J Comb Optim
DOI 10.1007/s10878-015-9893-7

Robustness of power-law networks: its assessment and optimization

Huiling Zhang[1] · Yilin Shen[1] · My T. Thai[1,2]

Abstract Many practical complex networks, such as the Internet, WWW and social networks, are discovered to follow power-law distribution in their degree sequences, i.e., the number of nodes with degree i in these networks is proportional to $i^{-\beta}$ for some exponential factor $\beta > 0$. However, these networks also expose their vulnerabilities to a great number of threats such as adversarial attacks on the Internet, cyber-crimes on the WWW or malware propagations on social networks. Although power-law networks have been found robust under random attacks and vulnerable to intentional attacks via experimental observations, how to better understand their vulnerabilities from a theoretical point of view still remains open. In this paper, we study the vulnerability of power-law networks under random attacks and adversarial attacks using the in-depth probabilistic analysis on the theory of random power-law graph models. Our results indicate that power-law networks are able to tolerate random failures if their exponential factor β is $<$2.9, and they are more robust against intentional attacks if β is smaller. Furthermore, we reveal the best range [1.8, 2.5] for the exponential factor β by optimizing the complex networks in terms of both their vulnerabilities and costs.

Huiling Zhang and Yilin Shen are co-first authors.

✉ My T. Thai
thaitramy@tdt.edu.vn; mythai@cise.ufl.edu

Huiling Zhang
huiling@cise.ufl.edu

Yilin Shen
yshen@cise.ufl.edu

[1] Department of Computer & Information Science and Engineering, University of Florida, Gainesville, FL 32611, USA

[2] Division of Algorithms and Technologies for Networks Analysis, Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam

Published online: 14 May 2015

When $\beta < 1.8$, the network maintenance cost is very expensive, and when $\beta > 2.5$ the network robustness is unpredictable since it depends on the specific attacking strategy.

Keywords Power-law networks · Robustness · Probabilistic analysis · Optimization

1 Introduction

One of the most remarkable discoveries in many real-world networks is the power-law distribution in their degree sequences, ranging from the Internet, WWW, biological networks to social networks (Faloutsos et al. 1999; Mayo et al. 2015). In particular, the number of nodes with degree i in these complex networks is observed to be proportional to $i^{-\beta}$ for some exponential factor $\beta > 0$. Thus, β is the key parameter that will determine the degree distributions and consequently, will decide how the networks would look like and how they would function. However, recent studies have revealed that power-law distribution appears to be vulnerable to a great number of threats such as adversarial attacks on the Internet, cybercrimes on the WWW or malware propagations on social networks (Albert et al. 2004; Yan et al. 2011). Therefore, a better knowledge about this power-law property would not only guarantee the normal network function but also reduce the maintenance cost (i.e., the cost to construct links, the cost to protect critical nodes, etc.), which are essential factors of networking problems.

Most studies investigating this power-law property have been focused on how such degree heterogeneity nature can impact the robustness of networks (Albert et al. 2004, 2000; Holme et al. 2002), or how one can quickly and efficiently generate an ideal power-law network with a given degree sequence (Aiello et al. 2000). Focusing on the security factor, the works (Albert et al. 2000; Cohen et al. 2000; Holme et al. 2002; Satorras and Vespignani 2002) have empirically shown that power-law networks appear robust under random attacks and vulnerable to intentional attacks via experimental observations. Nevertheless, there are several important security aspects of this property that are left untouched. For instance, are power-law networks surely more vulnerable to intentional attacks than random failures? How can we accurately assess the robustness of power-law networks under various kinds of threat, e.g., random failure and adversarial attack? Can we design more stable and robust power-law networks by adjusting the parameter β?

Regarding these issues, one of the most crucial question is which *measure* coping with the network vulnerability the best? There have been many studies proposing different metrics to account for the network vulnerability (Albert et al. 2000, 2004; Luciano et al. 2007; Matisziw and Murray 2009; Ventresca and Aleman 2015), among which the degree of suspected nodes or edges, the average shortest path length, the global clustering coefficients, the available number of compromised $s{-}t$ flows, the diameters, the variance among the size of connected components, the relative size of the largest cluster and the average size of the isolated clusters appear to be the most popular and effective. Unfortunately, these measures do not seem to cast well for some particular kinds of network vulnerabilities, especially when network fragmentation is of high priority, as depicted in Fig. 1.

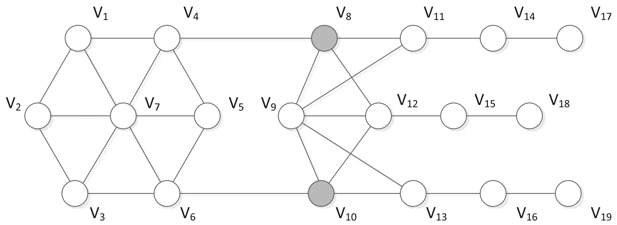

Fig. 1 An example of internet: the removal of v_8 and v_{10} (*grey nodes*) is sufficient to destroy the function of the whole network such that only <40 % nodes connect each other

Let us consider a simple example in Fig. 1 illustrating a small portion of the Internet, where nodes $v_1, v_2, \ldots, v_7$ are ISPs and the rest are consumers or transmission nodes. As revealed in this figure, any successful corruptive attacks to nodes v_8 and v_{10} are sufficient to bring the whole network down to its knees with no satisfied customers. In a different attacking strategy, the removal of node v_7 or v_9, if the adversary was to use maximum degree centrality as the metric, does not appear to harm the network function because all customers are still satisfied. These removals also reduce the global clustering coefficients to 0 and increase the average shortest path to nearly 3. Besides, if the attacker uses the available number of compromised flows from v_1 to v_2, the destructions of nodes v_4 and v_7 will drop the flow to 1, and they still, unfortunately, cannot destroy the existence of the giant ISP component providing services to the (almost) whole network.

This example illustrates an important point that the other metrics are lack of. In order to break down the network, we need to somehow control the balance among disconnected components while ensuring the nonexistence of giant components. One possible and effective way to do so is to measure the *total pairwise connectivity* ($\mathbb{P}$), i.e. the number of connected node-pairs (Dinh et al. 2011) in the network. Back to our example, a scrutiny look into the destructions of nodes v_8 and v_{10}, which we know can break down the network's function, reveals that they, indeed, reduce the network total pairwise connectivity to its greatest extent (more than 60 %). This great reduction, as a result, significantly malfunctions the whole network. The measure $\mathbb{P}$ also lends itself effectively a lot of practical network applications. As we discussed above, since many large-scale networks have been shown to be power-law networks, the removal of critical nodes and links regarding this metric not only reduces the network performance but also can possibly disconnect those networks from the outside world. Another application of this metric can be found in destroying terrorist networks, e.g. to breakdown the communication between any two terrorist individuals to the greatest extent, as well as protecting the functionality in communication networks.

Another limitation of these prior works is their heavy dependence on the experiments and failures to optimize the power-law networks. In other words, we cannot apply them to enhance the robustness of power-law networks, and in the meanwhile reduce their costs. To our best knowledge, this work is the first attempt from a theoretical point of view targeting in the two objectives mentioned above: (1) assess the impact of random and intentional attacks on power-law networks; (2) optimize power-law networks based on their toleration on threats and maintenance costs, which are used to guarantee the network functionality and reliability.

1.1 Our contributions

Using the well-accepted power-law random graph model (Aiello et al. 2000), we did an in-depth analysis using probability theory with respect to different kinds of threats: random failures, preferential attacks and degree-centrality attacks. Our significant conclusions are (1) a complex network can tolerate random failures if its exponential factor is <2.9, (2) power-law networks are more robust under preferential node attacks and degree-centrality node attacks when they have smaller exponential factor β, and (3) in order to maintain a reliable complex system, we optimize the power-law networks by investigating on the optimal range of exponential factor β beforehand. For both communication networks and social networks, the best β is illustrated to be lying in the interval $[1.8, 2.5]$, which gives a decent explanation to the structures of real-world networks (Faloutsos et al. 1999). When $\beta < 1.8$, the maintenance of network is very costly, and when $\beta > 2.5$, the network vulnerability is unpredictable due to its dependence on the specific attacking strategy.

1.2 Organizations

In Sect. 2, we introduce the random power-law model, and attack taxonomy. Some results in the literature and our fundamental results on intact power-law networks are presented in Sect. 3. The analysis of the robustness of power-law networks under random failure, preferential attacks and degree-centrality attacks are proposed in Sects. 4 and 5 and Sect. 6 respectively. Section 7 focuses on the optimization of power-law networks. Sections 4, 5, 6, 7 also provide subsections to illustrate the good range of exponential factor β under different attacks and with respect to the network optimizations. Related works are introduced in Sects. 8 and 9 concludes the whole paper.

