

False Positive and False Negative Effects on Network Attacks

Yilun Shang¹ 

Received: 19 June 2017 / Accepted: 11 November 2017 / Published online: 20 November 2017
© Springer Science+Business Media, LLC, part of Springer Nature 2017

Abstract Robustness against attacks serves as evidence for complex network structures and failure mechanisms that lie behind them. Most often, due to detection capability limitation or good disguises, attacks on networks are subject to false positives and false negatives, meaning that functional nodes may be falsely regarded as compromised by the attacker and vice versa. In this work, we initiate a study of false positive/negative effects on network robustness against three fundamental types of attack strategies, namely, random attacks (RA), localized attacks (LA), and targeted attack (TA). By developing a general mathematical framework based upon the percolation model, we investigate analytically and by numerical simulations of attack robustness with false positive/negative rate (FPR/FNR) on three benchmark models including Erdős-Rényi (ER) networks, random regular (RR) networks, and scale-free (SF) networks. We show that ER networks are equivalently robust against RA and LA only when FPR equals zero or the initial network is intact. We find several interesting crossovers in RR and SF networks when FPR is taken into consideration. By defining the cost of attack, we observe diminishing marginal attack efficiency for RA, LA, and TA. Our finding highlights the potential risk of underestimating or ignoring FPR in understanding attack robustness. The results may provide insights into ways of enhancing robustness of network architecture and improve the level of protection of critical infrastructures.

Keywords Complex networks · False positive · False negative · Percolation · Attack robustness

1 Introduction

Complex systems can often be characterized by networks in which the nodes of the network are the components of the system and the links connecting the nodes represent the interactions

✉ Yilun Shang
shyl@tongji.edu.cn

¹ School of Mathematical Sciences, Tongji University, Shanghai 200092, China

between the components. The past two decades or so have witnessed numerous advances in network science because many real-world networks such as the Internet, the World Wide Web, protein-protein interaction networks, metabolic networks, food webs, and social networks are found to have a variety of topological and dynamical properties [1–4]. The function and stability of networks rely crucially on the interconnections between nodes in which failed nodes will disable others connecting through them to the network and may destroy or cripple the entire network. As such, it is of great importance to understand network robustness, i.e., how the structure of a network changes as the nodes in it are removed either through random or malicious attacks [5]. In theoretical studies, percolation model [3, 6–8] plays a prominent role in understanding attacks on networks where the percolation phase transition occurs at a certain critical occupation probability p_c , above which a giant component (proportional to the size of the original network) exists showing the robustness of a macroscopic cluster. Using the giant component as the relevant order parameter, some remarkable robustness characteristics of complex networks are observed. These include the Achilles' heel phenomenon [5, 9, 10], namely, heterogeneous networks are highly robust against random attacks but are extremely fragile to attacks targeted at hubs, and cascading processes in interdependent networks [11–13].

In most existing studies on attack robustness, the initial network is assumed to be intact and all nodes are functional for simplicity. Moreover, the tacit assumption has often been made that the states of the nodes are accurate to the attacker. However, in real scenarios, the initial network mostly contains a mixture of functional and dysfunctional (i.e., compromised) nodes, and some observation errors are prevalent. In particular, a functional node has a chance, referred to as false positive rate (FPR) [14], to be regarded as compromised by the attacker (so that attacks will not be launched against it). This may be due to the limitation of the attacker's detection capability and/or the disguise of the nodes. For proactive cyber defense, for example, IP mutation techniques have been proposed to disguise the identity of hosts from sniffers and scanners based on software-defined networking [15]. To protect distributed software systems, message forwarding and traffic padding [16, 17] are exploited to camouflage the real traffic flow so that the functional components are obscured from attack. Similarly, false negative rate (FNR) [14] indicates the probability of falsely viewing an already dysfunctional node as functional (which may incur subsequent vain attacks). In military actions, false targets can be deployed to lure the attacker away from genuine elements and increase the cost of the attacker.

The aim of this paper is to investigate how the existence of FPR and FNR influences the robustness of complex networks under attacks in terms of the critical percolation threshold and the giant component size. Here, three typical classes of attack strategies are considered:

- Random attack (RA), where randomly chosen nodes are removed from the network, meaning that each node in the network is attacked with the same probability. RA consists of the simplest percolation model, which naturally describes system decay, random errors or attacks without prior knowledge of the network topology; see e.g. [3, 5–7, 10, 11].
- Localized attack (LA), where nodes surrounding a seed node are removed layer by layer, causing aggregated damage of adjacent components limited to a specific area. LA can be induced by natural disasters such as earthquakes, floods, and tsunamis, as well as mass attacks including hazardous chemicals, bomb blasts, and computer viruses; see e.g. [8, 18–22].
- Targeted attack (TA), where nodes with a higher degree (such as hubs) are more vulnerable, meaning that nodes are attacked in decreasing order of their degree. TA reflects

real-world situations like malicious attacks against transportation hubs or important power stations, sabotage on the Internet, and actions on key figures of terrorist organizations; see e.g. [5–7, 9, 23–25].

We develop a mathematical framework for understanding the structure transition of a set of common complex networks benchmarks under RA, LA, and TA, respectively. They include Erdős-Rényi (ER) networks [26] with a Poisson degree distribution $P(k) = e^{-\lambda} \lambda^k / k!$ and average degree $\langle k \rangle = \lambda$, random regular (RR) networks following a degenerated degree distribution $P(k) = \delta_{k,k_0}$ meaning that each node has exactly k_0 links, and scale-free (SF) networks [2, 3, 5] characterized by a power-law degree distribution $P(k) \sim k^{-\gamma}$ ($\gamma > 0$) with a lower and upper cutoff, k_{\min} and k_{\max} . Our extensive simulations are in good agreement with analytical calculations. In addition to the model networks, we perform simulations on large-scale real-world networks, including a communication network and an infrastructure network, to demonstrate the obtained results. Our method is shown to be capable of predicting attack robustness and uncovering network structural characteristics.

1.1 Related Work on Attack Robustness of Complex Networks

Due to practical significance of network robustness against deliberate attacks and failures, a considerable amount of research effort has been devoted to understanding attack robustness in the past decades. The pioneering work of Albert et al. [5] reveals an important property of scale-free networks that they display a surprisingly high degree of tolerance against random decay. Most of prior works dealing with the effect of removing nodes uniformly at random or in decreasing order of their degrees are reviewed in [3, 7]. Drawing upon numerical simulations, the recent work [23] systematically explores the structure transition of targeting nodes for removal based on a number of measures including degree, betweenness, closeness, and eigenvector centrality. A multi-strategy targeted attack launched sequentially based on several centrality measures is studied numerically in [27]. Some other recent works focus on random and targeted attacks on interdependent networks and competing networks; see, e.g., [11–13, 28]. More recently, there has been some interest in the newly introduced localized attack partly due to its analytical tractability [8, 18, 22]. However, as mentioned above, none of these works take real situations involving FPR or FNR into account.

We mention that a concept similar in spirit to FPR is the acquaintance immunization [29], in which one neighbor of a randomly selected node is vaccinated so that it will not be removed from the network. However, acquaintance immunization and its variants are mostly studied in the context of epidemic spreading on networks for a different purpose (emphasizing virus propagation and epidemic threshold) using different approaches [30, 31].

1.2 Contributions

Some of our main findings are summarized in Tables 1, 2, and 3.

The rest of the paper is organized as follows. In Sect. 2, we describe the models and present the analytical frameworks geared towards attack robustness in the presence of false positives and false negatives under RA, LA, and TA, respectively. In Sect. 3, we perform numerical experiments on synthetic networks. In Sect. 4, we illustrate the obtained results on real-world networks of diverse nature. Some discussions regarding benefit-cost ratio and possible extensions are provided in Sect. 5.

Table 1 Attack robustness of ER networks with average degree $\langle k \rangle = 4$

ER networks		
	Results	Meanings
p_c	p_c decreases as α increases	More robust with larger FPR
	RA=LA when $\alpha = 0$ or $q = 1$	Equally robust against RA and LA when no FPR or the initial network is intact
	TA \geq LA \geq RA when $\alpha > 0^a$	LA is between TA ($\delta = 0$) and TA ($\delta = 1$) in terms of attack robustness
P_∞	P_∞ increases as α increases	The same as above
	RA=LA when $\alpha = 0$ or $q = 1$	
	RA \geq LA \geq TA when $\alpha > 0^a$	
PCR	PCR grows steadily ^b as p increases and soars when close to $p = 1$	Diminishing marginal attack efficiency for RA, LA, and TA
	PCR increases as α increases	More efficient with larger FPR
	PCR decreases as β decreases	Less efficient with larger FNR

^aTA indicates targeted attack with $\delta = 1$; see (15) for the definition of δ

^bShowing small peaks at p_c

2 Theoretical Framework on Attacks Involving False Positives and False Negatives

We consider a random network characterized by an arbitrary degree distribution $P(k)$, which is the probability that a randomly chosen node has k links. The generating function of the degree distribution is defined as $G_0(x) = \sum_{k=0}^{\infty} P(k)x^k$ [3,32]. We assume that there are a fraction q of functional nodes and a fraction $1 - q$ of failed nodes in the initial network. Given a functional node, the probability of its being regarded as “failed” by the attacker is denoted by α , i.e., FPR. To avoid confusion, we shall use quotation marks to emphasize the states of nodes with respect to the attacker, which can be real or unreal. Analogously, FNR is signified by β , which is the probability that a failed node is regarded as “functional” by the attacker. As a rule, only “functional” nodes will be attacked and then become failed.

We assume that attack is launched against the network until a fraction $1 - p$ of “functional” nodes in the entire network are attacked. A major characteristic of network functionality is the relative size of the giant component, denoted by P_∞ , which consists of all functional nodes after attack. The critical threshold at which the giant component first collapses, i.e. $P_\infty \sim 0$, is denoted by p_c . In the following, we address three types of attacks, RA, LA, and TA, respectively.

