Percolation of attack with tunable limited knowledge

Yilun Shang
Department of Computer and Information Sciences,
Northumbria University, Newcastle upon Tyne, NE1 8ST, UK

Percolation models shed a light on network integrity and functionality and have numerous applications in network theory. This paper studies a targeted percolation (α-model) with incomplete knowledge, where the highest degree node in a randomly selected set of n nodes is removed at each step and the model features a tunable probability that the removed node is instead a random one. A 'mirror image' process (β-model), in which the target is the lowest degree node, is also investigated. We analytically calculate the giant component size, the critical occupation probability and the scaling law for percolation threshold with respect to the knowledge level n under both models. We also derive self-consistency equations to analyze the k-core organization including the size of k-core and its corona in the context of attacks under tunable limited knowledge. These percolation models are characterized by some interesting critical phenomena and reveal profound quantitative structure discrepancies between Erdős-Rényi networks and power-law networks.

PACS numbers:

I. INTRODUCTION

Network science has proven of incredible ability to study the structure, dynamics and function of networked systems across the sciences. Resilience (robustness) plays a crucial role in natural and man-made complex systems, where the interconnection topology underpinning their proper functionality is subject to varied kinds of failures and attacks [1–3]. Scale-free networks, for example, are shown to be fragile under attacks targeted at hub nodes while remain robust against random attacks in terms of the relative size P_∞ of giant component or cluster in the network [4]. Percolation models describing the structural and topological properties of clusters have been widely used for understanding and evaluating the network resilience. Single networks under ordinary node and bond percolation undergo a continuous phase transition at some critical occupation probability p_c [3, 5], while percolation over networks of networks often causes cascading failures yielding an abrupt first-order phase transition [6]. Many modified percolation strategies such as L-hop percolation [7], bootstrap percolation [8, 9] and localized percolation [10, 11], have been studied in complex networks motivated by different practical limitations and applicable scenarios.

Much work has focused on intentional attacks at nodes with highest degree or other centrality measures assuming complete knowledge of the network [4, 5, 12–14]. However, one realistic restriction that should be imposed on many attackers is their limited ability to observe full information of the entire network. For example, cyber criminals on the internet can only probe the status of some servers and routers and immunization decision-makers during an epidemic such as COVID-19 are not able to trace social contacts of each individual in a population [15]. Liu et al. [16] recently studied an attack strategy under limited knowledge, where only degrees of n nodes are observed at each time and the most connected one is removed. Under this framework, taking n as the network size recovers the traditional targeted attack and the other extremal case of $n = 1$ yields the random attack scenario.

Despite these advances, we still lack understanding of how incomplete knowledge of the network during an attack can affect the network robustness. Here, we develop theoretical frameworks for understanding attack strategies with limited knowledge covering not only the benefit-oriented scenario in [16] but a cost-oriented model, where the least connected node among the n randomly observed nodes is removed. Attack cost has been proposed to be positively correlated with node degree in [17]. In this context, aiming at low-degree nodes can effectively lower the attacker's cost. Network dismantling strategies minimizing attack cost have been extensively studied under a variety of practical considerations during the last years [18–21]. Moreover, our framework features a tunable extent of orientation, meaning that at each round of attack, the cost-oriented attacker removes the least connected node among the n observed ones with probability α and a random node otherwise (α-model), and similarly the benefit-oriented attacker can choose the most connected node among the observed ones with probability β and a random node otherwise (β-model). We investigate how percolation under tunable limited knowledge influences the giant component P_∞, the critical threshold p_c and derive analytical relations for the percolation threshold with respect to the knowledge level n in both models for networks with arbitrary degree distributions.

Apart from the giant component fraction P_∞, another important functional component of a network is its k-core [22]. The k-core of a network is defined as the largest subgraph with all node degrees no less than k. It can be obtained by recursively peeling off nodes with degree less than k from the remaining network. Unlike ordinary percolation, k-core structure emerges at the criticality discontinuously for $k \geq 3$ [23, 24], resembling features of a cascading failure. As an important centrality measure quantifying number of interconnections, k-core has found numerous applications in ecological, social, and financial systems [22, 25, 26] and some generalizations have been proposed [27, 28]. We here further investigate the k-core subgraph organization of our α- and β-models under attack with limited knowledge. The studied properties include the relative size of k-core N_{kc}, the average degree D_{kc} of k-core, as well as the relative size of corona $N_k(k)$, which is a subgraph of k-core and has degree exactly k within the k-core [24]. In all cases, our theory has good agreement with simulation results, and we observe distinct qualitative characteristics of percolation phase transition and k-core organization for Erdős-Rényi (ER) and power-law (PL) networks.

II. PERCOLATION UNDER TUNABLE LIMITED KNOWLEDGE

We consider a random network $G(V,E)$, where V represents the node set and E is the edge set. Let $|V|=N$ be the number of nodes in the network, which has a degree distribution $P(q)$ following the configuration model [3]. At time step $t=0$, we denote by $P(q;0)=P(q)$ the initial degree distribution for a random node. The generating function for the degree distribution is defined as $G_0(x)=\sum_{q=0}^{\infty}P(q)x^q$ and the corresponding generating function for the excess degrees is $G_1(x)=G_0'(x)G_0'(1)^{-1}$, where $G_0'(1)=\langle q\rangle$ represents the mean degree of a node in $G(V,E)$ [5]. Moreover, given probability $P(q)$, the cumulative distribution, namely the probability that a randomly chosen node has degree at most q, is seen to be given by $F(q;0)=F(q)=\sum_{s=0}^{q}P(s)$ for $q\geq 0$.

A. Analytical framework for α-model

Given $\alpha\in[0,1]$, the cost-oriented attacker at each time step is assumed to observe n random nodes in our α-model. We remove the one with lowest degree with probability α and remove a random node of them with probability $1-\alpha$. This process continues until a fraction of $1-p$ nodes are removed from $G(V,E)$. Clearly, the parameter α quantifies the degree of cost-orientation of the attacker. The case of $\alpha=1$ corresponds to the most conservative attack and the case of $\alpha=0$ is the same as random attack because removing a random node from the n randomly observed nodes is the same as attacking a random node from the whole network.

To determine the relative size P_∞ of the giant component and the critical value p_c, we assume that only nodes are deleted but the edges linking the removed nodes to the remaining network are kept intact. We will delete these edges in a later stage. Formally, let $P(q;t)$ be the degree distribution of a random node in the remaining network at step $t\geq 0$. Following the idea of [16], we will make use of the associated cumulative distribution $F(q;t)=\sum_{s=0}^{q}P(s;t)$, namely, the probability that a randomly chosen remaining node at step t has degree no more than q.