2 Models and threat taxonomy

Notations used in the paper are given as follows:

β	Exponential factor of PLRG
$\mathbb{P}$	Total pairwise connectivity
$G_{(\alpha,\beta)}$	(α, β) graph
n	The number of nodes of $G_{(\alpha,\beta)}$
m	The number of edges of $G_{(\alpha,\beta)}$
Δ	Maximum degree of $G_{(\alpha,\beta)}$
$\mathbf{d}$	Degree sequence of $G_{(\alpha,\beta)}$
$\overline{d}$	Expected average degree of $G_{(\alpha,\beta)}$
$\tilde{d}$	Second-order average degree of $G_{(\alpha,\beta)}$
G_r, G_p, G_c	The residual graph of $G_{(\alpha,\beta)}$ after random failure, preferential attack and degree-centrality attack respectively
$\mathbf{d}_r, \mathbf{d}_p, \mathbf{d}_c$	Expected degree sequences of G_r, G_p, G_c
y_i^r, y_i^p, y_i^c	Expected number of nodes of degree i in G_r, G_p, G_c
p	The probability of random failure

p_i	The probability that a node of degree i is attacked in preferential attack
c	The expected number of failure nodes in expected preferential attack
β'	Exponential factor in interactive preferential attack
G_{cp}	The residual graph of $G_{(\alpha,\beta)}$ only consisting of the protected degree-centrality nodes
G_s	The residual graph of $G_{(\alpha,\beta)}$ after immunizing individuals in S
$\overline{d}_p, \overline{d}_c, \overline{d}_s$	Eexpected average degree of G_p, G_c, G_s
$\tilde{d}_p, \tilde{d}_c, \tilde{d}_s$	Second-order average degree of G_p, G_c, G_s

This following consists of two parts: (1) the power-law random graph (PLRG) model, on which our analysis is based throughout the whole paper and (2) threat taxonomy, including random failures and intentional attacks.

2.1 Power-law random graph (PLRG) model

There are two main categories of random graph models to generate graphs with skewed degree sequences, evolutionary and structural. *Evolutionary* models lead to the skewed degree distributions by identifying growth primitives, including multi-objective optimization and statistical preferential attachment (Barabasi and Albert 1999). Despite its advantages to explore additional network semantics, the tight dependencies between iterations in evolutionary models bring the biggest obstacle in the probabilistic analysis (Bollobás et al. 2001). *Structural models*, on the other hand, start with a given skewed degree distribution (e.g., a power-law distribution based on the degree sequences of a real-world network) and generate a graph with the degree sequence, satisfying certain randomness properties (Aiello et al. 2000; Gkantsidis et al. 2003). The greatest advantage of such structural models is their tractability to theoretical analysis, due to its discard of dependencies in evolutionary models by taking skewed degree sequences (Aiello et al. 2000; Chung and Lu 2002). Although the term configuration is used, a lot of mathematicians also noted this advantage by exploiting several properties in structural random graph models (Molloy and Reed 1995, 1998). Therefore, in this paper, we use the well-accepted structural PLRG model in (Aiello et al. 2000) in order to explore the robustness of power-law networks from an in-depth theoretical perspective.

First, we consider the following random graph, (α, β) *graph* $G_{(\alpha,\beta)}$, with its power-law degree distribution depending on two given values α and β.

Definition 1 ((α, β) **Graph** $G_{(\alpha,\beta)}$) Given an undirected graph $G = (V, E)$ having $|V| = n$ nodes and $|E| = m$ edges, it is called a (α, β) power-law graph if its maximum degree is $\varDelta = \left\lfloor e^{\alpha/\beta} \right\rfloor$ and the number of nodes with degree i is

$$y_i = \left\lfloor \frac{e^{\alpha}}{i^{\beta}} \right\rfloor . \tag{1}$$

For simplicity, we assume that there is no isolated nodes. Note that the number of nodes $n = e^{\alpha}\zeta(\beta) + O(n^{\frac{1}{\beta}} - 1)$ and the number of edges $m = \frac{1}{2}e^{\alpha}\zeta(\beta - 1) + O(n^{\frac{2}{\beta}} - 1)$, where $\zeta(\beta) = \sum_{i=1}^{\infty} \frac{1}{i^{\beta}}$ is the *Riemann Zeta function*. Since there is only a very small error $o(1)$ when $\beta > 2$ when counting the number of both nodes and edges, we simply denote them as $n \doteq e^{\alpha}\zeta(\beta)$ and edges $m \doteq \frac{1}{2}e^{\alpha}\zeta(\beta - 1)$. If the sum of degrees is odd, a node of degree 1 will be added.

Next, given the parameters α and β, the structural PLRG model is to construct a (α, β) power-law graph according to its degree sequence $\mathbf{d}$, which consists of a sequence of integers $(1, \ldots, 1, 2, \ldots, 2, \ldots, \Delta)$ where the number of i is equal to y_i defined in the above Definition 1.

Definition 2 (PLRG Model) Given $\mathbf{d} = (d_1, d_2, \ldots, d_n)$ be a sequence of integers

$$(1, \ldots, 1, 2, \ldots, 2, \ldots, \Delta)$$

where the number of i is equal to y_i, the PLRG model can be applied to generate a random power-law graph as follows: First, we introduce $D = \sum_{i=1}^{n} d_i$ mini-nodes lying in n clusters of each size d_i where $1 \leq i \leq n$. Then, a random perfect matching is constructed among the mini-nodes such that a graph on the n original nodes is constructed as suggested by this perfect matching in the natural way: two original nodes are connected by an edge if and only if at least one edge in the random perfect matching connects the mini-nodes of their corresponding clusters.

2.2 Threat taxonomy

In this paper, we focus on investigating the robustness of power-law networks under random failure and two types of intentional attacks, i.e. preferential attack and degree-centrality attack.

Definition 3 (Random failure) Each node in $G_{(\alpha,\beta)}$ fails randomly with the same probability.

Definition 4 (Preferential attack) Each node in $G_{(\alpha,\beta)}$ is attacked with higher probability if it has a higher degree.

Definition 5 (Degree-centrality attack) The adversary only attacks the set of degree-centrality nodes in $G_{(\alpha,\beta)}$.

2.3 Notation explanation

With respect to each threat, we define the residual networks of the power-law network $G_{(\alpha,\beta)}$ as G_r, G_p and G_c after the occurrence of random failure, preferential attack and degree-centrality attack respectively. Their corresponding expected degree sequences are denoted as $\mathbf{d}_r$, $\mathbf{d}_p$ and $\mathbf{d}_c$, where the number of d_i^r, d_i^p and d_i^c are referred to as y_i^r, y_i^p and y_i^c.

In addition, we define a power-law network under certain threats to be *highly-connected* if a.s. its pairwise connectivity $\mathbb{P} = \Theta(n^2)$ and *lowly-connected* otherwise.

3 Preliminaries

In this section, we first present some useful results in the literature, which illustrate the important relations between the size of largest connected components and the degree sequence in random networks. Based on them, we then derive some fundamental results to evaluate the robustness of power-law networks. In this paper, the size of a connected component $S \subseteq G$ is the total number of nodes in S and the connected component S is called giant component if its size is $\Theta(n)$.

3.1 Previous works

Lemma 1 (Molloy and Reed 1995) *In a random graph G with the maximum degree Δ, $\lambda_i n$ nodes are of degree i where $\sum_{i=1}^{\Delta} \lambda_i = 1$. The giant components exist in G when*

$$Q = \sum_{i=1}^{n} i(i-2)\lambda_i > 0 \tag{2}$$

and there exists $\epsilon > 0$ such that $\Delta < n^{1/4-\epsilon}$. Otherwise, there is a.s. no giant component in G when $Q < 0$ and $\Delta < n^{1/8-\epsilon}$.

Lemma 2 (Chung and Lu 2002) *In a random graph G with degree sequence $\mathbf{d} = (d_1, d_2, \ldots, d_n)$, the giant component a.s. exists if its expected average degree $\overline{d}$ is at least 1, and there is a.s. no giant component if its expected second-order average degree $\tilde{d}$ is at most 1. Furthermore, all connected components have volume (the sum of degrees in a connected component) at most $\sqrt{n}\log n$ with probability at least $1 - o(1)$ if $\tilde{d} < 1$. Here the expected average degree $\overline{d}$ and second-order average degree $\tilde{d}$ are defined as*

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i, \quad \tilde{d} = \frac{\sum_{i=1}^{n} d_i^2}{\sum_{i=1}^{n} d_i} \tag{3}$$

where d_i is the elements in the degree sequence.