Table 2 Attack robustness of RR networks with degree 4

RR networks		
	Results	Meanings
p_c	p_c decreases as α increases	More robust with larger FPR
	$RA = TA \geq LA$ when $q = 1$	Equally robust against RA and TA when the initial network is intact
	$TA \geq RA \geq LA$ when $q < 1$	More robust against RA than against TA when the initial network is not intact Most robust against LA in terms of p_c^a
P_∞	P_∞ increases as α increases	Similar as above (but with some exceptions on robustness of LA^a)
	$LA \geq RA = TA$ when $q = 1$	
	$LA \geq RA \geq TA$ when $q < 1$ (with some exceptions ^a)	
PCR	PCR grows steadily ^b as p increases and soars when close to $p = 1$	Diminishing marginal attack efficiency for RA, LA, and TA
	PCR increases as α increases	More efficient with larger FPR
	PCR decreases as β decreases	Less efficient with larger FNR

^aIt happens that $P_\infty(RA) > P_\infty(TA) > P_\infty(LA)$ when $q < 1$, α and p are relatively large (see Fig. 6a), indicating that the network can be the least robust against LA in terms of P_∞

^bShowing small peaks at p_c

2.1 Random Attack

In a random attack, each “functional” node is attacked with probability $1 - p$. It is easy to see that the RA process can be accommodated by the classical node percolation [3,6] with occupation probability $\alpha q + p(1 - \alpha)q$ using law of total probability. Recall that the generating function of the degree distribution is $G_0(x) = \sum_{k=0}^\infty P(k)x^k$. The size distribution of the clusters that can be reached following a randomly selected edge is generated in a self-consistent equation [6]

$$H_1(x) = 1 - \alpha q - p(1 - \alpha)q + [\alpha q + p(1 - \alpha)q]xG_1(H_1(x)), \tag{1}$$

where $G_1(x) = G'_0(x)/G'_0(1)$. Then the probability generating function for the size of the cluster to which a randomly selected node belongs is generated by

Table 3 Attack robustness of SF networks with $\gamma = 2.47$ and average degree 4.01

SF networks		
	Results	Meanings
p_c	p_c decreases as α increases	More robust with larger FPR
	$LA \geq RA^a$ for all α and q	Always more robust against RA than against LA and TA
	$TA \geq RA$ for all α and q	LA and TA are comparable with appropriate δ
P_∞	P_∞ increases as α increases	The same as above
	$RA \geq LA^a$ for all $\alpha, q,$ and p	
	$RA \geq TA$ for all α, q, p	
PCR	PCR grows steadily ^b as p increases and soars when close to $p = 1$	Diminishing marginal attack efficiency for RA, LA, and TA
	PCR increases as α increases	More efficient with larger FPR
	PCR decreases as β decreases	Less efficient with larger FNR
	$TA \geq RA$	TA and LA are more efficient than RA
	$LA \geq RA^a$	

^aThis is the case when γ is relatively small; otherwise, the inequality reverses

^b Showing small peaks at p_c

$$H_0(x) = 1 - \alpha q - p(1 - \alpha)q + [\alpha q + p(1 - \alpha)q]xG_0(H_1(x)). \tag{2}$$

Hence, the mean size of small clusters is

$$H'_0(1) = [\alpha q + p(1 - \alpha)q] \cdot \left\{ 1 + \frac{[\alpha q + p(1 - \alpha)q]G'_0(1)}{1 - [\alpha q + p(1 - \alpha)q]G'_1(1)} \right\}, \tag{3}$$

which diverges when $1 = [\alpha q + p(1 - \alpha)q]G'_1(1)$ marking the critical value p_c at which the giant component collapses. Therefore, for $q > 0$ and $\alpha < 1$ we have

$$p_c(RA) = \min \left\{ \max \left\{ 0, \frac{\frac{1}{G'_1(1)} - \alpha q}{(1 - \alpha)q} \right\}, 1 \right\}. \tag{4}$$

Note that we recover the critical occupation probability $p_c = 1/G'_1(1)$ in [6] for an initially intact network with no false positives; i.e., $q = 1$ and $\alpha = 0$. When $q > 1/G'_1(1)$, the initial network has a giant component, and we observe that $p_c(RA)$ is a decreasing function with respect to α . This agrees with our intuition that the network becomes more robust when α

increases, namely, more nodes are protected from attack. Moreover, it is easily seen from (4) that $p_c(\text{RA}) = 1$ when $q \leq 1/G'_1(1)$, and that $p_c(\text{RA})$ decreases with respect to q .

The fraction $S(\text{RA})$ of the giant component in the network is given by

$$S(\text{RA}) = 1 - H_0(1) = [\alpha q + p(1 - \alpha)q][1 - G_0(H_1(1))], \tag{5}$$

where $H_1(1)$ satisfies $H_1(1) = 1 - \alpha q - p(1 - \alpha)q + [\alpha q + p(1 - \alpha)q]G_1(H_1(1))$. By definition, we have $P_\infty(\text{RA}) = S(\text{RA})$. From (4) and (5) we observe that p_c and P_∞ do not change with β because a failed node remains so regardless of the attacker’s perspective. However, the introduction of FNR is not only for the generality of the work. Recall in the beginning of this section that the considered attack is launched against the network until a fraction $1 - p$ of “functional” nodes are attacked. Despite not changing the static quantities such as p_c and P_∞ , β clearly plays a role in this rule and affects the percolation process. The influence of FNR is manifested in the cost of attack (see Sect. 2.4 below) and the dynamical performance-cost ratio in the process of attack (c.f. Figs. 4, 7, and 10).

2.2 Localized Attack

We next consider the localized attack on the network by attacking a fraction $1 - p$ of “functional” nodes starting from a randomly chosen root node, its nearest neighbors (shell 1), next nearest neighbors (shell 2), and so forth [18]. Formally, shell l is defined as the set of nodes that are at distance l from the root.

In the network of N nodes (with $N \rightarrow \infty$ in the thermodynamic limit), we divide the whole attack process into three regimes (c.f. Fig. 1): (i) We first attack a fraction $1 - p$ of “functional” nodes starting from the root node as follows. The attacker checks nodes in the ascending order of their distance from the root. If a node is “functional”, then it is removed from the network, otherwise it maintains the same state. The nodes in the same shell of the root are checked randomly, and after nodes in shell l are fully checked, the attacker begins checking nodes in shell $l + 1$. The process continues until a fraction $1 - p$ of “functional” nodes in the entire network are removed. We define the *attacked area* as the set consisting of all nodes that are checked and are removed after the attack. After the attack, we suppose that the links connecting the attacked area to outside (i.e., the rest of the nodes) are still left in place and that all links connecting two nodes outside the attacked area are also present (see Fig. 1b); (ii) We remove these links connecting the attacked area to outside (see Fig. 1c); (iii) We remove all failed nodes (together with their incident links) outside the attacked area (see Fig. 1d).

Since there are qN random functional nodes and $(1 - q)N$ random failed nodes in the initial network on average, the attacked area is composed of $(1 - p)(1 - \alpha)qN + (1 - p)(1 - q)N$ nodes. Hence, there are a fraction $s := p + \alpha q(1 - p)$ of nodes outside the the attacked area. We first consider the regime (i). Let $A_s(k)$ be the number of nodes with degree k out of the attacked area. In the limit of $N \rightarrow \infty$, following the approach introduced in [18, 22, 33], we find that the probability to have a node with degree k outside the attacked area, $P_s(k) = A_s(k)/(sN)$, is given by

$$P_s(k) = P(k) \frac{f^k}{G_0(f)}, \tag{6}$$

and the average degree outside the attacked area is $\langle k(s) \rangle = \sum_{k=0}^\infty P_s(k)k = fG'_0(f)/G_0(f)$, where $f \equiv G_0^{-1}(s)$. We write $G_s(x) = \sum_{k=0}^\infty P_s(k)x^k = G_0(fx)/G_0(f)$ for the generating function of $P_s(k)$.

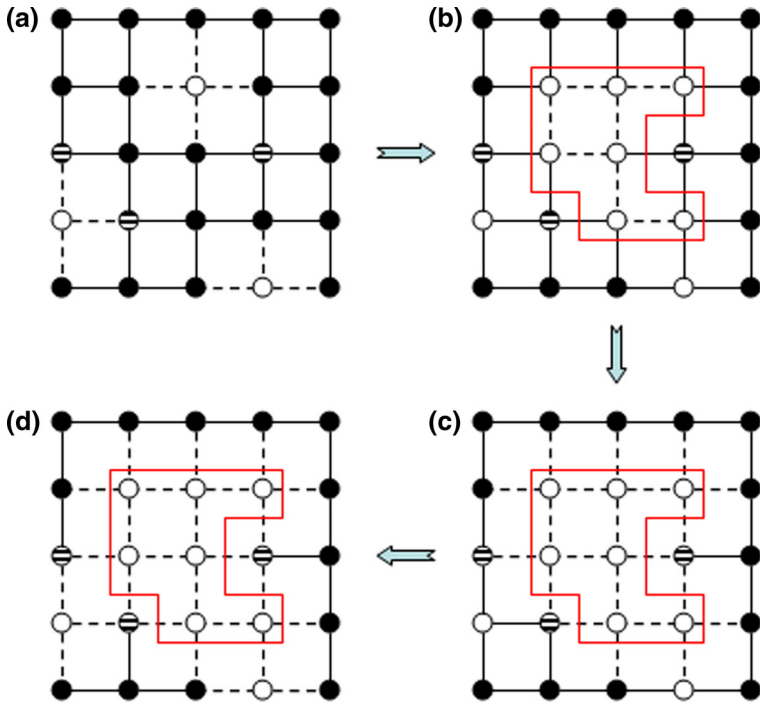


Fig. 1 Schematic illustration of the LA process. **a** A fraction q of nodes are initially functional (black represents the functional nodes, white the failed nodes, and shadowed the functional nodes that are not known to the attacker due to FPR). **b** A fraction $1 - p$ of “functional” nodes are removed starting from the central node (root). The attacked area is indicated by a red contour. All links connecting to nodes outside the attacked area are envisaged to be present. **c** Links connecting the attacked area to outside are removed. **d** Failed nodes outside the attacked area are removed

In the regime (ii), the number of links connecting the attacked area to outside can be expressed as

$$\tilde{L}(s) = \frac{Ns\langle k(s) \rangle L(s)}{Ns\langle k(s) \rangle + L(s)} = N \left[fG'_0(f) - \frac{G'_0(f)^2}{G'_0(1)} \right], \tag{7}$$

where $L(s) = N[G'_0(1)f^2 - G'_0(f)f]$ means the number of open links belonging to the outer shell of the attacked area [33]. Due to the randomness of interconnections, the resultant network outside the attacked area can be thought of as the outcome of a bond percolation with occupation probability given by $\tilde{q} = 1 - \tilde{L}(s)/(sN\langle k(s) \rangle) = G'_0(f)/[G'_0(1)f]$. At this stage, the probability generating function of nodes’ degree distribution, signified by $\tilde{G}_0(x)$, becomes [18,22]

$$\begin{aligned} \tilde{G}_0(x) &= G_s(1 - \tilde{q} + \tilde{q}x) \\ &= \frac{1}{G_0(f)} G_0 \left[f + \frac{G'_0(f)}{G'_0(1)} (x - 1) \right]. \end{aligned} \tag{8}$$

