Generalized k-cores of networks under attack with limited knowledge

Yilun Shang[1]

[1]*Department of Computer and Information Sciences, Northumbria University*
Newcastle upon Tyne, NE1 8ST, United Kingdom
Email: shylmath@hotmail.com

Abstract

Network theory has been used as an effective approach for understanding and controlling many real-world large-scale systems. A significant aspect of network operation is its robustness against failures and attacks. Here, we develop a theoretical framework for two classes of network attack with limited knowledge, namely, min-n and max-n attacks, where only n nodes are observed and a node with smallest or largest degree is removed at a time until a fraction $1-p$ of nodes are attacked. We study the effect of these attacks on the generalized k-core (Gk-core) of the network, which is obtained by implementing a k-leaf pruning process, removing progressively nodes with degree smaller than k alongside their nearest neighbors. This removal process can be viewed as a generation of the ordinary k-core decomposition. It is found that the $G2$-core undergoes a continuous phase transition with respect to p while Gk-core shows a first-order percolation transition for $k \geq 3$ under both types of attacks for all n. We reveal that knowing one more node during attacks, improving from $n = 1$ to $n = 2$, turns out to be most beneficial in terms of changing the robustness of Gk-core in both directions. Moreover, it is shown that degree heterogeneity plays a role in robustness as prioritizing attack on small-degree nodes in heterogeneous networks may help consolidate the Gk-core, but also in stability where hub nodes act like anchors stabilizing the Gk-core structure. Our results offer insight into the design of resilient complex systems and evaluation of network robustness and stability.

Keywords: Network, core, percolation, attack, robustness, stability

1. Introduction

Networks are an increasingly crucial model of diverse complex systems such as the Internet, brain, infrastructures and social activities [1, 2]. The connectivity of such networks when a fraction of nodes are attacked plays a critical role in maintaining network robustness and survivability, which sustain their operation and performance [3, 4]. Network robustness is often studied using the percolation theory borrowed from statistics physics, where the size of the giant component P_∞ is measured as the order parameter of the system when a fraction $1-p$ of nodes are knocked out from the network. Malicious attacks targeting at the most connected nodes (i.e. nodes with highest degrees) and random attacks are two of the most intensively studied attack strategies in complex network research [5, 6, 4, 7, 8].

The targeted attack assuming full knowledge of network structure and random attack assuming no knowledge, however, are not a good approximation in many realistic large-scale complex

systems. In the Internet, for example, complete information of the network may not be available but experienced cyber criminals are often able to probe the status of some routers and servers when launching a potential attack. Another example is the immunization problem in social networks [9, 10]. In most cases it is impossible to obtain complete social contacts of each individual and immunization doses are usually expensive. A feasible strategy would be to obtain the interaction information of a sample group of people and immunize or remove the key individuals. Recently, a new model of immunization under limited knowledge has been proposed where a number n nodes are observed at a time and the node with highest degree is removed to prevent spreading infection in a population of N agents [11]. The model links the two extremes, namely random attack ($n = 1$) and targeted attack ($n = N$). It is found that the knowledge of the order of $\ln N$ nodes in scale-free networks represents a critical value in effective suppressing epidemic propagation.

The disintegration of networks based on partial information has also been studied from the perspective of imprecise and uncertain observation, where the information of all nodes are observed but may be imprecise. Gallos et al. [12] have considered attack strategies in which a node of degree q is removed with probability proportional to q^γ, where γ is associated with uncertain attack information giving rise to more vulnerable higher degree nodes when $\gamma > 0$ or more vulnerable lower degree nodes when $\gamma < 0$. In [13], the observed degrees of nodes are modified by random perturbations. An analogous edge version accommodating perturbed edge weights is examined in [14].

Most of the existing literatures on network robustness including the above mentioned works consider the giant component size P_∞ as the primary indicator of functional component of a network. Other important functional subgraph structures include core [15, 16] and k-core [17, 18]. In the recent work [19], a k-leaf removal process for $k \geq 2$ is introduced to produce a generalized k-core structure (Gk-core), where a k-leaf is defined as a node with degree less than k. In this pruning process, k-leaves together with their nearest neighbors are removed from the network progressively. The resulting subgraph, i.e. Gk-core, is equivalent to the classical core structure when $k = 2$ [16], and it is found to be an effective measure of network robustness against virus like attacks deactivating weak nodes (k-leaves) and their first nearest neighbors [20, 21, 22]. Some interesting phenomena have been reported regarding Gk-core. For instance, it is shown that [21] $G2$-core undergoes a second-order phase transition fulfilling Widom scaling whereas Gk-core with $k \geq 3$ undergoes an abrupt percolation transition failing this identity in modular networks with Erdős-Rényi communities.

In this paper, we develop a mathematical framework for understanding network robustness in terms of the numbers of nodes and edges in Gk-core under two types of general attacks with limited knowledge. We consider the min-n (and max-n resp.) attack where n random nodes in the network are observed at each stage and the node with minimum (and maximum resp.) degree is removed. We refer to n as the knowledge index. The max-n attack features an active immunization strategy [11] while the analogous min-n attack sheds an important insight on the mild depreciation process with cost taken into consideration. In addition to robustness, we study the stability of Gk-core by looking into the number of same nodes that are retained in the Gk-core over multiple independent attacks [20]. The allows us to quantify the extent to which a core functional component can sustain irrespective of specific damage caused by a attack under limited knowledge.

We apply the derived frameworks to random networks with arbitrary degree distributions including Erdős-Rényi (ER) random graphs and log-normal random networks. It is found that $G2$-core undergoes a second-order phase transition as the attack carries out while Gk-core for

$k \geq 3$ emerges discontinuously for all attack strategies considered. The effect of growth in knowledge index turns out to have a diminishing marginal utility, leaving the transition from $n = 1$ to $n = 2$ most beneficial, in both min-n and max-n attacks. Further more, we show that degree heterogeneity plays an essential role in robustness where attacking small-degree nodes in heterogeneous networks may help build the Gk-core, but also in stability where hub nodes act like anchors stabilizing the Gk-core structure. In addition to synthetic networks, two real-world networks including a mathematicians coauthor network and an electronic circuit have been examined and compared with our theoretical predictions.

