Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

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

Estimating multilateral trade behaviors on the world trade web with limited information

Cangfeng Ding [a,b], Kan Li [a,*]

[a] School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China
[b] School of Mathematics and Computer Science, Yanan University, Yanan, 716000, China

ARTICLE INFO

Article history:
Received 4 May 2015
Received in revised form
12 November 2015
Accepted 12 November 2015

Keywords:
Network reconstruction
Configuration model
Link prediction
Limited information
Socioeconomic networks

ABSTRACT

The privacy and the confidentiality limit the amount of the accessible information in investigating socioeconomic systems. Reconstructing network with limited information is of great application importance. Most existing network reconstruction methods concentrate on completely connected networks, undirected or unweighed networks. In the paper, we present a novel approach to reconstruct the topological structure of directed weighted network (DWN) for estimating multilateral trade behaviors on the World Trade Web using limited information: the values of out- and in-strengths of nodes, and the total number of links and nodes. Our approach uses first out- and in-strength fitnesses to estimate the unknown out- and in-degrees. Then we propose a directed enhanced configuration model (DECM) and reconstruct the topological structure of real DWN. This approach is experimentally verified on real-world networks, using the significantly topological properties: assortativity and clustering. The results show a good agreement between these quantities calculated on real DWN and its DECM-induced ensemble averages. Furthermore, we compare DECM algorithm with weighted node similarity algorithms for link prediction, demonstrating that DECM outperforms similarity-based link prediction algorithms. Finally, the comparison between DECM and directed weighted configuration model (DWCM) using Information-theoretic criteria rigorously confirms that DECM evidently outperforms DWCM for reconstructing network topological structure. Accordingly, this approach can be employed as an important tool for providing deeper insights into the privacy protection of socioeconomic networks.

1. Introduction

Over the last few years a large number of studies have largely contributed to the analysis and understanding of topological structure of socioeconomic systems in terms of networks [1,2]. The socioeconomic systems can typically be represented by the directed weighted networks (DWNs) whose nodes are economic institutions/actors and links are a variety of economic behaviors/relationships such as the trade (import and export) behaviors between world countries on the World Trade Web (WTW) and follower/followee behaviors/relationships on social network. Network reconstruction approaches are to predict/reproduce topological structure of network. Due to privacy and confidentiality issues, the complete information of network among economic institutions/actors is unobservable, making it incredibly difficult to exactly reconstruct the topological structure of network as well as to effectively carry out the analysis of network. Various concrete applications originate from the solution to the issue. For example, Reconstructing the topological structure of WTW can be used to estimate the trade volume and trade partners of a country. Current studies on the network reconstruction with limited information include dense reconstruction methods and sparse reconstruction methods. The majority of dense reconstruction methods, such as Maximum Entropy, hypothesize the network is completely connected. In fact, real-world networks typically exhibit a heterogeneous distribution of connectedness, and such approaches easily lead to underestimate topological properties of networks. While spare reconstruction methods are highly computational and incapable to set an appropriate value for such heterogeneity of connectivity, such as message-passage approach; or are only to reconstruct such networks that are undirected or unweighted. In addition, link prediction methods can predict existent yet unknown links or spurious links or future links related to the network evolution, but they can rarely predict completely topological structure of networks, or rarely take the weights of undirected/directed links into account [3–6]. For the WTW, we need predict all the link interactions and import/export trade volumes among

* Corresponding author.
E-mail addresses: dfeng808@sina.com (C. Ding), likan@bit.edu.cn (K. Li).

http://dx.doi.org/10.1016/j.neucom.2015.11.127
0925-2312/

countries in order to get multilateral trade behaviors/relationships, but current link prediction methods are difficult to estimate its topological structure.

To resolve the mentioned issues, we propose a novel approach that is appropriate for reconstructing the directed weighted networks using limited information: the total number of nodes and links as well as out- and in-strength values of all nodes. Our method mainly involves in proposing a novel directed enhanced configuration model (DECM) for DWN reconstruction, taking the intrinsic node properties as fitness to estimate out- and in-degrees of all nodes, and using these ingredients to generate the ensembles of DWN through DECM. In a nutshell, our approach will be fallen into two specific procedures. Firstly, we utilize the fitness to allocate the probabilities of directed links, and then use the computed probabilities to generate an ensemble of the directed binary network with specified total link number. Accordingly, we are capable to estimate the out-degree and in-degree of each node. Secondly, we employ the estimated values of node degrees as well as the real values of out- and in-strengths of all nodes to generate an ensemble of DWN through DECM, and estimate the important topological properties. To verify the effectiveness of our approach, we make use of real-world networks as the empirical cases of study. In those examples, we have known the entire information about networks in advance, so we compare the values of properties between the real-world DWNs and their DECM-reconstructed counterparts to validate the results of reconstruction with the limited information. In our approach, experiments show a good agreement between the values of topological properties of the real-world DWNs and their counterparts of DECM ensemble, revealing the validity of our method in reconstructing the topological structure of DWNs by means of limited information. Finally, we use the information-theoretic criteria [7] to compare the DECM with the directed weighted configuration model(DWCM) [8,9], rigorously demonstrating that the joint specification of the out-strengths,in-strengths,out-degrees and in-degrees cannot be reduced to that of the out- and in-strengths separately.