We finally consider the regime (iii), which consists of a node percolation with occupation probability q . As the random connection process can be modeled by a branching process, we define $\tilde{G}_1(x) = \tilde{G}'_0(x)/\tilde{G}'_0(1)$ to be the generating function of the underlying branching

process. Accordingly, the size distributions of the clusters that can be reached from a randomly chosen link, and the clusters that can be traversed by randomly following a starting node are generated, respectively, by [3,32]

$$\tilde{H}_1(x) = 1 - q + qx\tilde{G}_1[\tilde{H}_1(x)] \tag{9}$$

and

$$\tilde{H}_0(x) = 1 - q + qx\tilde{G}_0[\tilde{H}_1(x)]. \tag{10}$$

The mean size of small clusters equals

$$\tilde{H}'_0(1) = q \left[1 + \frac{q\tilde{G}'_0(1)}{1 - q\tilde{G}'_1(1)} \right]. \tag{11}$$

The diverging point of (11) indicates the critical occupation probability $p_c(\text{LA})$, at which a giant component first forms. A direct calculation shows that

$$p_c(\text{LA}) = \min \left\{ \max \left\{ 0, \frac{s_c - \alpha q}{1 - \alpha q} \right\}, 1 \right\} \tag{12}$$

is determined by

$$G'_0(1) = qG''_0(f), \tag{13}$$

where $f \equiv G_0^{-1}(s)$. Clearly, $p_c(\text{LA}) = 1$ when $q \leq 1/G'_1(1)$,¹ and similarly as in the RA case, $p_c(\text{LA})$ is a decreasing function with respect to both α and q .

Using (9) and (10), the fraction S of the giant component in the resultant network outside the attacked area is given by

$$S(\text{LA}) = 1 - \tilde{H}_0(1) = q[1 - \tilde{G}_0(\tilde{H}_1(1))], \tag{14}$$

where $\tilde{H}_1(1)$ satisfies $\tilde{H}_1(1) = 1 - q + q\tilde{G}_1(\tilde{H}_1(1))$. The relative size of the giant component as a fraction of the original network is given by $P_\infty(\text{LA}) = sS(\text{LA})$.

From (12), (13), and (14), we can easily reproduce the critical values $p_c(\text{LA})$ and $P_\infty(\text{LA})$ in [18] when the initial network is intact and FPR is absent, namely, when $q = 1$ and $\alpha = 0$. Again, FNR, β , does not affect p_c and P_∞ .

2.3 Targeted Attack

In a targeted attack, a fraction $1 - p$ of “functional” nodes are attacked and removed based on their degrees. Following [25,34], we assign to each node a value

$$W_\delta(k_i) = \frac{k_i^\delta}{\sum_{i=1}^N k_i^\delta}, \tag{15}$$

to represent the probability that a node i with degree k_i is attacked if it is “functional”, where δ is a real and N is the number of nodes in the network as before. When $\delta > 0$, nodes with higher degree have a higher probability to be removed; pushing it to the limit $\delta \rightarrow \infty$ yields the attack strategy that nodes are removed strictly in the decreasing order of degrees. The case $\delta < 0$ implies the opposite strategies. Note that the case $\delta = 0$ is equivalent to the random attack with equal probability. In fact, we can show that $p_c(\text{TA}) = p_c(\text{RA})$ and $P_\infty(\text{TA}) = P_\infty(\text{RA})$ when $\delta = 0$; see below and Appendix A.

¹ When $q = 1/G'_1(1)$, it follows from (13) that $s_c = 1$ and hence $p_c = 1$. However, if $q < 1/G'_1(1)$, the solution s_c of (13) can be larger than one, which is an extraneous root as s by definition cannot exceed one. This is why we formally impose an upper bound in (12).

Given a δ , we divide the targeted attack process into three regimes: (i) We remove the failed nodes in the initial network. Namely, $(1 - q)N$ nodes are removed in this regime; (ii) We attack a fraction $1 - p$ of “functional” nodes according to (15) in the remaining network. The attacked area now consists of the failed nodes in the remaining network. Then we assume that all links connecting the attacked area and the remaining nodes are still left in place; (iii) We remove those links connecting the attacked area and the remaining nodes.

Since a fraction $1 - q$ of nodes are removed randomly from the initial network in the regime (i), the probability generating function for the node degree of the remaining network is [3,34]

$$\bar{G}_0(x) := G_0(1 - q + qx), \tag{16}$$

and the corresponding degree distribution is denoted by $\bar{P}(k) := \frac{1}{k!} \frac{d^k \bar{G}_0}{dx^k} \Big|_{x=0}$. Let $\bar{G}_1(x) = \bar{G}'_0(x)/\bar{G}'_0(1) = G_1(1 - q + qx)$.

In the regime (ii), there are $(1 - p)(1 - \alpha)qN$ nodes in the attacked area. Let $t = [qN - (1 - p)(1 - \alpha)qN]/qN = p + \alpha(1 - p)$. Therefore, there are a fraction t of nodes in the resulting network. Let $A_t(k)$ be the number of nodes with degree k out of the attacked area. Let $G_\delta(x) = \sum_{k=0}^\infty \bar{P}(k)x^{k^\delta}$ [34,35]. Following the method introduced in [34], we find that the probability to have a node with degree k outside the attacked area, $P_t(k) = A_t(k)/(tqN)$, can be expressed by

$$P_t(k) = \bar{P}(k) \frac{g^{k^\delta}}{G_\delta(g)}, \tag{17}$$

and the average degree outside the attacked area is $\langle k^\delta(t) \rangle := \sum_{k=0}^\infty P_t(k)k^\delta = gG'_\delta(g)/G_\delta(g)$, where $g \equiv G_\delta^{-1}(t)$. We write $G_t(x) = \sum_{k=0}^\infty P_t(k)x^k = \frac{1}{t} \sum_{k=0}^\infty \bar{P}(k)g^{k^\delta}x^k$ for the generating function of $P_t(k)$.

In the regime (iii), since the network is randomly connected, the resultant network outside the attacked area can be viewed as the outcome of a bond percolation with occupation probability given by

$$\hat{q} = \frac{tqN \langle k(t) \rangle}{qN \langle k \rangle} = \frac{tG'_t(1)}{\bar{G}'_0(1)}, \tag{18}$$

where $\langle k(t) \rangle$ is the average degree of remaining nodes, and $\langle k \rangle$ is the average degree of the remaining network after the regime (i). Using the same approach as in [3], the probability generating function of remaining nodes’ degree distribution, signified by $\hat{G}_0(x)$, becomes

$$\begin{aligned} \hat{G}_0(x) &= G_t(1 - \hat{q} + \hat{q}x) \\ &= \frac{1}{G_\delta(g)} \sum_{k=0}^\infty \bar{P}(k)g^{k^\delta} [1 + \hat{q}(x - 1)]^k. \end{aligned} \tag{19}$$

As the random connection process can be modeled by a branching process, we define

$$\hat{G}_1(x) = \frac{\hat{G}'_0(x)}{\hat{G}'_0(1)} = \frac{G'_t(1 + \hat{q}(x - 1))}{G'_t(1)} \tag{20}$$

to be the generating function of the underlying branching process. By combining (20) and the criterion for the network to collapse [32], $1 = \hat{G}'_1(1)$, we find that t_c satisfies

$$G'_t(1) = \hat{q}G''_t(1) \tag{21}$$

where $g \equiv G_\delta^{-1}(t)$. Therefore,

$$p_c(\text{TA}) = \min \left\{ \max \left\{ 0, \frac{t_c - \alpha}{1 - \alpha} \right\}, 1 \right\}. \tag{22}$$

The fraction S of the giant component in the resultant network outside the attacked area satisfies

$$S(\text{TA}) = 1 - \hat{G}_0(u), \tag{23}$$

where u satisfies the transcendental equation $u = \hat{G}_1(u)$, and $g = G_\delta^{-1}(p + \alpha(1 - p))$. The relative size of the giant component as a fraction of the original network is given by $P_\infty(\text{TA}) = tqS(\text{TA})$.

As mentioned before, we will show the equivalence of RA and TA in the case $\delta = 0$ in Appendix A. Also, note that when $p = 1$, i.e., no attacks are launched, we have $P_\infty(\text{RA}) = P_\infty(\text{LA}) = P_\infty(\text{TA})$ for all q (clearly, α and β play no role herein). This can be shown straightforwardly by comparing (5), (14), (23) and a similar argument as in Appendix A. The details are left to the interested reader.

2.4 Cost of Attack

To appropriately evaluate the effectiveness of attack, the attack cost should be factored in. In [28], the authors analyzed the attack cost for two competing networks, in which the stronger network may take over inactive nodes from the weaker network at a price of reduction of its own robustness. Here, with the attacker as an external object and in the presence of FPR and FNR, it is natural to define the attack cost, $c = c(p)$, as

$$c = [(1 - \alpha)q + \beta(1 - q)](1 - p), \tag{24}$$

which measures the fraction $1 - p$ of “functional” nodes that are attacked. The reason behind is that the action of attack incurs costs [36]. Moreover, we define the attack performance-cost ratio (PCR) when a fraction $1 - p$ of “functional” nodes are attacked as

$$\text{PCR} = \frac{1 - P_\infty}{c}, \tag{25}$$

which quantifies the efficiency of attacks. The quantity PCR in (25) measures the change of remaining network size excluding the giant component per “functional” node attacked. Hence, a higher value of PCR indicates a more efficient attack, which may cause more severe damages and is likely exploited by malicious attackers. See also Sect. 5 for a discussion regarding PCR.

3 Results

In this section, we conduct analytical analysis and numerical calculations of the theoretical expressions obtained above to better appreciate and test the effect of FPR and FNR on attack robustness for some model networks. All the simulation results are obtained for networks with $N = 10^6$ nodes.

The networks are generated via configuration model [3,32] with a prescribed degree distribution detailed below. A pair of labels (L_1, L_2) is associated with each node in the network. We initially label each node as failed with probability $1 - q$ and functional with probability q in L_1 independently. Each functional(L_1) node is labeled as failed with probability α and

functional with probability $1 - \alpha$ in L_2 independently. Similarly, each failed(L_1) node is labeled as failed with probability $1 - \beta$ and functional with probability β in L_2 independently. To obtain the percolation threshold p_c , we begin with $p = 1$ and a list of functional(L_2) nodes. We take nodes progressively from the list according to RA, LA, TA strategies respectively, and change its state in L_2 to failed with probability $1 - p$; each failed(L_2) node is then deleted together with its incident links. After checking the whole list, we then remove all remaining failed(L_1) nodes and their incident links in the network, calculate the fraction P_∞ of the giant component. We reduce p and repeat the process until $P_\infty < 10^{-3}$.