2. Set-up and model formulation

2.1. Networks under attack with limited knowledge

We consider a random network $G(V, E)$ with the node set V and edge set E. There are $|V| = N$ nodes in the network and they have an arbitrary degree distribution following the configuration model [2, 23]. Specifically, let $P(q) = P(q; 0)$ be the probability that a randomly chosen node has degree q initially at time $t = 0$. The generating function for the degree distribution of the network is defined as $G_0(x) = \sum_{q=0}^{\infty} P(q)x^q$ and the outgoing degrees of nodes reached by following a randomly chosen edge can be generated by the so-called excess degree generating function [24, 23] $G_1(x) = G_0'(x)G_0'(1)^{-1}$, where $G_0'(1) = \langle q \rangle$ is the mean degree of a node in $G(V, E)$.

The attacker is assumed to have limited knowledge over the network structure, namely, a number n of nodes and their degrees. Here, $n \in [1, N]$ indicates the extent of knowledge. At each time step, the attacker will remove the node with highest degree among the randomly selected n nodes in the max-n strategy. The process is repeated until a fraction of $1 - p$ nodes are deleted from the network $G(V, E)$. The max-n strategy is effective when immunizing individuals against epidemics like COVID-19 [25, 11], where testers can be sent out to stores collecting contact information of customers. Those with highest tracked contacts during a certain period of time may be immunized or quarantined. Analogously, we also consider the min-n strategy, where the node with smallest degree among the n randomly selected nodes will be removed at each step. This scenario gives us an estimate for the maximal achievable network integrity. It is also of interest when practical limitations such as cost is taken into consideration as the attack cost has often found to be positively correlated to the node degree [26]. Practical network dismantling strategies minimizing attack cost have attracted considerable research attention recently [27, 28, 29, 30].

2.2. Robustness and stability of Gk-cores

Given an integer $k \geq 2$, recall that a k-leaf is a node with degree less than k. The k-leaf pruning process starts from randomly removing a k-leaf with all its nearest neighbors and their incident edges. We repeat this process until no k-leaves remain in the network. The resulting network is called the Gk-core. It is shown in [21] that Gk-core is not a function of the network under consideration but rather may vary with the deletion order of k-leaves. However, the removal process is self-averaging in the thermodynamic limit of $N \to \infty$, namely, almost all resulting networks admit the same degree distribution irrespective of the deletion order [19, 20].

To characterize the network robustness in terms of Gk-core under the two types of attacks, we denote by $N_{kc}^{\min}$, $L_{kc}^{\min}$, $N_{kc}^{\max}$, and $L_{kc}^{\max}$, respectively, the average relative (i.e., normalized

by N) numbers of nodes and edges in the resulting Gk-core after launching min-n and max-n attacks over the network $G(V, E)$. Here, the average is taken over the ensemble of possible graphs created in the random network $G(V, E)$ with arbitrary degree distributions. By convention in statistics physics, we will resort to mean-field theory [2, 23], which allows us to appreciate the average network structure characteristics in the large graph size limit.

An orthogonal dimension to network robustness is the stability, which looks into the common nodes shared by a functional component such as giant component [31, 32] and Gk-core [20] when a network undergoes repeated independent percolation processes. A higher number of common nodes indicates a stronger stability of the network as these nodes are likely to be retained regardless of a specific adverse event. Under min-n attack, we can measure the stability of Gk-core by computing the relative number of nodes inside all Gk-cores for ℓ independent realization of min-n attacks on the network, that is,

$$S_{kc}^{\min}(\ell) := \frac{1}{N} \left| \cap_{l=1}^{\ell} C_l \right|, \tag{1}$$

where C_l represents the Gk-core in the l-th realization of min-n attack where a fraction $1 - p$ of nodes are deleted from $G(V, E)$, and $|C|$ is the cardinality of set C as before. Under max-n attacks, we can similarly denote the stability of Gk-core by $S_{kc}^{\max}(\ell)$. When $\ell = 1$, we have $S_{kc}^{\min}(1) = N_{kc}^{\min}$ and $S_{kc}^{\max}(1) = N_{kc}^{\max}$ by definition. Notice that the "stability" of a randomly chosen subset of size M from the network $G(V, E)$ would decay exponentially fast as $S(\ell) \propto (M/N)^{\ell}$.

3. Analytical solutions for attacks under limited knowledge

In this section, we develop a theoretical framework for deriving the numbers of nodes and edges as well as the stability of Gk-core under the two types of attacks with limited knowledge. Recall that the initial degree distribution of the network $G(V, E)$ is $P(q) = P(q; 0)$. The corresponding cumulative distribution, i.e. the probability that a randomly chosen node has degree at most q, is given by $F(q) = F(q; 0) = \sum_{r=0}^{q} P(r)$ for $q \geq 0$.

3.1. Gk-cores under min-*n attack*

In the min-n attack, recall that we at each step scrutinize n random nodes and remove the one with lowest degree. This process runs until a fraction of $1 - p$ nodes are removed. Assuming that we only delete the node but keep the edges connecting the removed node with the remaining nodes, we denote by $P(q; t)$ the degree distribution of a randomly selected remaining node at time $t \geq 0$. The corresponding cumulative distribution is given by $F(q; t) = \sum_{r=0}^{q} P(r; t)$, i.e., the probability that a randomly selected node in the remaining network at time t has degree at most q.

Using the minimum order statistics for independent random variables [33], the degree distribution of the attacked node at time $t \geq 0$ can be expressed as

$$\begin{aligned} & [1 - (1 - F(q; t))^n] - [1 - (1 - F(q - 1; t))^n] \\ =& \Delta \left(1 - (1 - F(q; t))^n\right) \\ =& - \Delta \left((1 - F(q; t))^n\right), \end{aligned} \tag{2}$$

for $q \geq 0$, where Δ is the difference operator with respect to q and we set $F(-1; t) = 0$ for all t. Since the quantity (2) gives the probability that the attacked node at time t has degree q, with one

more node being removed we obtain

$$N(q;t+1) = N(q;t) + \Delta\left((1-F(q;t))^n\right), \tag{3}$$

where $N(q;t)$ is the number of nodes with degree q in the remaining network at time t. Noticing that there are $N-t$ nodes in the network at time t, we have $P(q;t) = N(q;t)/(N-t)$. Hence, it follows from (3) in the continuous limit that