Corollary 1 *All connected components a.s. have sizes at most $\frac{1}{2}\sqrt{n}\log n + 1$ if $\tilde{d} < 1$.*

Proof Consider a connected component S, the volume of S is defined as $\mathsf{Vol}(S) = \sum_{v_i \in S} d_i$. Since there are at least $|S| - 1$ edges in a connected component of size $|S|$, we have $2(|S| - 1) \leq \mathsf{Vol}(S) \leq \sqrt{n}\log n$. Therefore, the size of S is upper bounded by $\frac{1}{2}\sqrt{n}\log n + 1$. □

3.2 Robustness of intact power-law networks

Theorem 1 *For a power-law network represented as a (α, β) graph $G_{(\alpha,\beta)}$,*

- *If $\beta < 3.47875$, the pairwise connectivity $\mathbb{P}$ is $\Theta(n^2)$;*
- *If $\beta \geq 3.47875$, the range of pairwise connectivity $\mathbb{P}$*

is a.s. at most $\frac{1}{2}n\left(c(\beta)n^{\frac{2}{\beta}}\log n - 1\right)$.

where $c(\beta) = 16/\left[\zeta(\beta)\left(2 - \frac{\zeta(\beta-2)}{\zeta(\beta-1)}\right)\right]^2$ *is a constant on any given* β.

To prove Theorem 1, we first show the relation between the largest component and our metric, the total pairwise connectivity, in the following lemma.

Lemma 3 *Suppose that the maximum size of a connected component in the graph* $G = (V, E)$ *is* ℓ*, the pairwise connectivity* $\mathbb{P}$ *is then at most* $\frac{n(\ell-1)}{2}$.

Proof To prove the upper bound, we consider the worst case that the whole network consists of all connected components of size ℓ except some leftover nodes. Suppose that there are c_1 connect components of size ℓ and the number of leftover nodes is c_2, we have $n = c_1\ell + c_2$. Therefore, the pairwise connectivity $\mathbb{P}$ is

$$\mathbb{P} \le c_1\binom{\ell}{2} + \binom{c_2}{2} \le c_1\binom{\ell}{2} + \frac{c_2}{\ell}\binom{\ell}{2} = \frac{c_1\ell + c_2}{\ell}\binom{\ell}{2} = \frac{n(\ell-1)}{2}.$$

□

Proof of Theorem 1:

Proof First, according to F. Chung *et al.* (Aiello et al. 2000), we can find the threshold 3.47875 of β such that $Q > 0$ when $\beta < 3.47875$ and $Q < 0$ when $\beta > 3.47875$.

When $\beta < 3.47875$, according to Lemma 1, since $Q > 0$, there exists one giant component of size $\Theta(n)$. Therefore, the pairwise connectivity $\mathbb{P}$ is $\Theta(n^2)$.

When $\beta > 3.47875$, according to Aiello *et al.* (Aiello et al. 2000), a connected component S in the (α, β) graph a.s. has the size at most $c(\beta)n^{\frac{2}{\beta}} \log n$. Then the upper bound of $\mathbb{P}$ follows directly from Lemma 3. □

In the following three sections, since the power-law networks with β at least 3.47875 are lowly-connected even if they are not attacked, we will focus on exploiting the robustness of power-law networks with $\beta < 3.47875$ under random failures, preferential attacks and degree-centrality attacks respectively.

Note that these results are very important in the following analysis of pairwise connectivity. In particular, if there exist giant components in residual networks after attacks, we know that residual networks are with pairwise connectivity $\mathbb{P} = \Theta(n^2)$ and thus are able to conclude that corresponding original graphs are robust under attacks. We also derive upper bounds on pairwise connectivity based on the expected degree sequence in residual networks where there are no giant components.

4 Random failures

In this section, we focus on the robustness of power-law networks after random failures, in which each node has the same probability p $(0 < p < 1)$ to fail. The total pairwise connectivity $\mathbb{P}$ in the residual graph G_r is proven as in the following Theorem 2. Based on this theorem, we further investigate the good range of exponential factor β.

4.1 Robustness under random failures

Theorem 2 *In a residual graph G_r of $G_{(\alpha,\beta)}$ after random failures,*

- *If $\beta < \beta_p$, the expected pairwise connectivity $E(\mathbb{P})$ is a.s. $\Theta(n^2)$;*
- *If $\beta \geq \beta_p$, the pairwise connectivity $\mathbb{P}$ is a.s. at most $\frac{1}{2}n\left(c_r(\beta)n^{\frac{2}{\beta}}\log n - 1\right)$.*

where β_p satisfies that $(1-p)\zeta(\beta_p - 2) - (2-p)\zeta(\beta_p - 1) = 0$ and

$$c_r(\beta) = 16/\left[\zeta(\beta)\left(2 - p - (1-p)\frac{\zeta(\beta-2)}{\zeta(\beta-1)}\right)\right]^2.$$

To prove Theorem 2, we first show the expected degree distribution in G_r as follows.

Lemma 4 *The expected degree distribution of graph G_r is*

$$E(y_i^r) = (1-p)^{i+1}\sum_{k=i}^{\Delta}\binom{k}{i}\frac{e^\alpha}{k^\beta}p^{k-i},$$

where degree i is $1 \leq i \leq \Delta$.

Proof Let p_k^i be the probability that a node v of degree k in $G_{(\alpha,\beta)}$ has its degree to be i in G_r. When $k < i$, it is clear that $p_k^i = 0$; otherwise when $k \geq i$, v will become a node of of degree i in G_r if and only if v itself does not fail but $k-i$ of its neighbors fail. Hence, the probability p_k^i is $\binom{k}{i}(1-p)[p^{k-i}(1-p)^i]$, i.e. $\binom{k}{i}p^{k-i}(1-p)^{i+1}$.

Thus, according to the basic definition of expected value, the expected number of nodes of degree i in G_r is

$$\mathsf{E}(y_i^r) = \sum_{k=1}^{\Delta}p_k^i\frac{e^\alpha}{k^\beta} = (1-p)^{i+1}\sum_{k=i}^{\Delta}\binom{k}{i}\frac{e^\alpha}{k^\beta}p^{k-i}.$$

□

Proof of Theorem 2

Proof First of all, we show that Lemma 2 cannot be applied here. Consider the expected second-order average degree $\tilde{d}_r$ of G_r, we have

$$\tilde{d} = p + (1-p)\frac{\zeta(\beta-2)}{\zeta(\beta-1)}.$$

It is easy to see that $\tilde{d} > 1$ for any p and β.

In an alternative way, we use Lemma 1 and branching process method to prove our theorem. *The basic idea is as follows*: according to the expected degree of G_r, we first find a threshold β_p using Lemma 1, which determines whether the total pairwise connectivity $\mathbb{P}$ of the residual network G_r is a.s. $\Theta(n^2)$ or not. If not, that is,

$\beta > \beta_p$, we further use branching process method to prove that $\mathbb{P}$ in G_r is a.s. at most $\frac{1}{2}n\left(c_r(\beta)n^{\frac{2}{\beta}}\log n - 1\right)$. First, we compute β_p for G_r as

$$Q_r = \sum_{i=1}^{\Delta} i(i-2)(1-p)^{i+1} \sum_{k=i}^{\Delta} \binom{k}{i} \frac{e^{\alpha}}{k^{\beta}} p^{k-i} \tag{4a}$$

$$= e^{\alpha}(1-p) \sum_{i=1}^{\Delta} \frac{1}{i^{\beta}} \sum_{j=1}^{i} j(j-2) \binom{i}{j} p^{i-j}(1-p)^{j} \tag{4b}$$

$$= e^{\alpha}(1-p)^2 \sum_{i=1}^{\Delta} \frac{i^2(1-p) - i(2-p)}{i^{\beta}} \tag{4c}$$

$$\doteq e^{\alpha}(1-p)^2 \left[(1-p)\zeta(\beta-2) - (2-p)\zeta(\beta-1)\right], \tag{4d}$$

where step (4c) follows similarly from the calculation of expected value and variance in binomial distribution.

Let us consider the case that the threshold β_p satisfies $(1-p)\zeta(\beta-2) - (2-p)\zeta(\beta-1) = 0$. When $\beta < \beta_p$, we have $Q_r > 0$. Thus, the expected pairwise connectivity $E(\mathbb{P})$ is a.s. $\Theta(n^2)$ according to Lemma 1.

```
$i \leftarrow 0$;
$E_0 = L_0 = \{v\}$ by picking an arbitrary node $v$;
while $|L_i| \neq 0$ do
    $i \leftarrow i + 1$;
    Choose an arbitrary $u$ from $L_{i-1}$ and expose all its neighbors $N(u)$;
    $E_i = E_{i-1} \cup N(u)$;
    $L_i = (L_i \setminus (\{u\}) \cup (N(u) \setminus E_{i-1})$;
end
```

Algorithm 1: Branching process method

When $\beta > \beta_p$, we use the following branching process method (Algorithm 1) on G_r according to its expected degree sequence $E(y_i^r)$. In the algorithm, we define E_i and L_i as the set of exposed nodes and live nodes in iteration i respectively, where live nodes are referred to as the subset of exposed nodes whose neighbors have not been exposed. Note that $|L_i| = 0$ if and only if the entire component is exposed. For simplicity, we define random variables $\mathcal{E}_i = |E_i|$ and $\mathcal{L}_i = |L_i|$ as the number of exposed nodes and live nodes. Let $\mathcal{T}$ denote the whole number of iterations in branching process, that is, $\mathcal{T}$ also measures the size of connected component since exactly one node is exposed in each iteration. We further define an edge to be a "backedge" if it connects u and some node in E_{i-1}. We denote $D_i = |N(u)|$ and $B_i = |N(u) \cap E_{i-1}| - 1$ measuring the degree of the node exposed in iteration i and the number of "backedge". By definition, we have $\mathcal{L}_i - \mathcal{L}_{i-1} = D_i - B_i - 2$ immediately.

