Chaos, Solitons and Fractals 114 (2018) 69–80

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Density centrality: identifying influential nodes based on area density formula

Ahmed Ibnoulouafi [a,*], Mohamed El Haziti [b]

[a] LRIT Laboratory, Associated Unit to CNRST (URAC No 29) IT Rabat Center - Faculty of Sciences In Rabat, Mohammed V University In Rabat, B.P.1014 RP, Rabat, Morocco

[b] Higher School of Technology (E.S.T), SALE, Morocco

ARTICLE INFO

Article history:
Received 29 September 2017
Revised 14 June 2018
Accepted 16 June 2018

Keywords:
Complex networks
Centrality measures
Influential nodes

ABSTRACT

Identifying central nodes in a network is crucial to accelerate or contain the spreading of information such as diseases and rumors. The problem is formulated as follows, given a complex network, which node(s) is (are) the more important ? The idea of centrality was initially introduced in the context of sociology, to look whether there is a relation between the location of an individual in the network and its influence in group processes. Since then, a plethora of centrality measures has emerged over the years and were employed in a multitude of contexts to rank nodes according to their topological importance. Each centrality targets the problem of influence from its own perspective.

In this article we introduce a new centrality that takes inspiration from Area density formula to define the density of each node by considering the degree and the distance between two nodes in a neighborhood of order r = 1, 2, 3, etc... To examine the performances of the proposed measure, we conduct our experiments on synthetic as well as real-world networks by comparing the monotonicity, correlation, the network damage caused by deleting important nodes and the spreading capabilities of nodes using the classical Susceptible-Infected-Recovered (SIR) epidemic model. According to the empirical results, the proposed measure can effectively evaluate the importance of nodes.

1. Introduction

Complex networks is a term that designate networks whose behaviors in a large scale level cannot be predicted from the behavior of their individual entities (nodes and links). This is due to the complexity of such networks (very large amount of data) and the lack of knowledge of the phenomena occurring in these networks. Complex networks are present everywhere, social networks [1], traffic networks [2,3], power grid [4,5], P2P networks [6], etc. The major progresses made in network science in the past years had a great impact on the understanding and description of complex networks [7]. Recently, with the explosion of BIG DATA, the application of complex networks is more popular than ever, especially due to the fact that they constitute a great support to information spreading. One of the research areas included in the field of information spreading that has recently attracted a lot of attention is the task of identifying influential nodes in a network. This topic follows two lines of research: (a) identification of individual influential nodes [8] and (b) identification of a group of nodes that, by acting all together, generate the largest spread of information. In other words, we search the most influential community in a network [9]. In this paper, we focus on the first problem.

Identifying key nodes in complex networks has attracted increasing attention in recent years due to its many applications. In the spread of diseases, being able to locate and immunize the most influential individuals can prevent further spread of the virus [10]. In a large-scale computer network, it is crucial to design a robust and secured architectures by creating backup servers and redundant links according to the importance of servers. In marketing, the promotion of product passes through locating central individuals that can accelerate the diffusion of information and thus optimize the sales of products [11].

To address this problem, many centrality measures have been proposed over the two last decades, some describe the local environment around a node (e.g., Degree centrality [12], PageRank [13] and Local Centrality [14]) others the more global position of a node in the network (e.g., Closeness [12] and Betweenness [12,15], two of the most widely used centrality measures based on a model of non splitting information transmission along shortest paths). In the category of global measures, Kitsak et al. [16] proposed a

* Corresponding author.
E-mail address: ahmedibnoulouafi@gmail.com (A. Ibnoulouafi).

https://doi.org/10.1016/j.chaos.2018.06.022
0960-0779/

K-core decomposition based on the assumption that nodes located in the core of the network have higher spreading capabilities than the ones located in the periphery. K-core demonstrate a better ranking than Degree centrality in many real networks. However, recent research pointed out that nodes located within the same core often have distinct influences, which exposes the lack of K-core in term of distinguishing between nodes spreading capabilities. In order to surpass this failure and improve the K-core method, many alternatives were proposed. The first one is the mixed degree decomposition (MDD) that incorporate the residual and exhausted degree [17]. The MDD method shows better performance in identifying node position and can further distinguish some of the nodes belonging to the same core. However, it gives equal importance to the removed nodes regardless of their positions in social networks, leading to the limited improvement in performance [18]. Lin et al. presented an improved neighbors K-core (INK), that ranks nodes by taking into account the shortest path distance between a target node and the node set with the highest K-core value [19]. Recently, Bae et al. defined a new variant of K-core, called Neighborhood coreness centrality (C_{nc}), which is given by summing all neighbors K-core values [18]. Another interesting centrality, called DIL, was proposed in[20] as a great alternative to Betweenness centrality. Instead of working on a global level, DIL centrality ranks nodes based on local information (degree value and the importance of lines) to identify network bridges. This measure showed great performances and is adapted to large-scale networks due to its low complexity. Recently, in [21], an interesting method called node information dimension (labeled NID) that exploits local and global characteristics of complex networks was proposed. NID centrality characterizes the importance of nodes by aggregating the local dimensions at different topological scales. Although the method has a high computational complexity, its hybrid nature makes it a better method to identify influential nodes.

