



Identification of influential spreaders based on classified neighbors in real-world complex networks



Chao Li^{a,b}, Li Wang^{a,b,*}, Shiwen Sun^{a,b}, Chengyi Xia^{a,b,*}

^a Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, PR China

^b Key Laboratory of Computer Vision and System (Ministry of Education), Tianjin University of Technology, Tianjin 300384, PR China

ARTICLE INFO

Keywords:

Influential spreaders
Identification algorithms
Classified neighbors
Complex networks

ABSTRACT

Identifying the influential spreaders in complex network is a very important topic, which is conducive to deeply understanding the role of nodes in the information diffusion and epidemic spreading among a population. To this end, in this paper, we propose a novel classified neighbors algorithm to quantify the nodal spreading capability and further to differentiate the influence of various nodes. Here, we believe that the contribution of different neighbors to their focal node is different, and then classify the neighbors of the focal node according to the removal order of the neighbor in the process of k -shell decomposition. By assigning different weights for each class of neighbors and summing up the neighbors' contributions, the spreading capacity of the focal node can be accurately characterized. Through extensive simulation experiments over 9 real-world networks, the weight distribution of different types of neighbors has been optimized, and the results strongly indicate that the current algorithm has the higher ranking accuracy and differentiation extent when compared to other algorithms, such as degree centrality, k -shell decomposition method and mixed degree decomposition approach. Current results can help to greatly reduce the cost of sales promotion, considerably suppress the rumor dissemination and effectively control the outbreak of epidemics within many real-world systems.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Over the past two decades, various dynamics taking place upon complex networks [1–10] receive a great deal of concern, such as synchronization, consensus, evolutionary game, epidemics and opinion dynamics. Among them, the spreading process over networks [11] is an important subject in the field of physics, chemistry, public hygienics, biology and even sociology. As an example, influenza spreading, rumor diffusion, cascading failure, or information dissemination can be all incorporated into the range of spreading process. To be particularly worth mentioning, the epidemic spreading in complex networks has been extensively investigated and many studies have clearly indicated that this process is strongly correlated with the topology of networks [11,12]. Among them, Pastor-Satorras and Vespignani [13] defined a dynamical model for the epidemic spreading in scale-free networks, and found a striking phenomenon which declared the absence of an epidemic threshold and its associated critical behavior under the thermodynamic limit. Starting from this work, many models have

* Corresponding author at: Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, PR China.

E-mail addresses: wjtut08@126.com, xialooking@163.com (C. Xia).

been developed to analyze the spreading properties on networks and then to increase the spreading threshold for a specific disease [14–23]. As a further step, several efficient immunization schemes [24–28] have also been put forward to inhibit the outbreaks of infectious diseases.

Although large quantities of previous works try to macroscopically understand and control the spreading process, in the very recent years more and more attention has been paid to microscopically studying the individual spreading capability, which characterizes the number of nodes covered within a specified process when a single agent initiates an epidemic [29–32]. In fact, knowing the spreading capability for each node is vital for us to devise efficient methods to control the guidance of public opinion [33,34], promote the popularity of novel products [35–37] and even refrain the outbreaks of diseases within the population [38,39]. In addition, it will be also beneficial to search for the spreading origin (i.e., initial spreader) of a certain disease or information [40].

At present, several classic topology indicators can be used to quantitatively characterize the individual spreading capabilities [41–52], such as degree, closeness, betweenness, clustering coefficient, Katz centrality and so on. Despite the intuition that the most connected nodes (hubs) or nodes with high betweenness centrality are thought to be very influential spreaders inside networks, many works have also revealed that the real-world spreading scenarios cannot well agree with these ideas. Thus, exploring the novel index to denote the individual spreading capacity becomes an intriguing topic. Among them, the k -shell decomposition algorithm [48] takes advantage of the debarking method to rank the nodes. In this method, the algorithm starts by iteratively removing all nodes with degree 1 until no such these nodes remain, and we assign their k -shell value to be 1. In a similar fashion, we will obtain the 2-shell nodes by recursively removing the remaining nodes with degree 2, and this procedure continues until all nodes are assigned to a k -shell value. It is obvious that the nodes with high k -shell value will tend to locate at the center of the network, and hence the covered range originating from these nodes is likely to be larger, which means the higher spreading capability. Based on some real infectious diseases, this method was found to perform better than other methods in identifying the influential spreaders.