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

Plugging $P(q;t) = \Delta F(q;t)$ into (4), we have

$$\Delta\left((N-t)\frac{\partial F(q;t)}{\partial t} - F(q;t) - (1-F(q;t))^n\right) = 0. \tag{5}$$

Since $F(-1;t) = 0$ for all $t \geq 0$, we have from (5) that for any $q \geq 0$,

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

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

$$F(q;t) = 1 - \left(1 + ((1-F(q))^{1-n} - 1)e^{(n-1)\ln\left(\frac{N-t}{N}\right)}\right)^{\frac{1}{1-n}}. \tag{7}$$

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

$$F_p^{\min}(q) = 1 - \left(1 + ((1-F(q))^{1-n} - 1)p^{n-1}\right)^{\frac{1}{1-n}}, \tag{8}$$

which is the cumulative distribution of the degree of a random remaining node after removing a $1-p$ fraction of nodes (while keeping all edges) under the min-n attack. When $n = 1$, it is easy to check that the solution of the system (6) is $F_p^{\min}(q) = F(q)$, which agrees with the limit value by taking $n \to 1^+$ in (8). Therefore, the probability that a randomly chosen node in the remaining network when a fraction of $1-p$ nodes are removed (but assuming edges are kept intact) can be expressed as

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

We define the corresponding generating function as $\hat{G}_0(x) = \sum_{q=0}^{\infty} P_p(q)x^q$ and the mean degree is $\langle q(p)\rangle = \sum_{q=0}^{\infty} qP_p(q) = \hat{G}_0'(1)$.

Next, we consider to remove the edges linking the removed nodes to the remaining nodes. Since the configuration model network $G(V,E)$ is randomly connected, the probability of a random edge leading to a remaining node is on a par with the ratio of the number of edges leaving the remaining nodes to the total number of edges leaving all nodes in the original network $G(V,E)$:

$$\hat{p} = \frac{pN\langle q(p)\rangle}{N\langle q\rangle} = \frac{p\,\hat{G}_0'(1)}{G_0'(1)}. \tag{10}$$

Hence, deleting the edges leading to a removed node is equivalent to deleting randomly a $1 - \hat{p}$ fraction of edges of the remaining nodes in a randomly connected network. Following the percolation approach in [24], the generating function of the remaining nodes after min-n attack is

$$\tilde{G}_0(x) = \hat{G}_0(1 - \hat{p} + \hat{p}x) = \sum_{q=0}^{\infty} \tilde{P}(q)x^q, \tag{11}$$

where $\hat{p}$ is given by (10) and $\tilde{P}(q)$ means the probability of a random node with degree q in the remaining network, which we will denote by $\tilde{G}(\tilde{V}, \tilde{E})$. Clearly, $|\tilde{V}| = pN$. Recall that $F_p^{\min}(q) = F(q)$ when $n = 1$. Hence, $P_p(q) = P(q)$ by (9), $\hat{G}_0(x) = G_0(x)$, and $\hat{p} = p$ by (10). The generating function (11) gives $\tilde{G}_0(x) = G_0(1-p+px)$, which is consistent with [24] under the random attack scenario.

With (11) we are now in the position to consider the Gk-core percolation process (over the remaining network $\tilde{G}(\tilde{V}, \tilde{E})$) and derive the numbers of nodes and edges in Gk-core for $k \geq 2$. Similarly as in [21], we consider specifically three categories of nodes in $\tilde{G}$ during the k-leaf pruning algorithm: a node is α-removable if it can become a $(k-1)$-leaf; a node is β-removable if it can become a neighbor of a k-leaf; a node is non-removable if it is in the Gk-core. Given comments in Section 2, the category of a node may vary with the order of deletion. However, a node cannot be in both α-removable and β-removable categories at the same time [16, 21]. Moreover, with some abuse of notation, let α and β be the probabilities that the end node of a randomly chosen edge emanating from a random node i is α-removable and β-removable in $\tilde{G}\backslash\{i\}$, respectively. Note that the definition of these probabilities is different from [19].

An end node i reached by following a random edge emanating from the other end node j is not α-removable or β-removable in $\tilde{G}\backslash\{j\}$ if i has no less than $k-1$ neighbors which are not α-removable or β-removable in $\tilde{G}\backslash\{i\}$ and any other neighbors of i are β-removable in $\tilde{G}\backslash\{i\}$. Hence,

$$1 - \alpha - \beta = \sum_{q=k}^{\infty} \frac{q\tilde{P}(q)}{\tilde{G}_0'(1)} \sum_{s=k-1}^{q-1} \binom{q-1}{s} (1-\alpha-\beta)^s \beta^{q-1-s}, \tag{12}$$

where $\frac{q\tilde{P}(q)}{\tilde{G}_0'(1)}$ is the probability that i has degree q by (11). The node i has $q-1$ edges leaving from it, s of which link to nodes not of category α or β in $\tilde{G}\backslash\{i\}$.

Moreover, an end node i reached by following a random edge emanating from the other end node j is β-removable if it has an α-removable neighbor in $\tilde{G}\backslash\{i\}$. Therefore,

$$1 - \beta = \sum_{q=1}^{\infty} \frac{q\tilde{P}(q)}{\tilde{G}_0'(1)} (1-\alpha)^{q-1}. \tag{13}$$

Employing (11), (12), (13) and the binomial expansion, we obtain

$$\begin{aligned}\alpha =& \sum_{q=1}^{\infty} \frac{q\tilde{P}(q)}{\tilde{G}_0'(1)} \sum_{s=0}^{k-2} \binom{q-1}{s} (1-\alpha-\beta)^s \beta^{q-1-s} \\ =& \frac{1}{\tilde{G}_0'(1)} \sum_{s=0}^{k-2} \frac{(1-\alpha-\beta)^s}{s!} \tilde{G}_0^{(s+1)}(\beta) \\ =& \frac{1}{\hat{G}_0'(1)} \sum_{s=0}^{k-2} \frac{(1-\alpha-\beta)^s \hat{p}^s}{s!} \hat{G}_0^{(s+1)}(1-\hat{p}+\hat{p}\beta) \end{aligned} \tag{14}$$

and

$$\beta = 1 - \frac{\tilde{G}_0'(1-\alpha)}{\tilde{G}_0'(1)} = 1 - \frac{\hat{G}_0'(1-\alpha\hat{p})}{\hat{G}_0'(1)}, \tag{15}$$

where $\hat{p}$ is given by (10), $\hat{G}_0^{(s)}(x) = \sum_{q=0}^{\infty} P_p(q) \frac{q!}{(q-s)!} x^{q-s}$ is the s-th derivative of $\hat{G}_0(x)$, and $P_p(q)$ is given by (9).