Due to the possible randomness in the node labeling/selection operation of the attack strategies, the plotted results are averaged over 100 independent simulation runs for each of the randomly generated networks. The error bars of the 100 simulation runs with respect to their average are less than 0.01 in all of the simulation cases. We do not plot the error bars for the sake of having a better visual effect.

3.1 Erdős-Rényi Networks

An ER network follows a Poisson degree distribution $P(k) = e^{-\lambda} \lambda^k / k!$ ($k \geq 0$) with average degree λ . Therefore, $G_0(x) = G_1(x) = e^{\lambda(x-1)}$. For RA, it follows easily from (4) and (5) that

$$p_c(\text{RA}) = \min \left\{ \max \left\{ 0, \frac{1 - \alpha \lambda q}{(1 - \alpha) \lambda q} \right\}, 1 \right\}, \tag{26}$$

and

$$P_\infty(\text{RA}) = [\alpha q + p(1 - \alpha)q] \left\{ 1 - e^{\lambda[H_1(1)-1]} \right\}, \tag{27}$$

where $H_1(1)$ satisfies $H_1(1) = 1 - \alpha q - p(1 - \alpha)q + [\alpha q + p(1 - \alpha)q]e^{\lambda[H_1(1)-1]}$. For LA, we calculate that $\tilde{G}_0(x) = \tilde{G}_1(x) = e^{\lambda s(x-1)}$. Using (12) and (14), we obtain

$$p_c(\text{LA}) = \min \left\{ \max \left\{ 0, \frac{1 - \alpha \lambda q^2}{(1 - \alpha q) \lambda q} \right\}, 1 \right\}, \tag{28}$$

and

$$P_\infty(\text{LA}) = [p + \alpha q(1 - p)]q \cdot \left\{ 1 - e^{\lambda[p + \alpha q(1 - p)][\tilde{H}_1(1)-1]} \right\}, \tag{29}$$

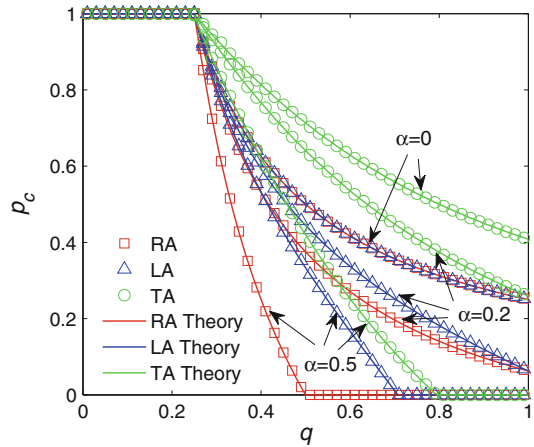
where $\tilde{H}_1(1)$ satisfies the transcendental equation $\tilde{H}_1(1) = 1 - q + q \cdot e^{\lambda[p + \alpha q(1 - p)][\tilde{H}_1(1)-1]}$. For TA with $\delta = 1$, using (18) and the fact that $\tilde{G}_0(x) = \tilde{G}_1(x) = e^{\lambda q(x-1)}$ we obtain $\hat{q} = q e^{\lambda q(g-1)}$. Hence, $p_c(\text{TA})$ is given by (22), where t_c is determined by $1 = \lambda q g^2 e^{\lambda q(g-1)}$ and $t = e^{\lambda q(g-1)}$ using (21). It follows from (23) that

$$P_\infty(\text{TA}) = pq + \alpha q(1 - p) - q e^{\lambda q(g-1) + \lambda q g^2 e^{\lambda q(g-1)}(u-1)}, \tag{30}$$

where u satisfies $u = \exp\{\lambda q(u - 1)(p + \alpha(1 - p))\{1 + \frac{1}{\lambda q} \ln[p + \alpha(1 - p)]\}^2\}$ and g is determined by $p + \alpha(1 - p) = e^{\lambda q(g-1)}$.

When $\alpha = 0$, i.e., no false positives exist, it can be proved that $p_c(\text{RA}) = p_c(\text{LA})$ and $P_\infty(\text{RA}) = P_\infty(\text{LA})$ by using a transformation $H_1(1) - 1 = p[\tilde{H}_1(1) - 1]$ (c.f. Figs. 2, 3). When $q = 1$, i.e., all nodes are functional in the initial network, we can show in a similar way that $p_c(\text{RA}) = p_c(\text{LA})$ and $P_\infty(\text{RA}) = P_\infty(\text{LA})$ (c.f. Figs. 2, 3). Interestingly, these mean

Fig. 2 Percolation threshold p_c as a function of initial functional probability q for ER networks with size $N = 10^6$ and $\lambda = 4$ under different FPR $\alpha = 0, 0.2,$ and 0.5 . Theoretical predictions (solid lines) and simulations (symbols) for RA, LA, and TA (with $\delta = 1$), respectively, agree well with each other



that RA and LA have exactly the same attack power for ER networks when either no false positives exist or no initial failed nodes exist. The special case of having both $\alpha = 0$ and $q = 1$ is previously observed in [18], and the equivalence of the two strategies is interpreted by a close competition between two factors behind LA. We will see below that this interpretation can be largely extended in two directions.

Figure 2 shows the behavior of the critical value p_c under RA, LA, and TA (with $\delta = 1$) for a variety of values of α . The results gathered in Fig. 2 allow us to draw several interesting comments. First, as expected from the above theoretical results, an increase in the initial functional probability q systematically yields an decrease in p_c for all attack strategies and values of α considered. The shared transition point of $p_c = 1$ at $q = 1/G'_1(1) = 1/\lambda$ marks the percolation threshold of the initial network, where a giant component first forms [6]. Second, for any given value of q , the threshold p_c is found to decrease with α for all attack strategies because more functional nodes are prevented from being attacked when α becomes larger. Third, when $\alpha = 0$, namely, all functional nodes are exposed to the attacker, we have $p_c(\text{RA}) = p_c(\text{LA})$ as commented above. Note that the initial failure is random; in other words the functional nodes in the initial network still constitutes an ER network with average degree λq . Therefore, the two competitive factors behind LA, namely, the factor due to heterogeneity that hubs are more likely within the attacked area accelerating the network fragmentation and the factor due to localization that only nodes on the surface of the attacked area contribute to the breakdown mitigating the fragmentation process, compensate for each other in ER networks as observed in [18]. In a like manner, when $q = 1$, the network facing the attacker is again an ER network, as each node is regarded as “failed” with probability α at random. Fourth, for $\alpha > 0$, we observe in general that $p_c(\text{TA}) > p_c(\text{LA}) > p_c(\text{RA})$ indicating that TA (with $\delta=1$ here) is the most powerful attack while RA (i.e., TA with $\delta = 0$) is the least. The interesting observation that LA is in-between can be explained as follows. The high degree nodes are more likely to be attacked under TA as compared to LA;² while the functional nodes in the attacked hole³ under LA, i.e., those nodes that are checked by

² In general, for $q < 1$ and $\alpha > 0$, there exists a $\delta^* = \delta^*(\lambda)$ such that the inequality $p_c(\text{TA}(\delta)) > p_c(\text{LA})$ holds for all $\delta > \delta^*$. Note that δ^* increases with λ since there is a higher probability that hubs in a denser network are attacked under LA. In principle, the value of δ^* can be analytically determined in our framework.

³ Here, “attacked hole” means those nodes that are checked by the attacker during the LA process. It is slightly different from the attacked area defined in Sect. 2.2.

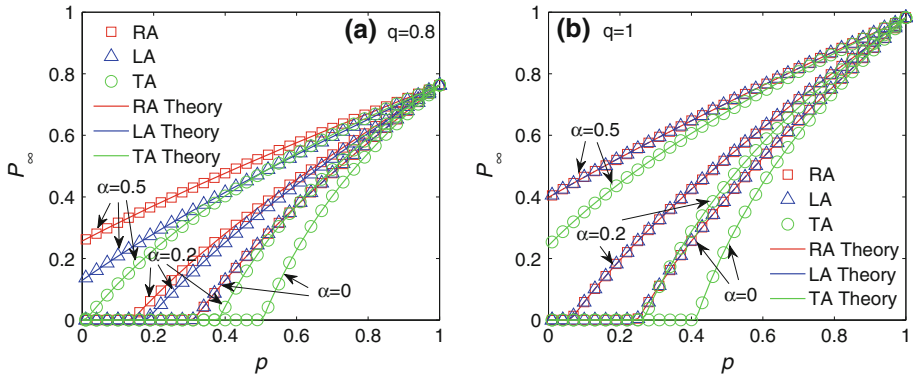


Fig. 3 Fraction of giant component P_∞ as a function of p for ER networks with size $N = 10^6$ and $\lambda = 4$ under different FPR $\alpha = 0, 0.2$, and 0.5 . **a** is for $q = 0.8$ and **b** is for $q = 1$. Theoretical predictions (solid lines) and simulations (symbols) for RA, LA, and TA (with $\delta = 1$), respectively, agree well with each other

the attacker but remain functional due to FPR, generally do not contribute to the integrity of the remaining network, which is hence more vulnerable than that under RA.⁴ Finally, all these attack strategies, in terms of p_c , turn out to be sensitive with respect to the change of α ; for example, RA with $\alpha = 0$ is actually more powerful than TA with $\alpha = 0.5$ for, say, all $q > 0.4$, highlighting the risk of ignoring FPR in understanding attack robustness.

We display in Fig. 3 the fraction of giant component P_∞ as a function of p under RA, LA, and TA for a variety of values of α and q . First, note that there are second-order percolation transition behaviors and the critical threshold at $P_\infty = 0$ (if exists) coincides with the critical probability p_c in Fig. 2 for all attack strategies and all α and q considered. When $q = 1$ and $\alpha = 0.5$, for example, $P_\infty(\text{RA})$, $P_\infty(\text{LA})$, and $P_\infty(\text{TA})$ are always positive (Fig. 3b). This means that even all “functional” nodes are attacked, the network maintains a giant component. This is in line with the results in Fig. 2, where $p_c(\text{RA}) = p_c(\text{LA}) = p_c(\text{TA}) = 0$ for $\alpha = 0.5$ at $q = 1$. Second, when $\alpha = 0$ (see Fig. 3a) or $q = 1$ (see Fig. 3b), we observe that $P_\infty(\text{RA}) = P_\infty(\text{LA})$ for all $p \in [0, 1]$ again due to the fact that in either case the attacker confronts an ER network generalizing the results of [18] along two directions. Third, for a given p we observe from Fig. 3a that $P_\infty(\text{RA}) > P_\infty(\text{LA}) > P_\infty(\text{TA}(\delta = 1))$ when $q < 1$ and $\alpha > 0$. In general, similarly as commented above, the inequality $P_\infty(\text{LA}) > P_\infty(\text{TA}(\delta))$ holds for a sufficiently large δ depending on λ .

