SIAM J. APPL. MATH.
Vol. 80, No. 3, pp. 1272–1289

GENERALIZED K-CORE PERCOLATION IN NETWORKS WITH COMMUNITY STRUCTURE*

YILUN SHANG†

Abstract. Community structure underpins many complex networked systems and plays a vital role when components in some modules of the network come under attack or failure. Here, we study the generalized k-core (Gk-core) percolation over a modular random network model. Unlike the archetypal giant component based quantities, Gk-core can be viewed as a resilience metric tailored to gauge the network robustness subject to spreading virus or epidemics paralyzing weak nodes, i.e., nodes of degree less than k, and their nearest neighbors. We develop two complementary frameworks, namely, the generating function formalism and the rate equation approach, to characterize the Gk-core of modular networks. Through extensive numerical calculations and simulations, it is found that $G2$-core percolation undergoes a continuous phase transition while Gk-core percolation for $k \geq 3$ displays a first-order phase transition for any fraction of interconnecting nodes. The influence of interconnecting nodes tends to be more visible nearer the percolation threshold. We find by studying modular networks with two Erdős–Rényi modules that the interconnections between modules affect the $G2$-core percolation phase transition in a way similar to an external field in a spin system, where Widom's identity regulating the critical exponents of the system is fulfilled. However, this analogy does not seem to exist for Gk-core with $k \geq 3$ in general.

Key words. complex network, community structure, resilience, core, percolation

AMS subject classifications. 91D30, 60K35, 82B26

DOI. 10.1137/19M1290607

1. Introduction. Complex networks have become a popular theoretical and applicable tool to the modeling and analysis of interaction phenomena that occur in a broad range of real-world systems [23, 7, 9, 24]. One of the most common and fascinating structure properties in many real-world systems is the modularity or community structure, where nodes are more tightly joined together in mesoscopic groups (called modules or communities) with intraconnections than between groups with interconnections. These interconnections often run between a small fraction of interconnecting nodes in each group [12, 34]. For example, the world wide web has more connections within pages on closely related topics than between distinct topics. Metabolic networks have more connections within functional modules than between modules. Transportation networks have more connections within cities than between cities. Due to numerous applications across disciplines such as technology, biology, and social sciences, detecting and characterizing community structure has been a very prolific research area for many years; see, e.g., [11, 28, 29, 35, 27] and references therein.

Despite enormous interest in community detection algorithms and models in data mining and applications, nevertheless, there have been relatively few efforts toward understanding the implication of community structure to the resilience or stability, which plays a crucial role in maintaining the connectivity and functionality of networks [5]. In particular, how the small fraction of nodes sustaining interconnections

*Received by the editors September 30, 2019; accepted for publication (in revised form) March 23, 2020; published electronically May 20, 2020.
https://doi.org/10.1137/19M1290607
Funding: This work was supported by the UoA Flexible Fund under grant 201920A1001.
†Department of Computer and Information Sciences, Northumbria Univeristy, Newcastle, NE1 8ST, UK (yilun.shang@northumbria.ac.uk).

affect modular network robustness is not entirely understood. Based on the multivariate generating function approach, Shai et al. [30] showed that the number of modules undergoes a tipping point separating two regimes in modular networks: in one regime, deletion of interconnecting nodes breaks the entire network abruptly as modules are disconnected from each other while in the other regime, the network fragments continuously as the modules collapse internally with the removal of interconnecting nodes. A recent study by Dong et al. [8] reveals that the role of interconnections in networks with community structure can be viewed as an external field in a ferromagnetic-paramagnetic spin system near the percolation phase transition of giant connected component. This unique relationship for modular networks has also found to be responsible for structural resilience for modules embedded in two dimensional spaces [10] as well as the integrity of giant secure component in the presence of different classes of vulnerabilities [33].

In this article, we consider the resilience of modular networks in terms of a recently introduced metric, Generalized k-core (or Gk-core) [1], instead of the previous giant connected component based metrics (c.f. [30, 8, 10, 33]). Analytically, it turns out to be challenging to tackle since Gk-core percolation has high dependency and community structure greatly adds to its complexity. In the Gk-core percolation, we recursively remove a k-leaf (i.e., a node with degree less than k) together with all its nearest neighbors and their incident edges. When the process stops, the resulting subgraph is called Gk-core, which can be thought of as a generalization of the usual k-core (the maximum subgraph having degrees at least k) [3, 22, 18, 6] and core (equivalent to $G2$-core) [20, 19, 37]. Core has its origin in the Karp–Sipser algorithm for finding vertex cover and is closely related to maximum matching problem and network controllability [36, 21, 38]. Note that core has been studied under different names in graph theory [16] and that Gk-core is not a direct generalization of the ordinary k-core. We choose to use this name following the previous works [1, 31, 32] to avoid adding confusion. When assessing network resilience, unlike the connected components, Gk-core is ideal to quantify the robustness of networks against epidemics, in which a virus compromises weak nodes, i.e., k-leaves, and their nearest neighbors. The spreading dynamics of information has found rich applications within and beyond complexity science [15, 17].