Recall that $N_{kc}^{\min}$ means the probability that a randomly selected node $i \in V$ belongs to the Gk-core of $\tilde{G}$ generated by (11). The node i is in Gk-core if it is in $\tilde{G}$, has no less than k neighbors that are not α-removable or β-removable, and all other neighbors of i are of category β in $\tilde{G}\backslash\{i\}$. Consequently,

$$\begin{aligned} N_{kc}^{\min} =& p \sum_{q=k}^{\infty} \tilde{P}(q) \sum_{s=k}^{q} \binom{q}{s} (1-\alpha-\beta)^s \beta^{q-s} \\ =& p\tilde{G}_0(1-\alpha) - p \sum_{s=0}^{k-1} \frac{(1-\alpha-\beta)^s}{s!} \tilde{G}_0^{(s)}(\beta) \\ =& p\hat{G}_0(1-\alpha\hat{p}) - p \sum_{s=0}^{k-1} \frac{(1-\alpha-\beta)^s \hat{p}^s}{s!} \hat{G}_0^{(s)}(1-\hat{p}+\hat{p}\beta), \end{aligned} \tag{16}$$

where $P_p(q)$ and $\hat{p}$ are given by (9) and (10), respectively.

The expected normalized number of edges $L_{kc}^{\min}$ in the Gk-core of $\tilde{G}$ can be calculated as follows. Since an edge is in the Gk-core if the two end nodes are in the Gk-core, we obtain

$$L_{kc}^{\min} = p(1-\alpha-\beta)^2 \frac{\tilde{G}_0'(1)}{2} = p(1-\alpha-\beta)^2 \hat{p} \frac{\hat{G}_0'(1)}{2}, \tag{17}$$

where we used $|\tilde{V}| \frac{\tilde{G}_0'(1)}{2} = |\tilde{E}|$ and $pN = |\tilde{V}|$ similarly as in [20].

Finally, we derive the stability $S_{kc}^{\min}(\ell)$ of Gk-core for any $\ell \geq 1$. A random node is in the Gk-core of $\tilde{G}$ if it is in $\tilde{G}$ and has no less than k neighbors that are also within the Gk-core. Hence,

in view of (1) and (11) we obtain

$$\begin{aligned} S_{kc}^{\min}(\ell) =& p\sum_{q=k}^{\infty}\tilde{P}(q)\left[\sum_{s=k}^{q}\binom{q}{s}(1-\alpha-\beta)^s\beta^{q-s}\right]^{\ell} \\ =& p\sum_{q=k}^{\infty}\frac{\hat{p}^q\hat{G}_0^{(q)}(1-\hat{p})}{q!} \\ & \cdot\left[(1-\alpha)^q-\sum_{s=0}^{k-1}\binom{q}{s}(1-\alpha-\beta)^s\beta^{q-s}\right]^{\ell} \end{aligned} \tag{18}$$

where $\hat{p}$ is given by (10) and the term in the square brackets is the probability that a randomly chosen node belongs to the Gk-core given it is in $\tilde{G}$ and has degree q. When $n = 1$, we have $\hat{p} = p$ and $\hat{G}_0(x) = G_0(x)$ by (9) and (10). Hence, the stability (18) is in line with the result in [20] under random attack.

3.2. Gk-cores under max-*n attack*

When max-n attack is performed on $G(V, E)$, we at each step observe the degrees of n randomly chosen nodes and delete the one with highest degree. The attack stops when a fraction of $1 - p$ nodes are ditched. Similarly as in Section 3.1, we obtain the degree distribution of the attacked node at time t by using the maximum order statistics as

$$F(q;t)^n - F(q-1;t)^n = \Delta(F(q;t)^n), \tag{19}$$

for $q \geq 0$, where $F(-1;t) = 0$ for all t. With this replacing (2), along the same line of [11] we have the cumulative distribution of the degree of a random remaining node after deleting a $1 - p$ fraction of nodes (while keeping all edges) under the max-n attack

$$F_p^{\max} = \left(1 + (F(q)^{1-n} - 1)p^{n-1}\right)^{\frac{1}{1-n}} \tag{20}$$

and the corresponding degree distribution of a random remaining node

$$P_p(q) = \Delta F_q^{\max}(q) = F_p^{\max}(q) - F_p^{\max}(q-1), \tag{21}$$

which redefines the expression (9).

The generating function (when all edges are kept intact) is defined as $\hat{G}_0(x) = \sum_{q=0}^{\infty} P_p(q)x^q$. Similarly as in Section 3.1, when edges leading to the removed nodes under max-n attack are deleted, the generating function of the remaining nodes can be obtained as (11), where we will use (21) here to feed into $P_p(q)$ and $\tilde{P}(q)$. When $n = 1$, we also have $P_p(q) = P(q)$ by (21) and reproduce the random attack scenario with $\tilde{G}_0(x) = G_0(1 - p + px)$.

Again, denote by $\tilde{G}(\tilde{V}, \tilde{E})$ the resulting network after max-n attack. Following the method in Section 3.1, we can analogously calculate the two probabilities α and β as in (14) and (15), respectively. Likewise, the expected relative size of Gk-core is given by

$$N_{kc}^{\max} = p\hat{G}_0(1-\alpha\hat{p}) - p\sum_{s=0}^{k-1}\frac{(1-\alpha-\beta)^s\hat{p}^s}{s!}\hat{G}_0^{(s)}(1-\hat{p}+\hat{p}\beta) \tag{22}$$

and the expected normalized number of edges of Gk-core is

$$L_{kc}^{\max} = p(1-\alpha-\beta)^2\hat{p}\frac{\hat{G}_0'(1)}{2}, \tag{23}$$

where $\hat{G}_0(x)$ and $\hat{p}$ require the input of (21).