We also find in literature measures inspired from fields such as physics. Estrada in [22] uses the physical concept of vibrations to account for the node vulnerability. To do so, he submerges the network in question in a thermal bath. The thermal fluctuation works as the “perturbations” acting on the network. As a measure of vulnerability, he uses the displacement of a node from its “equilibrium” position due to small ”oscillations” in the network. Rossi et al. [23] introduced a centrality measurement based on continuous-time quantum walk. Using the quantum Jensen-Shannon divergence, they related the importance of a vertex to the influence that its initial phase has on the interference patterns that emerge during the quantum walk evolution. They also showed that under particular settings, their centrality is almost linearly correlated with Degree centrality. Relying on Kirchhoff's law for electric circuits, Avrachenkov et al.[24] gave a new concept of betweenness centrality (called beta current flow centrality) for weighted network. In [25], Ma et al. proposed a new centrality (labeled Gravity) based on the Isaac Newton classical gravity formula. Gravity centrality considers the K-core value of a node as its mass, and the shortest path distance between two nodes in a network is viewed as their distance. In a word, designing an effective method to identify and rank the node importance is still an open issue.

In this paper, we take inspiration from the Area density formula to propose a new centrality measure we called Density centrality. For each node, the Density centrality is computed by considering the degree and the distance between two nodes in a neighborhood of order r = 1, 2, 3, etc... To evaluate the proposed centrality measure, we report a series of experiments on synthetic and real-world networks. Extensive comparisons with the most recent alternative measures are performed. Results show that Density Centrality provides a more accurate ranking list.

The remainder of this article is organized as follows. In Section 2, we review the necessary background on recent centrality measures used in this work and the evaluation metrics. In Section 3 we introduce the Density Centrality measure. The datasets, the experimental setup and results are presented in Section 4. Finally, Section 5 concludes the paper.

2. Background

In this section, we review briefly existing centrality measures that will be discussed in this work. The list of these centralities comprises:

- Gravity centrality: Combines a global measure (Coreness) and geodesic distance.
- DIL centrality: A local measure.
- Closeness centrality: A global measure based geodesic distance.

Additionally, we recall the definition of the evaluation metrics that are used to compare centralities. Monotonicity quantify the ability of a centrality measure to distinguish the nodes in a network, Kendall's Tau coefficient (τ) quantify the correlation between the nodes influence rankings of two measures and network efficiency expresses the connectivity of the network after removing the most important nodes.

2.1. Existing centrality measures

A network can be represented by a graph G=(V, E), where V is the set of network nodes V = $\{v_1, v_2, \dots, v_n\}$ and E the set of links connecting nodes E = $\{e_1, e_2, \dots, e_m\}$. In term of the adjacency matrix it is defined by:

$$A_{n,n} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix}$$

where $A_{ij} = 1$ if $\exists$ a link between i and j and 0 otherwise.

Degree centrality: The degree of a node is the number of edges incident with it as it is defined as:

$$k_i = \sum_{j=0}^{n} A_{ij}, \quad (1)$$

where A is the adjacency matrix.

Gravity centrality: Enlighten by the idea of classical gravity formula proposed by Isaac Newton, the authors in [25] view the K-core value of node i as its mass, and the shortest path distance between two nodes in a network is viewed as their distance. In this way, they defined a centrality (labeled as G) that measure the influence of node i as follow:

$$G(i) = \sum_{j \in \psi_i} \frac{K_s(i) K_s(j)}{d_{ij}^2}, \quad (2)$$

where d_{ij} is the shortest path distance between node i and j. ψ_i is the neighborhood set whose distance to node i is less than or equal to a given value r. In their paper, the r parameter is set to r = 3 (only nearest neighbors, next nearest neighbors and the next-next nearest neighbors are considered).

DIL centrality: A new ranking method [20] based on local information (degree value and the importance of lines) to identify the importance of bridge nodes. It represents by its low complexity a great alternative to the Betweenness centrality. The importance of the line between two nodes i and j is defined as:

$$I_{e_{ij}} = \frac{U}{\lambda}, \quad (3)$$

where e_{ij} is the link between nodes i and j, U = $(k_i - p - 1).(k_j - p - 1)$ reflects the connectivity ability of line e_{ij}, k_i is the degree of