Then, we calculate $\mathsf{E}(D_i)$, $\mathsf{E}(B_i)$ and $\mathsf{E}(\mathcal{L}_i)$ respectively. Consider one edge in original graph $G_{(\alpha,\beta)}$. It still exists iff both endpoints are not failed, that is, the expected number of edges in G_r is $(1-p)^2 m$. Therefore,

$$\begin{aligned}\mathsf{E}(D_i) &= \sum_{i=1}^{\Delta} i \frac{i(1-p)^{i+1 \sum_{k=i}^{\Delta} \binom{k}{i} \frac{e^{\alpha}}{k^{\beta}} p^{k-i}}}{(1-p)^2 m} \\ &= \frac{1}{\zeta(\beta-1)} \sum_{i=1}^{\Delta} \frac{i^2(1-p)+ip}{i^{\beta}} \\ &\doteq (1-p)\frac{\zeta(\beta-2)}{\zeta(\beta-1)} + p.\end{aligned}$$

Since $|N(u) \cap E_{i-1}| \geq 1$, we have $\mathsf{E}(B_i) \geq 0$. By substituting $\mathsf{E}(D_i)$ and $\mathsf{E}(B_i)$ into $\mathcal{L}_i - \mathcal{L}_{i-1} = D_i - B_i - 2$, we have

$$\begin{aligned}\mathsf{E}(\mathcal{L}_i) &= \mathcal{L}_1 + \sum_{j=2}^{i} \mathsf{E}(\mathcal{L}_j - \mathcal{L}_{j-1}) \\ &= d_0 + \sum_{j=2}^{i} \mathsf{E}(D_j - B_j - 2) \\ &\leq d_0 + (i-1)\left((1-p)\frac{\zeta(\beta-2)}{\zeta(\beta-1)} + p - 2\right) \\ &= d_0 - \lambda(p,\beta)(i-1)\end{aligned}$$

where $\lambda(p,\beta) = 2 - p - (1-p)\frac{\zeta(\beta-2)}{\zeta(\beta-1)}$ and the initial node is assumed to have degree d_0.

Since $|\mathcal{L}_j - \mathcal{L}_{j-1}| \leq \Delta = e^{\frac{\alpha}{\beta}}$, according to *Azuma's Martingale Inequality* (Chung and Lu 2006),

$$\mathsf{Pr}\left[|\mathcal{L}_i - E(\mathcal{L}_i)| > \mathcal{T}\right] \leq 2e^{\frac{-\mathcal{T}^2}{2ie^{\frac{2\alpha}{\beta}}}}$$

where $i = \frac{16}{(\lambda(p,\beta))^2} e^{\frac{2\alpha}{\beta}} \log n = c_r(\beta) n^{\frac{2}{\beta}} \log n$ and $\mathcal{T} = \lambda(p,\beta)i/2$. Since we know

$$\mathsf{E}(\mathcal{L}_i) + \mathcal{T} \leq d_0 - \lambda(p,\beta)(i-1) + \frac{\lambda(p,\beta)}{2} i < 0$$

for any d_0. Therefore,

$$\begin{aligned}&\mathsf{Pr}\left[\mathcal{T} > \frac{16}{(\lambda(p,\beta))^2} e^{\frac{2\alpha}{\beta}} \log n\right] = \mathsf{Pr}[\mathcal{T} > i] \\ &\leq \mathsf{Pr}[\mathcal{L}_i > 0] \leq \mathsf{Pr}[\mathcal{L}_i > \mathsf{E}(\mathcal{L}_i) + \mathcal{T})] \\ &\leq 2e^{\frac{-\mathcal{T}^2}{2ie^{\frac{2\alpha}{\beta}}}} = \frac{2}{n^2}.\end{aligned}$$

Thus, the probability that, in graph G_r, there is a non-failure node v belonging to a connected component of size larger than $c_r(\beta) n^{\frac{2}{\beta}} \log n$ is at most $n\frac{2}{n^2} = o(1)$, i.e.

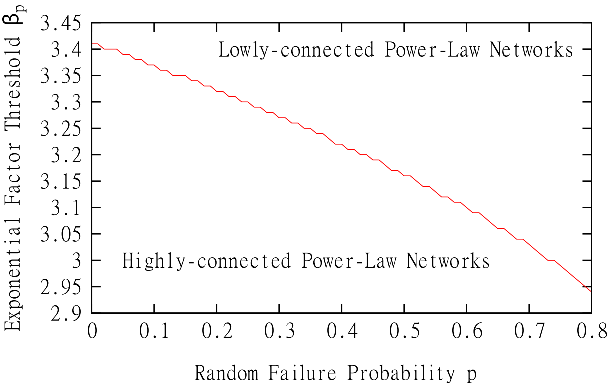

Fig. 2 Relation between threshold β_p and failure probability p

G_r has the largest connected component of size a.s. $c_r(\beta)n^{\frac{2}{\beta}} \log n$. Hence, the upper bound of pairwise connectivity in G_r follows from Lemma 3 directly. □

4.2 Good range of β under random failures

According to Theorem 2, we exploit the good range of exponential factor β in terms of the pairwise connectivity $\mathbb{P}$ of power-law networks. By obtaining the threshold β_p from $(1-p)\zeta(\beta_p-2)-(2-p)\zeta(\beta_p-1)=0$, the relation between threshold β_p and failure probability p can be revealed in the following Fig. 2.

Based on Theorem 2, power-law networks are highly-connected when $\beta > \beta_p$ and lowly-connected otherwise. As one can see from Fig. 2, power-law networks of exponential factor $\beta > 2.9$ will still remain highly-connected under random failures even when the failure probability p is unrealistically 0.8. That is, we can confidently claim that power-law networks have an extremely high tolerance to random failures when its exponential factor $\beta < 2.9$.

5 Preferential attacks

As power-law networks are tolerable to random failures, one will question whether it can still tolerate intentional attacks if the intruders intend more to attack "hub" nodes. In this section, we focus on the robustness of power-law networks under preferential attacks. As we defined above, in preferential attacks, each node in the network is attacked with higher probability if it has larger degree. Therefore, consider the costs to attack for intruders, we focus on the following two preferential attack schemes:

- *Interactive Preferential Attacks:* one way to control the costs to attack is to attack a node w.r.t. its degree and a new parameter β'. That is, a node of degree i is attacked with probability $1-\frac{1}{i^{\beta'}}$;
- *Expected Preferential Attacks:* another way to control the costs to attack is based on the expected number of nodes c to attack. When the intruder decides

c, ranging between 0 and $e^\alpha \zeta(\beta)$, a node of degree i is attacked with probability $p_i = c\frac{i}{e^\alpha \zeta(\beta-1)}$ since the expected number of failure nodes is equal to c, namely $\sum_i \frac{e^\alpha}{i^\beta} p_i = c$.

As one can see, in both these schemes, a node of higher degree, often referred to as a "hub", is *more preferentially* attacked, that is, it has higher probability to be attacked. By denoting their corresponding residual graphs as G_p^I and G_p^E, their total pairwise connectivity are proven in Theorems 3 and 4 respectively.

5.1 Interactive preferential attacks $\left(p_i = 1 - \frac{1}{i^{\beta'}}\right)$

Theorem 3 *In a residual graph G_p^I of $G_{(\alpha,\beta)}$ after interactive preferential attacks,*

- *If $\beta + \beta' < 3.47875$, the expected pairwise connectivity $E(\mathbb{P})$ is $\Theta(n^2)$;*
- *If $\beta + \beta' \geq 3.47875$, the pairwise connectivity $\mathbb{P}$ is a.s. at most $\frac{1}{2}n\left(c(\beta)n^{\frac{2}{\beta}}\log n\right.$ $\left.-1\right)$,*

where $c(\beta) = 16/\left[\zeta(\beta)\left(2 - \frac{\zeta(\beta-2)}{\zeta(\beta-1)}\right)\right]^2$ is a constant on any given β.

Theorem 3 can be proven in the same way as in Theorem 1 after showing the expected degree in residual graph G_p^I as in Lemma 7, which is based on the following two lemmas.