With the help of minimum order statistics for n independent random variables (e.g. [29, Thm. 8.1]), the degree distribution $P_a^{\min}(q;t)$ of the observed lowest-degree node at time step t is

$$\begin{aligned}&[1-(1-F(q;t))^n]-[1-(1-F(q-1;t))^n]\\ =&\Delta\left(1-(1-F(q;t))^n\right)=-\Delta\left((1-F(q;t))^n\right),\end{aligned}\quad (1)$$

for $q\geq 0$, where Δ represents the difference operator with respect to q. We set $F(-1;t)=0$ for all t. By the definition of α-model, the degree distribution of the attacked node at step t is

$$P_a(q;t)=\alpha P_a^{\min}(q;t)+(1-\alpha)P_a^{\mathrm{ran}}(q;t),\quad (2)$$

where $P_a^{\mathrm{ran}}(q;t)=P(q;t)=\Delta F(q;t)$ is the probability of choosing a random node of degree q. Let $N(q;t)$ be the number of nodes of degree q in the remaining network at step t. With one more node being attached, we obtain

$$N(q;t+1)=N(q;t)-P_a(q;t)\quad (3)$$

by recalling the assumption that we only delete nodes but not edges.

It follows from (1), (2), and (3) in the continuous limit that

$$\begin{aligned}\frac{\partial N(q;t)}{\partial t}=&\alpha\Delta\left((1-F(q;t))^n\right)-(1-\alpha)\Delta F(q;t)\\ =&(N-t)\frac{\partial P(q;t)}{\partial t}-P(q;t)\end{aligned}\quad (4)$$

since $N(q;t)=(N-t)P(q;t)$. Hence,

$$\begin{aligned}\Delta\Bigg(&-F(q;t)+(N-t)\frac{\partial F(q;t)}{\partial t}\\ &-\alpha(1-F(q;t))^n+(1-\alpha)F(q;t)\Bigg)=0.\end{aligned}\quad (5)$$

Recalling $F(-1;t)=0$ for $t\geq 0$, we obtain from (5) that for $q\geq 0$,

$$\begin{cases}(N-t)\frac{\partial F(q;t)}{\partial t}=\alpha(1-F(q;t))^n+F(q;t)\\ \qquad\qquad -(1-\alpha)F(q;t)-\alpha,\quad t>0,\\ F(q;0)=F(q).\end{cases}\quad (6)$$

For $n>1$, by directly integrating (6) we obtain the solution

$$\begin{aligned}F(q;t)=&1-\Big(1+((1-F(q))^{1-n}-1)\\ &\cdot e^{\alpha(n-1)\ln\left(\frac{N-t}{N}\right)}\Big)^{-\frac{1}{n-1}}.\end{aligned}\quad (7)$$

Noticing that $(1-p)N=t$, we rewrite (7) as

$$F_p(q)=1-\Big(1+((1-F(q))^{1-n}-1)p^{\alpha(n-1)}\Big)^{-\frac{1}{n-1}},\quad (8)$$

which is the cumulative distribution of the degree of a random node when a $1-p$ fraction of nodes are removed (with all edges intact) in the α-model. In the case of $n=1$, it is easy to verify that the solution of the system (6) becomes $F_p(q)=F(q)$ for any $\alpha\in[0,1]$ as one would expect through taking the limit $n\to 1^+$ in (8). Moreover, when $\alpha=0$, we recover the random attack scenario from (8), namely $F_p(q)=F(q)$ for any $n\geq 1$.

Hence, the probability of a randomly chosen node in the remaining network having degree q, when a fraction of $1-p$ nodes are removed (but no edge is deleted), becomes

$$P_p(q)=\Delta F_p(q)=F_p(q)-F_p(q-1). \tag{9}$$

Let u be the probability that an edge does not connect to the giant component. Then

$$1-u=\sum_{q=0}^{\infty}\frac{qP(q)}{\langle q\rangle}R(q)(1-u^{q-1}), \tag{10}$$

where $R(q)$ is the probability that a node is in the remaining network given it has degree q. Since $P(q)R(q)=pP_p(q)$, thanks to (10) we obtain

$$1-u=\frac{p}{\langle q\rangle}\sum_{q=0}^{\infty}qP_p(q)(1-u^{q-1}). \tag{11}$$

Therefore, the relative size of giant component is calculated as

$$P_\infty=\sum_{q=0}^{\infty}P(q)R(q)(1-u^q)=p\sum_{q=0}^{\infty}P_p(q)(1-u^q), \tag{12}$$

where u is obtained by solving (11). The critical point p_c occurs when (11) starts to have solution $u<1$. Equating the derivatives of both sides of (11) at $u=1$ gives the critical occupation probability p_c as follows:

$$\langle q\rangle=p_c\sum_{q=2}^{\infty}q(q-1)P_{p_c}(q). \tag{13}$$

Next, we determine the scaling law of the critical value p_c in the limit of $n\to\infty$. Let $\hat{p_c}=p_c(n\to\infty)$. For $\alpha>0$, the critical occupation probability behaves as (see Appendix A)

$$p_c\sim\hat{p_c}+\frac{A}{n}e^{-\hat{\eta}n}. \tag{14}$$

Here, $\hat{p_c}$ is determined by the following equation

$$\frac{\langle q\rangle}{\hat{p_c}}=q_1q_2\Big(1-\frac{1-F(q_2)}{\hat{p_c}^{\,\alpha}}\Big)+\sum_{q=q_2+1}^{\infty}q(q-1)\frac{P(q)}{\hat{p_c}^{\,\alpha}}, \tag{15}$$

where q_2 is the smallest degree satisfying $F(q)\geq 1-\hat{p}_c^{\alpha}$ and $q_1=q_2-1$. The coefficient A in the higher-order analytical term in (14) is a constant given by $A=2\hat{p_c}^{\,\alpha}q_3/\left(\alpha\hat{p_c}^{\,\alpha-1}q_1q_2-\langle q\rangle(\alpha-1)\hat{p_c}^{\,\alpha-2}\right)$ and the decay rate $\hat{\eta}=\min_q\{|\alpha\ln\hat{p_c}-\ln(1-F(q))|\}$. Remembering the definitions of q_1 and q_2, it is clear that the minimum rate $\hat{\eta}$ is attained at $q=q_1$ or $q=q_2$ since $F(q)$ is monotonic. q_3 is taken as the q that attained the minimum rate $\hat{\eta}$.

When $\alpha=0$, we can readily recover the random attack scenario with p_c independent of n. In fact, taking $\alpha=0$ we have $F(q_2)=0$ with $q_2=q_{\min}-1$, where $q_{\min}$ means the minimum degree of $G(V,E)$. Hence, the formula (15) yields

$$\hat{p_c}=\frac{\langle q\rangle}{\sum_{q=q_{\min}}^{\infty}q(q-1)P(q)}=\frac{1}{G_1'(1)}, \tag{16}$$

which reproduces the classical result of node percolation [5, Eq. (12)].

B. Analytical framework for β-model

In the β-model we will instead consider a benefit-oriented attacker who will observe n random nodes at each step and remove the one with highest degree (with probability β) or a random one (with probability $1-\beta$). We run the process until a fraction of $1-p$ nodes are deleted from the network $G(V,E)$.