However, when the k -shell method is used to decompose the networks, we only consider the remaining links among the existent nodes (named as the residual links, correspondingly the concept of residual degree), but often neglect the links connecting to the removed nodes (called the exhausted links, accordingly exhausted degree). In order to hold the same spreading capability, all nodes within the same core should have the identical number of exhausted links, at least a very small fluctuation in the exhausted degree. In reality, the exhausted degree often exhibits a very heterogeneous distribution in real networks. Thus, all nodes with the same residual degree will be assigned the same rank even though their exhausted degree may fluctuate much more. As an example, the k -shell method may designate the same k -shell value for many nodes in tree-like network or Barabási-Albert networks. Therefore, it is found that k -shell method is efficient in identifying the most influential spreader for infectious disease, but it is not too suitable to sort the nodes regarding the information or rumor propagation, and thus there exists much room for us to improve the quality of k -shell method since k -shell method denotes a very coarse-grained decomposition. For instance, Zeng and Zhang [52] propose a mixed degree decomposition (MDD) procedure to predict the spreading capability by combining the residual degree and exhausted degree, which can effectively reduce the degeneracy of the k -shell algorithm. Furthermore, Liu et al [53,54] take use of optimization algorithm to remove the core-like groups and redundant links inside the networks and then adopt the k -shell method to identify the true core within networks, and henceforth the performance of k -shell method has been further enhanced.

As mentioned in Ref. [52], different neighbors may play a distinct role in measuring the individual spreading capability (SC). Thus, in this paper, we propose a well-refined method to further classify the neighbors according to their relative sequence of being removed during the k -shell decomposition. Then, we will set a different weight for each class of neighbors to characterize the spreading capability, and further simulate the Susceptible-Infected-Removed (SIR) epidemic model on several typical real-world network data sets. It is evidenced that the proposed algorithm is better than degree centrality, k -shell, and MDD method regarding the performance of measuring the spreading capability. Meanwhile, the time complexity of our method is also largely reduced when compared to that of MDD algorithm.

The rest of this paper is structured as follows. In Section 2, we introduce the neighbor classification in detail, and then we briefly describe the SIR epidemic model and the corresponding real-world network data sets adopted in this paper. After that, we will carry out extensive numerical simulations upon 9 real-world networks and compare the present method and other typical algorithms in Section 3. Finally, we end this paper with some concluding remarks in Section 4.

2. Model and algorithm

2.1. Classified neighbors

In order to present the idea on the node assortment, we firstly look at a typical scenario shown in Fig. 1, where the spreading capability of node d is usually different from those of nodes a , b and c although their k -shell value (k_S) is identical ($k_S = 1$). Here, node d is removed after nodes a , b and c have been deleted from the network, and the coarse-grained k -shell decomposition does not consider this sequence. Analogously, in the $k_S = 2$ layer, nodes j and l are also different from nodes k , m and n even though their k -shell values are all 2. Therefore, it is necessary to take the removal sequence within the same k_S layer into account and further classify the neighboring nodes for each node with the same k_S values during the network decomposition. Our method starts from the aforementioned assortment mechanism regarding the nodes with the same k_S value, and we will simultaneously record the k -shell value and the removing order for each node during the process

Download English Version:

<https://daneshyari.com/en/article/8901444>

Download Persian Version:

<https://daneshyari.com/article/8901444>

[Daneshyari.com](https://daneshyari.com)