Lemma 5 *In graph $G_{(\alpha,\beta)}$, the probability that a node v of degree i incident to another node u of degree x is $\frac{ix}{e^\alpha \zeta(\beta-1)}$.*

Proof Consider a node v of degree i, in the matching of mini-nodes, at least one of i mini-nodes for v connects to another one of x for node u of degree x. Thus, we have

$$\frac{\binom{i}{1}\binom{x}{1}f(N-2)}{f(N)} = \frac{ix}{N-1} = \frac{ix}{N} + O(\frac{1}{N^2}) \doteq \frac{ix}{e^\alpha \zeta(\beta-1)}$$

where $f(n) = (n-1)!!$ representing the number of perfect matchings for N nodes and $N = e^\alpha \zeta(\beta-1)$ denotes the number of mini-nodes. □

Lemma 6 *For a node v of degree i, the expected number of non-failure neighbors $E(N_p^I(i))$ of v is $i\frac{\zeta(\beta+\beta'-1)}{\zeta(\beta-1)}$.*

Proof According to Lemma 5, node v has probability $\frac{ix}{e^\alpha \zeta(\beta-1)}$ to connect to node u of degree x. Since node u of degree x has the non-failure probability $\frac{1}{x^{\beta'}}$, then we have the expected non-failure neighbor of v to be

$$\mathsf{E}(N_p^I(i)) \doteq \sum_{x=1}^{\Delta} \frac{ix}{e^\alpha \zeta(\beta-1)} \frac{1}{x^{\beta'}} \frac{e^\alpha}{x^\beta} \doteq i\frac{\zeta(\beta+\beta'-1)}{\zeta(\beta-1)}.$$

□

Lemma 7 *The expected degree distribution of graph G_p^I is*

$$\mathsf{E}(y_i^p) \doteq \frac{e^\alpha}{\left(i\frac{\zeta(\beta-1)}{\zeta(\beta+\beta'-1)}\right)^{\beta+\beta'}}$$

where $i \in \left\{\frac{\zeta(\beta+\beta'-1)}{\zeta(\beta-1)}, \ldots, \Delta\frac{\zeta(\beta+\beta'-1)}{\zeta(\beta-1)}\right\}$.

Proof Consider the set of nodes with degree i in G_p, they are correspondent to the nodes of degree x in the original graph $G_{(\alpha,\beta)}$. Hence, the expected unattacked nodes in this set is $\frac{e^\alpha}{x^\beta}\frac{1}{x^{\beta'}} = \frac{e^\alpha}{x^{\beta+\beta'}}$. According to Lemma 6, we know the relation between i and x is $i \doteq x\frac{\zeta(\beta+\beta'-1)}{\zeta(\beta-1)}$. Therefore, we have the expected number of nodes of degree i in G_p^I to be $\frac{e^\alpha}{\left(i\frac{\zeta(\beta-1)}{\zeta(\beta+\beta'-1)}\right)^{\beta+\beta'}}$. □

5.2 Expected preferential attacks $\left(p_i = c\frac{i}{e^\alpha\zeta(\beta-1)}\right)$

Theorem 4 *In a residual graph G_p^E of $G_{(\alpha,\beta)}$ after expected preferential attacks,*

- *The pairwise connectivity $\mathbb{P}$ is a.s. $\Theta(n^2)$ if* $c < \min\left\{c \middle| \frac{\sum_x \frac{e^\alpha}{x^\beta}\left(1-\frac{cx}{e^\alpha\zeta(\beta-1)}\right)\left(x\left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right)\right)}{n-c} > 1\right\}$*;*
- *The pairwise connectivity $\mathbb{P}$ is a.s. at most $\frac{1}{4}n^{\frac{3}{2}}\log n$ if* $c > \max\left\{c \middle| \left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right)\frac{\zeta(\beta-2)-\frac{c\zeta(\beta-3)}{e^\alpha\zeta(\beta-1)}}{\zeta(\beta-1)-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)}} < 1\right\}$.

To prove Theorem 4, we again first show the expected degree distribution in G_p^E as follows.

Lemma 8 *For a node v of degree i, the expected number of non-failure neighbors $E(N_p^E(i))$ of v is* $i\left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right)$.

Proof According to Lemma 5, the node v has probability $\frac{ix}{e^\alpha\zeta(\beta-1)}$ to connect node u of degree x. Since node u of degree x has the non-failure probability $1-c\frac{x}{e^\alpha\zeta(\beta-1)}$, then we have the expected non-failure neighbor of v to be

$$\mathsf{E}(N_p(i)) \doteq \sum_{x=1}^{\Delta}\frac{ix}{e^\alpha\zeta(\beta-1)}\left(1-\frac{cx}{e^\alpha\zeta(\beta-1)}\right)\frac{e^\alpha}{x^\beta} \doteq i\left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right).$$

The proof is complete. □

Corollary 2 *The expected degree distribution of graph G_p^E is*

$$\mathsf{E}(y_i^p) \doteq \frac{e^\alpha}{i^\beta}\left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right)^\beta\left(1-\frac{ci}{(e^\alpha\zeta(\beta-1))\left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right)}\right)$$

where $i \in \left\{\left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right), \ldots, \Delta\left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right)\right\}$.

Proof of Theorem 4

Proof In the proof, we first calculate the expected average degree $\overline{y}_p^E$ based on Corollary 2 as

$$\overline{y}_p^E \doteq \frac{\sum_{x=1}^{\Delta}\frac{e^\alpha}{x^\beta}\left(1-\frac{cx}{e^\alpha\zeta(\beta-1)}\right)\left(x\left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right)\right)}{n-c}$$

and second-order average degree $\tilde{y}_p^E$ as

$$\begin{aligned}\tilde{y}_p^E &\doteq \frac{\sum_{x=1}^{\Delta}\frac{e^\alpha}{x^\beta}\left(1-\frac{cx}{e^\alpha\zeta(\beta-1)}\right)\left(x\left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right)\right)^2}{\sum_{x=1}^{\Delta}\frac{e^\alpha}{x^\beta}\left(1-\frac{cx}{e^\alpha\zeta(\beta-1)}\right)\left(x\left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right)\right)}\\ &\doteq \left(1-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)^2}\right)\frac{\zeta(\beta-2)-\frac{c\zeta(\beta-3)}{e^\alpha\zeta(\beta-1)}}{\zeta(\beta-1)-\frac{c\zeta(\beta-2)}{e^\alpha\zeta(\beta-1)}}.\end{aligned}$$

According to Lemma 2 and Corollary 1, there exists one giant component if $\overline{y}_p^E > 1$ and all components have size at most $\frac{1}{2}\sqrt{n}\log n + 1$ if $\tilde{y}_p^E < 1$, then the proof follows from Lemma 3 directly. □

5.3 Relations between β and expected attacked nodes

In interactive preferential attacks, according to Theorem 3, a power-law networks with exponential factor β will be lowly-connected if the intruder select a β' such that $\beta + \beta' \geq 3.47875$. Since a node of degree i is attacked with probability $1 - \frac{1}{i^{\beta'}}$ in this scheme, this node can survive with probability $\frac{1}{i^{\beta'}}$. Therefore, we have the expected number of survived nodes as

$$\sum_i \frac{e^\alpha}{i^\beta}\frac{1}{i^{\beta'}} \doteq e^\alpha\zeta(\beta+\beta'),$$

that is, the expected percentage of attacked nodes can be obtained by calculating $1 - \frac{\zeta(\beta+\beta')}{\zeta(\beta)}$.

Figure 3 reports the relation between β and expected attacked nodes under iterative preferential attacks. We observed that the expected number of attacked nodes decreases

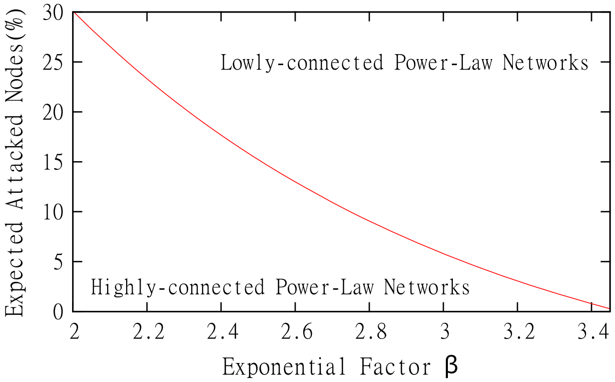

Fig. 3 Relation between β and attacked nodes under iterative preferential attacks

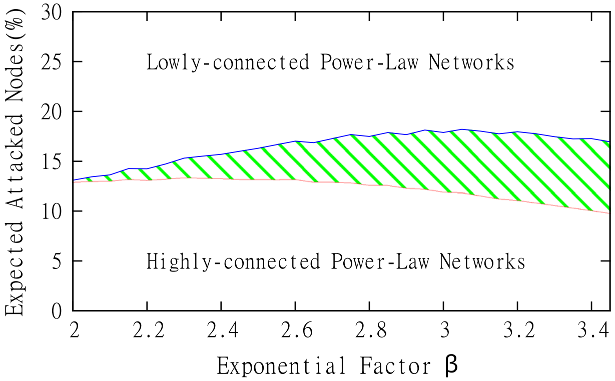

Fig. 4 Relation between β and attacked nodes under expected preferential attacks

sharply with the increase of β. Clearly, smaller β leads to a more robust power-law network.

