Physica A 506 (2018) 802–815

Contents lists available at ScienceDirect

Physica A

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

Power iteration ranking via hybrid diffusion for vital nodes identification

Tao Wu[a,*], Xingping Xian[b,*], Linfeng Zhong[c,e], Xi Xiong[d], H. Eugene Stanley[e]

[a] *School of Cyber Security and Information Law, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*
[b] *College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*
[c] *Department of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China*
[d] *School of Cybersecurity, Chengdu University of Information Technology, Chengdu 610225, China*
[e] *Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA*

HIGHLIGHTS

- Combine mass diffusion and heat conduction process for node ranking.
- Design a nonlinear hybrid mechanism for node state updating.
- The proposed PIRank method can capture different structural features.
- PIRank is immune to the localization transition of leading eigenvector.

ARTICLE INFO

Article history:
Received 17 November 2017
Received in revised form 8 March 2018
Available online 4 May 2018

Keywords:
Node ranking
Centrality measure
Physical process
Network disruption

ABSTRACT

One of the most interesting challenges in network science is to understand the relation between network structure and dynamics on it, and many topological properties, including degree distribution, community strength and clustering coefficient, have been proposed in the last decade. Prominent in this context is the centrality measures, which aim at quantifying the relative importance of individual nodes in the overall topology with regard to network organization and function. However, most of the previous centrality measures have been proposed based on different concepts and each of them focuses on a specific structural feature of networks. Thus, the straightforward and standard methods may lead to some bias against node importance measure. In this paper, we introduce two physical processes with potential complementarity between them. Then we propose to combine them as an elegant integration with the classic eigenvector centrality framework to improve the accuracy of node ranking. To test the produced power iteration ranking (PIRank) algorithm, we apply it to the selection of attack targets in network optimal attack problem. Extensive experimental results on synthetic networks and real-world networks suggest that the proposed centrality performs better than other well-known measures. Moreover, comparing with the eigenvector centrality, the PIRank algorithm can achieve about thirty percent performance improvement while keeping similar running time. Our experiment on random networks also shows that PIRank algorithm can avoid the localization phenomenon of eigenvector centrality, in particular for the networks with high-degree hubs.

* Corresponding authors.
E-mail addresses: wutaoadeny@gmail.com (T. Wu), xxp0213@gmail.com (X. Xian).

https://doi.org/10.1016/j.physa.2018.04.012

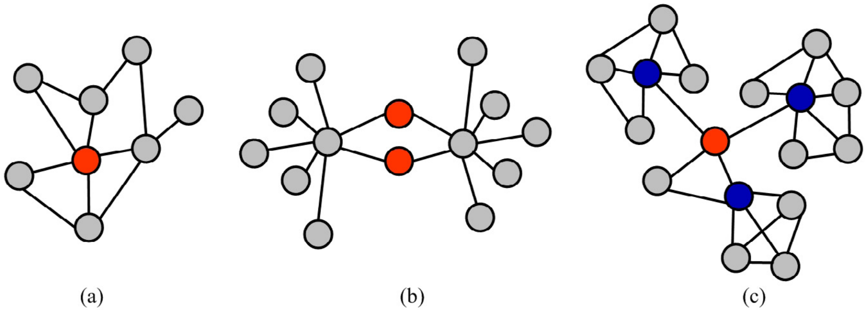

Fig. 1. Illustration of vital nodes with different structural features. (a) Local dominant node. (b) Intermediary nodes. (c) Network including vital nodes with complex structural features. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

1. Introduction

Networked data have become ubiquitous, and networks describing physical systems, protein interactions, computer communications, and people relationships are all becoming increasingly important in our day-to-day life. Complex network analysis has proven to be a successful tool for modeling and mining enormous networked data. Because of the heterogeneous of complex networks, some network nodes are more important to network function than others. Measuring the relative importance of individual nodes is important in both theoretical research and practical application. For example, identifying and protecting the crucial elements of the Internet so that the functioning of the system can be maintained, vaccinating influential individuals in contact networks so that the spread of an epidemic can be suppressed, and identifying and removing key vertices in a molecular network so that the bacteria can be eliminated. This context raises a fundamental question: given the data, how should one construct the node centrality measure such that it can capture structural features effectively?

In the last decade, node ranking problem has been particularly proposed to measure the importance of nodes within a given network [1], and a variety of centrality measures have been suggested based on different interpretations, including neighbor-based local centralities (degree centrality, local centrality [2], and collective influence [3]), location-based global centralities (closeness centrality [4], betweens centrality [5], and k-shell centrality [6,7]), and path counting centralities (eigenvector centrality [8], Katz's centrality [9], PageRank [10], and RA centrality [11,12]). Because of the variety of networked systems, the vital nodes in every task are concretized from different perspective. For example, the vital nodes in influence maximization problem are the influential sources that spread information [13], and the vital nodes in network disruption problem are the intermediate nodes that maintain network integrity [14]. Specifically, influence maximization problem has attracted attention recently [15–17], and many single node ranking methods [18–20] and multiple spreaders identification methods [21–23] been proposed for the problem. Meanwhile, researchers begin to be concerned about node ranking method for temporal networks [24]. Therefore, vital nodes may mean different things in various applications, and there is no general consensus on the definition. This paper focus on the fundamental problem, i.e. vital nodes identification based on network structure analysis, and explore the method to optimize the ranking accuracy. In face of such a problem, most of the existing centrality measures focus on a specific structural feature and have limits in node ranking. For example, the degree correlated local centralities (degree centrality, local centrality), global centralities (closeness centrality, betweens centrality) and random walk based centralities (PageRank, LeaderRank [25]) are all towards to high degree nodes. However, researchers have unveiled that low degree nodes always be very important and meaningful in many complex systems [26–28]. Therefore, to identify vital nodes accurately, centrality measures should be optimized to capture network nodes' structural information as comprehensively as possible. To illustrate this idea, Fig. 1 gives an example of vital nodes with different structural features. In detail, the local dominant node with red color in Fig. 1(a) preserves connections with most of the network nodes, the intermediary nodes with red color in Fig. 1(b) maintain the network's integrality and have more control on communications between network components, and in Fig. 1(c) the nodes with blue color have features of local dominant node and the nodes with red color have features of local dominant node and intermediary node simultaneously. It is easy to see that the nodes with colors play critical role in network structure and accurate node ranking should consider all the strucutral features.

As representative of the class of spectral centralities, eigenvector centrality measures the importance of a node based on the influence of its neighbors. High-influence neighbors contribute more to central node's influence than low-influence ones, and a node is influential if it has many influential neighbors. Eigenvector centrality calculates based only on local information in each step but can utilize global network information through successive iteration. It assigns a relative importance score v_i for node i that is proportional to the sum of the scores of the neighbors of node i. Mathematically, this can be written as $v_i = \lambda^{-1}\sum_j A_{ij}v_j$, where λ is a constant of proportionality and A_{ij} is an element of the adjacency matrix $\mathbf{A}$ of a network having value one if there is an edge between node i and j and zero otherwise. In the matrix form, we have $\mathbf{Av} = \lambda\mathbf{v}$, which means that the vector $\mathbf{v}$ of centralities v_i is an eigenvector of the adjacency matrix $\mathbf{A}$. Because centralities are all nonnegative, the Perron–Frobenius theorem [29] guarantees that the vector $\mathbf{v}$ of centralities v_i must be the leading eigenvector of non-negative real square matrix $\mathbf{A}$. Meanwhile, physical diffusion process mass diffusion (MD) and heat conduction (HC) have