Physica A 514 (2019) 620–630

Contents lists available at ScienceDirect

Physica A

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

An universal algorithm for source location in complex networks

Wang Hongjue

School of Systems Science, Beijing Normal University, Beijing, 100875, PR China

HIGHLIGHTS

- We proposed a universal source localization method for different propagation dynamics.
- The proposed method can accurately locate source without knowing what kind of propagation dynamics.
- We study the influence of parameters of propagation dynamics and properties of networks on locating accuracy.
- The proposed method has low complexity.
- We proposed a method for inferring initial time of source without knowing what kind of propagation dynamics and any parameters.

ARTICLE INFO

Article history:
Received 3 February 2018
Received in revised form 10 August 2018
Available online xxxx

Keywords:
Propagation dynamics
Source
Complex networks
Spearman's correlation
Initial time

ABSTRACT

We study locating propagation source in complex networks. We proposed an universal algorithm, which can accurately locate source of different propagation dynamics by using sparse observations. Without knowing the propagation dynamics and any parameters, we calculate Spearman centrality based on the character that positive correlation between inform time of nodes and geodesic distance between nodes and source. The algorithm have high location accuracy with low time complexity and can accurately infer initial time of source. All simulations on both model and real-world networks proved the feasibility and validity of this algorithm.

1. Introduction

In our real world, many complex systems can be described as complex networks [1–6]. In recent years, scientists have made a great progress in network science [7–9]. In this paper, we are interested in locating propagation source based on sparse observations in complex networks. In our modern society, propagation dynamics are ubiquitous [10–12]. For instance, infectious diseases [13] and rumor propagating in social networks [14], computer virus [15] and malware spreading on the Internet [16,17], cascading effects diffusing in power grids [16]. Locating propagation source problem has a wide range of applications [18]. In epidemiology, locating the source of epidemics can provide important information about the disease. In online social networks, we can find out who start a rumor by identifying the rumor source [19], and prevent many baleful effects by locating the sources of online computer virus.

So far, there have been a large number of research on locating propagation source in complex networks [19–22]. Some algorithms based on node centrality [19,20]. Node centrality is a very important property for analyzing networks and has a wide range of applications [23–26]. Jordan centrality is defined as the maximum geodesic distance from a node to any other infected node in the network [19], the Rumor centrality of a node is the number of permitted propagating permutations rooted at the node on a given informed graph [19], the unbiased betweenness was defined by dividing betweenness by

E-mail address: 448721577@qq.com.

https://doi.org/10.1016/j.physa.2018.09.114
0378-4371/

the degree of the node [27]. Some of these methods have low time complexity, but they need more observations, that is they need a snapshot of all nodes at one point, and have low location accuracy in most cases, they also be confined to only one propagation dynamics, such as SIR. For diffusion dynamics, a maximum probability method was proposed [22] which based on the idea that propagation delays obey Gaussian distribution, this method have higher location accuracy on the tree structure networks but have unsatisfactory performance on general networks. Shen proposed a method by time-reversal backward spreading which have high location accuracy and low time complexity, while this method is sensitive to noisy propagation of diffusion dynamics [21]. Furthermore, the latter two methods only applicable to diffusion dynamics and need to know all parameters of distribution of propagation delays. Based on controllability and observability theory, Hu presented an algorithm to locate source in diffusion dynamics by using compressive sensing [28]. The method have high location accuracy high robustness but is limited to diffusion dynamics.

On the problem of most source location methods are limited to only one propagation dynamics, we proposed a universal algorithm, which can accurately locate source in different propagation dynamics include Diffusion [21], SI [29], IC [27], etc. In reference [30], for a global perspective of demographics and mobility between different communities, Dirk Brockmann and Dirk Helbing replaced conventional geographic distance between communities by effective distance, and they proposed a source location algorithm based on the discovered character that the correlation of epidemic arrival time and effective distances is linear. Inspired by reference [30], in community [31], because the time of nodes receiving message from source are proportional to the geodesic distance between nodes and source, we define Spearman centrality by calculating Spearman's correlation coefficient between vector of geodesic distance between nodes and source and vector of nodes informed time. We use area under the receiver operating characteristic curve (AUC) to quantify the accuracy of our algorithm. For different propagation models, we can obtain high location accuracy with low time complexity on different model and real static networks. In addition, we also study the effect of propagation dynamics parameters, observer placement strategies and network structures which include average degree, network size and degree distribution on the accuracy of our algorithm. At last, we proposed a novel method for inferring initial time of source without knowing any parameters of dynamics.

The rest of this paper is arranged as follows. We first introduces the propagation model including diffusion and SI dynamics. Then we define Spearman centrality and discusses the proposed locating algorithm, including numerical simulation results and relevant analysis. Finally, we conclude this paper.

2. Propagation model

Our goal is to locate the source of propagating message which taking place on an undirected complex network using only limited observed knowledge. The topological structure of the complex network $G = (V, E)$,with $n = |V|$ nodes and $m = |E|$ edges is assumed to be known, where V is the set of nodes and E is the set of edges. The connection of the network G can be represented as an adjacency matrix A,and its element $a_{i,j}$ is 1 when a link between nodes v_i and v_j exists and 0 otherwise. If $a_{i,j} = 1$,nodes v_i and v_j are neighbors to each other. The degree k_i is the number of neighbors of node v_i .

2.1. Diffusion dynamics

Epidemic spreading, rumor propagation and financial crises cascading can be modeled as diffusion-like dynamics [21]. During diffusion dynamics, there are two possible states for any node: informed (the node has received the message) or uninformed (the node has not yet received the message). We suppose message propagating through the shortest paths in complex network. At first, source s begin to spread the message at initial time t_s,which also called its informed time. After an arbitrary node v_i received message for the first time at t_{v_i}, it will transfer the message to all its neighbors. And the uninformed neighbor v_j of node v_i receive the message at time $t_{v_i} + \theta$, where θ is the propagation time delay of the edge between v_i and v_j. The details of diffusion dynamics are shown in Fig. 1a.

2.2. SI dynamics

This dynamics is a type of epidemic model, which is employed to describe the infection processes of nodes in networks [29, 32]. In SI dynamics, nodes are initially uninformed and can receive message from neighbors along with the propagation of message. Once a node receive message, it remains informed forever. The state of an arbitrary node v_i is denoted as C_i, where

$$C_i = \begin{cases} 0, & uninformed \\ 1, & informed \end{cases}$$

the probability of an arbitrary node v_i being informed by its neighbors at time t is

$$p_i^{0\to 1}(t) = 1 - (1 - \lambda_i)^{\sum_{j=1,j\neq i}^{N} a_{ij} C_j(t)}$$

where λ_i is the informed rate of v_i, $C_i(t)$ stands for the state of node v_i at time t, and the superscript $0 \to 1$ denotes the change from uninformed state (0) to informed state (1). The details of SI dynamics are shown in Fig. 1b. During propagation dynamics, we monitor observer nodes $O = \{o_1, o_2, \ldots, o_r\}$ and record their informed time $T^O = \left[t_{o_1}, t_{o_2}, \ldots, t_{o_r}\right]$, where r is the number of observer nodes. After all nodes of V have been informed, the observed knowledge (O, T^O) is used to infer source.