The expected stability $S_{kc}^{\max}(\ell)$ of Gk-core for $\ell \geq 1$ similarly is

$$S_{kc}^{\max}(\ell) = p\sum_{q=k}^{\infty}\frac{\hat{p}^q\hat{G}_0^{(q)}(1-\hat{p})}{q!} \cdot\left[(1-\alpha)^q - \sum_{s=0}^{k-1}\binom{q}{s}(1-\alpha-\beta)^s\beta^{q-s}\right]^{\ell}. \tag{24}$$

The special case of $n = 1$ corresponds to $\hat{p} = p$ and $\hat{G}_0(x) = G_0(x)$ by (21), and therefore agrees with the stability under the random attack scenario [20].

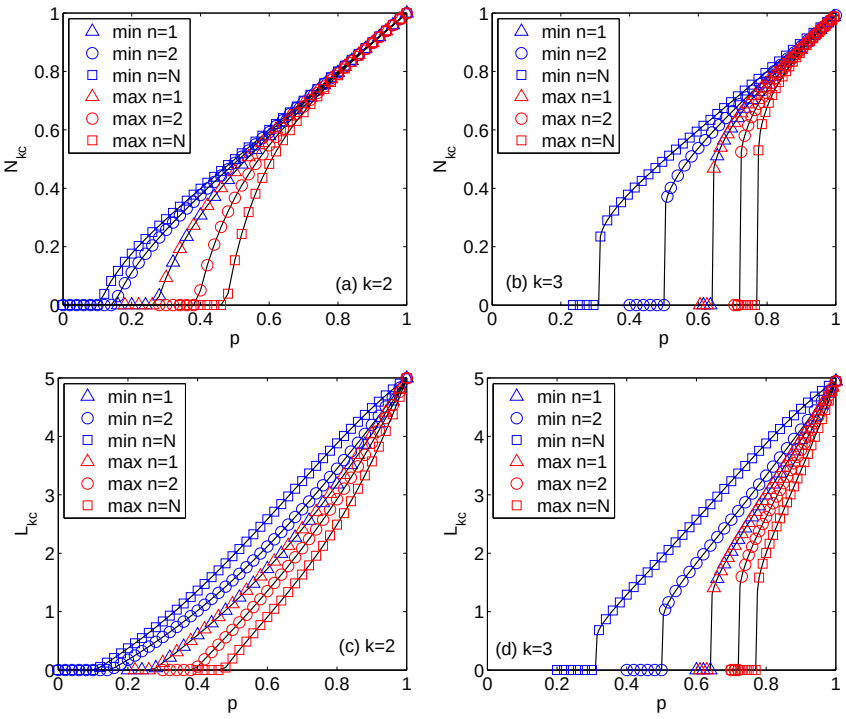

Figure 1: The fraction N_{kc} of Gk-core is shown in (a) for $k = 2$ and (b) for $k = 3$. The corresponding normalized number L_{kc} of edges in Gk-core is shown in (c) for $k = 2$ and (d) for $k = 3$. Analytical results (black solid curves) and simulations (symbols) under min-n and max-n attacks with different n agree well with each other. Blue symbols are for min-n attack and red symbols are for max-n attack with $n = 1$ (triangles), $n = 2$ (circles), and $n = N$ (squares). Simulations are averaged over 50 independent realizations for ER networks with size $N = 10^7$ and mean degree $\lambda = 10$.

4. Synthetic networks

We conduct numerical simulations in this section to test analytical results for Gk-cores derived in Section 3 on homogeneous networks with Poisson degree distributions and heavy-tailed log-normal networks. Simulations are based on networks with $N = 10^7$ nodes. Note that we here do not consider the well-known scale-free networks because they admit merely a trivial Gk-core [19] for all $k \geq 2$.

4.1. Erdős-Rényi random networks

For an Erdős-Rényi (ER) random network, its degree distribution obeys $P(q) = e^{-\lambda}\lambda^q/q!$ for $q \geq 0$ with mean degree λ. The generating function for degrees is $G_0(x) = e^{\lambda(x-1)}$. We exhibit in Fig. 1 the fraction N_{kc} of Gk-core under min-n and max-n attacks and the corresponding normalized number of edges L_{kc} in Gk-core for ER networks with $\lambda = 10$. Several interesting observations are in order.

Firstly, for the full spectra of min-n and max-n attacks, both N_{kc} and L_{kc} display continuous phase transition for $G2$-core while Gk-core for $k \geq 3$ emerges discontinuously at the percolation threshold. This agrees with the random attack situation (i.e. $n = 1$) discovered in [20]. We note that even the most harmful targeted attack max-N yields continuous phase transition for $G2$-core, and the mildest attack min-N is powerful enough to show a first-order phase transition for $G3$-core. In all cases, theoretical results show excellent consistency with extensive simulations (which have also been checked for $k \geq 4$). Secondly, the effect of knowledge growth (the increase of knowledge index n from 1 to N) has a diminishing marginal utility for both groups of measures $\{N_{kc}^{\min}, L_{kc}^{\min}\}$ and $\{N_{kc}^{\max}, L_{kc}^{\max}\}$. This phenomenon of Gk-core echoes the observation in [19] for the giant component P_∞ under max-n attack.

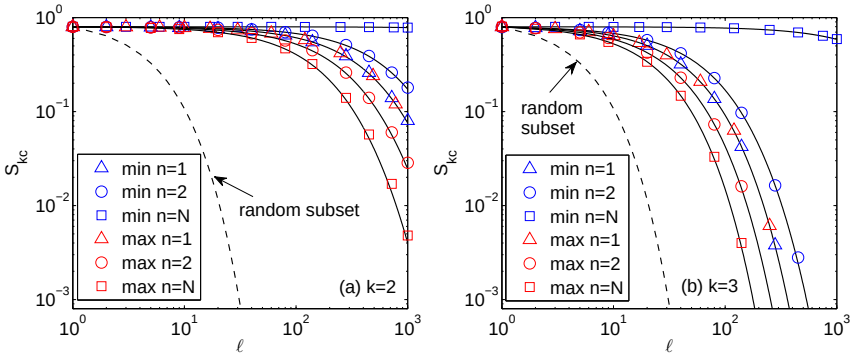

Figure 2: The stability $S_{kc}(\ell)$ of Gk-core with fraction 0.8 is shown in (a) for $k = 2$ and (b) for $k = 3$. Analytical results (black solid curves) and simulations (symbols) under min-n and max-n attacks with different n agree well with each other. Blue symbols are for min-n attack and red symbols are for max-n attack with $n = 1$ (triangles), $n = 2$ (circles), and $n = N$ (squares). Simulations are averaged over 50 independent realizations for ER networks with size $N = 10^7$ and mean degree $\lambda = 10$. Dashed curve means the stability of a randomly chosen subset with fraction 0.8.

