Reliability Engineering and System Safety 133 (2015) 157–166

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Characterizing and predicting the robustness of power-law networks

Sarah LaRocca, Seth D. Guikema*

Department of Geography and Environmental Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

ARTICLE INFO

Article history:
Received 26 July 2013
Received in revised form
21 July 2014
Accepted 26 July 2014
Available online 21 August 2014

Keywords:
Networks
Scale-free
Robustness

ABSTRACT

Power-law networks such as the Internet, terrorist cells, species relationships, and cellular metabolic interactions are susceptible to node failures, yet maintaining network connectivity is essential for network functionality. Disconnection of the network leads to fragmentation and, in some cases, collapse of the underlying system. However, the influences of the topology of networks on their ability to withstand node failures are poorly understood. Based on a study of the response of 2000 randomly-generated power-law networks to node failures, we find that networks with higher nodal degree and clustering coefficient, lower betweenness centrality, and lower variability in path length and clustering coefficient maintain their cohesion better during such events. We also find that network robustness, i.e., the ability to withstand node failures, can be accurately predicted a priori for power-law networks across many fields. These results provide a basis for designing new, more robust networks, improving the robustness of existing networks such as the Internet and cellular metabolic pathways, and efficiently degrading networks such as terrorist cells.

1. Introduction

Properly functioning networks are critical to modern life and economies. Communications networks, power systems, and transportation networks form the basis on which economic growth and security are built. The natural environment too is built largely of networks, from cellular metabolic pathways to large-scale ecological networks. In all cases, these networks are subject to failures of critical nodes and links. Communication hubs may be attacked or experience technical failures, bridge failures may lead to large-scale disruption in a transportation network as in the I-35 bridge failure [1], power networks may fail due to loss of lines and generation nodes, and ecological networks are subject to severe disruption as species become less common in the network. Being able to quickly and efficiently estimate the ability of a given network to withstand node failures, that is, its robustness, is central to being able to manage critical networks and increase their robustness. At the same time, being able to quickly and efficiently estimate robustness enables more efficient attacks on networks, such as terrorist networks, that we wish to degrade. However, there does not yet exist a method for estimating the robustness of networks quickly and accurately based on the topological characteristics of the network, and the existing understanding of the influence of topological characteristics on network robustness if limited. In this paper we focus on scale-free networks and develop such a model.

Scale-free networks exhibit a power-law nodal degree distribution where the probability that a given node is connected to k other nodes is described by $P(k) \sim k^{-\gamma}$ [2]. Empirical evidence indicates that nodal degree in many real networks is limited by the physical costs of adding links to a node. Such networks can be described by adding an exponential cutoff to the power-law distribution $P(k) \sim k^{-\gamma} e^{-(k/\kappa)}$, where κ is the cutoff above which it becomes physically very costly to add links to a node [3–6]. Scale-free networks have been demonstrated to be tolerant to random failures [7]. However, the combined influence of individual measures of network topology on failure tolerance has not been studied. Without an understanding of the relationship between topology and robustness to node failures, we are limited in our ability to design failure-tolerant networks across many different domains and in our ability to efficiently degrade networks that we wish to attack. Here, we present a systematic study of the effects of topological characteristics on power-law network fault tolerance, and we develop a topology-based statistical approach for estimating the ability of a network to tolerate node failures.

Our work helps to address the gap in current network robustness modeling in two ways. First, we develop a statistical model for quickly estimating the robustness of a network after node failure events for networks containing up to 1000 nodes. This model estimates robustness for up to 75% of the original nodes failing, making it useful not only for small failure events but also

* Corresponding author.
E-mail address: sguikema@jhu.edu (S.D. Guikema).

http://dx.doi.org/10.1016/j.ress.2014.07.023
0951-8320/

large-scale failure events induced by common-cause failures such as natural disasters in which large portions of networks fail [8–11]. Second, we use our statistical model to gain insight into the topological characteristics of networks that influence their robustness. This, together with rapid estimation of robustness, provides a strong basis on which robustness can be included in network design optimization.

2. Network topology

A network, or graph, is described by $\mathcal{G}=\{\mathcal{V},\mathcal{E}\}$, where $\mathcal{V}$ is the set of vertices, or nodes, and $\mathcal{E}$ is the set of edges, or links. For directed graphs, the elements of $\mathcal{E}$ are ordered pairs of distinct vertices, while for undirected graphs, the elements of $\mathcal{E}$ are unordered pairs of distinct vertices. The total number of nodes in a graph is equal to the number of elements in $\mathcal{V}$, that is, $N=|\mathcal{V}|$. Correspondingly, the number of edges in a graph is equal to the number of elements in $\mathcal{E}$, that is, $M=|\mathcal{E}|$ [12]. As an example, consider the network representation of an electric power system. The nodes would consist of the electric power generation plants, substations, line junctions, and demand nodes, and the edges would consist of the power lines connecting the nodes.