In Fig. 4 we plot the behavior of PCR as a function of p for various combinations of α and β . We observe that PCR is quantitatively similar for all attack strategies considered; it grows steadily all the way to $p \approx 0.9$, and then jumps dramatically when p gets to approach 1.⁵ This is due to the fact that only a few nodes are attacked in the short beginning period but single nodes start to be peeled off pushing to a high PCR; when the attack continues, the giant component size does not shrink as fast as the linear increase of the number of attacked nodes, reducing the PCR. This is the case even without FPR or FNR [7,26]; and the similarity between the three attack strategies may find its origin in the structural homogeneity of ER networks.

⁴ For a given $q < 1$, the difference between $p_c(\text{LA})$ and $p_c(\text{RA})$ becomes more and more prominent when α increases from 0 to $\alpha^*, 1 := \min_{p_c(\text{RA})=0} \{\alpha\}$; it then decreases and reaches zero at the point $\alpha = \alpha^*, 2 := \min_{p_c(\text{LA})=0} \{\alpha\}$.

⁵ This is remarkably reminiscent of the celebrated law of diminishing marginal returns in economy. Here, the marginal attack efficiency in terms of PCR decreases rapidly as the attack progresses, namely, as p decreases.

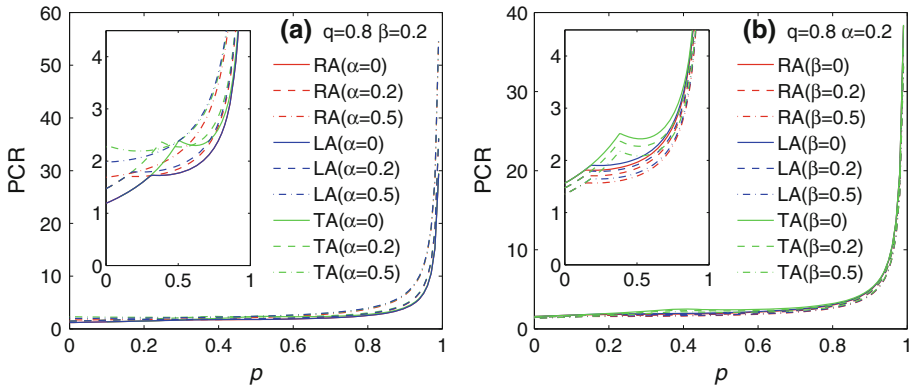


Fig. 4 Main panels: Performance-cost ratio PCR as a function of p for ER networks with size $N = 10^6$ and $\lambda = 4$ under RA, LA, and TA (with $\delta = 1$). **a** is for $q = 0.8, \beta = 0.2$; and **b** is for $q = 0.8, \alpha = 0.2$. Insets: Zoom-in view of the main panels

Despite the similarity, we note that for a given attack strategy, PCR is typically higher for larger α (Fig. 4a) but lower for larger β (Fig. 4b). The first observation is nontrivial as a larger α leads to a larger P_∞ (c.f. Fig. 3) but a smaller attack cost c (c.f. Eq.(24)). We performed extensive numerical calculations with different parameter combinations to confirm this. It can be understood as the effect of “protection” of α is not as marked as that of reducing the cost of the attacker. This also indicates that ignoring FPR would underestimate the harmfulness of a range of attacks including RA, LA, and TA. On the other hand, the influence of β is clear according to (25) since it does not affect P_∞ . In real situations, increasing β could exhaust the attacker and reduce the harm.

Furthermore, we observe interestingly that the PCR curves display small peaks or turning points, where the first derivatives are discontinuous. By comparing Fig. 4 with Figs. 2 and 3, we find that the peaks occur exactly at the corresponding percolation thresholds p_c . For example, when $q = 0.8$ and $\alpha = 0.2$, the peaks for TA (see the three green curves in Fig. 4b) appear at around 0.38, which is in consistent with the corresponding p_c . This phenomenon can be explained as follows. At p_c the giant component collapses into small components contributing substantially to PCR. We performed extensive numerical calculations for different combinations of q, α, β and λ , which produce quantitatively similar phenomena.

3.2 Random Regular Networks

A RR network has a degenerated degree distribution with each node linking to k_0 neighbors. Accordingly, $G_0(x) = xG_1(x) = x^{k_0}$. For RA, it follows from (4) and (5) that

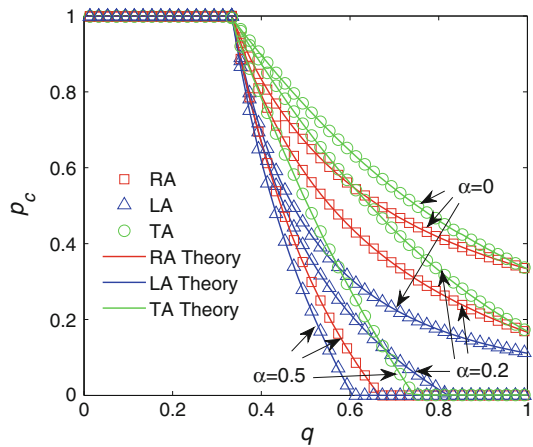
$$p_c(\text{RA}) = \min \left\{ \max \left\{ 0, \frac{1 - \alpha q(k_0 - 1)}{(1 - \alpha)q(k_0 - 1)} \right\}, 1 \right\}, \tag{31}$$

and

$$P_\infty(\text{RA}) = [\alpha q + p(1 - \alpha)q][1 - H_1(1)^{k_0}], \tag{32}$$

where $H_1(1)$ satisfies $H_1(1) = 1 - \alpha q - p(1 - \alpha)q + [\alpha q + p(1 - \alpha)q]H_1(1)^{k_0-1}$. For LA, we calculate that $\tilde{G}_0(x) = \left\{ 1 + [p + \alpha q(1 - p)] \frac{k_0-2}{k_0} (x - 1) \right\}^{k_0}$ and $\tilde{G}_1(x) = \tilde{G}_0(x)^{\frac{k_0-1}{k_0}}$.

Fig. 5 Percolation threshold p_c as a function of initial functional probability q for RR networks with size $N = 10^6$ and $k_0 = 4$ under different FPR $\alpha = 0, 0.2,$ and 0.5 . Theoretical predictions (solid lines) and simulations (symbols) for RA, LA, and TA (with $\delta = 1$), respectively, agree well with each other



Employing (12) and (14), we obtain

$$p_c(\text{LA}) = \min \left\{ \max \left\{ 0, \frac{[q(k_0 - 1)]^{-\frac{k_0}{k_0-2}} - \alpha q}{1 - \alpha q} \right\}, 1 \right\}, \tag{33}$$

and

$$P_\infty(\text{LA}) = [p + \alpha q(1 - p)]q \cdot \left\{ 1 - \left[1 + [p + \alpha q(1 - p)]^{\frac{k_0-2}{k_0}} [\tilde{H}_1(1) - 1] \right]^{k_0} \right\}, \tag{34}$$

where $\tilde{H}_1(1)$ satisfies the transcendental equation $\tilde{H}_1(1) = 1 - q + q \cdot \left\{ 1 + [p + \alpha q(1 - p)]^{\frac{k_0-2}{k_0}} [\tilde{H}_1(1) - 1] \right\}^{k_0-1}$. For TA with $\delta = 1$, using (18) and the fact that $\bar{G}_0(x) = (1 - q + qx)^{k_0}$ and $\bar{G}_1(x) = \bar{G}_0(x)^{\frac{k_0-1}{k_0}}$ we obtain $\hat{q} = g(1 - q + qg)^{k_0-1}$. Hence, $p_c(\text{TA})$ is given by (22), where t_c is determined by $1 = g^2 q(k_0 - 1)(1 - q + qg)^{k_0-2}$ and $t = (1 - q + qg)^{k_0}$ via (21). It follows from (23) that

$$P_\infty(\text{TA}) = [p + \alpha(1 - p)]q \cdot \left\{ 1 - \frac{(1 - q + qg)^{k_0} [1 + qg^2(1 - q + qg)^{k_0-2}(u - 1)]^{k_0}}{p + \alpha(1 - p)} \right\}, \tag{35}$$

where u satisfies $u = [1 + qg^2(u - 1)(1 - q + qg)^{k_0-2}]^{k_0-1}$ and g is determined by $p + \alpha(1 - p) = (1 - q + qg)^{k_0}$.

It can be easily checked that $p_c(\text{RA}) = p_c(\text{TA})$ and $P_\infty(\text{RA}) = P_\infty(\text{TA})$ for all α when $q = 1$ (c.f. Figs. 5 and 6b). This is because RA and TA have the same difference for RR networks when all nodes are functional at the beginning. However, this is not the case when $q < 1$ since the remaining network is no longer regular after an initial fraction $1 - q$ of nodes fail randomly.

