

DISTRIBUTED NETWORK TOPOLOGY RECONSTRUCTION IN PRESENCE OF ANONYMOUS NODES.

Thi-Minh-Dung Tran and Alain Y. Kibangou

Univ. Grenoble Alpes, CNRS, Inria, GIPSA-Lab, F-38000 Grenoble, France

ABSTRACT

This paper concerns the problem of reconstructing the network topology from data propagated through the network by means of an average consensus protocol. The proposed method is based on the distributed estimation of graph Laplacian spectral properties. Precisely, the identification of the network topology is implemented by estimating both eigenvalues and eigenvectors of the consensus matrix, which is related to the graph Laplacian matrix. In this paper, we focus the exposition on the estimation of the eigenvectors since the eigenvalues estimation can be achieved based on recent results of the literature using the same kind of data. We show how the topology can be reconstructed in presence of anonymous nodes, i.e. nodes that do not disclose their ID.

Index Terms— Network topology Reconstruction, Graph Laplacian spectrum, Eigenvectors, Anonymous nodes, Average Consensus.

1. INTRODUCTION

Network topology is usually a schematic description of the arrangement of a network, including its nodes and connecting edges. Inferring the network topology from data circulating in the network is, in general, a crucial issue in different science and engineering domains including biology, finance, computer science, transportation (delivery and distribution network), and electrical engineering. In computer networks, for example, the architecture of an overlay network - how it allocates addresses, etc.- may be significantly optimized by the knowledge of the distribution and connectivity of the nodes on the underlay network that actually carries the traffic. Several important systems, such as P4P and RMTTP, utilize information about the topology of the underlay network for optimization as well as management of the network [1].

In the literature, various methods have been proposed for reconstructing the network topology. Most approaches are based on an excite and observe paradigm. Centralized methods have been first introduced. In [2], a grounding procedure where a node broadcasts a zero state to its neighbors without being removed from a consensus network was introduced. This node knockout approach follows the same spirit than the gene-knockout procedure used in experimental biology for identifying interactions in gene networks. In [3], for

directed networks of linear time-invariant systems, a wide-sense stationary noise of unknown power spectral density was considered as stimulating input. The authors proposed several reconstruction algorithms based on the power spectral properties of the network response to that noise. In [4], the identifiability of the structure and dynamics of an unknown network driven by unknown noise was assessed based on factorizations of the output spectral density. The authors claimed that for networks with closed-loop transfer matrices that are minimum phase, the network reconstruction problem can have a unique solution from its output spectral density. In [5], a link between dynamical correlation and topology was established for coupled oscillators. For an undirected network with symmetric coupling matrix, it was stated that, in presence of noise, the dynamical correlation matrix is proportional to the pseudo-inverse of the graph Laplacian matrix. Based on this method, in [6] a fully distributed algorithm for reconstructing the topology of an undirected consensus network by using a distributed least-squares algorithm was presented. An efficient method for separating the off-diagonal entries of the reconstructed graph Laplacian matrix based on κ -means clustering was introduced.

In all the aforementioned methods, exogenous input are necessary to excite the network and in general ID or labels of nodes are assumed to be known. In nowadays networks, for privacy issues for instance, some nodes can prefer to not disclose their identity (or label). In such a case the reconstruction of the network topology, in a distributed way, becomes particularly challenging. In this paper, we study the reconstruction of the network topology when there is no exogenous input and in presence of anonymous nodes. We assume that the nodes are running an average consensus protocol and then make use of the data produced by the network during the transient to carry out the spectral decomposition of a Laplacian-based consensus matrix. For this purpose, we assume that the eigenvalues of the consensus matrix can be obtained from our previous works [7, 8]. Then, we show how to use these eigenvalues in order to estimate the eigenvectors that are used to infer the network topology. It is worth noting that in [9], knowing IDs of all the nodes, the authors resorted to an algebraic method using observability properties of a network. Both eigenvalues and eigenvectors of the Laplacian matrix were estimated with the price of significant data storage and

multiple initializations of the consensus protocol. In [10] a decentralized spectral analysis was proposed with the aim of estimating some top eigenvectors of a symmetric adjacency matrix. This approach is as complex as the one in [9].