Here, we propose two complementary methods, namely, the generating function formalism and the rate equation method, for deriving the numbers of nodes and edges of Gk-core in complex networks with community structure. The theoretical frameworks developed here are verified by extensive numerical calculations and simulations over modular networks with Erdős–Rényi network communities. Our results unravel that $G2$-core is essentially different from Gk-core with $k \geq 3$ over networks having community structure due to the following findings: (i) The emergence of $G2$-core undergoes a continuous phase transition as the network density changes while $G3$-core displays a first-order percolation phase transition, regardless of the intensity of interconnections; (ii) The Widom's identity relating critical exponents in physical phase transition [26, 14, 13] is only satisfied by $G2$-core percolation indicating that the effect of interconnections among modules can be analogized to an external field in a spin system. Moreover, our results uncover that adding a small fraction of interconnecting nodes boosts the size of Gk-core prominently only around the percolation threshold, which is in sharp contrast with the no-community case [31], where adding edges randomly or intentionally invariably gives rise to a uniform growth of Gk-core over a wide range of average degrees. The interconnecting nodes in networks with community structure play a unique part in bolstering the network robustness in terms of Gk-core.

The rest of the paper is organized as follows. In section 2, we present the interacting modular network model and Gk-core percolation process. The generating function formalism and rate equation method are developed in section 3 to gauge the numbers of nodes and edges in Gk-core. In section 4, we apply our theoretical framework to a benchmark random network model with Erdős–Rényi modules and investigate physical phase transitions relating critical exponents in Gk-core percolation numerically and by simulation. The paper is concluded in section 5 with some further remarks.

2. Model formulation. We consider a modular network $\mathcal{G}$ with m modules motivated by real-world networks where a certain (usually small) portion of nodes in each module maintain interconnections to other modules [8]. For $i \in \{1, 2, \dots, m\} := [m]$, module i has N_i nodes, and the intradegree distribution is defined as $P_i(q_i)$ meaning the probability of having q_i neighbors within module i. For each module i, there is a random subset of interconnecting nodes, which is taken as a random subset containing a fraction r_i of all the nodes of module i. For $i \neq j$, edges are randomly connected between these interconnecting nodes following interdegree distributions $P_{ij}(q_{ij})$ and $P_{ji}(q_{ji})$ as shown in Figure 1(a). In other words, the probability that an interconnecting node in module i has q_{ij} neighbors in module j is given by $P_{ij}(q_{ij})$. For ease of presentation, we often suppress the subscripts in the notations $P_i(\cdot)$ and $P_{ij}(\cdot)$ simply as $P(\cdot)$ when the meaning is clear from the independent variable. The interconnections between interconnecting nodes have the similar flavor as the bipartite version of configuration model [24, 25]. We will be interested in the thermodynamic limit as the network size tends to infinity by borrowing techniques from statistical physics [26, 14].

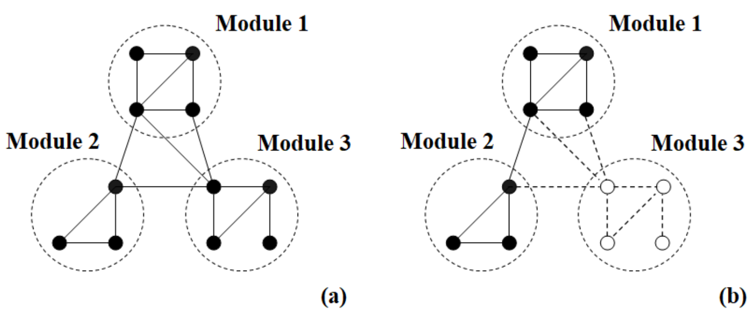

FIG. 1. *Schematic illustration of the modular network $\mathcal{G}$.* (a) *Three modules are interconnected via interconnecting nodes having fractions $r_1 = 1/2$, $r_2 = 1/3$, and $r_3 = 1/4$.* (b) *$G2$-core of the network (indicated by solid nodes), where module 3 has been removed as per the 2-leaf removal algorithm. Note that $G3$-core in this example is a null graph.*

Recall that a k-leaf in the network is defined as a node with degree less than k. The Gk-core percolation is implemented by a k-leaf removal algorithm, where at each time step a randomly chosen k-leaf is removed together with all nearest neighbors and their incident edges. The process is iterated until there is no k-leaves in the remaining network. The residual network is called Gk-core. Figure 1(b) shows the $G2$-core of the network in Figure 1(a).

It is worth noting that Gk-core defined above is not a function of the network, but depends on the order in which the k-leaves are deleted in the given network. This point has not been made explicitly in the previous works, e.g., [1, 31]. For instance, the network shown in Figure 2 has two different versions of $G3$-core. However, what's important is that the leaf removal process is self-averaging in the thermodynamic limit as the network size tends to infinity. In other words, almost all random residual networks have the same degree distribution, which is independent of the specific order of removal [1, 31]. Hence, we are able to tackle Gk-core using mean field analysis.