In expected preferential attacks, again Fig. 4 reveals the smaller β the better. According to Theorem 4, except of uncertain areas (shadow areas), we can see that the percentage of attacked nodes (under the red line) reduces when β increases.

6 Degree-centrality attacks

As power-law networks is quite vulnerable under preferential attacks, their toleration to the deterministic intentional attacks attracts more attentions. Also, one can also question whether it is still true under deterministic intentional attacks that power-law networks with smaller β can better maintain their functionalities. In this section, we consider the degree-centrality attack, in which the intruders intentionally attack the "hubs", the set of nodes of highest degrees. When all nodes of degree larger than x_0 are attacked simultaneously, we have the following Theorem 5.

6.1 Robustness under degree-centrality attacks

Theorem 5 *In a residual graph G_c of $G_{(\alpha,\beta)}$ after degree-centrality attacks,*

- *The pairwise connectivity $\mathbb{P}$ is a.s. $\Theta(n^2)$ if $x_0 > \min\left\{x_0 \middle| \frac{1}{\zeta(\beta-1)} \frac{\left(\sum_{x=1}^{x_0} \frac{1}{x^{\beta-1}}\right)^2}{\sum_{x=1}^{x_0} \frac{1}{x^\beta}} > 1\right\}$;*
- *The pairwise connectivity $\mathbb{P}$ is a.s. at most $\frac{1}{4}n^{\frac{3}{2}} \log n$ if $x_0 < \max\left\{x_0 \middle| \frac{1}{\zeta(\beta-1)} \sum_{x=1}^{x_0} \frac{1}{x^{\beta-2}} < 1\right\}$.*

To prove Theorem 5, we again first show the expected degree distribution in G_c as follows.

Lemma 9 *For a node v of degree i in original graph $G_{(\alpha,\beta)}$, the expected number of neighbors of degree larger than x_0 is $\frac{i}{\zeta(\beta-1)} \sum_{i=x_0+1}^{\Delta} \frac{1}{i^{\beta-1}}$.*

Proof According to Lemma 6, the probability that a node v of degree i incident to a node u of degree x is $\frac{ix}{e^\alpha \zeta(\beta-1)}$. Therefore, we have the expected number of neighbors of degree larger than x_0 to be

$$\mathsf{E}(N_c(i)) \doteq \sum_{x=x_0+1}^{\Delta} \frac{ix}{e^\alpha \zeta(\beta-1)} \frac{e^\alpha}{x^\beta} = \frac{i}{\zeta(\beta-1)} \sum_{x=x_0+1}^{\Delta} \frac{1}{x^{\beta-1}}.$$

□

Corollary 3 *The expected degree sequence in G_c is*

$$E(y_i^c) \doteq \frac{e^\alpha}{i^\beta} \left(\frac{1}{\zeta(\beta-1)} \sum_{x=1}^{x_0} \frac{1}{x^{\beta-1}} \right)^\beta$$

where $i \in \left\{\frac{1}{\zeta(\beta-1)} \sum_{x=1}^{x_0} \frac{1}{x^{\beta-1}}, \ldots, \frac{x_0}{\zeta(\beta-1)} \sum_{x=1}^{x_0} \frac{1}{x^{\beta-1}}\right\}$.

Proof of Theorem 5

Proof With the expected average degree $\overline{y}_c$ as

$$\overline{y}_c = \frac{\sum_{x=1}^{x_0} \frac{e^\alpha}{x^\beta} x \left(\frac{1}{\zeta(\beta-1)} \sum_{i=1}^{x_0} \frac{1}{i^{\beta-1}}\right)}{n - \sum_{i=x_0+1}^{\Delta} \frac{e^\alpha}{i^\beta}} = \frac{1}{\zeta(\beta-1)} \frac{\left(\sum_{x=1}^{x_0} \frac{1}{x^{\beta-1}}\right)^2}{\sum_{x=1}^{x_0} \frac{1}{x^\beta}}$$

and second-order average degree $\tilde{y}_c$ as

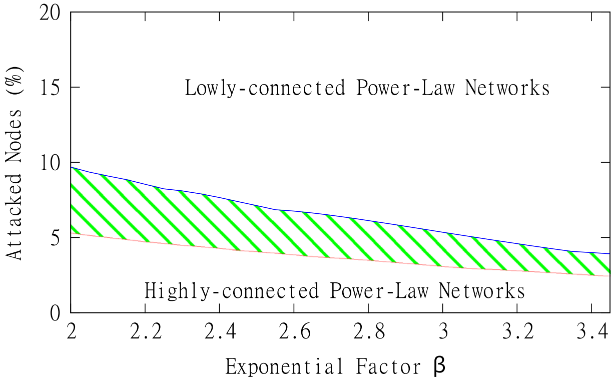

Fig. 5 Relation between β and attacked nodes under degree-centrality attacks

$$\tilde{y}_c = \frac{\sum_{x=1}^{x_0} \frac{e^\alpha}{x^\beta}\left[x\left(\frac{1}{\zeta(\beta-1)}\sum_{i=1}^{x_0}\frac{1}{i^{\beta-1}}\right)\right]^2}{\sum_{x=1}^{x_0}\frac{e^\alpha}{x^\beta}x\left(\frac{1}{\zeta(\beta-1)}\sum_{i=1}^{x_0}\frac{1}{i^{\beta-1}}\right)} = \frac{1}{\zeta(\beta-1)}\sum_{x=1}^{x_0}\frac{1}{x^{\beta-2}}.$$

The rest of proof is the same as Theorem 4. □

6.2 Relations between β and attacked nodes

Figure 5 illustrates the relations between β and attacked nodes under degree-centrality attacks based on Theorem 5. On the one hand, it is similar to expected preferential attacks that the percentage of attacked nodes (under the red line) reduces when β increases except of uncertain areas (shadow areas). On the other hand, under degree-centrality attacks, the intruder only needs to attack roughly 8 % less number of nodes to lower down the pairwise connectivity of power-law networks than under expected preferential attacks.

7 Optimization of power-law networks

The above vulnerability assessments give us a belief that power-law networks are more robust when β is smaller. However, a majority of real-world networks usually have their exponential factor β ranging from 2 to 2.5 rather than some small β approaching 1 or even less. The questions are intuitively raised: Is it better if real-world networks are denser such that they can be more robust? What causes them to be sparser than our expectation? Does there exist some potential optimization factors?

To address these questions, in this section, we investigate the tradeoff impact of maintenance costs and robustness guarantee on the power-law networks. In particular, we focus on the power-law networks with $\beta < 2.9$, which have been discovered to tolerate random failures to an extreme high degree. In addition, since Figs. 3, 4 and 5 already revealed that power-law networks can tolerate preferential attacks if they can tolerate degree-centrality attacks when $\beta < 2.9$, we focus on the guarantee of their functionality under degree-centrality attacks. We study the practical communication networks and social networks respectively to explore the underlying reasons of their real-world network topologies.

7.1 Communication networks

In the design of communication networks, such as the Internet, telecommunication networks and so on, we are required not only to guarantee their functionality but reduce the maintenance costs as well. Among various network performance metric, i.e., delay, packet loss, throughput, etc., the guarantee of its connectivity is of the high priority. That is, a real-world network only needs sufficient number of links to guarantee its functionality, and its other performance metric can be guaranteed by adjusting its capacity planning (Lakhina et al. 2004). In particular, we consider the costs including the link costs and the protection costs for critical nodes. Since the nodes with degree and betweenness centrality are closely correlated in non-fractal power-law networks (Kitsak et al. 2007), here we consider the critical nodes to be degree-centrality nodes.

To formulate the optimization function for power-law networks in communication networks, we first prove the following Lemma 10 by considering the worst case with respect to the robustness of power-law networks. That is, as mentioned above, after protecting the degree-centrality nodes, power-law networks a.s. remains highly-connected (its total pairwise connectivity is a.s. $\Theta(n^2)$) even though all other nodes are failed.

