Physica A 497 (2018) 118–125

Contents lists available at ScienceDirect

Physica A

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

Reconstruction of networks from one-step data by matching positions

Jianshe Wu *, Ni Dang, Yang Jiao

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Joint International Research Laboratory of Intelligent Perception and Computation, Xidian University, Xi'an, Shaanxi Province, 710071, China

HIGHLIGHTS

- A method for reconstructing the topology of networks with only one step data.
- This algorithm utilizes the natural sparsity and degree distribution of real world networks.
- The problem of matching position is defined and a simple algorithm is developed.
- The matching position can be used in network reconstruction and other problems.
- The proposed method is verified in networks with ten thousands of vertices.

ARTICLE INFO

Article history:
Received 2 March 2017
Received in revised form 2 July 2017
Available online 11 January 2018

Keywords:
Network reconstruction
Topology of network
Time series data
Reconstruction algorithm
Matching positions

ABSTRACT

It is a challenge in estimating the topology of a network from short time series data. In this paper, matching positions is developed to reconstruct the topology of a network from only one-step data. We consider a general network model of coupled agents, in which the phase transformation of each node is determined by its neighbors. From the phase transformation information from one step to the next, the connections of the tail vertices are reconstructed firstly by the matching positions. Removing the already reconstructed vertices, and repeatedly reconstructing the connections of tail vertices, the topology of the entire network is reconstructed. For sparse scale-free networks with more than ten thousands nodes, we almost obtain the actual topology using only the one-step data in simulations.

1. Introduction

Usually, the phenomena (e.g., phase information) of coupled agents occurred in a network are observed, but the topology of the network is unknown. Estimating the topology of the network from those phenomena becomes a key in a multidisciplinary research field [1–3]. Network reconstruction is a reverse engineering problem and a number of methods have been proposed to address this problem [4–9]. Since the network function and dynamics are various [10–12], the data measured for networks reconstructions are various, e.g., response dynamics [10], avalanche dynamics [11], random phase-resetting [12], knock-out data [13], noise [9], etc. In network reconstruction, a collection of unprecedented amount data measured in series time points, named as time series data is needed, especially for large scale networks. However, it is often

* Corresponding author.
E-mail address: jshwu@mail.xidian.edu.cn (J. Wu).

https://doi.org/10.1016/j.physa.2018.01.011
0378-4371/

that only limited observable data are available. In Ref. [14], the difficulty of reconstruction using short time series is tackled, and it is still a challenge.

When the time series is short, a convenient method is using the natural sparsity of networks [7,15,16]. Natural sparsity of networks indicates that the number of connections of a network is much less than the number of all possible connections in real world. Natural sparsity makes it possible to approximately reconstruct the topology with a short time series data, much less than the network size. But these methods are impossible to reconstruct with only one-step data. New method is required.

In this paper, the matching positions problems (MPPs) are defined firstly, and then the matching positions (MP) algorithm is developed to reconstruct the topology of a network with only one-step data. Besides the natural sparsity, the degree distribution of real world networks is approximately power-law [17,18], that is called scale-free. This indicates most of the vertices have few, even one or two edges, but small number of vertices has large amount of connections. The MP algorithm reconstructs the connections of the vertices with the smallest degree firstly. Removing the reconstructed vertices, the remained network is still scale-free. By repeatedly reconstructing the connections of the vertices with the smallest degree, the topology of entire network is reconstructed.

2. Problem formulation

We consider a general un-weighted network model of coupled agents, in which the phase of vertex i at time step k is determined by [19,20]:

$$\theta_i(k+1)=\sum_{j=1}^{N}\frac{a_{ij}}{k_m}\theta_j(k), \quad i=1,\ldots,N, \tag{1}$$

where N is the number of vertices and a_{ij} is the elements of the network adjacency matrix $A\in R^{N\times N}$. If there is an edge between vertices (agents) i and j, $a_{ij}=a_{ji}=1$; otherwise $a_{ij}=0$. k_m is normalization parameter, which may has different values in different applications. In the communication network model [15], k_m equals the number of neighbors of vertex i. In the original Vicsek model [19], k_m equals the number of neighbors of vertex i plus one (including i itself). In this paper, without loss generalization, we assume $k_m=1$ for clarity:

$$\theta_i(k+1)=\sum_{j=1}^{N}a_{ij}\theta_j(k), \quad i=1,\ldots,N. \tag{2}$$

In matrix form, Eq. (2) is written as Eq. (3) if the length of time series data is K.

$$\theta(k+1)=A\theta(k), \quad k=1,\ldots,K, \tag{3}$$

where $\theta(k)=(\theta_1(k),\ldots,\theta_N(k))^T$. Network reconstruction means estimating A from time series data $\theta(k)$, $k=1,\ldots,K$. If $K=1$, only $\theta(1)$ and $\theta(2)$ are available, that is the one-step data.

In mathematics, when $K\geq N$, we may obtain A by direct solving matrix equation (3). When $K<N$, the equation is underdetermined, and there are a set of possible solutions for A. Since the natural sparsity, we can choose the sparsest one from those possible solutions for the adjacency of the network [7,15,16]. These methods perform well when K is not too small and are impossible when $K\to 1$. In this paper, we address the challenge problem of network reconstruction when $K=1$.

Real world networks are formulated by two ingredients: growth and preferential attachment [17,18]. In the BA scale-free network model, from an initial network of m_0 vertices, a new vertex with m edges is added to the network at each time step [18]. The new vertex links to m vertices with probability $\prod(k_i)=k_i/\sum k_i, i=1,\ldots,N$, where k_i is the degree of vertex i.

Inspired by the natural networks formulation, it is reasonable to reconstruct the topology of a network by sequentially reconstructing the connections of the vertices. As a reverse engineering problem [1,4], we treat the network reconstruction as a reverse process of the network formulation. The edges of the latest added vertex which has the smallest degree, are reconstructed firstly. Reconstructing the connections of a vertex from one-step data is one of the MPPs that will be defined and analyzed in the next section. Removing the constructed vertex and its edges, the remained unconstructed network is still scale-free. Repeat the above process until the edges of all vertices are obtained (see Fig. 1).

3. Matching positions with single vertex

For clarity in representation, suppose vertex i has the smallest degree k_s in an N-vertex network. The k_s edges may have $l=C_N^{k_s}=\frac{N!}{k_s!(N-k_s)!}$ possible combinations; each of them is denoted as a 0–1 row vector $d_j\in R^{1\times N}$. The dictionary matrix $D=(d_1,\ldots,d_l)^T\in R^{l\times N}$ is formulated by all those possible row vectors. An example dictionary for a 4-vertex network with $k_s=2$ is