Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Correlation of cascade failures and centrality measures in complex networks

Ryan Ghanbari, Mahdi Jalili *, Xinghuo Yu

School of Engineering, RMIT University, Melbourne, Australia

HIGHLIGHTS

- Cascade depth is correlated with nodes' centrality measures in this paper.
- As degree of node increases, their cascade depth decreases.
- As betweenness and local rank of nodes increase, their cascade depth increases.

ARTICLE INFO

Article history:
Received 9 November 2016
Received in revised form 29 August 2017
Accepted 3 September 2017
Available online xxxx

Keywords:
Complex networks
Dynamics on networks
Cascade failures
Centrality measures
Protection
Resiliency

ABSTRACT

In complex networks, different nodes have distinct impact on overall functionality and resiliency against failures. Hence, identifying vital nodes is crucial to limit the size of the damage during a cascade failure process, enabling us to identify the most vulnerable nodes and to take solid protection measures to deter them from failure. In this manuscript, we study the correlation between cascade depth, i.e. the number of failed nodes as a consequence of single failure in one of the nodes, and centrality measures including degree, betweenness, closeness, clustering coefficient, local rank, eigenvector centrality, lobby index and information index. Networks behave dissimilarly against cascade failure due to their different structures. Interestingly, we find that node degree is negatively correlated with the cascade depth, meaning that failing a high-degree node has less severe effect than the case when lower-degree nodes fail. Betweenness centrality and local rank show positive correlation with the cascade depth. In order to make networks more resilient against cascade failures, one can remove nodes that ranked high in terms of those centrality measures showing negative correlation with the cascade depth.

1. Introduction

Power grids, road networks, airports, water distribution infrastructures and the Internet are examples of critical networks playing essential roles in modern communities. Their proper functioning and resiliency are of extreme importance. Such critical networks might be subject to failure in their components (nodes/edges). Errors (i.e., random failure) and attacks (i.e., intentional failures) have been studied on complex networks [1–4]. It has been shown that the resiliency of complex networks against errors and attacks depends on their structure. For example, scale-free networks – which are characterized by heavy-tailed degree distribution – are fragile to attacks, while being resilient against errors [1]. Networks with more homogeneous degree distribution (e.g., random networks) behave similarly against degree-based errors and attacks.

In some cases, a more catastrophic situation can happen and a failure in one component may propagate through the network resulting in failing some other components. This is referred to as cascade failure in the literature [5–8]. Such cascade failures were responsible for large-scale blackouts in power grids [9]. Power grids can be modeled as networked structures with generators, loads and transformers as nodes and wirings as links [10]. Any failures in one of the network components may give rise to an overload in other components and as a result a failure in them. Such failures can pervade through the network and result in a cascade failure process [5,11,12], which can in turn lead the whole or a fundamental part of the network to crash. To avoid this catastrophic collapse from happening, one way is to identify the critical capacity for edges, where setting the edge capacity parameter higher than

* Corresponding author.
E-mail addresses: ryan.ghanbari@rmit.edu.au (R. Ghanbari), mahdi.jalili@rmit.edu.au (M. Jalili), x.yu@rmit.edu.au (X. Yu).

http://dx.doi.org/10.1016/j.future.2017.09.007
0167-739X/

that makes the networks resilient against failure [10]. But, it is costly or even impractical in some cases. An alternative solution is to determine the most critical components in networks.

In this work, we examine a number of node centrality metrics on synthetic and real networks to study the role of each centrality measure in the cascade failure. To this end, we first obtain a number of centrality measures for the nodes, including degree, betweenness, closeness, clustering coefficient, eigenvalue, information index, lobbying index and local rank. We also consider cascade depth, which is defined as the number of failed nodes as a result of initial failure in a node. In other words, as a node fails, the loads are redistributed and the edges with load higher than their capacity (which is obtained from the initial load) fail. The redistribution-failure procedure is iterated until steady state is obtained, and finally, the number of the failed nodes is interpreted as the cascade depth of the initially failed node. Cascade depth is indeed a measure indicating vitality of nodes with regard to cascaded failures and the larger values of cascade depth imply the higher criticality of node in cascade failure. Although the role of various node centrality measures on network functions has been studied in the literature (see reviews in [13,14]), there is no existing work to study the relationship between centrality measure and cascade failure. The major contribution of this work is to study correlation of nodes' vitality (i.e., their centrality) and their influence on the cascade failures measured by the cascade depth.