Similarly as the above section, we will first assume that only nodes are removed but the edges connecting the removed nodes to the remaining network are kept intact. Recall the cumulative distribution at time t is $F(q;t)=\sum_{s=0}^{q}P(s;t)$. Using maximum order statistics for n independent random variables (e.g. [29, Thm. 8.1]), the degree distribution of the observed highest-degree node at step t is

$$P_a^{\max}(q;t)=\Delta(F(q;t)^n), \tag{17}$$

for $q\geq 0$. An alternative method for deriving (17) is also presented in [16]. We set $F(-1;t)=0$ for all t as before. By the definition of β-model, the degree distribution of the attacked node at step t is

$$P_a(q;t)=\beta P_a^{\max}(q;t)+(1-\beta)P_a^{\text{ran}}(q;t), \tag{18}$$

where $P_a^{\text{ran}}(q;t)=P(q;t)=\Delta F(q;t)$ is the probability of choosing a random node of degree q. Following the similar line in Section II.A, we obtain the dynamical system of the cumulative distribution as

$$\begin{cases}(N-t)\frac{\partial F(q;t)}{\partial t}=-\beta F(q;t)^n+F(q;t)\\ \qquad\qquad\qquad\quad -(1-\beta)F(q;t),\quad t>0,\\ F(q;0)=F(q).\end{cases} \tag{19}$$

for all $q\geq 0$.

When $n>1$, by directly integrating (19) we derive the solution

$$F(q;t)=\Big(1+(F(q)^{1-n}-1)\cdot e^{\beta(n-1)\ln\left(\frac{N-t}{N}\right)}\Big)^{-\frac{1}{n-1}}. \tag{20}$$

Inserting $(1-p)N = t$ into (20) gives rise to

$$F_p(q) = \left(1 + (F(q)^{1-n} - 1)p^{\beta(n-1)}\right)^{-\frac{1}{n-1}}, \quad (21)$$

which characterizes the cumulative distribution of the degree of a randomly chosen node when a $1-p$ fraction of nodes are removed (with all edges intact) in the β-model. When $n = 1$, by taking the limit $n \to 1^+$ in (21), we reproduce $F_p(q) = F(q)$ for any $\beta \in [0, 1]$. The special case of $\beta = 0$ also leads to $F_p(q) = F(q)$ for any n, namely the ordinary random attack scenario.

With the same reasoning as in Section II.A, we derive the self-consistency equation (11) for u, the probability of a random edge not connecting to the giant component, and the formula (12) for P_∞, the relative size of giant component. In (11) and (12) here, the probability of a node to be occupied in the remaining network $P_p(q) = \Delta F_p(q)$ will be fed by (21). The critical occupation probability p_c is then determined by (13).

Next, we examine the behavior of the critical value p_c in the limit of $n \to \infty$. Likewise, let $\hat{p_c} = p_c(n \to \infty)$. For $\beta > 0$, the critical occupation probability evolves as (see Appendix B)

$$p_c \sim \hat{p_c} - \frac{B}{n} e^{-\hat{\theta} n}, \quad (22)$$

where $\hat{p_c}$ is determined by the following equation

$$\frac{\langle q \rangle}{\hat{p_c}} = q_1 q_2 \left(1 - \frac{F(q_1)}{\hat{p_c}^{\beta}}\right) + \sum_{q=2}^{q_1} q(q-1) \frac{P(q)}{\hat{p_c}^{\beta}}, \quad (23)$$

where q_1 is the largest degree satisfying $F(q) \leq \hat{p}_c^{\beta}$ and $q_2 = q_1 + 1$. The constant coefficient B in the higher-order analytical term in (22) is given by $B = 2\hat{p_c}^{\beta} q_3 / \left(\beta \hat{p_c}^{\beta-1} q_1 q_2 - \langle q \rangle (\beta - 1) \hat{p_c}^{\beta-2}\right)$ and the decay rate $\hat{\theta} = \min_q \{|\beta \ln \hat{p_c} - \ln F(q)|\}$. In view of the monotonicity of the $F(q)$, the minimum rate $\hat{\theta}$ is attained at $q = q_1$ or $q = q_2$. q_3 is taken as the q that attained the minimum rate $\hat{\theta}$.

In the special case of $\beta = 0$, we reproduce the random attack scenario with p_c independent of n. Since $q_1 = q_{\max}$, namely, the largest degree of $G(V, E)$, it follows from (23) that

$$\hat{p_c} = \frac{\langle q \rangle}{\sum_{q=2}^{q_{\max}} q(q-1)P(q)} = \frac{1}{G_1'(1)}, \quad (24)$$

which again reproduces the classical node percolation formula [5, Eq. (12)] similarly as in our α-model with $\alpha = 0$.

C. Results for synthetic networks

We test our theory in Section II.A and II.B numerically for Erdős-Rényi (ER) networks and power-law (PL) networks over $N = 10^7$ nodes [3]. We consider ER networks featuring a Poisson degree distribution $P(q) = e^{-\lambda}\lambda^q/q!$ for $q \geq 0$ and mean degree $\langle q \rangle = \lambda = 5$. The PL networks considered here follow $P(q) \propto q^{-\gamma}$ with the degree exponent $\gamma = 2.2$, minimum degree $q_{\min} = 1$ and a cutoff $q_{\text{cut}} = 3000$.