Lemma 10 *Let G_{cp} be the residual graph of $G_{(\alpha,\beta)}$ only consisting of the protected degree-centrality nodes (the nodes of degree larger than x_0), we have*

- *The pairwise connectivity $\mathbb{P}$ is a.s. $\Theta(n^2)$ if $x_0 < \max\left\{x_0 \middle| \frac{1}{\zeta(\beta-1)} \frac{\left(\sum_{x=x_0+1}^{\Delta} \frac{1}{x^{\beta-1}}\right)^2}{\sum_{x=x_0+1}^{\Delta} \frac{1}{x^{\beta}}} > 1\right\}$;*
- *The pairwise connectivity $\mathbb{P}$ is a.s. at most $\frac{1}{4}n^{\frac{3}{2}}\log n$ if $x_0 > \min\left\{x_0 \middle| \frac{1}{\zeta(\beta-1)} \sum_{x=x_0+1}^{\Delta} \frac{1}{x^{\beta-2}} < 1\right\}$.*

Proof Consider that we protect only all nodes of degree larger than x_0 and all other nodes are failed. Similar as in Corollary 3, the expected degree sequence can be written as

$$\mathsf{E}(y_i^{cp}) \doteq \frac{e^{\alpha}}{i^{\beta}} \left(\frac{1}{\zeta(\beta-1)} \sum_{x=x_0+1}^{\Delta} \frac{1}{x^{\beta-1}}\right)^{\beta}.$$

The rest of proof is the same as Theorem 4. □

In order to guarantee the functionality of a power-law network, we take the above Lemma 10 as the condition. In the meanwhile, we aim to minimize the maintenance costs, which include the link costs and critical node protection costs. In detail, we consider the following cost functions:

- *Link Costs:* Consider a link (u, v) in $G_{(\alpha,\beta)}$, its link cost is heavily dependent on the number of messages it transmits according to (Marin-Perianu et al. 2008). Another crucial factor for the link cost is its capacity flow (Smirnov et al. 2003). Since the nodes with degree and betweenness centrality are closely correlated in non-fractal power-law networks (Kitsak et al. 2007), we consider the link cost proportional to the average of the degrees of its two endpoints.
- *Critical Node Protection Costs:* In terms of the critical nodes in $G_{(\alpha,\beta)}$, apart from their degrees, their protection costs are also closely related with the network density. According to (Smirnov et al. 2003), the costs will rise with the increase of density since it enlarges the demand of message exchanges. In addition, as investigated in (Marin-Perianu et al. 2008), the chain reaction leads to the roughly exponential increase of costs, we consider the cost $\gamma(x)$ to protect a node of degree x as $a * x^{b/\beta}$ for some constant a and b.

Therefore, we can confidently formulate the following *Mixed Linear Programming* (MIP), with two variables x_0 and β, as

$$\begin{aligned}
\min\ & \tfrac{1}{2}\textstyle\sum_{x=1}^{\Delta} \frac{e^\alpha}{x^\beta} x + \sum_{x=x_0+1}^{\Delta} \frac{e^\alpha}{x^\beta}\gamma(x) \\
\text{s.t.}\ & \frac{1}{\zeta(\beta-1)} \frac{\left(\sum_{x=x_0+1}^{\Delta} \frac{1}{x^{\beta-1}}\right)^2}{\sum_{x=x_0+1}^{\Delta} \frac{1}{x^\beta}} > 1 \\
& x_0 \in Z^+, x_0 \leq \Delta \\
& \beta > 0
\end{aligned} \tag{5}$$

Note that we omit the proportional constant of link cost since it does not affect the optimization of total maintenance costs.

7.2 Social networks

As we mentioned at the beginning of this paper, one of the main threats in social networks is the malware propagations (Yan et al. 2011). Thus, apart from the factors in (Barabasi and Albert 1999), the containments of these malicious spreading become another crucial factor of the sparsification of social networks. In other words, when an individual is infected, we want to minimize the expected number of total infected users, which can be realized by immunizing critical users beforehand. Therefore, the minimization of immunization costs becomes an urgent need.

Thus, in order to formulate the optimization function for power-law networks in social networks, we first investigate the upper bound of expected size of a connected component after protecting the critical users, which are again referred to as the degree-centrality nodes. That is, we focus on the size of connected components on residual network after removing such immunized users. By defining the residual graph G_s to be the residual power-law graph $G[V \setminus S]$ after immunizing individuals in S, the following Theorem 6 gives the bound of expected size of a connected component on G_s.

Theorem 6 *In the residual graph G_s of $G_{(\alpha,\beta)}$, the expected size of a connected component $\bar{c}$ is a.s. upper bounded by $O\left(n^{\frac{1}{4}}\right)$ when $\tilde{d}_s < 1$, that is, $x_0 < \max\left\{x_0 \middle| \frac{1}{\zeta(\beta-1)}\sum_{x=1}^{x_0}\frac{1}{x^{\beta-2}} < 1\right\}$.*

Proof Consider the connected components $c_1, c_2, \ldots, c_k$ in G_s, their expected size $\bar{c}$ can be written as $\frac{1}{k}\sum_{1\le i\le k} c_i$. According to (Chung and Lu 2002), all connected components a.s. have volume at most $C'\sqrt{n}$ for some constant C' when $\tilde{d} < 1$. Therefore, the number of connected components is at least $C\sqrt{n}$ where $C = 1/C'$. Supposing that $\bar{c} \ge \gamma n^{\frac{1}{4}}$ for some constant γ with probability ρ, the probability that any random pair of nodes are in the same component can be lower bounded by

$$\frac{1}{n^2\overline{d}_s^2}\rho\sum_{1\le i\le k} c_i^2 \ge \frac{1}{n^2\overline{d}_s^2}\rho\bar{c}^2 k \ge \frac{1}{n^2\overline{d}_s^2}\rho C\gamma^2 n.$$

On the other hand, according to F. Chung *et al.* (Chung and Lu 2002), we know that the probability for any random pair of nodes belonging to the same component is upper bounded by

$$\frac{\tilde{d}_s^2}{(1-\tilde{d}_s)n\overline{d}_s}.$$

Combining the above two bounds, we know

$$\frac{1}{n^2\overline{d}_s^2}\rho C\gamma^2 n \le \frac{\tilde{d}_s^2}{(1-\tilde{d}_s)n\overline{d}_s},$$

which implies that

$$\rho \le \frac{\overline{d}_s\tilde{d}_s^2}{C\gamma^2(1-\tilde{d}_s)}.$$

That is, by choosing C to be $\log n$, with probability at least $1-o(1)$, the expected size $\bar{c}$ of connected components is a.s. at most $O\left(n^{\frac{1}{4}}\right)$. □

Again, consider the above lemma as the condition and the same protection cost function of critical users $\gamma(x) = a * x^{b/\beta}$, we formulate the following mixed linear programming, with two variables x_0 and β, in order to make sure that the expected size of connected components in the residual power-law networks is no larger than $O\left(n^{1/4}\right)$.

$$\begin{array}{ll} \min & \sum_{x=x_0+1}^{\Delta}\frac{e^{\alpha}}{x^{\beta}}\gamma(x) \\ \text{s.t.} & \frac{1}{\zeta(\beta-1)}\sum_{x=1}^{x_0}\frac{1}{x^{\beta-2}} < 1 \\ & x_0 \in Z^+, x_0 \le \Delta \\ & \beta > 0 \end{array} \tag{6}$$

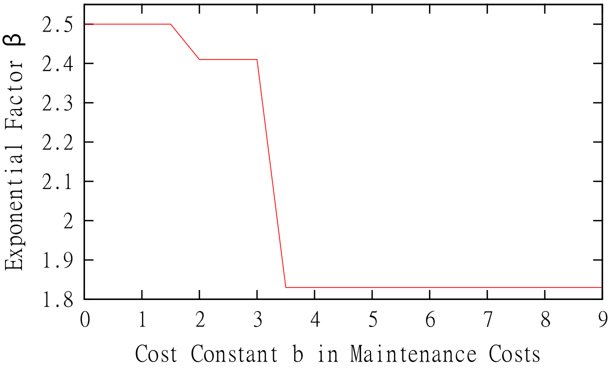

Fig. 6 Optimal robust communication networks

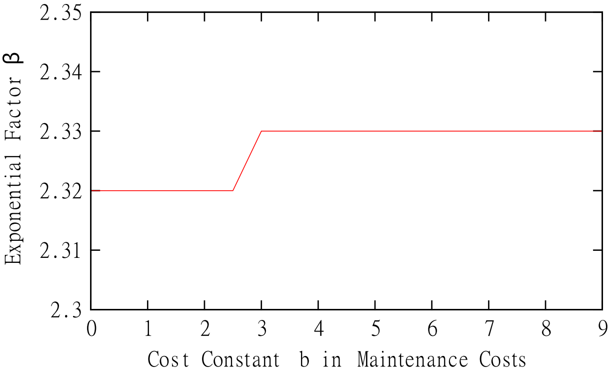

Fig. 7 Optimal robust social networks

7.3 Optimal range of exponential factor β

For the sake of communication networks, consider the practical range of protection costs from 0 to x^9/β for a node of degree x (that is $b \in [0, 3]$), Fig. 6 reveals the relation between maintenance costs and optimal β according to MIP (5). As one can see, the optimal β is from 1.8 to 2.5 no matter how large the constant b is, the exponential factor β is no $<$1.8. (Note that the curve is invariant for distinct network sizes since the effect of light-tailed elements in riemann zeta function can be neglected.)

Figure 7 reports that the optimal range of β is from 2.3 to 2.4 in social networks according to MIP (6). We observe that the increase of b does not really affect the range of β and the curve also remains invariant with respect to different network sizes.

