Systems & Control Letters 70 (2014) 85–91

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

A distributed solution to the network reconstruction problem

Fabio Morbidi [a,*], Alain Y. Kibangou [b]

[a] Inria, NeCS team, 655 Avenue de l'Europe, 38334 Montbonnot Saint Martin, France

[b] Univ. Grenoble Alpes, CNRS, Inria, Gipsa-Lab, F-38000 Grenoble, France

ARTICLE INFO

Article history:
Received 31 January 2014
Accepted 9 May 2014

Keywords:
Network reconstruction
Consensus algorithms
Distributed estimation
Networked systems

ABSTRACT

It has been recently shown in Ren et al. (2010) that by collecting noise-contaminated time series generated by a coupled-oscillator system at each node of a network, it is possible to robustly reconstruct its topology, i.e. determine the graph Laplacian. Restricting ourselves to linear consensus dynamics over undirected communication networks, in this paper we introduce a new dynamic average consensus least-squares algorithm to *locally* estimate these time series at each node, thus making the reconstruction process *fully distributed* and more easily applicable in the real world. We also propose a novel efficient method for separating the off-diagonal entries of the reconstructed Laplacian, and examine several concepts related to the trace of the dynamic correlation matrix of the coupled single integrators, which is a distinctive element of our network reconstruction method. The theory is illustrated with examples from computer, power and transportation systems.

1. Introduction

In graph theory the "basic inverse problem" consists in determining the topology of a graph from its adjacency spectrum [1]. It is well known that: (i) graphs with few distinct eigenvalues tend to have some kind of regularity; (ii) cospectral graphs have the same number of closed walks of a given length; (iii) there exist cospectral graphs which are *not* isomorphic. Therefore, the adjacency (and analogously the Laplacian) spectrum of a graph is not sufficient, in general, to unambiguously identify the graph topology.

Recently, the interest in the basic inverse problem has been revitalized by several works in the control and mathematical physics literatures, dealing with the reconstruction of the topology of a network of *dynamical systems* (this is also sometimes referred to as network "identification" or "exploration" problem). Uncovering the relationship between dynamics and network structure has indeed relevant applications in biology (biochemical, neural and ecological networks), finance, computer science (Internet and World Wide Web), transportation (delivery and distribution networks), and electrical engineering (power grids).

In [2], the interaction geometry among a known number of agents adopting (weighted) consensus-type algorithms for their coordination, is reconstructed using a grounding procedure inspired by experimental biology, called "node knockout" (see also [3]), while in [4] the consensus matrix is determined from its eigenstructure estimated in a distributed fashion.

Stochastic methods have lately emerged as powerful alternatives to the deterministic ones in [2–4]: the main idea here is to reconstruct the network topology from noise-contaminated time series collected at each node of the graph. Noise is ubiquitous, e.g., in biological networks and relying on such a natural variability as a non-invasive network-identification tool appears extremely promising [5]. Several stochastic reconstruction methods have been proposed in the recent literature. In [6], the topology of a directed weighted network of LTI systems is estimated via power spectral analysis, while in [7] the authors formulate the problem of network reconstruction as a compressing sensing problem. Other approaches have exploited the sparsity of the network and used the Bayesian information criterion to measure the graph structure from stationary time series [8] or optimization-based methods [9]: however, although the sparsity assumption is well justified in some applications (e.g. in biological networks), it may lead to poor results in other cases as shown in [10]. More related to the standard system-identification literature is the work in [5], where the authors formulated the network reconstruction problem as a variant of the spectral factorization problem, and [11] where the classical "direct method" of closed-loop identification is utilized. However, in these works unknown noise sources are applied only to the states that are measured, which is an unrealistic assumption in many applications. Finally, in [12], an interesting connection between dynamic correlation and topology in

* Corresponding author. Tel.: +33 (0)4 76 61 53 22.
E-mail addresses: fabio.morbidi@inria.fr (F. Morbidi), alain.kibangou@ujf-grenoble.fr (A.Y. Kibangou).