Next, we calculate the stability measures $S_{kc}^{\min}(\ell)$ and $S_{kc}^{\max}(\ell)$, as shown in Fig. 2, by testing the numbers of common nodes in ℓ independent realizations of min-n and max-n attacks, respectively. We observe that Gk-core is extremely stable under the mildest attack min-N. This can be understood as the nodes with smallest degrees are not likely to be present in the Gk-core. The stability $S_{kc}^{\min}$ declines as the knowledge index n decreases for min-n attack, while $S_{kc}^{\max}$ declines

as n increases for max-n attack. This phenomenon is non-trivial as stability is not necessarily correlated with robustness in general. For example, it is shown in [31] that ER networks are robust against link percolation but turn out to be unstable with respect to link failure due to lack of anchor nodes to maintain the structure of giant component. In our context here, more knowledge for max-n attack (and analogously, less knowledge for min-n attack) tends to damage hub nodes, which markedly fluctuates the Gk-core, and hence leads to the decline of $S_{kc}(\ell)$ as ℓ increases.

Comparing Fig. 2(a) and Fig. 2(b), it can be seen that the inner Gk-cores with larger k are less stable. This agrees with our intuition that the removal of more nodes, namely the k-leaves and their neighbors, tends to volatilize the Gk-core structure more seriously. However, in all cases considered in Fig. 2, the decay of stability is much pronouncedly slower than the random subset scenario, indicating the existence of non-trivial cohesive architecture of Gk-core.

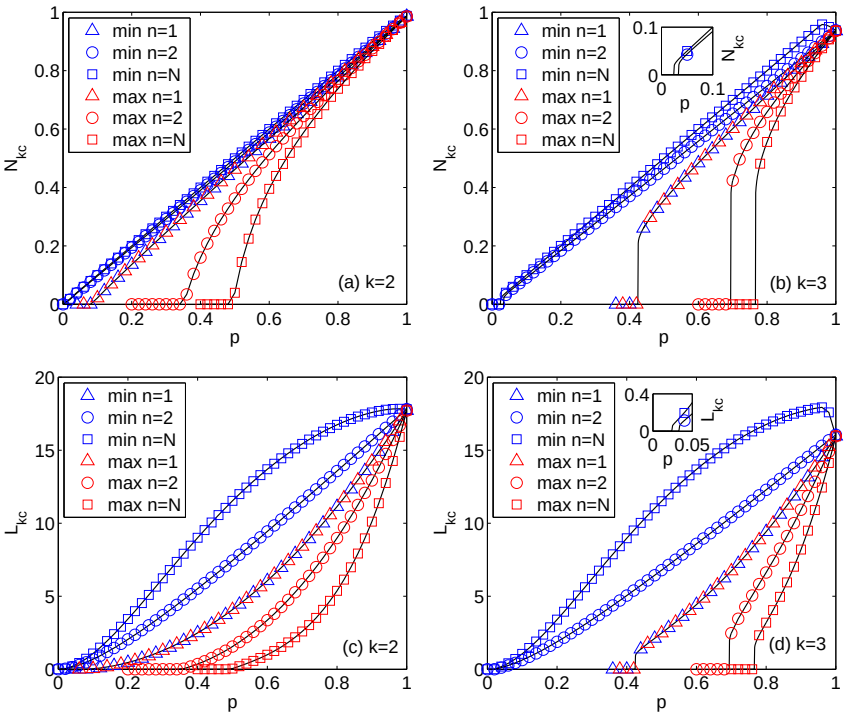

Figure 3: The fraction N_{kc} of Gk-core is shown in (a) for $k = 2$ and (b) for $k = 3$. The corresponding normalized number L_{kc} of edges in Gk-core is shown in (c) for $k = 2$ and (d) for $k = 3$. Analytical results (black solid curves) and simulations (symbols) under min-n and max-n attacks with different n agree well with each other. Blue symbols are for min-n attack and red symbols are for max-n attack with $n = 1$ (triangles), $n = 2$ (circles), and $n = N$ (squares). Simulations are averaged over 50 independent realizations for log-normal networks with size $N = 10^7$ and $\sigma = 2$. Insets in (b) and (d) show a magnified view of $N_{kc}^{\min}$ and $L_{kc}^{\min}$, respectively, for $n = 2$ and $n = N$ for small p.

4.2. Log-normal random networks

We consider a network with asymptotic log-normal degree distribution follows $P(q) \propto \exp(-(\ln q - \sigma)^2/4)$ for $q \geq 1$. Compared with scale-free distributions, log-normal distributions often fit data better in real-world growing networks [34, 35, 36], which are heavy-tailed and skewed with

mode $e^{\sigma-2}$ and mean $e^{\sigma+1}$ in our context. Fig. 3 shows the behavior of N_{kc} and L_{kc} as functions of occupation fraction p for log-normal networks with $\sigma = 2$ under min-n and max-n attacks.

We observe that $\{N_{kc}^{\min}, L_{kc}^{\min}\}$ and $\{N_{kc}^{\max}, L_{kc}^{\max}\}$ show continuous percolation thresholds for $G2$-core but have first-order phase transition as p evolves from 0 to 1 for Gk-core with $k \geq 3$ over the whole range of knowledge index n. This is qualitatively similar to the ER networks. However, there are two remarkable differences when comparing Fig. 3 with Fig. 1. Firstly, the effect of min-n attack on $N_{kc}^{\min}$ is much closer to that of random attack for log-normal networks than for ER networks. This phenomenon can be attributed to the heterogeneity of degree distribution in the log-normal network, where targeting at low-degree nodes could barely affect its Gk-core.

