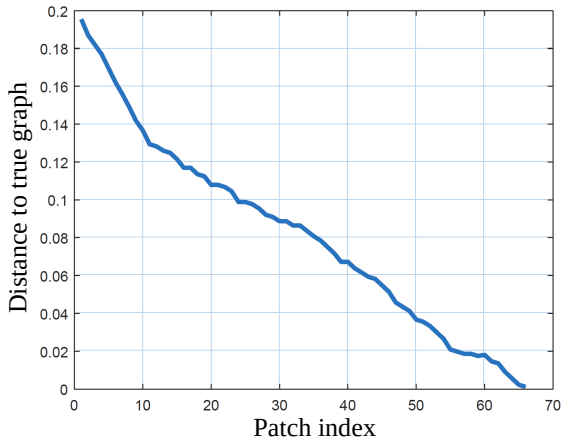

Fig. 4. Illustration of the monotonic decrease of the distance between the subnetwork G_S and the estimator as more patches are probed. We consider $K = 60$ nodes in S with probing limit $M = 10$. Each patch has $|S_i| = 5$ nodes. We have $P = 2K/M = 12$ patches. At each experiment two patches are probed, yielding a total of $P(P-1)/2 = 66$ experiments. The graph displays the distance between G_S and the estimated graph at the experiment $\ell = 1, 2, \ldots, 66$. Only two pairs were misclassified.

between two undirected graphs (or their corresponding adjacency matrices) $G^{(1)}$ and $G^{(2)}$ on K nodes as

$$\text{dist}\left(G^{(1)}, G^{(2)}\right) = \frac{\sum_{i<j} \left| g_{ij}^{(1)} - g_{ij}^{(2)} \right|}{K(K-1)/2}. \qquad (10)$$

REFERENCES

[1] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks*. London, UK: Cambridge University Press, November 2012.

[2] T. Liggett, *Interacting Particle Systems*, 1st ed. Springer-Verlag Berlin Heidelberg, 2005.

[3] P. Robert, *Stochastic Networks and Queues*, ser. Stochastic Modelling and Applied Probability. Springer-Verlag, 2003, vol. 52.

[4] M. Porter and J. Gleeson, *Dynamical Systems on Networks: A Tutorial*, ser. Frontiers in Applied Dynamical Systems: Reviews and Tutorials. Springer International Publishing, 2016. [Online]. Available: https://books.google.com/books?id=uzDuCwAAQBAJ

[5] A. Moneta, N. Chlass, D. Entner, and P. O. Hoyer, "Causal search in structural vector autoregressive models," in *NIPS Mini-Symposium on Causality in Time Series*, 2011.

[6] B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and M. Rabbat, "Characterization and inference of graph diffusion processes from observations of stationary signals," *IEEE Transactions on Signal and Information Processing over Networks,* available as IEEE early access article, 2017.

[7] S. Segarra, M. T. Schaub, and A. Jadbabaie, "Network inference from consensus dynamics," 2017. [Online]. Available: http://arxiv.org/abs/1708.05329

[8] J. Mei and J. M. F. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2077–2092, April 2017.

[9] D. Thanou, X. Dong, D. Kressner, and P. Frossard, "Learning heat diffusion graphs," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 484–499, September 2017.

[10] D. Materassi and M. V. Salapaka, "On the problem of reconstructing an unknown topology via locality properties of the Wiener filter," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1765–1777, July 2012.

[11] E. S. C. Ching and H. C. Tam, "Reconstructing links in directed networks from noisy dynamics," *Phys. Rev. E*, vol. 95, p. 010301, Jan 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.95.010301

[12] D. Napoletani and T. D. Sauer, "Reconstructing the topology of sparsely connected dynamical networks," *Physical review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 77, p. 026103, 2008.

[13] J. Ren, W.-X. Wang, B. Li, and Y.-C. Lai, "Noise bridges dynamical correlation and topology in coupled oscillator networks," *Phys. Rev. Lett.*, vol. 104, p. 058701, Feb 2010. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.104.058701

[14] Y. Yang, T. Luo, Z. Li, X. Zhang, and P. S. Yu, "A robust method for inferring network structures," in *Scientific Reports*, vol. 7, no. 5221, July 2017.

[15] A. Mauroy and J. Goncalves, "Linear identification of nonlinear systems: A lifting technique based on the koopman operator," *CoRR*, vol. abs/1605.04457, 2016. [Online]. Available: http://arxiv.org/abs/1605.04457

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

[17] P. Geiger, K. Zhang, B. Schölkopf, M. Gong, and D. Janzing, "Causal inference by identification of vector autoregressive processes with hidden components," in *Proceedings of the 32nd International Conference on Machine Learning*, vol. 37, Lille, France, July 2015, pp. 1917–1925.

[18] V. Matta and A. H. Sayed, "Consistent Tomography under Partial Observations over Adaptive Networks," July 2017. [Online]. Available: http://arxiv.org/abs/1707.06444

[19] A. H. Sayed, "Adaptation, Learning, and Optimization over Networks," *Found. Trends Mach. Learn.*, vol. 7, no. 4-5, pp. 311–801, 2014. [Online]. Available: http://dx.doi.org/10.1561/2200000051

[20] A. Santos, V. Matta, and A. H. Sayed, "Consistent tomography over diffusion networks under the low-observability regime," in *2018 IEEE International Symposium on Information Theory*, Colorado, USA, June 2018, accepted.