The paper is organized as follows: in Section 2, we formulate the problem under study. Then, the distributed solution of the network topology reconstruction problem is presented in Section 3. Before concluding the paper, the performance of the proposed method is illustrated in Section 4 by means of simulation results.

2. PROBLEM STATEMENT

Consider a network of nodes modeled with a connected undirected graph $G = (V, E)$, where $V = \{1, 2, \dots, N\}$ is the set of vertices of graph G (set of nodes), and $E \subset V \times V$ is the set of edges (set of links between nodes). The set of neighbors of node i is denoted by $N_i = \{j \in V : (i, j) \in E\}$ whereas $d_i = |N_i|$ stands for its degree. The network topology is completely captured by the Laplacian matrix \mathbf{L} whose entries are given by $l_{ii} = d_i$, $l_{ij} = -1$ if $(i, j) \in E$, and $l_{ij} = 0$ elsewhere.

If each node knows its label (ID) and those of its neighbors then it knows exactly the corresponding row and column of the Laplacian matrix. In what follows we assume that such a knowledge is only partial, meaning that some nodes do not disclose their ID to their neighbors. As a consequence, the nodes knowledge is up to permutations.

Let consider that the nodes perform a consensus protocol. For each node $i \in V$, let $x_i(k)$ denotes the value of node i at time-step k . At each time-step each node updates its state as

$$x_i(k) = x_i(k-1) + \epsilon \sum_{j \in N_i} (x_j(k-1) - x_i(k-1)), \quad (1)$$

in order to reach an agreement. Defining the state of the network as $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_N(k)]^T$, its time evolution is given by:

$$\mathbf{x}(k) = (\mathbf{I}_N - \epsilon \mathbf{L}) \mathbf{x}(k-1) = \mathbf{W} \mathbf{x}(k-1), \quad k = 1, 2, \dots, \quad (2)$$

where $\mathbf{W} = \mathbf{I}_N - \epsilon \mathbf{L}$ stands for the consensus matrix. With an appropriately chosen value of ϵ , the nodes reach asymptotically an agreement corresponding to the average of the initial values [11]:

$$\lim_{k \rightarrow \infty} \mathbf{x}(k) = \frac{1}{N} \mathbf{1} \mathbf{1}^T \mathbf{x}(0).$$

It has been shown in [12] that the averaging matrix $\frac{1}{N} \mathbf{1} \mathbf{1}^T$ can be factored as

$$\frac{1}{N} \mathbf{1} \mathbf{1}^T = \prod_{q=1}^D (\mathbf{I}_N - \gamma_q \mathbf{L}), \quad (3)$$

where $\{\gamma_q\}_{q=1}^D$ denotes the set of the inverse of the D distinct nonzero Laplacian eigenvalues. Given the consensus value

and the transient measurements of the network state, distributed solutions to factorization (3) of the averaging matrix have been recently proposed by the authors in [7, 8]. Therefore, the problem under study, in this paper, is the following: *Given the estimated Laplacian eigenvalues and the transient measurements of the network state, reconstruct the consensus matrix \mathbf{W} and then the Laplacian matrix \mathbf{L} knowing that some nodes ID are unknown.*

3. DISTRIBUTED SOLUTION FOR NETWORK TOPOLOGY RECONSTRUCTION

Assumption 1: All the nonzero Laplacian eigenvalues are distinct¹.