Secondly, the change of $N_{kc}^{\min}$ (Fig. 3(b)) and $L_{kc}^{\min}$ (Fig. 3(d)) for $G3$-core in the case of $n = N$ is non-monotonic. Both measures slightly climb as p decreases from 1, they reach the peak at around $p = 0.95$ and then decrease gradually. This means $G3$-core actually becomes stronger under initial min-N attack. The deletion of a small fraction of nodes with smallest degrees hinders the propagation effect in $G3$-core as the small-degree nodes are likely to connected to some larger-degree nodes which otherwise may be deleted in the later 3-leaf pruning procedure. A schematic scenario is illustrated in Fig. 4, which also indicates an explanation that $G2$-core does not show such non-monotonicity. This phenomenon is not observed in ER networks either due to its degree homogeneity.

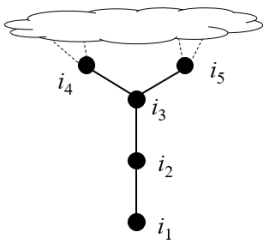

Figure 4: If i_1 is deleted under min-N attack, i_4 and i_5 will be in the $G3$-core; otherwise none of the five displayed nodes will be in the $G3$-core and the damage may further propagate. However, i_3, i_4, i_5 will be in the $G2$-core regardless of whether i_1 is deleted under min-N attack.

For log-normal networks, Fig. 5 shows the Gk-core stability S_{kc} under min-n and max-n attacks as a function of ℓ, the repeated rounds of attack, when the fraction of Gk-core is fixed at 0.8 as for ER networks shown in Fig. 2. Similarly as in the ER network case, the stability for Gk-core is highest under min-N attack and it is deteriorated as the attack becomes more harmful in terms of decreasing knowledge extent under min-n attack or increasing knowledge extent under max-n attack. This again is due to the fact that lower-degree nodes are less likely to contribute to the Gk-core. Nonetheless, comparing Fig. 5 with Fig. 2, it is found that $G2$-core is less stable for log-normal networks than for ER networks under all min-n and max-n attacks since the hub nodes in log-normal networks serve as anchors helping stabilize the $G2$-core. Inner cores such as $G3$-core, however, seem to be not sensitive to the degree heterogeneity difference in initial network topology.

5. Real-world networks

In this section, we compare the fraction of Gk-core under min-n and max-n attacks for two real-world networks. The first network ElecCirc is an electronic circuit [37], where logic gates

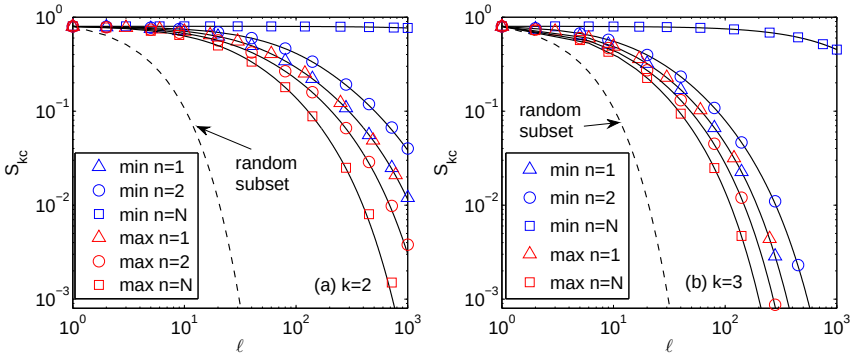

Figure 5: The stability $S_{kc}(\ell)$ of Gk-core with fraction 0.8 is shown in (a) for $k = 2$ and (b) for $k = 3$. Analytical results (black solid curves) and simulations (symbols) under min-n and max-n attacks with different n agree well with each other. Blue symbols are for min-n attack and red symbols are for max-n attack with $n = 1$ (triangles), $n = 2$ (circles), and $n = N$ (squares). Simulations are averaged over 50 independent realizations for log-normal networks with size $N = 10^7$ and $\sigma = 2$. Dashed curve means the stability of a randomly chosen subset with fraction 0.8.

in digital circuits form the nodes and the wires are edges. This network has 24097 nodes and 53248 edges showing a small-world distribution with mean degree 4.4. The second network CaMath describes collaboration among mathematicians [38], where nodes are authors and two authors are adjacent if they share the authorship of a paper. CaMath has a truncated power-law distribution with mean 3.9 over 253339 nodes and 496489 edges.

In Fig. 6, we show the relative numbers of nodes in the Gk-cores after launching min-n and max-n attacks over the two networks with $p = 0.9$. The theoretical predictions are generally in line with the real sizes of Gk-cores showing the effect of knowledge extent on different attack strategies. In some cases, for instance in $G3$-core of CaMath network, there are some notable discrepancies indicating that other structural features, in addition to degrees, such as correlation and clustering can strongly affect the Gk-core organization.

6. Conclusion

In this paper we have developed a theoretical framework to understand the influence on the Gk-core structure of a network under attacks with limited knowledge. Deviating from traditional giant components, Gk-core is revealed via a k-leaf removal process, which progressively deletes k-leaves together with their nearest neighbors and can be viewed as a generalization of the ordinary k-core decomposition. The robustness of Gk-core is characterized by the numbers of nodes and edges of it and the stability of Gk-core is framed as the same nodes that survive repeated realizations of attacks. To uncover the effect of realistic attacks on a network of size N, we have considered two types of attacks with limited knowledge extent regarding the network degrees, namely, min-n and max-n attacks, for the whole range of knowledge index $n = 1, 2, \cdots, N$. It is found that $G2$-core undergoes a second-order percolation transitions under both attacks while Gk-cores for $k \geq 3$ always display a discontinuous phase transition. Knowing one more node during attacks, improving from $n = 1$ to $n = 2$, turns out to be most beneficial in terms of changing the robustness of Gk-core. Some interesting non-monotonicity phenomena have also been uncovered exclusively for networks with heterogeneous degree distributions. The anchor nodes play a key role in stabilizing Gk-cores, reminiscent of those in network giant components under