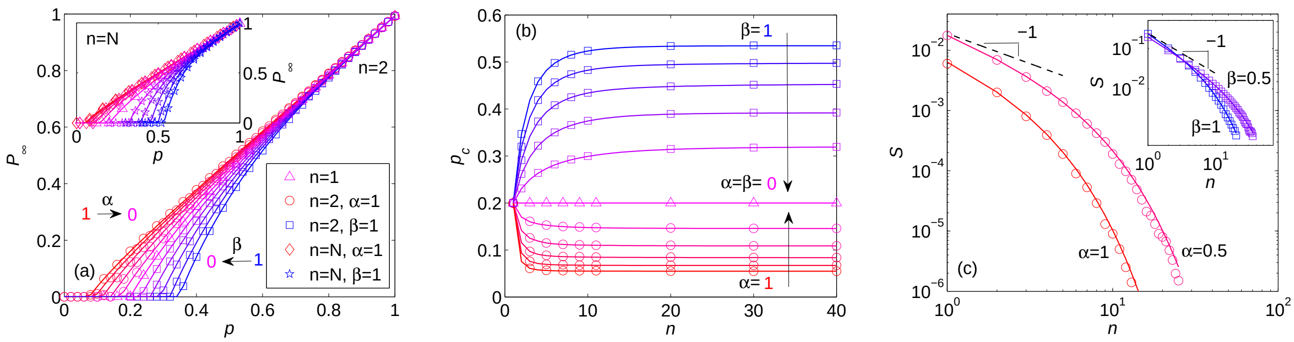

FIG. 1: (a) The relative size of giant component P_∞ for ER networks with $\langle q \rangle = 5$ and $N = 10^7$ as a function of occupation probability p. The points are numerical simulations for α-model with $n = 2$ (gradient red (upper) circles) and $n = N$ (gradient red (upper) diamonds) ranging from $\alpha = 1, 0.8, 0.6, 0.4, 0.2$; β-model with $n = 2$ (gradient (lower) blue squares) and $n = N$ (gradient blue (lower) pentagrams) ranging from $\beta = 1, 0.8, 0.6, 0.4, 0.2$. Magenta (middle) triangles are for the case of $\alpha = \beta = 0$ ($n = 1$). (b) The critical threshold p_c as a function of n similarly for α-model (gradient red (lower) circles) and β-model (gradient blue (upper) squares) ranging from $\alpha = 1, 0.8, 0.6, 0.4, 0.2$ and $\beta = 1, 0.8, 0.6, 0.4, 0.2$, respectively. (c) The scaling S as a function of n for α-model with $\alpha = 1, 0.5$ (gradient red circles) and β-model with $\beta = 1, 0.5$ (gradient blue squares). Solid curves are analytical results.

The results for ER networks and PL networks are shown in Fig. 1 and Fig. 2, respectively. As can be seen, the agreement between theory and simulations is excellent for both α-model and β-model. The degree of both cost-orientation (α) and benefit-orientation (β) shows a diminishing marginal effect when evolved from 0 to 1. For example, for $n = 40$ in Fig. 1(b), an increase from $\beta = 0$ to $\beta = 0.2$ boosts the critical threshold p_c by around 0.12, while an increase from $\beta = 0.8$ to $\beta = 1$ only pushes it up by around 0.05. Moreover, at any degree of cost-orientation or benefit-orientation, only a relatively small level of knowledge (n) is needed to achieve the effect close to a target attack (i.e. $n \sim N$); typically α-model enjoys a smaller n than β-model. For example, to approximate targeted attacks, n is required to be around 10 for β-model and just about a half for α-model in ER networks. For PL networks as shown in Fig. 2(b), the difference is much more prominent: n should be around 50 for β-model but virtually 1 for α-model since the random attack admits $p_c \sim 0$. It is found that the influence of cost-orientation (α) and benefit-orientation (β) is roughly comparable for ER networks due to the homogeneity of degree (see Fig. 1(a) and Fig. 1(b)). However, in PL networks the increase of β dramatically harms the network integrity especially for larger n (see the inset of Fig. 2(a) and Fig. 2(b)) as hubs in these networks play a cohesive role in connecting nodes to the giant component.

The behavior of scaling law for $S := |p_c - \hat{p_c}|$ with respect to n is shown in Fig. 1(c) for ER networks and in Fig. 2(c) for PL networks, respectively. As one would expect from (14) and (22) that S evolves as n^{-1} for small n. In Fig. 2(c) only β-model is displayed since $p_c = \hat{p_c} \sim 0$ for any α and $B = 0$ in (22) for the PL networks considered here.

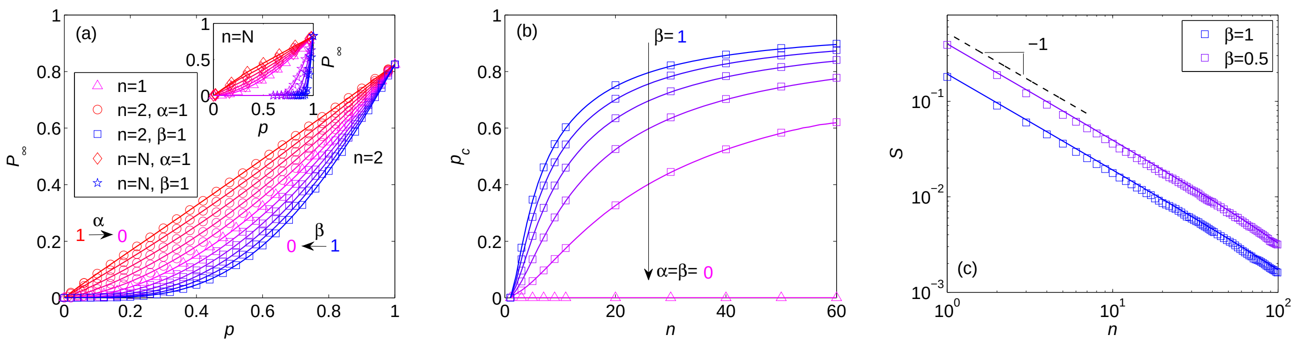

FIG. 2: (a) The relative size of giant component P_∞ for PL networks with $\gamma = 2.2$, $q_{\min} = 1$, $q_{\text{cut}} = 3000$ and $N = 10^7$ as a function of occupation probability p. The points are numerical simulations for α-model with $n = 2$ (gradient red (upper) circles) and $n = N$ (gradient red (upper) diamonds) ranging from $\alpha = 1, 0.8, 0.6, 0.4, 0.2$; β-model with $n = 2$ (gradient blue (lower) squares) and $n = N$ (gradient blue (lower) pentagrams) ranging from $\beta = 1, 0.8, 0.6, 0.4, 0.2$. Magenta (middle) triangles are for the case of $\alpha = \beta = 0$ ($n = 1$). (b) The critical threshold p_c as a function of n similarly for β-model (gradient blue squares) ranging from $\beta = 1, 0.8, 0.6, 0.4, 0.2$. For α-model, $p_c = 0$ constantly. (c) The scaling S as a function of n for β-model with $\beta = 1, 0.5$ (gradient blue squares). Solid curves are analytical results.

III. k-CORE OF A NETWORK UNDER TUNABLE LIMITED KNOWLEDGE

In this section, we develop a unified analytical framework to deal with the k-core of our α- and β-models. The k-core is viewed as the infinite $(k-1)$-ary subtree in the network, where a k-ary tree admits at least k branches at each node [23].

A. Theory

Recall in Section II we start with a configuration model $G(V, E)$ with degree distribution $P(q)$ and delete nodes according to α- and β-models but keeping edges connecting to remaining nodes intact. The resulting degree distribution is $P_p(q) = F_p(q) - F_p(q-1)$ with $F_p(q)$ given by (8) for α-model and (21) for β-model. Define the generating function as $\tilde{G}_0(x) = \sum_{q=0}^{\infty} P_p(q) x^q$ and the mean degree is calculated by $\langle q(p) \rangle = \sum_{q=0}^{\infty} q P_p(q) = \tilde{G}_0'(1)$.