Let us consider the following eigenvalue decomposition of the consensus matrix \mathbf{W} :

$$\mathbf{W} = \mathbf{U} \mathbf{D} \mathbf{U}^T \quad (4)$$

where \mathbf{U} and \mathbf{D} contain the eigenvectors and the eigenvalues of \mathbf{W} , respectively. Recall that the eigenvalues of \mathbf{W} and those of \mathbf{L} are related as $\mathbf{D} = \mathbf{I}_N - \epsilon \Delta$, Δ being the diagonal matrix of Laplacian eigenvalues. Note that for an arbitrary nonzero vector \mathbf{b} , we can alternatively write \mathbf{W} as:

$$\mathbf{W} = \tilde{\mathbf{U}} \mathbf{D} \text{diag}(\mathbf{b}^2)^{-1} \tilde{\mathbf{U}}^T, \quad (5)$$

where $\tilde{\mathbf{U}} = \mathbf{U} \text{diag}(\mathbf{b})$, and therefore $\tilde{\mathbf{U}} \tilde{\mathbf{U}}^T = \mathbf{U} \text{diag}(\mathbf{b}^2) \mathbf{U}^T$ and $\tilde{\mathbf{U}}^T \tilde{\mathbf{U}} = \text{diag}(\mathbf{b}^2)$. We can then estimate scaled eigenvectors $\tilde{\mathbf{U}}$ instead of \mathbf{U} .

At a given node j , our purpose is to reconstruct the matrix \mathbf{W} , knowing $x_j(k)$, $k = 0, 1, \dots, N-1$, and the matrix \mathbf{D} . We will first estimate the unknown scaled eigenvectors.

3.1. Estimation of scaled eigenvectors of the Laplacian-based consensus matrix

From (2), we know that:

$$\mathbf{x}(k) = \mathbf{W}^k \mathbf{x}(0) = \mathbf{U} \mathbf{D}^k \mathbf{U}^T \mathbf{x}(0).$$

Since the Laplacian eigenvectors form a basis of \mathbb{R}^N , the initial condition $\mathbf{x}(0)$ can be expanded such that $\mathbf{x}(0) = \sum_{i=1}^N \beta_i \mathbf{U}_{\cdot i} = \mathbf{U} \mathbf{b}$, where $\mathbf{b} = (\beta_1 \ \beta_2 \ \dots \ \beta_N)^T$ contains the expansion coefficients, and $\mathbf{U}_{\cdot i}$ stands for the i -th eigenvector, i.e. the i -th column of \mathbf{U} . Therefore:

$$\mathbf{x}(k) = \mathbf{U} \mathbf{D}^k \mathbf{U}^T \mathbf{U} \mathbf{b} = \mathbf{U} \mathbf{D}^k \mathbf{b} = \mathbf{U} \text{diag}(\mathbf{b}) \text{vecd}(\mathbf{D}^k),$$

where $\text{vecd}(\cdot)$ is the column vector built with the diagonal entries of the matrix in argument. We can equivalently write the above equation as:

$$\mathbf{x}(k) = \tilde{\mathbf{U}} \text{vecd}(\mathbf{D}^k), \quad \text{with } \tilde{\mathbf{U}} = \mathbf{U} \text{diag}(\mathbf{b})$$

¹This assumption is not restrictive. Random graphs fulfill this assumption with high probability

At node j , the state at instant k is then given by:

$$x_j(k) = \sum_{i=1}^N \lambda_i^k(\mathbf{W}) \tilde{u}_{j,i} = \sum_{i=1}^N \lambda_i^k(\mathbf{W}) \beta_i u_{j,i},$$

$\lambda_i(\mathbf{W})$ being the i th eigenvalue of \mathbf{W} , $u_{j,i}$ the j th entry of the eigenvector $\mathbf{U}_{\cdot,i}$, and $\tilde{u}_{j,i} = u_{j,i}\beta_i$.