http://dx.doi.org/10.1016/j.sysconle.2014.05.008
0167-6911/

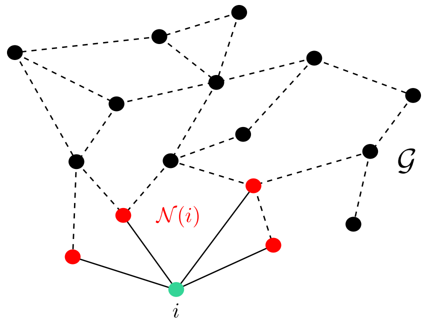

Fig. 1. *Distributed network reconstruction*: node i must reconstruct the topology of the network $\mathcal{G}$ by only exchanging information with its set of neighbors $\mathcal{N}(i)$ (red nodes). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

noisy coupled-oscillator networks (where the noise magnitude is known) has been unveiled. The reconstruction algorithm originally proposed in [12] has been recently improved in [13] by defining a different thresholding mechanism based on the largest eigenvalue of the Laplacian matrix estimated in a distributed fashion. However, all the aforementioned stochastic approaches are *centralized* or *semi-centralized* (as [13]), since they require the knowledge of time series at each node of the network, and thus impractical for real-world applications.

This paper builds upon [12] and extends it in several new directions. In particular, the original contributions of this work are threefold:

- The approach in [12] is made *fully distributed* by leveraging a new dynamic average consensus least-squares algorithm for the *local* estimation of the noisy time series at each node of the graph. In this way, each node is able to infer the topology of the overall undirected network (cf. Fig. 1): to the best of our knowledge, together with [4], this is the only completely-decentralized reconstruction algorithm available in the literature for networks of dynamical systems. An original stability analysis of the proposed dynamic average consensus least-squares estimator is also performed.
- A new robust and effective mechanism for separating the -1 and 0 off-diagonal entries of the reconstructed graph Laplacian is proposed. The method is based on *κ-means clustering* [14] and it is simpler than the thresholding criteria utilized in [12,13].
- Interesting connections are shown between the trace of the dynamic correlation matrix of the coupled single-integrator nodes (which plays a crucial role in our reconstruction mechanism), the *total effective resistance* of the network [15], and the *H_2 norm of noisy reduced consensus dynamics* [16,17].

Note that the distributed network-reconstruction algorithm proposed in this paper may represent a valid alternative (at least at small scales) to "web crawlers" for the World Wide Web and to "traceroute sampling" for the Internet, where there does not exist a vantage point with complete information about the overall structure of the system. Moreover, it may be useful for probing the structure of dynamically-changing networks, e.g. road networks, where links can appear/disappear over time because of accidents or works on the carriageway.

The rest of this paper is organized as follows. Section 2 presents some background material. The main theoretical results of the work are introduced in Sections 3 and 4, and the theory is illustrated with numerical simulations on realistic networks in Section 5. Finally, Section 6 summarizes the main contributions of the paper and outlines some promising future research directions.

2. Preliminaries

In this section we recall some notions of algebraic graph theory and robust control, and introduce the notation.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph (or network), where $\mathcal{V} = \{1, \ldots, n\}$ is the set of nodes and $\mathcal{E}$ the set of links. $\mathcal{N}(i)$ will indicate the set of nodes adjacent to node i in the graph $\mathcal{G}$. All graphs in this paper are finite, connected, with no self-loops and multiple links.

The adjacency matrix $\mathbf{A} = [a_{ij}]$ of graph $\mathcal{G}$ is an $n \times n$ matrix defined as

$$a_{ij} = \begin{cases} 1 & \text{if } \{i, j\} \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