In Fig. 5 we plot the behavior of the critical value p_c under RA, LA, and TA (with $\delta = 1$) for different values of α . First, similarly as in Fig. 2, an increase in the initial functional probability q systematically yields an decrease in p_c for all attack strategies and values of

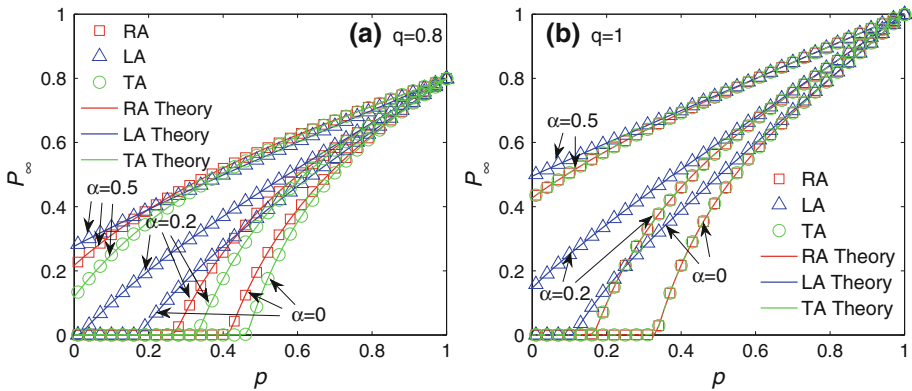


Fig. 6 Fraction of giant component P_∞ as a function of p for RR networks with size $N = 10^6$ and $k_0 = 4$ under different FPR $\alpha = 0, 0.2,$ and 0.5 . **a** is for $q = 0.8$ and **b** is for $q = 1$. Theoretical predictions (solid lines) and simulations (symbols) for RA, LA, and TA (with $\delta = 1$), respectively, agree well with each other

α . Here, the shared transition point of $p_c = 1$ at $q = 1/G'_1(1) = 1/(k_0 - 1)$ marks the percolation threshold of the initial network. Moreover, for any given attack strategy and value of q , the threshold p_c decreases as α increases because more functional nodes are prevented from being attacked when α becomes larger. Second, we have $p_c(\text{RA}) > p_c(\text{LA})$ for the whole spectrum of α and q (unless both are equal to zero when the network cannot be disintegrated by attacking even all “functional” nodes). This can be explained as follows. Since a fraction $1 - q$ of node are dysfunctional at random initially, the remaining network is still relatively homogeneous. Thus, the factor of localization in LA is more dominant and the underlying network becomes more robust against LA than against RA.⁶ The similar phenomenon in the special case of $q = 1$ and $\alpha = 0$ was observed in [18]. Obviously, TA (with $\delta > 0$) is the most powerful attack as expected. Finally, as in the ER case, all attack strategies considered seem to be responsive with respect to the change of α .

We next turn to the relative size of giant component P_∞ shown in Fig. 6. Note that, as in the case of ER networks, the critical threshold at $P_\infty = 0$ (if exists) coincides with the critical probability p_c in Fig. 5 for all attack strategies and all α and q considered. Moreover, when $q = 1$ we observe from Fig. 6b that $P_\infty(\text{TA}) = P_\infty(\text{RA})$ for all α and p as predicted by the mathematical analysis. Obviously, LA is the least powerful attack strategy since $P_\infty(\text{LA})$ is the highest curve for all α and p (see Fig. 6b) in line with the observation in Fig. 5. However, it is remarkable to note that $P_\infty(\text{LA}) < P_\infty(\text{TA}) < P_\infty(\text{RA})$ for $\alpha = 0.5$ and $p > 0.4$ in Fig. 6a. It is somewhat unexpected as LA is the least powerful attack in terms of the percolation threshold p_c (c.f. Fig. 5). This phenomenon can be explained as follows. When $q < 1$ and $\alpha > 0$, the network facing the attacker is no longer regular, the factor of localization in LA does not always play the leading role as compared to that of heterogeneity in LA. It is especially so during the early period of attack (namely, when p is large) because high degree nodes are prone to be attacked accelerating the fragmentation process. The factor of heterogeneity in LA even exceeds that in TA with $\delta = 1$ for $p > 0.4$ herein. This further highlights the subtlety of the competition between the two factors behind LA, which is not observed before.

⁶ However, such dominance is not universal. In terms of the giant component size, we may have $P_\infty(\text{RA}) > P_\infty(\text{TA})$ in some circumstances (c.f. Fig. 6a).

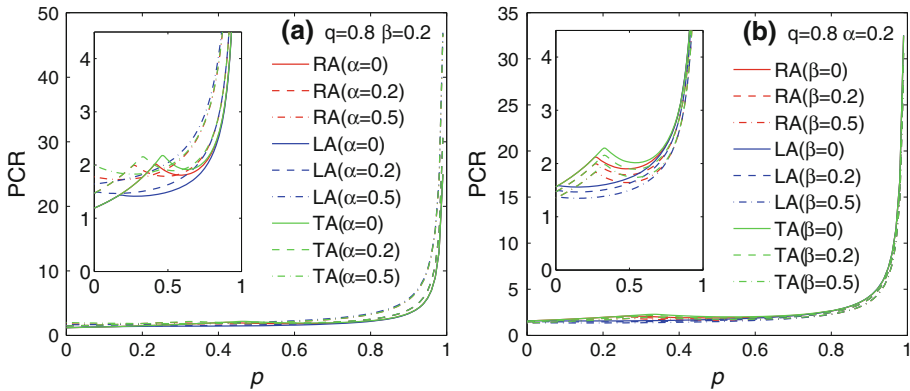


Fig. 7 Main panels: Performance-cost ratio PCR as a function of p for RR networks with size $N = 10^6$ and $k_0 = 4$ under RA, LA, and TA (with $\delta = 1$). **a** is for $q = 0.8, \beta = 0.2$; and **b** is for $q = 0.8, \alpha = 0.2$. Insets: Zoom-in view of the main panels

Fig. 8 Percolation threshold p_c as a function of initial functional probability q for SF networks with size $N = 10^6, \gamma = 2.47, k_{\min} = 2$, and $\langle k \rangle = 4.01$ under different FPR $\alpha = 0, 0.2$, and 0.5 . Theoretical predictions (solid lines) for RA, LA, and TA (with $\delta = 1$), respectively, agree well with each other

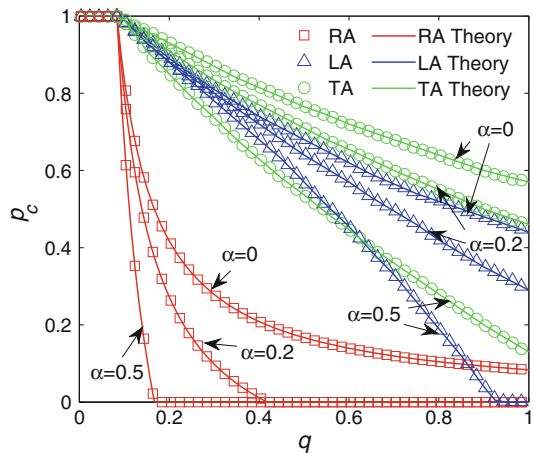


Figure 7 illustrates the behavior of PCR for various combinations of α and β . As in the case of ER networks, the PRC curves are quantitatively similar for all attack strategies considered. For a given attack strategy, PCR is typically higher for larger α (Fig. 7a) but lower for larger β (Fig. 7b). Similar comments for ER networks in Sect. 3.1 can also be applied here. Moreover, we observe that the PCR curves display turning points at the corresponding percolation thresholds p_c (c.f. Fig. 5). They are caused by the disintegration of the giant component. As in the case of ER networks, we performed extensive numerical calculations for different values of q, α, β and k_0 to verify these phenomena. Finally, comparing Fig. 7 with Fig. 4, we observe that the peak values for RR networks at around $p = 1$ are slightly lower than those for ER networks. This is reasonable because RR networks often have a lattice-like architecture, which makes the attacks less effective [3, 18].

3.3 Scale-Free Networks

A SF network follows a power-law degree distribution $P(k) \sim k^{-\gamma}$ ($k_{\min} \leq k \leq k_{\max}$), where $\gamma > 0$ is the scaling exponent, k_{\min} and k_{\max} represent the minimum and maximum

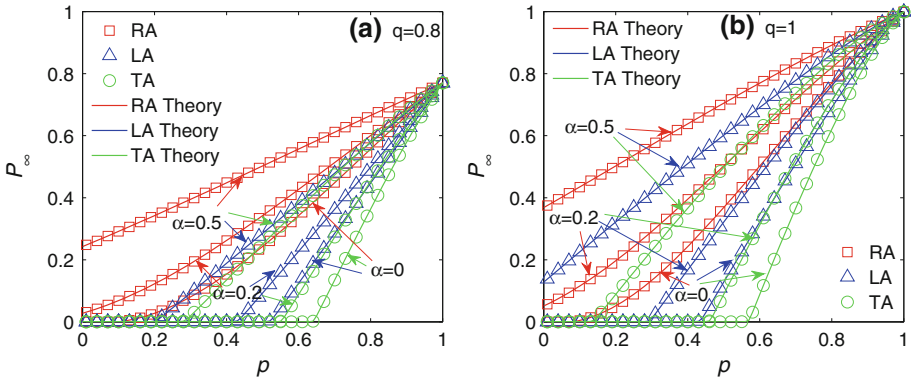


Fig. 9 Fraction of giant component P_∞ as a function of p for SF networks with size $N = 10^6$, $\gamma = 2.47$, $k_{\min} = 2$, and $\langle k \rangle = 4.01$ under different FPR $\alpha = 0, 0.2$, and 0.5 . **a** is for $q = 0.8$ and **b** is for $q = 1$. Theoretical predictions (solid lines) and simulations (symbols) for RA, LA, and TA (with $\delta = 1$), respectively, agree well with each other

degrees, respectively. We show in Fig. 8 the behavior of the critical value p_c for SF networks under RA, LA, and TA (with $\delta = 1$) with different values of α . Note that we use networks with approximately the same average degree, i.e., $\langle k \rangle \approx 4$, in all simulations. Comparing Fig. 8 with the results reported in Figs. 2 and 5, we are led to several interesting conclusions.

First, an increase in the initial functional probability q systematically yields an decrease in p_c for all attack strategies and values of α . The shared transition point of $p_c = 1$ at $q = 1/G'_1(1) \approx 0.08$ marks the percolation threshold of the initial network. Again, for any given attack strategy and value of q , the threshold p_c decreases as α increases. Second, for any given α and q , we observe that $p_c(\text{RA})$ is much smaller than $p_c(\text{LA})$ in sharp contrast to the cases of ER and RR networks. This phenomenon is due to the fact that SF networks are highly heterogeneous. The factor of heterogeneity in LA turns dominant; hubs are more likely to be attacked, which accelerate the fragmentation.⁷ Moreover, as in ER networks, the strategies LA and TA are generally comparable. For given α, q and a SF network, we may determine a critical value $\delta^* \geq 0$ so that $p_c(\text{LA}) = p_c(\text{TA}(\delta = \delta^*))$ holds. For example, we observe from Fig. 8 that $p_c(\text{LA}) < p_c(\text{TA}(\delta = 1))$ at $q = 1$ for all α considered, which indicates that the corresponding δ^* in question must be smaller than 1. Third, as in the ER and RR cases, all attack strategies considered are quite sensitive with respect to the change of α and q . For example, the curve $p_c(\text{RA})$ with $\alpha = 0$ intersects the curve $p_c(\text{LA})$ with $\alpha = 0.5$ (see Fig. 8), which implies that the attack robustness of the SF network against RA and LA may switch as α and q change.⁸ These phenomena are noteworthy in attack robustness of real-world networked systems when FPR are uncertain or unknown.