Any given graph can be uniquely represented by an $N\times N$ adjacency matrix, A. If there exists an edge from some vertex i to some vertex j, then the element a_{ij} is 1; otherwise, it is 0. Network topology can be described by a variety of measures which can be calculated from an adjacency matrix. Four such measures are particularly useful for characterizing the structure of a network: degree distribution, betweenness centrality, clustering coefficient, and path length [13]. Table 1 presents the expected effect of a change in the mean of each of these measures on network robustness.

2.1. Degree distribution

In undirected networks, the nodal degree, k, of a given node is defined as the number of edges that are incident the node; the mean degree of a network, $\langle k\rangle$, is defined as

$$\langle k\rangle=\frac{1}{N}\sum_{i\in V}k_i. \tag{1}$$

Typically, the nodes in a given network do not all have the same degree; rather, the distribution of nodal degrees in the network can be described by some probability density function, $P(k)$, which gives the probability that a randomly selected node has exactly k edges [13]. Many networks have degree distributions that follow a power-law, described by

$$P(k)\sim k^{-\gamma}, \tag{2}$$

where γ is a constant. Additionally, empirical evidence indicates that nodal degree in many real networks is limited by the physical costs of adding links to a node. Such networks can be described by adding an exponential cutoff to the power-law distribution, that is,

$$P(k)\sim k^{-\gamma}e^{-(k/\kappa)}, \tag{3}$$

where κ is the cutoff above which it becomes physically very costly to add links to a node [3–6].

Table 1
Expected effect of change in mean network measure on network robustness.

Network measure	Change in mean	Change in network robustness
Degree	+	+
Betweenness	+	−
Path length	+	−
Clustering coefficient	+	+

2.2. Betweenness centrality

Another important measure of network topology is the betweenness coefficient, which is defined as the total number of shortest paths passing through a given node. Relatedly, the betweenness centrality of a node is defined as follows:

$$Cb_k=\sum_i\sum_j\frac{\rho_{ikj}}{\rho_{ij}},\quad i\neq j\neq k, \tag{4}$$

where ρ_{ij} is the number of shortest paths (i.e., minimal sequences of edges) from node i to node j and ρ_{ikj} is the number of these paths that pass through node k [12]. Betweenness and betweenness centrality are useful measures of the importance of a node because they quantify the number of shortest paths that will become longer if the node is removed from the graph.

2.3. Path length

Path length, d_{ij}, describes the length of the shortest path between a given pair of nodes. Then, average path length describes the mean of the shortest distance between all pairs of nodes in a network. That is,

$$\ell=\frac{1}{N(N-1)_i}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{V}}d_{ij}, \tag{5}$$

where d_{ij} is the length of the shortest path (i.e., number of edges) between node i and node j.

2.4. Clustering coefficient

The clustering coefficient was introduced by [14] as a means of quantifying the degree to which nodes are clustered in a graph. Suppose a node i is connected to k_i other nodes, or neighbors. Then the clustering coefficient for a given node i is defined as follows:

$$C_i=\frac{2\mathcal{E}_i}{k_i(k_i-1)}, \tag{6}$$

where $\mathcal{E}_i$ is the actual number of edges that exist between each of the neighbors.

3. Methods

3.1. Simulation

Prior work on network robustness focuses on relatively small numbers of networks due to the limited number of real networks for which data is available [12,15–20]. However, this significantly limits the statistical strength of the insights that can be drawn from the analysis. To overcome this limitation, we begin by randomly generating 2000 networks with degree distributions following a power-law with exponential cutoff and distribution parameters representative of scale-free networks in a variety of domains [13]. Our algorithm is a variation on preferential attachment and is provided in the Appendix.

This algorithm is not guaranteed to produce a connected network, so after generating a network we check to see if it is fully connected using a breadth-first search. If the network is not connected, we discard it and try again. For most degree distributions and parameters we are able to generate a connected network in a very small number (<5) of attempts. We select five pairs of distribution parameters (Table 3) to represent realistic networks based on the network data presented in [13] (Table 2). We generate 400 random networks for each parameter combination: 20 networks for each of 20 sizes (Table 4). The network sizes between 100 and 1000 are generated from a uniform random distribution.