2. Related works

Recently, a number of studies have been carried out to discover the role of the components in different dynamical processes spanning from epidemic management, innovation dispersion, viral marketing, and social movement to idea dissemination [15,16]. Different centrality measures (e.g., degree, betweenness, closeness, eigenvalue and information theory based) have been proposed to identify critical components in networks. By discovering the critical nodes, one can control the propagation of information in social networks, spreading epidemics in a society, and prevent disastrous cascade failures leading to blackouts in power grids or internet outages [17,18]. However, finding vital nodes in a network is not trivial [13]. At first, the term 'vital node' has a vast meaning. Sometimes, we are interested in finding the small set of people whose vaccination can stop the disease from spreading while sometimes we need to find the most critical buses which if removed, the power grid will experience a cascade of failures. Secondly, to establish a reasonable balance between local and global parameters is challenging.

Cascade failure is a catastrophic type of failure, where a random/intentional failure starting from a (set of) node(s) may quickly spread through the network, which may result in failing a large portion of network components [5–8,10–12,19]. Power grids are one of real systems where cascade failures can have dramatic socioeconomic consequences, by leaving many without power. Cascade failures have been studied in the literature including on power grids [20–23]. However, the previous works have mainly focused on modeling cascade failures [24], designing efficient protection strategies [25,26] or studying dependence of cascade failures on the structural properties [9,27] of networks. The behavior of a cascade failure depends on the location of initial failure, i.e., the node(s)/edge(s) that are initially fail. Such a study is missing in the literature, which will be addressed in this work.

Information cascade is a similar topic to cascade failure in the literature of network science. A piece of information, starting from a set of seed nodes, disseminates through the network. Influence maximization is a well-known problem in this field, which is to find the optimal set of seed nodes such that initially activating them has the maximum influence, i.e., the largest number of finally activated nodes. Corley et al. [28] defined the most vital nodes in a network as those whose failure engenders the highest decline in maximum flow between a particular node pair. In another work, Corley et al. [29] found the k-most vital nodes whose failure makes the shortest distance between two arbitrary nodes the highest possible. Real networks often have community structure. He et al. [30] introduced an approach to find the top-k influential spreaders in networks with community structure. Recently, Jalili and Perc studied the correlation between influence range, i.e., the number of activated nodes as a result of initially activating a node, and centrality measure [14]. They used independent cascade model and identified the centrality measures showing strong positive correlations.

3. Network centrality measures

In this work, we consider unweighted and undirected networks. Let us denote the network $G = \{V, E\}$, with V and E being the set of nodes and links, respectively. The network can be represented by its adjacency matrix $A = \{a_{ij}\}$, where $a_{ij} = 1$ if an edge exists between nodes v_i and v_j, and $a_{ij} = 0$ otherwise. We consider networks without any self-loops, i.e., $a_{ii} = 0$. Centrality of a node (or edge) in a network determines its importance in a certain functionality of the network. A centrality index assigns a score to all nodes indicating their vitality in the network. Due to the implicit meaning of 'importance', disparate indices have been introduced to cover the concept. The most trivial method to evaluate the centrality of nodes accounts for the number of immediate neighbors of a node which is presented as degree centrality. The degree d_i of node v_i is the number of edges incident on that node:

$$d_i = \sum_j a_{ij} \tag{1}$$

Node degree is a simple centrality measure that needs only local information on the nodes. Degree is indeed the simplest centrality measure, and has been shown to control many of network functions. In order to obtain an effective ranking parameter to overcome computationally complex calculations in large-scale networks, Chen et al. [31] proposed another local centrality measure, called Local Rank, which considers information on nodes' fourth order neighbors. Local rank is a compromise between the degree centrality and other time-consuming measures. The Local Rank of each node v_i is computed as follows:

$$LR\,(i) = \sum_{j \in \Gamma_i} Q\,(j) \tag{2}$$

$$Q\,(j) = \sum_{k \in \Gamma_j} R\,(k) \tag{3}$$

where Γ_i is the set of immediate neighbors of v_i and $R\,(k)$ accounts for the number of immediate and the next immediate neighbors of v_k. Another measure which quantifies the interconnection in the network is clustering coefficient. It measures the local connectivity in the network (i.e., indicating to what extent the neighbors of a node are interconnected). For a node with degree d_i, the maximum number of possible edges among its neighbors is $d_i(d_i - 1)$. The clustering coefficient is the portion of these possible edges that indeed exist, and is computed as:

$$C_i = \frac{|\{r, s\}|}{d_i\,(d_i - 1)} \mathit{if}\ r, s \in \mathit{Neighbors}\ (i)\ \&\ \{r, s\} \in \mathit{Edges} \tag{4}$$

Please cite this article in press as: R. Ghanbari, et al., Correlation of cascade failures and centrality measures in complex networks, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.09.007.