In Fig. 9, we display the fraction of giant component P_∞ as a function of p for a range of values of α and q . As in ER and RR cases, we observe that there are second-order percolation

⁷ Since a SF network becomes less heterogeneous when the scaling exponent γ gets larger [3], we may determine, as in [18], a critical $\gamma_c = \gamma_c(\alpha, q)$ with which $p_c(\text{RA}) = p_c(\text{LA})$ holds. When $\gamma > \gamma_c$, the SF networks, similar to RR networks, become more robust against LA compared to RA. However, note that γ_c is no longer unique in our situation due to the existence of α and q . For example, when $\alpha = 0.5$ and $q = 0.98$, we know from Fig. 8 that $\gamma = 2.47$ is a candidate for $\gamma_c(0.5, 0.98)$, but values sufficiently close to 2.47 are candidates as well.

⁸ Note that there is another switch of attack power between RA and LA in terms of the power-law exponent γ . SF networks behave similarly as RR networks when γ is large (c.f. Fig. 11b and Ref. [18])

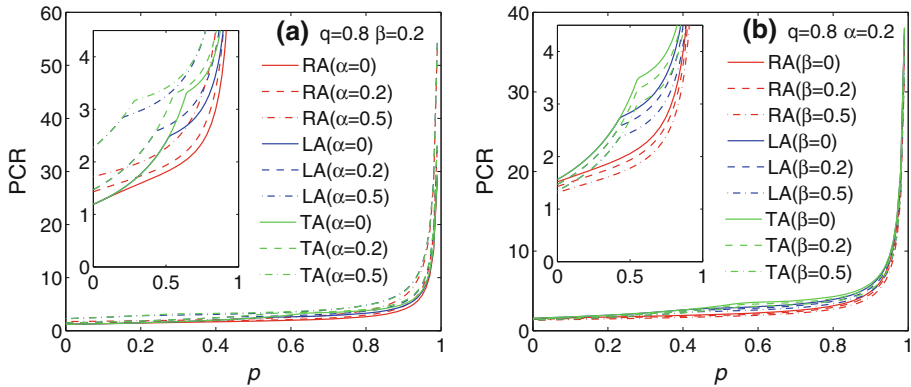


Fig. 10 Main panels: Performance-cost ratio PCR as a function of p for RR networks with size $N = 10^6$, $\gamma = 2.47$, $k_{\min} = 2$, and $\langle k \rangle = 4.01$ under RA, LA, and TA (with $\delta = 1$). **a** is for $q = 0.8$, $\beta = 0.2$; and **b** is for $q = 0.8$, $\alpha = 0.2$. Insets: Zoom-in view of the main panels

transition behaviors and the critical threshold at $P_\infty = 0$ (if exists) coincides with the critical probability p_c shown in Fig. 8 for all attack strategies and all α and q considered. Moreover, for a given p we observe that $P_\infty(\text{RA}) > P_\infty(\text{LA})$ and $P_\infty(\text{RA}) > P_\infty(\text{TA})$ for all q and α considered. This agrees with the results reported in Fig. 8 (in terms of p_c), indicating that the SF network is universally and strictly more resilient against RA than against LA and TA (with $\delta > 0$), namely, RA is the least powerful attack strategy.⁹ Recall that this is not true for either ER networks (c.f. Fig. 3) or RR networks (c.f. Fig. 6).

The PCR curves shown in Fig. 10 share some common patterns as compared to ER networks (Fig. 4) and RR networks (Fig. 7). Namely, (i) PCR grows steadily all the way to p around 0.9, and then increases rapidly when p is close to 1; (ii) PCR is higher for larger α but lower for larger β ; (iii) PCR curves display peaks/turning points at the corresponding percolation thresholds p_c for all attack strategies. However, from the insets of Fig. 10 we discern a major difference between TA(LA) and RA. We observe from Fig. 10b that PCRs for LA and TA (with $\delta = 1$) are obviously higher than PCR for RA; the height, for example, of the turning point for $\text{TA}(\beta = 0)$ is about 3.5, which is larger than the counterparts in ER networks (around 2.5) and RR networks (around 2.2). This is due to the heterogeneity of SF networks, rendering TA the most powerful attack among these attack strategies. We have performed extensive simulations for various values of δ , which shows that the height of PCR curves increases with δ as expected. Moreover, the theoretical framework provided here allows us to identify the detailed level of attack power (in terms of PCR) of LA ranging from RA, i.e., TA with $\delta = 0$, to TA with any large δ . For instance, Fig. 10b indicates that the attack power of LA is between RA and TA with $\delta = 1$ for all α , β , and q considered. These information can be instrumental for both attackers and defenders in choosing appropriate attack and defense strategies.

4 Applications

In this section we investigate the attack robustness of two real-world large-scale networks under RA, LA, and TA with false positives. The first network is an e-mail network (E-mail)

⁹ Again, this is true only for small γ . When γ is large, SF networks behave similarly as RR networks.

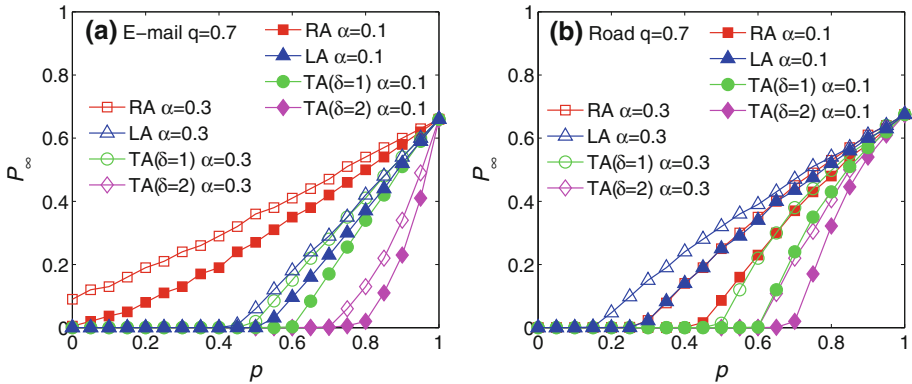


Fig. 11 Fraction of giant component P_∞ as a function of p for **a** E-mail network and **b** Road network under RA, LA, TA($\delta = 1$), and TA($\delta = 2$) with different FPR $\alpha = 0.1$ and $\alpha = 0.3$. Each data point is obtained by an ensemble averaging of 100 realizations

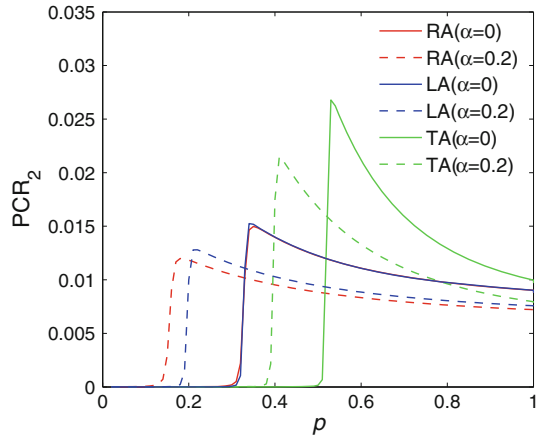
which contains $N = 56969$ nodes and average degree 2.96 [37,38]. The nodes of the e-mail network corresponds to e-mail addresses and a link between two addresses is established if an e-mail is exchanged between them. This network is collected at Kiel University in Germany over a period of 113 days, which is shown to be a scale-free network with scaling exponent $\gamma = 1.81$. The second network is a road network (Road) of Pennsylvania in USA, which contains $N = 1.087 \times 10^6$ nodes [39]. The nodes of the network are the intersections between roads and the edges are road segments between intersections. This network has a power-law exponent 8.99 with average degree 2.83, which means that it is no longer a scale-free network but falls in some class of random planar network [40,41].

In Fig. 11, we plot the behavior of relative size of giant component P_∞ as a function of p under various attack strategies and false positives. Comparing Fig. 11a with Fig. 9a, we observe that E-mail has the signature of a typical scale-free network, in which $P_\infty(\text{RA})$ is evidently higher than $P_\infty(\text{TA})$ and $P_\infty(\text{LA})$ for all α considered indicating that E-mail network is more robust against RA than against TA an LA. LA is seen to be comparable with TA with $\delta = 1$. Moreover, E-mail network becomes more vulnerable when we increase δ from 1 to 2, as one would expect, due to the heterogeneity nature of the network. For example, when $\alpha = 0.1$ and $p = 0.8$, i.e., when 20 percent of “functional” nodes are attacked, the network almost collapses under TA with $\delta = 2$ but maintains one-half of the giant component under TA with $\delta = 1$. On the other hand, Fig. 11b shows that Road network is more robust against LA than against RA and TA regardless of α , which is quantitatively similar to the behavior of RR networks (c.f. Fig. 6). In addition, we observe interestingly that the curve $P_\infty(\text{RA})$ with $\alpha = 0.3$ is nearly overlapped with $P_\infty(\text{LA})$ with $\alpha = 0.1$. This means that LA launched on the Road network may be mimicked perfectly by RA with appropriate false positives. A sophisticated attacker could take advantage of this “bug” for disrupting infrastructure systems such as the one considered here.

5 Discussion

We have presented a general framework for coping with attack robustness in the presence of false positives and false negatives, within which three important classes of attacks, namely,

Fig. 12 Transformed performance-cost ratio PCR_2 as a function of p for ER networks with size $N = 10^6$ and $\lambda = 4$ under RA, LA, and TA (with $\delta = 1$). We set $q = 0.8$ and $\beta = 0.2$



RA, LA, and TA are studied. We show how false positives and false negatives can influence the network robustness in terms of percolation thresholds, giant component sizes, as well as the performance-cost ratios. In general, ER, RR, and SF networks have their respective features in response to these attacks, showing a good sensitivity with respect to false positives.

As shown in Figs. 4, 7, and 10, the PCR value reaches local maximum when the giant connected component vanishes at p_c and fails to clearly differentiate varieties of attacks in the initial stage of the attack (i.e., in the regime of p close to 1). An intuitive quantification of the benefit-cost ratio is that it should be maximized at p_c in the whole range of $p \in [0, 1]$. By applying a linear operator on PCR, we may define a new measure as

$$PCR_2 = \frac{d}{dp} \left(\frac{P_\infty}{1 - c} \right), \tag{36}$$

which quantifies the change rate of the ratio of the giant component size to the number of unattacked nodes with respect to p . Although the physical meaning of PCR_2 is not as straightforward as PCR, a higher value of PCR_2 intuitively implies a more harmful attack. The PCR_2 values for ER networks are shown in Fig. 12, which very well agree with our above intuition of benefit-cost ratio; PCR_2 is maximized when the network is broken down into pieces at the critical threshold p_c . However, from (36) and (24), we see that a weakness of this new measure is that it is not sensitive to the FNR, i.e., β , when q is close to 1 (in fact, the curves for other values of β are very close to those shown in Fig. 12). But this is the most interesting regime in practice. It would be highly desirable to construct other viable options.