In the next stage, we will remove these edges. Note that the configuration model is randomly connected, and the probability of a random edge leading to a remaining node is equivalent to the ratio of the number of edges leaving the remaining nodes to the total number of edges leaving all nodes in $G(V, E)$:

$$\tilde{p} = \frac{pN \langle q(p) \rangle}{N \langle q \rangle} = \frac{p\, \tilde{G}_0'(1)}{G_0'(1)}. \tag{25}$$

Removing the edges leading to a deleted node is on a par with removing randomly a $1 - \tilde{p}$ fraction of edges of the remaining nodes in a randomly connected network. Mimicking the bond percolation in [30], we obtain the generating function of the remaining nodes under attack with tunable limited knowledge as

$$\overline{G}_0(x) = \tilde{G}_0(1 - \tilde{p} + \tilde{p}x) = \sum_{q=0}^{\infty} \overline{P}(q) x^q, \tag{26}$$

where $\tilde{p}$ is given in (25) and $\overline{P}(q)$ characterizes the probability of a randomly chosen node having degree q in the resulting network, which we will denote by $\overline{G}(\overline{V}, \overline{E})$. Apparently, we have $|\overline{V}| = pN$. Moreover, in the case of $n = 1$, we know that $F_p(q) = F(q)$ and $P_p(q) = P(q)$

for any α or β (again depending on whether we consider α-model via (8) or β-model via (21)). Accordingly, $\tilde{G}_0(x) = G_0(x)$, and $\tilde{p} = p$ by (25). The generating function (26) reduces to $\overline{G}_0(x) = G_0(1-p+px)$, which is in line with the ordinary random attack framework [30].

With (26) at hand, we now study the k-core structure of $\overline{G}(\overline{V}, \overline{E})$ following the idea of [23, 24, 31]. We first consider the relative size, N_{kc}, of k-core. To this end, let z be the probability that an end node of a random edge is the root of an infinite $(k-1)$-ary tree. The self-consistency equation reads as

$$z = \sum_{q=k}^{\infty} \frac{q\overline{P}(q)}{\langle q \rangle} \Big(\sum_{s=k-1}^{q-1} \binom{q-1}{s} z^s (1-z)^{q-1-s} \Big), \qquad (27)$$

where $q\overline{P}(q)/\langle q \rangle$ is the probability of $q-1$ out-reaching edges of the end node and the binomial coefficient thingy explains that s out of these $q-1$ neighbors must belong to the infinite subtree. A node is in the k-core indicates that it has at least k neighbors in the infinite subtree. Hence, the probability that a randomly chosen node is in the k-core becomes

$$N_{kc} = p \sum_{q=k}^{\infty} \overline{P}(q) \Big(\sum_{s=k}^{q} \binom{q}{s} z^s (1-z)^{q-s} \Big). \qquad (28)$$

Employing the generating function formalism with the s-th order derivative $\overline{G}_0^{(s)}(x) = \sum_{q=s}^{\infty} \overline{P}(q)(q!/(q-s)!)x^{q-s}$ [3, 5], we rewrite (27) and (28) as

$$z = 1 - \frac{1}{\langle q \rangle_{\overline{G}}} \sum_{s=0}^{k-2} \frac{z^s \overline{G}_0^{(s+1)}(1-z)}{s!} \qquad (29)$$

and

$$N_{kc} = p - p \sum_{s=0}^{k-1} \frac{z^s \overline{G}_0^{(s)}(1-z)}{s!}, \qquad (30)$$

where we use $\langle q \rangle_{\overline{G}} = \overline{G}_0'(1)$, the mean degree of $\overline{G}$, to differentiate from the mean degree of G, $\langle q \rangle$.

An important subgraph of k-core is called corona, which is a subset of nodes having degree k in the k-core [23, 24]. The relative size of corona is denoted by $N_k(k)$. More generally, for $q \geq k$, we define $N_k(q)$ as the fraction of nodes with degree q in the k-core. Therefore,

$$N_k(q) = p \sum_{s=q}^{\infty} \overline{P}(s) \binom{s}{q} z^q (1-z)^{s-q}, \qquad (31)$$

where s represents the degree of such a node in $\overline{G}$. With the help of generating functions, we obtain

$$N_k(k) = p \frac{z^k \overline{G}_0^{(k)}(1-z)}{k!}. \qquad (32)$$

As a byproduct, we can also derive the mean degree, D_{kc}, of the k-core as follows.

$$D_{kc} = \sum_{q=k}^{\infty} q \frac{N_k(q)}{N_{kc}} = \frac{1}{N_{kc}} \sum_{q=k}^{\infty} \frac{z^q \overline{G}_0^{(q)}(1-z)}{(q-1)!}. \qquad (33)$$

B. Results for synthetic networks

The same ER and PL networks as those in Section II.C are considered here for the study of k-core in Fig. 3 and Fig. 4, respectively. Fig. 3 shows the relative size of k-core N_{kc} and the relative size of corona $N_k(k)$ for ER networks for the two cases of $k=2$ and $k=3$ with various parameters α, β, and n. The counterpart results for PL networks are shown in Fig. 4. An overarching observation is that for both α- and β-models with general n, $k=3$ remains the watershed for k-core percolation [22, 23], namely, N_{kc} shows a continuous phase transition for $k<3$ while k-core for $k \geq 3$ emerges discontinuously.

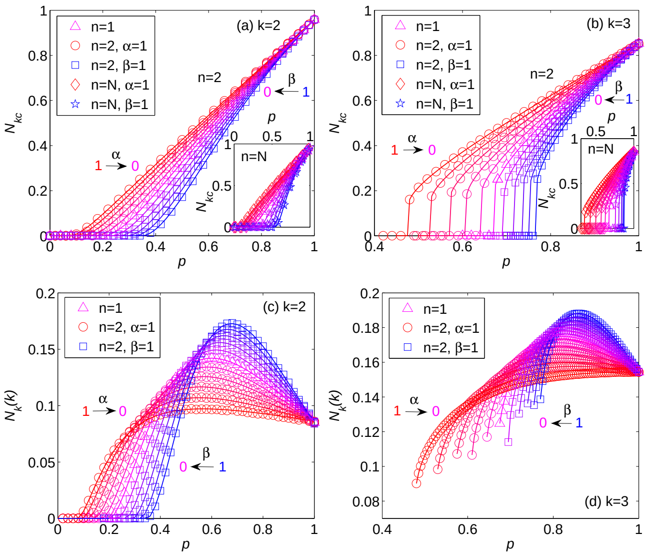

FIG. 3: The relative size of k-core N_{kc} with (a) $k=2$ and (b) $k=3$ for ER networks with $\langle q \rangle = 5$ and $N = 10^7$ as a function of occupation probability p. The points are numerical simulations for α-model with $n=2$ (gradient red (upper) circles) and $n=N$ (gradient red (upper) diamonds) ranging from $\alpha = 1, 0.8, 0.6, 0.4, 0.2$; β-model with $n=2$ (gradient blue (lower) squares) and $n=N$ (gradient blue (lower) pentagrams) ranging from $\beta = 1, 0.8, 0.6, 0.4, 0.2$. The corresponding relative size of corona $N_k(k)$ similarly with (c) $k=2$ and (d) $k=3$ as a function of p. Magenta (middle) triangles are for the case of $\alpha = \beta = 0$ ($n=1$). Solid curves are analytical results.