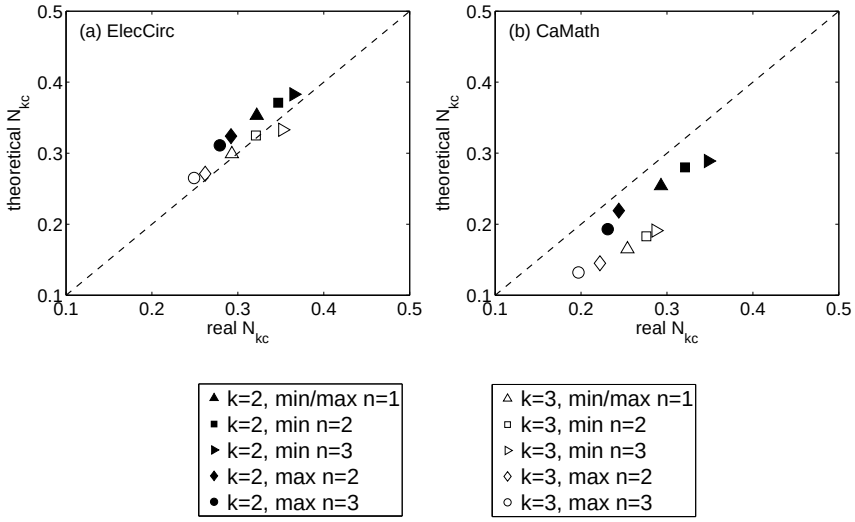

Figure 6: The fraction N_{kc} of Gk-core at $p = 0.9$ in (a) the electronic circuit network and (b) the mathematicians coauthorship network versus analytical predictions for $k = 2$ (solid symbols) and $k = 3$ (hollow symbols) under min-n attack with $n = 1$ (upper triangles), $n = 2$ (squares), $n = 3$ (right triangles) and max-n attack with $n = 2$ (diamonds) and $n = 3$ (circles). Simulation results are averaged over 20 independent realizations.

link percolation. The insights brought here could help to design robust and stable networked systems.

References

[1] A.-L. Barabási, Network Science, Cambridge University Press, Cambridge, 2016.
[2] M. E. J. Newman, Networks, 2nd Edition, Oxford University Press, Oxford, 2018.
[3] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, D. J. Watts, Phys. Rev. Lett. 85 (2000) 5468–5471.
[4] R. Cohen, S. Havlin, Complex Networks: Structure, Robustness and Function, Cambridge University Press, Cambridge, 2010.
[5] R. Albert, H. Jeong, A.-L. Barabási, Nature 406 (2000) 378.
[6] R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, Phys. Rev. Lett. 86 (2001) 3682–3685.
[7] S. Iyer, T. Killingback, B. Sundaram, Z. Wang, PLoS ONE 8 (2013) e59613.
[8] Y. Shang, IEEE Trans. Syst. Man Cybern. Syst. 49 (2019) 821–832.
[9] G. S. Costa, S. C. Ferreira, Phys. Rev. E 101 (2020) 022311.
[10] Z. Wang, C. T. Bauch, S. Bhattacharyya, A. d'Onofrio, P. Manfredi, M. Perc, N. Perra, M. Salathé, D. Zhao, Phys. Rep. 664 (2016) 1–113.
[11] Y. Liu, H. Sanhedrai, G. Dong, L. M. Shekhtman, F. Wang, S. V. Buldyrev, S. Havlin, Natl. Sci. Rev. (2020) nwaa229.
[12] L. K. Gallos, R. Cohen, P. Argyrakis, A. Bunde, S. Havlin, Phys. Rev. Lett. 94 (2005) 188701.
[13] Y. Shang, EPL (Europhys. Lett.) 95 (2011) 28005.
[14] Y. Yin, Q. Liu, C. Zhang, J. Zhou, Physica A 531 (2019) 120957.
[15] B. C. Coutinho, A.-K. Wu, H.-J. Zhou, Y.-Y. Liu, Phys. Rev. Lett. 124 (2020) 248301.
[16] Y.-Y. Liu, E. Csóka, H. Zhou, M. Pósfai, Phys. Rev. Lett. 109 (2012) 205703.
[17] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, Phys. Rev. Lett. 96 (2006) 040601.
[18] Y.-X. Kong, G.-Y. Shi, R.-J. Wu, Y.-C. Zhang, Phys. Rep. 832 (2019) 1–32.
[19] N. Azimi-Tafreshi, S. Osat, S. N. Dorogovtsev, Phys. Rev. E 99 (2019) 022312.
[20] Y. Shang, New J. Phys. 21 (2019) 093013.
[21] Y. Shang, SIAM J. Appl. Math. 80 (2020) 1272–1289.

[22] Y. Shang, Phys. Rev. E 101 (2020) 042306.
[23] M. E. J. Newman, S. H. Strogatz, D. J. Watts, Phys. Rev. E 64 (2001) 026118.
[24] M. E. J. Newman, Phys. Rev. E 66 (2002) 016128.
[25] E. C. Dinleyici, R. Borrow, M. A. P. Safadi, P. van Damme, F. M. Munoz, Hum. Vaccinces Immunother. (2020). URL: doi:10.1080/21645515.2020.1804776.
[26] X.-L. Ren, N. Gleinig, D. Tolić, N. Antulov-Fantulin, Complexity 2018 (2018) 9826243.
[27] A. Patron, R. Cohen, D. Li, S. Havlin, Phys. Rev. E 95 (2017) 052305.
[28] X.-L. Ren, N. Gleinig, D. Helbing, N. Antulov-Fantulin, Proc. Natl. Acad. Sci. USA 116 (2019) 6554–6559.
[29] S. Wandelt, X. Sun, D. Feng, M. Zanin, S. Havlin, Sci. Rep. 8 (2018) 13513.
[30] C. Wang, Y. Xia, IEEE Access 8 (2020) 172398–172404.
[31] M. Kitsak, A. A. Ganin, D. A. Eisenberg, P. L. Krapivsky, D. Krioukov, D. L. Alderson, I. Linkov, Phys. Rev. E 97 (2018) 012309.
[32] R. Kühn, J. van Mourik, Phys. Rev. E 102 (2020) 032302.
[33] B. C. Arnold, N. Balakrishnan, H. N. Nagaraja, A First Course in Order Statistics, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2008.
[34] M. Seshadri, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos, J. Leskovek, in: Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Las Vegas, Nevada, 2008, pp. 596–604.
[35] A. D. Broido, A. Clauset, Nat. Commun. 10 (2019) 1017.
[36] M. Feng, L.-J. Deng, F. Chen, M. Perc, J. Kurths, Proc. R. Soc. A 476 (2020) 20200019.
[37] R. F. i Cancho, C. Janssen, R. V. Solé, Phys. Rev. E 64 (2001) 046119.
[38] R. D. Castro, J. W. Grossman, Math. Intell. 21 (1999) 51–53.