By stacking its N consecutive measurements, node j obtains:

$$\mathbf{x}_j = \mathbf{\Upsilon} \tilde{\mathbf{U}}_j^T. \quad (6)$$

Where:

$$\mathbf{x}_j = \begin{pmatrix} x_j(0) \\ x_j(1) \\ \vdots \\ x_j(N-1) \end{pmatrix}, \tilde{\mathbf{U}}_j^T = \begin{pmatrix} \tilde{u}_{j,1} \\ \tilde{u}_{j,2} \\ \vdots \\ \tilde{u}_{j,N} \end{pmatrix}, \quad (7)$$

$$\mathbf{\Upsilon} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \lambda_2(\mathbf{W}) & \cdots & \lambda_N(\mathbf{W}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_2^{N-1}(\mathbf{W}) & \cdots & \lambda_N^{N-1}(\mathbf{W}) \end{pmatrix}. \quad (8)$$

From Assumption 1, we can conclude that the Vandermonde matrix $\mathbf{\Upsilon}$ is full rank. Therefore node j can efficiently solve the above problem and get:

$$\tilde{\mathbf{U}}_j^T = \mathbf{\Upsilon}^{-1} \mathbf{x}_j. \quad (9)$$

In other words, from its own measurements and knowing the eigenvalues of \mathbf{W} , node j can efficiently compute the j th row of $\tilde{\mathbf{U}}$.

Collaboration with all the nodes is mandatory in order to reconstruct $\tilde{\mathbf{U}}$. For this purpose, let us consider a message passing scheme where the nodes send to their neighbors the rows they have. After a number of message exchanges, at least equal to the diameter of the graph, node j obtains all the rows of $\tilde{\mathbf{U}}$. However, since some of the nodes are assumed to be anonymous, they do not send the label of the row they have (or equivalently their label or ID). Two kinds of messages are transmitted through the network: the rows associated with non-anonymous nodes are correctly labeled while those associated with anonymous nodes are not labelled. As a consequence, node j only receives $\tilde{\mathbf{U}}$ up to rows permutation, i.e. it receives $\hat{\mathbf{U}} = \Pi \tilde{\mathbf{U}}$, Π being a permutation matrix.

3.2. Network topology reconstruction

Since $\tilde{\mathbf{U}}^T \tilde{\mathbf{U}} = \text{diag}(\mathbf{b}^2)$, we also have $\hat{\mathbf{U}}^T \hat{\mathbf{U}} = \text{diag}(\mathbf{b}^2)$. Therefore, from (5), we get:

$$\mathbf{W} = \Pi \hat{\mathbf{W}} \Pi^T, \quad \text{with} \quad \hat{\mathbf{W}} = \hat{\mathbf{U}} \mathbf{D} (\hat{\mathbf{U}}^T \hat{\mathbf{U}})^{-1} \hat{\mathbf{U}}^T. \quad (10)$$

We can conclude that from the estimated matrix $\hat{\mathbf{W}}$, we obtain a graph that is isomorphic to the original graph. In what follows we state two sufficient conditions to ensure correct

reconstruction of the graph. We first recall the following notions of observability [13]:

A network is said to be:

- node-observable from a given node if that node is able to reconstruct the entire network state from its own measurements. This issue has been studied for instance in [14]. It has been stated that a network with a state matrix having at least one non-simple eigenvalue is not node-observable.
- neighborhood observable from a given node if that node can reconstruct the entire network state from its own measurements and those of its neighbors.
- globally observable if it is neighborhood observable from any node.

Proposition 1 Assume that the graph is node-observable or neighborhood observable from node j . If all the entries of the initial condition are distinct then the network topology can exactly be reconstructed from node j .

Proof: From $\tilde{\mathbf{U}}$ and $\text{diag}(\mathbf{b}^2)$ the initial condition $\mathbf{x}(0)$ can be reconstructed as follows: $\mathbf{x}(0) = \tilde{\mathbf{U}} \text{diag}(\mathbf{b}^2) \beta$, with $\beta = (1/\beta_1^2 \ 1/\beta_2^2 \ \cdots \ 1/\beta_N^2)^T$. Since $\tilde{\mathbf{U}}$ is known only up to rows permutation through $\hat{\mathbf{U}}$, the reconstructed initial condition $\hat{\mathbf{x}}(0) = \hat{\mathbf{U}} \text{diag}(\mathbf{b}^2) \beta$ is related to the actual initial condition as follows: $\hat{\mathbf{x}}(0) = \Pi \mathbf{x}(0)$. The graph being node-observable or neighborhood-observable from node j , node j can reconstruct the initial condition from its consecutive measurements. It can therefore compare $\hat{\mathbf{x}}(0)$ with $\mathbf{x}(0)$. If all the entries of $\mathbf{x}(0)$ are distinct then the permutation matrix can be retrieved and the matrix \mathbf{W} exactly recovered from $\hat{\mathbf{W}}$. ■