Our work is a first step towards the systematic study of false positive and false negative effects, which assumes that the attack strategy is set once and for all. It remains an open problem to determine the FPR/FNR effect when adaptive attacks or combined attacks are launched. Such sophisticated tactics, which are extremely harmful to the network architectures, have recently been identified in a number of computer networks [42]. Furthermore, from the perspective of network theory, generalizations to interdependent networks as well as correlations within the networks (such as assortativity) are appealing because such correlations are known to affect the dynamics and percolation thresholds prominently; see e.g. [43–45]. In particular, the understanding of the interplay between FPR/FNR and the inter-layer dependence could shed light on a wealthy of attack robustness phenomena in the real world.

Acknowledgements The author is grateful to the anonymous referees for their valuable comments and suggestions that have greatly improved the presentation of the paper. The work is funded in part by the National Natural Science Foundation of China (11505127), the Shanghai Pujiang Program (15PJ1408300), and the Program for Young Excellent Talents in Tongji University (2014KJ036).

Appendix: Equivalence of RA and TA when $\delta = 0$

Fix $\delta = 0$. It follows from (18) that $\hat{q} = g$. Therefore, Eq. (21) reduces to

$$\begin{cases} \sum_{k=0}^{\infty} \bar{P}(k)k = g \sum_{k=2}^{\infty} \bar{P}(k)k(k-1), \\ g = \bar{G}_{\delta=0}^{-1}(t) = t, \end{cases} \quad (37)$$

which implies $t_c = 1/[qG'_1(1)]$. Comparing (4) with (22) we obtain $p_c(\text{RA}) = p_c(\text{TA})$.

From (16), (19) and (23) we obtain

$$\begin{aligned} P_{\infty}(\text{TA}) = tqS(\text{TA}) &= tq \left\{ 1 - \sum_{k=0}^{\infty} \bar{P}(k)[1 + g(u-1)]^k \right\} \\ &= tq[1 - G_0(1 + qg(u-1))], \end{aligned} \quad (38)$$

where u is determined by

$$u = \hat{G}_1(u) = \bar{G}_1(1 + g(u-1)) = G_1(1 + qg(u-1)). \quad (39)$$

Note that (39) can be recast as $1 + qg(u-1) = 1 - qg + qgG_1(1 + qg(u-1))$ and $g = t = p + \alpha(1-p) = \alpha + p(1-\alpha)$. Hence, comparing (5) with (38) we conclude $P_{\infty}(\text{RA}) = P_{\infty}(\text{TA})$.

References

1. Watts, D., Strogatz, S.: Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998)
2. Barabási, A., Albert, R.: Emergence of scaling in random networks. *Science* **286**, 509–512 (1999)
3. Newman, M.: *Networks: An Introduction*. Oxford University Press, New York (2010)
4. Dorogovtsev, S.N., Mendes, J.F.F.: *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, New York (2013)
5. Albert, R., Jeong, H., Barabási, A.: Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000)
6. Callaway, D.S., Newman, M., Strogatz, S.H., Watts, D.J.: Network robustness and fragility: percolation on random graphs. *Phys. Rev. Lett.* **85**, 5468–5471 (2000)
7. Magnien, C., Latapy, M., Guillaume, J.-L.: Impact of random failures and attacks on Poisson and power-law random graphs. *ACM Comput. Surv.* **43**, 13 (2011)
8. Shang, Y.: Subgraph robustness of complex networks under attacks. *IEEE Trans. Syst. Man Cybern. Syst.* <https://doi.org/10.1109/TSMC.2017.2733545>
9. Cohen, R., Erez, K., ben-Avraham, D., Havlin, S.: Breakdown of the internet under intentional attack. *Phys. Rev. Lett.* **86**, 3682–3685 (2001)
10. Shang, Y.: Vulnerability of networks: fractional percolation on random graphs. *Phys. Rev. E* **89**, 012813 (2014)
11. Buldyrev, S.V., Parshani, R., Paul, G., Stanley, H.E., Havlin, S.: Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028 (2010)
12. Radicchi, F.: Percolation in real interdependent networks. *Nat. Phys.* **11**, 597–602 (2015)
13. Liu, X., Stanley, H.E., Gao, J.: Breakdown of interdependent directed networks. *Proc. Nat. Acad. Sci.* **113**, 1138–1143 (2016)
14. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: *Bayesian Data Analysis*, 3rd edn. Chapman & Hall/CRC, Boca Raton (2013)
15. Kreutz, D., Ramos, F.M.V., Verissimo, P.E., Rothenberg, C.E., Azodolmolky, S., Uhlig, S.: Software-defined networking: a comprehensive survey. *Proc. IEEE* **103**, 14–76 (2015)

16. Yu, S., Zhao, G., Dou, W., James, S.: Predicted packet padding for anonymous Web browsing against traffic analysis attacks. *IEEE Trans. Inf. Forensics Secur.* **7**, 1381–1393 (2012)
17. Wang, L., Ren, S., Korel, B., Kwiat, K.A.: Improving system reliability against rational attacks under given resources. *IEEE Trans. Syst. Man, Cybern. A, Syst. Humans* **44**, 446–456 (2014)
18. Shao, S., Huang, X., Stanley, H.E., Havlin, S.: Percolation of localized attack on complex networks. *New J. Phys.* **17**, 023049 (2015)
19. Berezin, Y., Bashan, A., Danziger, M.M., Li, D., Havlin, S.: Localized attacks on spatially embedded networks with dependencies. *Sci. Rep.* **5**, 8934 (2015)
20. Yuan, X., Shao, S., Stanley, H.E., Havlin, S.: How breadth of degree distribution influences network robustness: comparing localized and random attacks. *Phys. Rev. E* **92**, 032122 (2015)
21. Yuan, X., Dai, Y., Stanley, H.E., Havlin, S.: k -core percolation on complex networks: comparing random, localized, and targeted attacks. *Phys. Rev. E* **93**, 062302 (2016)
22. Shang, Y.: Localized recovery of complex networks against failure. *Sci. Rep.* **6**, 30521 (2016)
23. Iyer, S., Killingback, T., Sundaram, B., Wang, Z.: Attack robustness and centrality of complex networks. *PLoS ONE* **8**, e59613 (2013)
24. Chen, G., Dong, Z.Y., Hill, D.J., Xue, Y.S.: Exploring reliable strategies for defending power systems against targeted attacks. *IEEE Trans. Power Syst.* **26**, 1000–1009 (2011)
25. Gallos, L.K., Cohen, R., Argyrakis, P., Bunde, A., Havlin, S.: Stability and topology of scale-free networks under attack and defense strategies. *Phys. Rev. Lett.* **94**, 188701 (2005)
26. Bollobás, B.: *Random Graphs*. Springer, Berlin (1998)
27. Ventresca, M., Aleman, D.: Network robustness versus multi-strategy sequential attack. *J. Comput. Netw.* **3**, 126–146 (2015)
28. Podobnik, B., Horvatic, D., Lipic, T., Perc, M., Buldú, J.M., Stanley, H.E.: The cost of attack in competing networks. *J. R. Soc. Interface* **12**, 20150770 (2015)
29. Cohen, R., Havlin, S., ben-Avraham, D.: Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.* **91**, 247901 (2003)
30. Pastor-Satorras, R., Castellano, C., Van Mieghem, P., Vespignani, A.: Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015)
31. Chen, C., Tong, H., Prakash, G.A., Tsourakakis, C.E., Eliassi-Rad, T., Faloutsos, C., Chau, D.H.: Node immunization on large graphs: theory and algorithms. *IEEE Trans. Know. Data Eng.* **28**, 113–126 (2016)
32. Newman, M., Strogatz, S.H., Watts, D.J.: Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118 (2001)
33. Shao, J., Buldyrev, S.V., Braunstein, L.A., Havlin, S., Stanley, H.E.: Structure of shells in complex networks. *Phys. Rev. E* **80**, 036105 (2009)
34. Huang, X., Gao, J., Buldyrev, S.V., Havlin, S., Stanley, H.E.: Robustness of interdependent networks under targeted attack. *Phys. Rev. E* **83**, 065101(R) (2011)
35. Shang, Y.: Degree distribution dynamics for disease spreading with individual awareness. *J. Syst. Sci. Complex.* **28**, 96–104 (2015)
36. Wang, S., Zhang, Z., Kadobayashi, Y.: Exploring attack graph for cost-benefit security hardening: a probabilistic approach. *Comput. Secur.* **32**, 158–169 (2013)
37. Ebel, H., Mielsch, L.-I., Bornholdt, S.: Scale-free topology of e-mail networks. *Phys. Rev. E* **66**, 035103(R) (2002)
38. Braha, D., Bar-Yam, Y.: From centrality to temporary fame: dynamic centrality in complex networks. *Complexity* **12**, 59–63 (2006)
39. Leskovec, J., Lang, K., Dasgupta, A., Mahoney, M.W.: Community structure in large networks: natural cluster sizes and the absence of larger well-defined clusters. *Internet Math.* **6**, 29–123 (2009)
40. Amaral, L.A., Scala, A., Barthélemy, M., Stanley, H.E.: Classes of small-world networks. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149–11152 (2000)
41. Barthélemy, M.: Spatial networks. *Phys. Rep.* **499**, 1–101 (2011)
42. Al-Hamadi, H., Chen, R.: Adaptive network defense management for countering smart attack and selective capture in wireless sensor networks. *IEEE Trans. Netw. Service Manag.* **12**, 451–466 (2015)
43. D’Agostino, G., Scala, A., Zlatić, V., Caldarelli, G.: Robustness and assortativity for diffusion-like processes in scale-free networks. *EPL* **97**, 68006 (2011)
44. Zhou, D., Stanley, H.E., D’Agostino, G., Scala, A.: Assortativity decreases the robustness of interdependent networks. *Phys. Rev. E* **86**, 066103 (2012)
45. Tyra, A., Li, J., Shang, Y., Jiang, S., Zhao, Y., Xu, S.: Robustness of non-interdependent and interdependent networks against dependent and adaptive attacks. *Phys. A* **482**, 713–727 (2017)