The relative sizes of corona and k-core display a somewhat unexpected inconsistency for ER networks with respect to α- and β-models. When the occupation probability p is relatively large ($p > 0.5$ for example in Fig. 3(c)), the corona takes up a larger proportion in the k-core for β-model than α-model although the k-core size in β-model is smaller than that in α-model. Interestingly, PL networks do not display such phenomena (c.f. Fig. 4). This raises a caveat that, in the early stage of attacks, a benefit-oriented attacker can make an ER network fairly 'unstable' (compared to a cost-oriented attacker) in the sense of producing a smaller k-core and a more significant portion of the k-core being its corona. The loss of central nodes under benefit-oriented attacks with limited knowledge in ER networks could be more detrimental to network structure than expected as the nodes in corona are the weakest in the k-core.

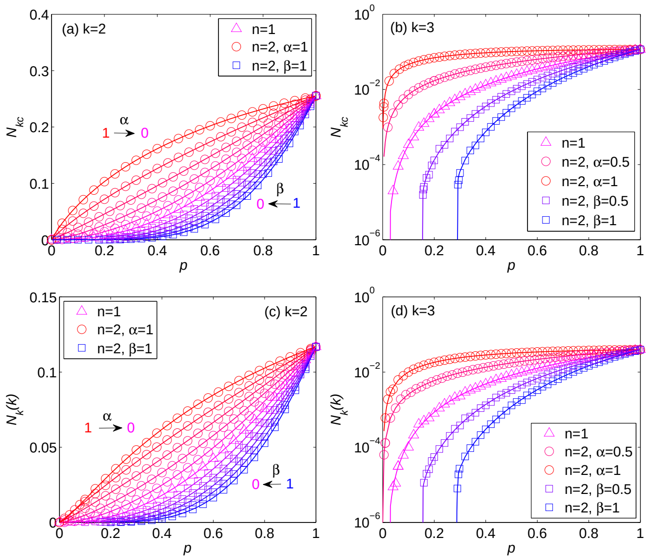

FIG. 4: The relative size of k-core N_{kc} with (a) $k = 2$ and (b) $k = 3$ for PL networks with $\gamma = 2.2$, $q_{\min} = 1$, $q_{\text{cut}} = 3000$ and $N = 10^7$ as a function of occupation probability p. The points are numerical simulations for α-model with $n = 2$ (gradient red (upper) circles) ranging from (a) $\alpha = 1, 0.8, 0.6, 0.4, 0.2$ and (b) $\alpha = 1, 0.5$; β-model with $n = 2$ (gradient blue (lower) squares) ranging from (a) $\beta = 1, 0.8, 0.6, 0.4, 0.2$ and (b) $\beta = 1, 0.5$. The corresponding relative size of corona $N_k(k)$ with (c) $k = 2$ and (d) $k = 3$ as a function of p. Magenta (middle) triangles are for the case of $\alpha = \beta = 0$ ($n = 1$). Solid curves are analytical results.

IV. CONCLUSION

In this paper we have developed the theory of percolation under tunable limited knowledge featuring both a cost-oriented attacker (α-model) and a benefit-oriented attacker (β-model). The limited ability to observe the network structure is relevant in many realistic scenarios and much coveted. We analytically solve the giant component size and critical percolation thresholds for uncorrelated networks with arbitrary degree distributions. We systematically derive the scaling law for the critical percolation threshold with respect to the level of knowledge n for both models. In addition, we study the k-core percolation in the context of tunable limited knowledge, and show how these attacks affect the k-core structure including the size of k-core and its corona. It is found that the structure discrepancy between ER and PL networks profoundly influences both the giant component based percolation phase transition and their k-core organization.

Appendix A: Asymptotic behavior of p_c for α-model as $n \to \infty$

To determine the asymptotic behavior of p_c from (13), we will first examine that of $F_p(q)$ from (8). Write $F_p(q) = \hat{F}_p(q) + \delta_p(q)$, where constant $\hat{F}_p(q)$ is the leading term and $\delta_p(q)$ is vanishing as $n \to \infty$.

For large q satisfying $1 - F(q) < p^\alpha$, we have

$$\begin{aligned} F_p(q) \sim & 1 - \Big(1 + \frac{p^{\alpha n}}{(1-F(q))^n} - p^{\alpha n}\Big)^{-\frac{1}{n}} \\ \sim & 1 - \Big(1 + \frac{p^{\alpha n}}{(1-F(q))^n}\Big)^{-\frac{1}{n}} \\ = & 1 - \frac{1-F(q)}{p^\alpha} \exp\Big\{ -\frac{1}{n} \ln\Big(1 + \frac{(1-F(q))^n}{p^{\alpha n}}\Big)\Big\} \\ \sim & 1 - \frac{1-F(q)}{p^\alpha} \cdot \Big(1 - \frac{(1-F(q))^n}{n p^{\alpha n}}\Big) \\ \sim & 1 - \frac{1-F(q)}{p^\alpha} + \frac{1}{n} e^{-n(\alpha \ln p - \ln(1-F(q)))}. \end{aligned} \tag{A1}$$

For $1 - F(q) \to p^\alpha$ satisfying $p^\alpha/(1-F(q)) \to 1$, set $\xi = 1 - p^\alpha/(1-F(q))$. Since $(p^\alpha/(1-F(q)))^n = (1-\xi)^n \sim e^{-n\xi} \sim 1 - n\xi$, we obtain

$$\begin{aligned} F_p(q) \sim & 1 - \Big(1 + 1 - n\Big(1 - \frac{p^\alpha}{1-F(q)}\Big) - p^{\alpha n}\Big)^{-\frac{1}{n}} \\ = & 1 - \exp\Big\{ -\frac{\ln 2}{n} - \frac{1}{n} \\ & \cdot \ln\Big(1 - \frac{n}{2}\Big(1 - \frac{p^\alpha}{1-F(q)}\Big) - \frac{p^{\alpha n}}{2}\Big)\Big\} \\ \sim & 1 - \exp\Big\{ -\frac{\ln 2}{n} + \frac{1}{2}\Big(1 - \frac{p^\alpha}{1-F(q)}\Big) + \frac{p^{\alpha n}}{2n}\Big\} \\ \sim & \frac{\ln 2}{n} - \frac{1}{2}\Big(1 - \frac{p^\alpha}{1-F(q)}\Big). \end{aligned} \tag{A2}$$

For small q satisfying $p^\alpha < 1 - F(q) < 1$,

$$\begin{aligned} F_p(q) \sim & 1 - \Big(1 + \frac{p^{\alpha n}}{(1-F(q))^n} - p^{\alpha n}\Big)^{-\frac{1}{n}} \\ \sim & 1 - \exp\Big\{ -\frac{p^{\alpha n}}{n(1-F(q))^n} + \frac{p^{\alpha n}}{n}\Big\} \\ \sim & \frac{p^{\alpha n}}{n(1-F(q))^n} \\ = & \frac{1}{n} e^{-n(\ln(1-F(q)) - \alpha \ln p)}. \end{aligned} \tag{A3}$$

When $F(q) = 0$, clearly we have $F_p(q) = 0$.