Proposition 2 Assume that the network contains a set \mathcal{A} of anonymous nodes and the graph is node-observable or neighborhood observable from node j . If all the entries of the initial condition associated with the anonymous nodes are all distinct then the network topology can be exactly reconstructed from node j .

The proof of this proposition is similar to the previous one. Here, the permutation ambiguity is restricted to the entries associated with the anonymous nodes. We can also note that if the initial condition is driven from a continuous joint probability distribution, then the reconstruction of the network topology is almost sure.

The proposed network topology reconstruction algorithm is described as follows:

Algorithm 1 (Network topology Reconstruction)

1. Inputs:

- Laplacian spectrum $\text{sp}(\mathbf{L}) = \{\lambda_1(\mathbf{L}), \dots, \lambda_N(\mathbf{L})\}$, N being the number of nodes.
- The stepsize ϵ of the consensus protocol with matrix $\mathbf{W} = \mathbf{I}_N - \epsilon \mathbf{L}$.
- The state measurements $x_i(k)$, $k = 0, 1, \dots, N-1$, $i = 1, \dots, N$.

- The estimated initial condition $\hat{\mathbf{x}}(0)$.
2. Compute:
- Set of eigenvalues $\lambda_i(\mathbf{W}) = 1 - \epsilon \lambda_i(\mathbf{L})$ for $i = 1, \dots, N$.
 - Vandermonde matrix \mathbf{T} (8)
3. Each node j :
- (a) estimates the j -th row of $\tilde{\mathbf{U}}$ using (9);
 - (b) sends to its neighbors the rows it has. After a number of message exchanges, it obtains $\hat{\mathbf{U}}$;
 - (c) computes $\text{diag}(\mathbf{b}^2) = \hat{\mathbf{U}}^T \hat{\mathbf{U}}$;
 - (d) reconstructs the initial condition $\hat{\mathbf{x}}(0) = \hat{\mathbf{U}} \text{diag}(\mathbf{b}^2) \beta$ and then finds a permutation matrix Π so that $\mathbf{x}(0) = \Pi \hat{\mathbf{x}}(0)$;
 - (e) estimates the consensus matrix as

$$\mathbf{W} = \Pi \left(\hat{\mathbf{U}} \mathbf{D} \text{diag}(\hat{\mathbf{U}}^T \hat{\mathbf{U}})^{-1} \hat{\mathbf{U}}^T \right) \Pi^T,$$

\mathbf{D} being the diagonal matrix with eigenvalues $\lambda_i(\mathbf{W})$, $i = 1, \dots, N$, as entries.

- (f) deduces the Laplacian matrix $\hat{\mathbf{L}} = \frac{1}{\epsilon} (\mathbf{I}_N - \mathbf{W})$ by rounding the entries to integer values.

4. NUMERICAL RESULTS

To evaluate the performance of the proposed method, we randomly generate graphs with different number nodes. For each considered size of graph, we generate 100 graphs. A random symmetric matrix is first generated from a uniform distribution. The weights above 0.5 are set to 1 and the others to zero. The obtained matrix is considered as the adjacency matrix of the graph. We considered the resulting graph as a good candidate to be evaluated only if the Laplacian matrix has distinct eigenvalues. The results below are averaged values over the 100 random graphs. The performance is evaluated using the success rate of existing links (SREL) and the success rate of non-existing links (SRNL), [5]. SREL(SRNL) is defined as the ratio of the number of successfully predicted existent (non-existent) links to the total number of existent (nonexistent) links. We assess the quality of network reconstruction with respect to the quality of the estimation of the Laplacian eigenvalues.