In summary, the analysis on both communication networks and social networks gives us a reasonable explanation of the topology in real-world power-law networks, that is, the best range of the exponential factor β is $[1.8, 2.5]$. In other cases, the network maintenance cost either becomes very expensive when $\beta < 1.8$, or the network robustness is unpredictable when $\beta > 2.5$ due to its dependence on the specific attacking strategy.

8 Related works

There are a great number of studies regarding the tolerance of real-world networks against failures and attacks using different metrics. Edge vulnerability in metabolic networks was studied by Kaiser et al. with respect to the average shortest path and the clustering coefficient (Kaiser and Hilgetag 2004). For the sake of power grid networks, (Albert et al. 2004) investigated their vulnerability by measuring the loss of connectivity under various threats, including random, cascading, load-based and degree-based nodal failures. The disruption of vital interstate systems was assessed by (Matisziw and Murray 2009) according to the available number of compromised s-t flows. The devastating consequences of link failures in networks are investigated in (Dinh et al. 2014). (Cohen et al. 2000) showed the resilience of the Internet to the random breakdown of the nodes based on percolation theory. In (Satorras and Vespignani 2002), Satorras *et al.* also revealed that the random uniform immunization of individuals cannot lead to the eradication of communications in complex social networks using the reduced prevalence rate. (Doyle et al. 2005), using a novel metric *ELASTICITY*, explored that Internet topologies are less affected by both random and targeted attacks than the power-law networks. In general, the robustness of other complex networks was studied in (Jamakovic and Van Mieghem 2008) using algebraic connectivity, i.e., the second-smallest eigenvalue of the Laplacian matrix of a graph. Recently, from a different perspective, (Alderson and Doyle 2010) focused on the role of organization and design in terms of the complexity in highly organized technological and biological systems.

More generally, (Albert et al. 2000) first compared the robustness of complex systems with the power-law and exponential properties. By measuring the diameters, the relative size of the largest cluster and the average size of the isolated clusters, the power-law networks are empirically observed to tolerate failures to a surprising degree but their survivability decreases rapidly under attacks after comparing them with exponential networks. Later on, (Holme et al. 2002) further investigated the degree of harms to power-law networks under different strategies of attacks. Unfortunately, all these observations are derived from experiments and lack their theoretical foundations.

9 Conclusion

In this paper, from a theoretical point of view, we study the robustness of power-law networks under various threats, i.e. random failures, preferential attacks and degree-centrality attacks. Essentially, the power-law networks are illustrated to extremely tolerate random failures. In the meanwhile, they are more robust under both preferential attacks and degree-centrality attacks if they have a smaller exponential factor β.

Moreover, from the other perspective, we further exploit the topologies of practical real-world networks by optimizing the costs and guaranteeing their robustness. The best range of the exponential factor β is illustrated to be $[1.8, 2.5]$, which gives a reasonable explanation for the topologies of most real-world networks. When $\beta < 1.8$, the network maintenance cost is very expensive, and when $\beta > 2.5$, the network robustness is unpredictable since it depends on the specific attacking strategy.

There are a few open issues worthy of discussion. First of all, based on the existence giant components, we get the pairwise connectivity threshold $\mathbb{P} = \Theta(n^2)$, which is the signature of highly-connected networks. It is interesting to obtain tighter bounds of the size of giant components and further better evaluate the pairwise connectivity in residual networks, which can be the extension of our work. Next, by adjusting parameters, the structural network model in this paper can represent many real power-law networks. The experimental results in this paper provide important insights into the design of optimal network with respect to different threats. However, to better assess the robustness of networks, experiments on real networks are necessary in the future work.

Acknowledgments This work is partially supported by NSF Career Award CNS-0953284 and NSF CCF-1422116.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Aiello W, Chung F, Lu L (2000) A random graph model for massive graphs. In Proceedings of the thirty-second annual ACM symposium on Theory of computing, STOC ’00, pp 171–180, New York, NY, USA. ACM

Albert R, Albert I, Nakarado GL (2004) Structural vulnerability of the north american power grid. Phys Rev E 69(2):025103

Albert R, Jeong H, Barabasi A (2000) Error and attack tolerance of complex networks. Nature 406(6794):378–382

Alderson DL, Doyle JC (2010) Contrasting views of complexity and their implications for network-centric infrastructures. Trans Syst Man Cybern Part A 40(4):839–852

Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512

Bollobás B, Riordan O, Spencer J, Tusnády G (2001) The degree sequence of a scale-free random graph process. Random Struct Algorithms 18(3):279–290

Chung F, Lu L (2002) Connected components in random graphs with given expected degree sequences. Ann Comb 6(2):125–145

Chung F, Lu L (2006) Concentration inequalities and martingale inequalities: a survey. Internet Math 3(1):79–127

Cohen R, Erez K, Ben-Avraham D, Havlin S (2000) Resilience of the internet to random breakdowns. Phys Rev Lett 85(21):4626

Dinh T, Thai M, Nguyen H (2014) Bound and exact methods for assessing link vulnerability in complex networks. J Comb Optim 28(1):3–24

Dinh TN, Xuan Y, Thai MT, Pardalos PM, Znati T (2011) IEEE/ACM Transactions on On new approaches of assessing network vulnerability: Hardness and approximation. Networking, PP(99):1

Doyle JC, Alderson DL, Li L, Low S, Roughan M, Shalunov S, Tanaka R, Willinger W (2005) The “robust yet fragile” nature of the Internet. Proc Natl Acad Sci USA, 102

Faloutsos M, Faloutsos P, Faloutsos C (1999) On power-law relationships of the internet topology. In Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication, SIGCOMM ’99, pp 251–262, New York, NY, USA. ACM

Gkantsidis C, Mihail M, Zegura E (2003) The markov chain simulation method for generating connected power law random graphs. In Proceedings of the 5th Workshop on Algorithm Engineering and Experiments (ALENEX). SIAM

Holme P, Kim BJ, Yoon CN, Han SK (2002) Attack vulnerability of complex networks. Phys Rev E 65(5):056109

Jamakovic A, Van Mieghem P (2008) On the robustness of complex networks by using the algebraic connectivity. In Proceedings of the 7th international IFIP-TC6 networking conference on AdHoc and sensor networks, wireless networks, next generation internet, NETWORKING'08, pp 183–194, Berlin, Heidelberg. Springer-Verlag

Kaiser M, Hilgetag CC (2004) Edge vulnerability in neural and metabolic networks. Biol Cybern 90:311–317. doi:10.1007/s00422-004-0479-1

Kitsak M, Havlin S, Paul G, Riccaboni M, Pammolli F, Stanley HE (2007) Betweenness centrality of fractal and nonfractal scale-free model networks and tests on real networks. Phys Rev E (Statistical, Nonlinear, and Soft Matter Physics) 75(5):056115

Lakhina A, Papagiannaki K, Crovella M, Diot C, Kolaczyk ED, Taft N (2004) Structural analysis of network traffic flows. In Proceedings of the joint international conference on Measurement and modeling of computer systems, SIGMETRICS '04/Performance '04, pp 61–72, New York NY, USA. ACM

Luciano, Rodrigues F, Travieso G, Boas VPR (2007) Characterization of complex networks: a survey of measurements. Adv Phys 56(1):167–242

Marin-Perianu RS, Scholten J, Havinga PJM, Hartel PH (2008) Cluster-based service discovery for heterogeneous wireless sensor networks. Int J Parallel Emerg Distrib Syst 23(4):325–346

Matisziw TC, Murray AT (2009) Modeling s-t path availability to support disaster vulnerability assessment of network infrastructure. Comput Oper Res 36(1):16–26 Part Special Issue: Operations Research Approaches for Disaster Recovery Planning

Mayo M, Abdelzaher A, Ghosh P (2015) Long-range degree correlations in complex networks. Comput Soc Netw 2(1):1–13

Molloy M, Reed B (1995) A critical point for random graphs with a given degree sequence. Random Struct Algorithms 6:161–179

Molloy M, Reed B (1998) The size of the giant component of a random graph with a given degree sequence. Comb Probab Comput 7(3):295–305

Satorras RP, Vespignani A (2002) Immunization of complex networks. Phys Rev E 65(3):036104

Smirnov M, Biersack EW, Blondia C, Bonaventure O, Casals O, Karlsson G, Pavlou G, Quoitin B, Roberts J, Stavrakakis I, Stiller B, Trimintzios P, Mieghem PV, editors (2003) Quality of Future Internet Services, COST Action 263 Final Report, volume 2856 of Lecture Notes in Computer Science. Springer

Ventresca M, Aleman D (2015) Efficiently identifying critical nodes in large complex networks. Comput Soc Netw 2(1):1–16

Yan G, Chen G, Eidenbenz S, Li N (2011) Malware propagation in online social networks: Nature, dynamics, and defense implications. In Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security (AsiaCCS'11), Hongkong, China