Therefore, the leading term of $F_p(q)$ is expressed as

$$\hat{F}_p(q) \sim \begin{cases} 0, & p^\alpha < 1 - F(q), \\ 1 - \frac{1-F(q)}{p^\alpha}, & 1 - F(q) < p^\alpha, \end{cases} \tag{A4}$$

and the vanishing term

$$\delta_p(q) = \begin{cases} 0, & F(q) = 0, \\ \frac{1}{n} e^{-\eta(q) n}, & F(q) > 0, \end{cases} \tag{A5}$$

where the decay rate is $\eta(q) = |\alpha \ln p - \ln(1 - F(q))|$.

With the above exponential convergence of the distribution $F_p(q)$, we then analyze the convergence of the critical occupation probability p_c. Define $q_1 = \max_{1-F(q)>p^\alpha} q$ and $q_2 = \min_{1-F(q)\leq p^\alpha}$. By (13), (A4) and (A5), we derive

$$\begin{aligned}\frac{\langle q\rangle}{p_c} =& \sum_{q=2}^{\infty} q(q-1)P_{p_c}(q) = \sum_{q=2}^{\infty} q(q-1)\Delta F_{p_c}(q) \\ =& \sum_{q=2}^{\infty} q(q-1)\Delta \hat{F_{p_c}}(q) - 2\sum_{q=1}^{\infty} q\delta_{p_c}(q) \\ \sim& q_1 q_2\Big(1 - \frac{1-F(q_2)}{p_c^\alpha}\Big) + \sum_{q=q_2+1}^{\infty} q(q-1)\frac{P(q)}{p_c^\alpha} \\ &- 2\sum_{q=1}^{\infty} q\delta_{p_c}(q). \end{aligned} \tag{A6}$$

Let $\hat{p_c} = \lim_{n\to\infty} p_c$ and decompose p_c as $p_c = \hat{p_c} + \delta_c$, where $\hat{p_c}$ is the constant leading term and the next term δ_c is vanishing as $n \to \infty$. We solve $\hat{p_c}$ through the following equation

$$\frac{\langle q\rangle}{\hat{p_c}} = q_1 q_2\Big(1 - \frac{1-F(q_2)}{\hat{p_c}^{\alpha}}\Big) + \sum_{q=q_2+1}^{\infty} q(q-1)\frac{P(q)}{\hat{p_c}^{\alpha}}. \tag{A7}$$

By using (A6), (A7) and Taylor's series at $\hat{p_c}$, we derive

$$\delta_c \sim \hat{A}\sum_{q=1}^{\infty} q\delta_{\hat{p_c}}(q), \tag{A8}$$

where $\hat{A} = 2\hat{p_c}^{\alpha}/\big(\alpha\hat{p_c}^{\alpha-1}q_1q_2 - \langle q\rangle(\alpha-1)\hat{p_c}^{\alpha-2}\big)$. In view of (A5) and the monotonicity of $F(q)$, we see that the dominant decay rate is given by $\hat{\eta} = \min_q\{|\alpha\ln\hat{p_c} - \ln(1-F(q))|\}$, where the optimum is attained at q_1 or q_2. Putting all bits together, we obtain for α-model

$$p_c \sim \hat{p_c} + \frac{\hat{A}q_3}{n}e^{-\hat{\eta}n} \tag{A9}$$

as $n \to \infty$, where q_3 is taken as the q that attained the minimum rate $\hat{\eta}$.

Appendix B: Asymptotic behavior of p_c for β-model as $n \to \infty$

In our β-model we will similarly start with the analysis of $F_p(q)$ from (21). Write $F_p(q) = \hat{F}_p(q) + \varepsilon_p(q)$, where constant $\hat{F}_p(q)$ is the leading term and $\varepsilon_p(q)$ is vanishing as $n \to \infty$.

For small q satisfying $F(q) < p^\beta$, we have

$$\begin{aligned}F_p(q) \sim& \Big(1 + \frac{p^{\beta n}}{F(q)^n} - p^{\beta n}\Big)^{-\frac{1}{n}} \\ \sim& \Big(1 + \frac{p^{\beta n}}{F(q)^n}\Big)^{-\frac{1}{n}} \\ =& \frac{F(q)}{p^\beta}\exp\Big\{-\frac{1}{n}\ln\Big(1 + \frac{F(q)^n}{p^{\beta n}}\Big)\Big\} \\ \sim& \frac{F(q)}{p^\beta}\cdot\Big(1 - \frac{F(q)^n}{np^{\beta n}}\Big) \\ \sim& \frac{F(q)}{p^\beta} - \frac{1}{n}e^{-n(\beta\ln p - \ln F(q))}. \end{aligned} \tag{B1}$$

For $F(q) \to p^\beta$ such that $p^\beta/F(q) \to 1$, set $\zeta = 1 - p^\beta/F(q)$. Since $(p^\beta/F(q))^n = (1-\zeta)^n \sim e^{-n\zeta} \sim 1 - n\zeta$, we obtain

$$\begin{aligned}F_p(q) \sim& \Big(1 + 1 - n\Big(1 - \frac{p^\beta}{F(q)}\Big) - p^{\beta n}\Big)^{-\frac{1}{n}} \\ =& \exp\Big\{-\frac{\ln 2}{n} - \frac{1}{n} \\ &\cdot\ln\Big(1 - \frac{n}{2}\Big(1 - \frac{p^\beta}{F(q)}\Big) - \frac{p^{\beta n}}{2}\Big)\Big\} \\ \sim& \exp\Big\{-\frac{\ln 2}{n} + \frac{1}{2}\Big(1 - \frac{p^\beta}{F(q)}\Big) + \frac{p^{\beta n}}{2n}\Big\} \\ \sim& 1 - \frac{\ln 2}{n} + \frac{1}{2}\Big(1 - \frac{p^\beta}{F(q)}\Big). \end{aligned} \tag{B2}$$

For large q satisfying $p^\beta < F(q) < 1$,

$$\begin{aligned}F_p(q) \sim& \Big(1 + \frac{p^{\beta n}}{F(q)^n} - p^{\beta n}\Big)^{-\frac{1}{n}} \\ \sim& \exp\Big\{-\frac{p^{\beta n}}{nF(q)^n} + \frac{p^{\beta n}}{n}\Big\} \\ \sim& 1 - \frac{p^{\beta n}}{nF(q)^n} = 1 - \frac{1}{n}e^{-n(\ln F(q) - \beta\ln p)}. \end{aligned} \tag{B3}$$

For $F(q) = 1$, we have $F_p(q) = 1$.