Figure 1 depicts an example of topology reconstruction. We can note that only three links have not been reconstructed and no non-existent link has been predicted. Figures 2 and 3 describe the average SREL and SRNL for graphs of different sizes with respect to the norm of the estimation error of Laplacian eigenvalues.

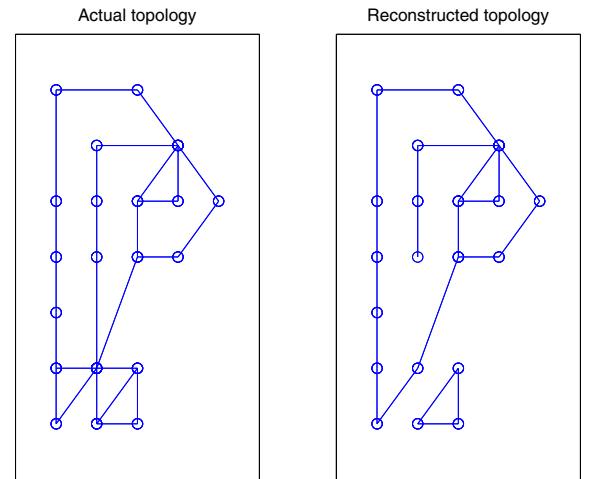


Fig. 1. Example of topology reconstruction with the proposed algorithm.

One can note that whatever the size of the graph, SREL and SRNL decrease with the quality of estimation of Laplacian eigenvalues. In addition, for a given estimation error, the performance also decreases when the size of the graph grows. This is due to the quality of estimation of the eigenvectors. Indeed, by increasing the size of the graph, the rows at the bottom of matrix \mathbf{T} becomes more and more small and then more prone to numerical errors. Therefore quality of estimation can be significantly reduced. It is then necessary to resort to more efficient methods to solve the eigenvectors estimation problem for large graphs.

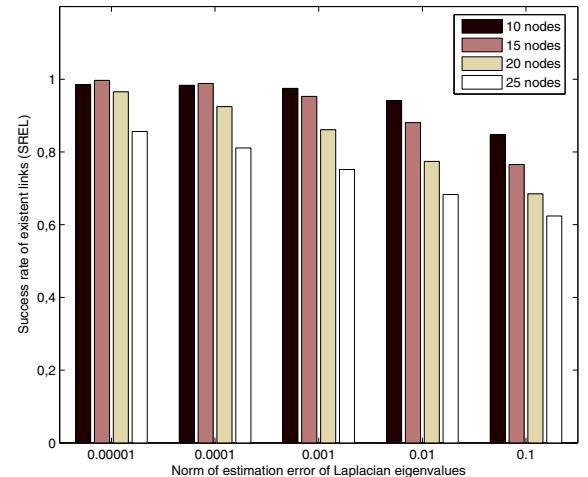


Fig. 2. SREL for graphs with different sizes according to the norm of estimation error of Laplacian eigenvalues.

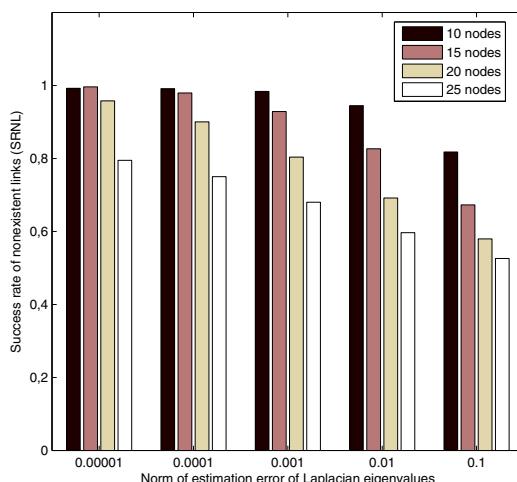


Fig. 3. SRNEL for graphs with different sizes according to the norm of estimation error of Laplacian eigenvalues.