The remainder of this paper is organized as follows. In Section 2, we introduce the related work. In Section 3, we utilize fitnesses to estimate the unknown degree of each node, propose the directed enhanced configuration model for DWN reconstruction, and then utilize these ingredients to reconstruct the DWN. Next, we experimentally verify the performance of our approach in Section 4. In Section 5, we compare DECM with DWCM using information-theoretic criteria. Finally we sum up this work and discuss future work in Section 6.

2. Related work

A vast variety of contributions have focused primarily on the investigations of topological structure of socioeconomic networks. One of the most important contributions to socioeconomic networks is that of reconstructing network topology structure with the limited available information. The network reconstruction is to generate a randomized ensemble of a given network with the specified constraints so as to detect patterns of complicated interaction arising from many systems in a variety of disciplines [10]. The reconstruction of topological structure of network with the limited available information still has various types of the widely unexplored and fully unsolved problems in the field of complex networks [11]. A large number of concrete applications originate from the solution to this issue.

The current studies on the reconstruction of networks with limited information can be roughly divided into two categories: dense reconstruction [12,13] and sparse reconstruction [11,14,15] approaches. The majority of dense reconstruction methods employ Maximum Entropy [16] to reconstruct the unknown links of the network, estimating link weights through a principle of maximum homogeneity. But, these methods hypothesize the network is completely connected. In fact, the real-world networks typically exhibit a heterogeneous distribution of connectedness, and such approaches easily lead to underestimate the topological properties of networks. The main approaches of spare reconstruction are message-passage algorithm [11] and bootstrapping algorithm [14,15]. The message-passage approach explores the space of possible network structures and obtains an arbitrarily heterogeneous network. Nevertheless, this method is highly computational and still incapable to set an appropriate value for such heterogeneity of connectivity. By contrast, the bootstrapping approach makes use of the Exponential Random Graph Model (ERGM) [17] where a fitness model [18] is used to interpret the Lagrange multipliers defining in configuration model, to generate a randomized ensemble of a given network and address the issue of heterogeneity. However, it is only to reconstruct such networks that are undirected or unweighed. In the paper, we propose DECM to topological structure of directed weighted networks.

3. Methodology

In what follows, we take the values of out- and in-strengths as the fitnesses to allocate the probabilities of directed links, estimate the unknown out- and in-degree of each node, and then reconstruct the topological structure of DWN through the proposed DECM.

3.1. Preliminaries

Given a DWN G_0 with N nodes, its topological structure is completely specified by the asymmetric, weighted and adjacent matrix W whose entries are $\{w_{ij}\}_{i,j=1}^{N}$, where w_{ij} represents the weight of a directed link from node i to node j. The value of w_{ij} is non-zeros when there exists a directed link between node i and node j; otherwise the value is 0. The value of out-strength and in-strength of node i are $s_i^{out} = \sum_{j\neq i} w_{ij}$ and $s_i^{in} = \sum_{j\neq i} w_{ji}$, respectively. In the corresponding directed binary representation of the network G_0, denoted as G_1, we can obtain the asymmetric, binary and adjacent matrix A whose entry a_{ij} equals $\varphi(w_{ij})$, where $\varphi(x)$ represents the step function whose value equals 1 if $x > 0$ and 0 otherwise. Consequently, the value of out-degree and in-degree of node i are $k_i^{out} = \sum_{j\neq i} a_{ij}$ and $k_i^{in} = \sum_{j\neq i} a_{ji}$ respectively. In the following description, we will always use the network representation, except for the special instruction.

3.2. The estimation of node degree

Given a real-world network G_0, we can generate a randomized ensemble of its corresponding directed binary network G_1, Ω_{DCM}, by the directed configuration model (DCM) (or ERGM) [17,19]. In the ensemble Ω_{DCM}, the ensemble average of the node out-degrees and in-degrees, denoted as $\{\langle k_i^{out}\rangle_{\Omega_{DCM}}\}$ and $\{\langle k_i^{in}\rangle_{\Omega_{DCM}}\}$, are restricted to the real values of out-degrees and in-degrees, denoted as $\{k_i^{out}\}$ and $\{k_i^{in}\}$, respectively, and the rest of the topological structure is maximally random. The DCM carries out a simply effective prescription that the distribution of probability for ensemble Ω_{DCM} is defined in the light of a set of Lagrange multipliers $\{x_i, y_i\}$. The ensemble probability p_{ij} of a directed link from node i to node j in DCM is defined as

$$p_{ij} = \frac{x_i y_j}{1 + x_i y_j}, \tag{1}$$