Hence, the leading term of $F_p(q)$ is expressed as

$$\hat{F}_p(q) \sim \begin{cases} \frac{F(q)}{p^\beta}, & F(q) < p^\beta, \\ 1, & F(q) > p^\beta, \end{cases} \tag{B4}$$

and the vanishing term

$$\varepsilon_p(q) = \begin{cases} -\frac{1}{n}e^{-\theta(q)n}, & F(q) < 1, \\ 0, & F(q) = 1, \end{cases} \tag{B5}$$

where the decay rate is $\theta(q) = |\beta\ln p - \ln F(q)|$.

Next, we examine the convergence of the critical occupation probability p_c. Define $q_1 = \max_{F(q)\leq p^\beta} q$ and $q_2 = \min_{F(q)>p^\beta}$. Employing (13), (B4) and (B5), we obtain

$$\begin{aligned}\frac{\langle q\rangle}{p_c} =&\sum_{q=2}^{\infty} q(q-1)P_{p_c}(q) = \sum_{q=2}^{\infty} q(q-1)\Delta F_{p_c}(q)\\ =&\sum_{q=2}^{\infty} q(q-1)\Delta \hat{F_{p_c}}(q) - 2\sum_{q=1}^{\infty} q\varepsilon_{p_c}(q)\\ \sim& q_1 q_2\Big(1-\frac{F(q_1)}{p_c^{\beta}}\Big) + \sum_{q=2}^{q_1} q(q-1)\frac{P(q)}{p_c^{\beta}}\\ &-2\sum_{q=1}^{\infty} q\varepsilon_{p_c}(q).\end{aligned} \tag{B6}$$

Let $\hat{p_c} = \lim_{n\to\infty} p_c$ and decompose p_c as $p_c = \hat{p_c} + \varepsilon_c$, where $\hat{p_c}$ is the constant leading term and the next term ε_c is vanishing as $n \to \infty$. We solve $\hat{p_c}$ through the following equation

$$\frac{\langle q\rangle}{\hat{p_c}} = q_1 q_2\Big(1-\frac{F(q_1)}{\hat{p_c}^{\beta}}\Big) + \sum_{q=2}^{q_1} q(q-1)\frac{P(q)}{\hat{p_c}^{\beta}}. \tag{B7}$$

Invoking (B6), (B7) and Taylor's series at $\hat{p_c}$, we derive

$$\varepsilon_c \sim \hat{B}\sum_{q=1}^{\infty} q\varepsilon_{\hat{p_c}}(q), \tag{B8}$$

where $\hat{B} = 2\hat{p_c}^{\beta}/\Big(\beta\hat{p_c}^{\beta-1}q_1q_2 - \langle q\rangle(\beta-1)\hat{p_c}^{\beta-2}\Big)$. In view of (B5) and the monotonicity of $F(q)$, we conclude that the dominant decay rate is given by $\hat{\theta} = \min_q\{|\beta\ln\hat{p_c} - \ln F(q)|\}$, where the optimum is attained at q_1 or q_2. Putting all pieces together, we have for β-model

$$p_c \sim \hat{p_c} - \frac{\hat{B}q_3}{n}e^{-\hat{\theta}n} \tag{B9}$$

as $n \to \infty$, where q_3 is taken as the q that attained the minimum rate $\hat{\theta}$.

[1] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, Cambridge, 2010).
[2] Y. Yang, T. Nishikawa, and A. E. Motter, Science **358**, eaan3184 (2017).
[3] M. E. J. Newman, *Networks, 2nd Edition* (Oxford University Press, Oxford, 2018).
[4] R. Albert, H. Jeong, and A.-L. Barabási, Nature **406**, 378 (2000).
[5] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[6] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, Nature **464**, 1025 (2010).
[7] Y. Shang, W. Luo, and S. Xu, Phys. Rev. E **84**, 031113 (2011).
[8] G. J. Baxter, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. E **82**, 011103 (2010).
[9] M. A. D. Muro, S. V. Buldyrev, and L. A. Braunstein, Phys. Rev. E **101**, 042307 (2020).
[10] S. Shao, X. Huang, H. E. Stanley, and S. Havlin, New J. Phys. **17**, 023049 (2015).
[11] W. Wang, S. Yang, H. E. Stanley, and J. Gao, Nat. Commun. **10**, 2114 (2019).
[12] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).
[13] S. Iyer, T. Killingback, B. Sundaram, and Z. Wang, PLoS ONE **8**, e59613 (2013).
[14] N. Almeira, O. V. Billoni, and J. I. Perotti, Phys. Rev. E **101**, 012306 (2020).
[15] E. C. Dinleyici, R. Borrow, M. A. P. Safadi, P. van Damme, and F. M. Munoz, Hum. Vaccinces Immunother. (2020), URL `doi:10.1080/21645515.2020.1804776`.
[16] Y. Liu, H. Sanhedrai, G. Dong, L. M. Shekhtman, F. Wang, S. V. Buldyrev, and S. Havlin, Natl. Sci. Rev. p. nwaa229 (2020).
[17] X.-L. Ren, N. Gleinig, D. Tolić, and N. Antulov-Fantulin, Complexity **2018**, 9826243 (2018).
[18] A. Patron, R. Cohen, D. Li, and S. Havlin, Phys. Rev. E **95**, 052305 (2017).
[19] S. Wandelt, X. Sun, D. Feng, M. Zanin, and S. Havlin, Sci. Rep. **8**, 13513 (2018).
[20] X.-L. Ren, N. Gleinig, D. Helbing, and N. Antulov-Fantulin, Proc. Natl. Acad. Sci. USA **116**, 6554 (2019).
[21] A. Smolyak, O. Levy, I. Vodenska, S. Buldyrev, and S. Havlin, Sci. Rep. **10**, 16124 (2020).
[22] Y.-X. Kong, G.-Y. Shi, R.-J. Wu, and Y.-C. Zhang, Phys. Rep. **832**, 1 (2019).
[23] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. Lett. **96**, 040601 (2006).
[24] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, Phys. Rev. E **73**, 056101 (2006).
[25] F. Morone, G. D. Ferraro, and H. A. Makse, Nat. Phys. **15**, 95 (2019).
[26] F. Zhang, Y. Zhang, L. Qin, W. Zhang, and X. Lin, in *Proc. 31st AAAI Conference on Artificial Intelligence* (AAAI Press, San Francisco, CA, 2017), pp. 245–251.
[27] N. K. Panduranga, J. Gao, X. Yuan, H. E. Stanley, and S. Havlin, Phys. Rev. E **96**, 032317 (2017).
[28] Y. Shang, Phys. Rev. E **101**, 042306 (2020).
[29] H. Pishro-Nik, *Introduction to Probability, Statistics, and Random Processes* (Kappa Research, Sunderland, MA, 2014).
[30] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
[31] N. Azimi-Tafreshi, J. G.-G. nes, and S. N. Dorogovtsev, Phys. Rev. E **90**, 032816 (2014).