5. CONCLUSION

In this paper, we have proposed a way for reconstructing the network topology in the presence of anonymous nodes from average consensus measurements. The consensus matrix is determined from its eigenstructure estimated in distributed way. Since the Laplacian spectrum can be obtained by means of methods proposed in the recent literature, each node can deduce a row-permuted matrix of eigenvectors because of anonymous nodes. The obtained graph is isomorphic to the original one. If the graph is node-observable or neighborhood observable from node j and if all the entries of $\mathbf{x}(0)$ are distinct then node j can exactly reconstruct the network topology. If the entries of the initial condition $\mathbf{x}(0)$ are independently generated from a continuous probability distribution, then node j can reconstruct the network topology almost surely. The main assumption in this paper is that all eigenvalues are distinct, that is the case of most random graphs. Future works encompass the design of the network reconstruction protocol that deals with spectrums in which the multiplicities of the eigenvalues can be higher than 1 and also directed graphs. In addition, numerical issues for large graphs are to be considered for making the proposed method scalable.

REFERENCES

- [1] H. B. Acharya and M. G. Gouda, "On the hardness of topology inference," in *In Proc. of 12th Int. Conf. on Distributed Computing and Networking (ICDCN 2011)*, Bangalore, India, January 2011, pp. 251–262.
- [2] M. Nabi-Abdolyousefi and M. Mesbahi, "Network identification via node knockout," *IEEE Trans. on Automatic Control*, vol. 57, no. 12, pp. 3214–3219, December 2012.
- [3] S. Shahrampour and V. M. Preciado, "Topology identification of directed dynamical networks via power spectral analysis," *IEEE Trans. on Automatic Control*, vol. 60, no. 8, August 2015.
- [4] D. Hayden, Ye Yuan, and J. Goncalves, "Network reconstruction from intrinsic noise: Minimum-phase systems," in *In proc. of American Control Conference (ACC)*, June 2014, pp. 4391–4396.
- [5] J. Ren, W.-X. Wang, B. Li, and Y.-C. Lai, "Noise bridges dynamical correlation and topology in coupled oscillator networks," *Physical Review Letters*, vol. 104, pp. 058701:1–058701:4, February 2010.
- [6] F. Morbidi and A. Y. Kibangou, "A distributed solution to the network reconstruction problem," *Systems & Control Letters*, vol. 70, pp. 85 – 91, 2014.
- [7] T.M.D. Tran and A.Y. Kibangou, "Distributed estimation of graph Laplacian eigenvalues by the alternating direction of multipliers method," in *Proc. 19th IFAC World Congress*, Cape Town, South Africa, 2014.
- [8] T.M.D. Tran and A.Y. Kibangou, "Distributed estimation of Laplacian eigenvalues via constrained consensus optimization problems," *Systems and Control Letters*, vol. 80, pp. 56–62, 2015.
- [9] A. Y. Kibangou and C. Commault, "Decentralized Laplacian eigenvalues estimation and collaborative network topology identification," in *3rd IFAC Workshop on Distributed Estimation and Control in Networked Systems (NecSys'12)*, Santa Barbara, USA, 2012, pp. 7–12.
- [10] D. Kempe and F. McSherry, "A decentralized algorithm for spectral analysis," *Journal of Computer and System Sciences*, vol. 74, pp. 70–83, 2008.
- [11] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems Control Letters*, vol. 53, pp. 65–78, 2004.
- [12] A.Y. Kibangou, "Graph laplacian based matrix design for finite-time distributed average consensus," in *Proc. of American Control Conference (ACC)*, 2012, June 2012, pp. 1901–1906.
- [13] A.Y. Kibangou and C. Commault, "Observability in connected strongly regular graphs and distance regular graphs," *IEEE Trans. on Control of Network Systems*, vol. 1, no. 4, pp. 360–369, December 2014.
- [14] N. O'Clery, Y. Yuan, G.B. Stan, and M. Barahona, "Observability and coarse graining of consensus dynamics through the external equitable partition," *Physical Review E*, vol. 88, pp. 042805:1–042805:13, October 2013.