The Laplacian matrix $\mathbf{L} = [\ell_{ij}]$ of graph $\mathcal{G}$ is an $n \times n$ symmetric positive semidefinite matrix defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$ where $\mathbf{D} = \text{diag}(\mathbf{A}\mathbb{1})$ is the degree matrix and $\mathbb{1}$ is a column vector of n ones. From this definition, it follows that:

$$\ell_{ij} = \begin{cases} \sum_{j=1}^{n} a_{ij} & \text{if } i = j, \\ -1 & \text{if } \{i, j\} \in \mathcal{E}, \quad i, j \in \{1, \ldots, n\}, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathbf{L}^{\dagger}$ be the (Moore–Penrose) pseudoinverse of the Laplacian matrix, $\mathbf{J} = \frac{1}{n}\mathbb{1}\mathbb{1}^T$ the $n \times n$ (rank one) averaging matrix, and $\mathbf{I}_n$ the $n \times n$ identity matrix. Then, we have that

$$\mathbf{L}^{\dagger}\,\mathbf{L} = \mathbf{I}_n - \mathbf{J}, \tag{1}$$

which is the projection matrix onto the image of $\mathbf{L}$ [15]. From (1), it can be verified that:

$$\mathbf{L}^{\dagger} = (\mathbf{L} + \mathbf{J})^{-1} - \mathbf{J}, \tag{2}$$

and that

$$\text{trace}(\mathbf{L}^{\dagger}) = \sum_{i=2}^{n} \frac{1}{\lambda_i(\mathbf{L})}, \tag{3}$$

where $0 = \lambda_1(\mathbf{L}) < \lambda_2(\mathbf{L}) \le \cdots \le \lambda_n(\mathbf{L})$ are the ordered eigenvalues of $\mathbf{L}$ [15]. Note that $\mathbf{L}^{\dagger}$ inherits from $\mathbf{L}$ the property of being symmetric and positive semidefinite. Moreover, $\mathbf{L}^{\dagger}$ and $\mathbf{L}$ share the same null space.

The H_2 norm of a general LTI system $\dot{\mathbf{x}} = \mathbf{A}\,\mathbf{x} + \mathbf{B}\,\mathbf{u}$, $\mathbf{y} = \mathbf{E}\,\mathbf{x}$, with $\mathbf{A}$ Hurwitz, is given by $(\text{trace}(\mathbf{E}\mathbf{X}\mathbf{E}^T))^{1/2}$ where the positive semidefinite matrix $\mathbf{X}$ (the controllability Gramian) solves the algebraic Lyapunov equation $\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0}$. In addition, if the pair $(\mathbf{A}, \mathbf{B})$ is controllable, $\mathbf{X}$ is positive definite.

Notation: $|\mathscr{S}|$ will denote the cardinality of the set $\mathscr{S}$, $\emptyset$ the empty set, $\text{Re}(z)$ and $|z|$ the real part and modulus of the complex number z, respectively, $\otimes$ the Kronecker product, $\text{blkdiag}(\cdot)$ a block-diagonal matrix, $\mathbb{E}[\,\cdot\,]$ the expectation operator, $\|\cdot\|_2$ the Euclidean norm of a vector and $\|\cdot\|_{\mathcal{L}_2}$ the total energy or $\mathcal{L}_2$ norm of a vector-valued signal.

3. Problem formulation

For the reader's convenience we briefly review here the main results in [12], which form the basis for our subsequent developments. We will start with a general formulation dealing with directed graphs, and then we will specialize our results to *undirected networks* with coupled single-integrator node dynamics.

Consider a directed network of n nonidentical coupled oscillators where $\mathbf{x}_i \in \mathbb{R}^m$, $i \in \{1, \ldots, n\}$, denotes the state variable of the ith oscillator and $\mathbf{x}_i(0)$ the initial state. In the presence of noise, the dynamics of the whole oscillator system can be expressed as

$$\dot{\mathbf{x}}_i = \mathbf{F}_i(\mathbf{x}_i) - \gamma \sum_{j=1}^{n} \ell_{ij}\,\mathbf{H}(\mathbf{x}_j) + \boldsymbol{\eta}_i, \tag{4}$$

where $\mathbf{F}_i : \mathbb{R}^m \to \mathbb{R}^m$ is the intrinsic dynamics of the ith oscillator, $\mathbf{H} : \mathbb{R}^m \to \mathbb{R}^m$ is the coupling function of the oscillators, $\gamma > 0$ is the coupling strength, and $\boldsymbol{\eta}_i \in \mathbb{R}^m$ is the zero-mean