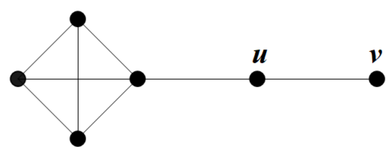

FIG. 2. *An example network with 6 nodes, which has different G3-cores. If v is deleted first, then G3-core is a complete graph over 4 nodes. If u is deleted first, then G3-core is a null graph.*

3. Theoretical framework. Given $k \geq 2$, let n_{kc} and l_{kc} represent the expected relative sizes of nodes and edges in Gk-core of the random network $\mathcal{G}$ with communities (with respect to the size of the entire network, i.e., $N := \sum_{i=1}^{m} N_i$). In other words, l_{kc} is the ratio of the number of edges surviving in the Gk-core to the total number of nodes N. We will first introduce the generating function method to derive the mean-field solutions for these quantities. Then we develop the rate equation method which can serve as an alternative approach to confirm the generating function formalism.

3.1. Generating function formalism. Unlike the relatively straightforward percolation theory for giant component based metrics [24], core based percolation benefits from a different route. During the k-leaf removal algorithm described in section 2, we scrutinize the nodes of $\mathcal{G}$ in three categories: (1) a-removable, if they can become a $(k-1)$-leaf; (2) b-removable, if they can become a neighbor of a k-leaf; and (3) nonremovable, if they belong to the Gk-core. As discussed above, the Gk-core depends on the removal order and the way a node is removed also depends on the order [1, 2]. However, a node can not be both a-removable and b-removable at the same time. This is similar to the paradigm established in [1, 20]; however, the introduction of community structure requires a nontrivial refined treatment of the probability partition to derive solvable self-consistency equations. It is tempting to consider a version of multivariate generating functions that have been used to deal with component based resilience in modular networks [30, 8, 10, 33]. However, we find it cumbersome with a monolithic multivariate version and choose to track the interconnections separately from intraconnections in Gk-core percolation.

To facilitate the discussion, we introduce the following definitions. Given $i \in [m] = \{1, 2, \dots, m\}$, let a_i and b_i be the probabilities that an end node of a randomly chosen intraedge (with the other end node, say v) in module i is a-removable and b-removable in $\mathcal{G}\backslash v$, respectively. We will refer to nodes in module i that are a-removable and b-removable as nodes of type a_i and type b_i, respectively. Furthermore, for $i \neq j$, let a_{ij} and b_{ij} be the probabilities that the end node in module i of a randomly chosen interedge (with the other end node in module j, say v) is a-removable and b-removable in $\mathcal{G}\backslash v$, respectively. Define the following generating functions for the degree distributions presented in section 2 as $G_{ii}(x_{ii}) = \sum_{q_i=0}^{\infty} P(q_i) x_{ii}^{q_i}$ for all i and $G_{ij}(x_{ij}) = \sum_{q_{ij}=0}^{\infty} P(q_{ij}) x_{ij}^{q_{ij}}$ for $i \neq j$. The sth derivative of $G_{ij}(x_{ij})$ is signified by $G_{ij}^{(s)}(x_{ij}) := \sum_{q_{ij}=0}^{\infty} P(q_{ij}) \frac{q_{ij}!}{(q_{ij}-s)!} x_{ij}^{q_{ij}-s}$ and the sth derivative of $G_{ii}(x_{ii})$ is signified by $G_{ii}^{(s)}(x_{ii})$ likewise. The average degrees are given by $\langle q_i \rangle := \sum_{q_i=0}^{\infty} q_i P(q_i) = G_{ii}^{(1)}(1)$ and $\langle q_{ij} \rangle := \sum_{q_{ij}=0}^{\infty} q_{ij} P(q_{ij}) = G_{ij}^{(1)}(1)$.

THEOREM 1. *Given $k \geq 2$, the mean-field solutions for Gk-core percolation over a network $\mathcal{G}$ with m communities are given by*

$$n_{kc} = \frac{\sum_{i=1}^m N_i n_{kc}^i}{\sum_{i=1}^m N_i} \quad \text{and} \quad l_{kc} = \frac{\sum_{i=1}^m N_i l_{kc}^i + \sum_{1\leq i<j\leq m} N_i r_i l_{kc}^{ij}}{\sum_{i=1}^m N_i}, \tag{3.1}$$

where

$$\begin{aligned} n_{kc}^i = {} & (1-r_i)\left[G_{ii}(1-a_i) - \sum_{s_i=0}^{k-1} \frac{(1-a_i-b_i)^{s_i}}{s_i!} G_{ii}^{(s_i)}(b_i)\right] \\ & + r_i\left[G_{ii}(1-a_i)\prod_{j\neq i}^m G_{ij}(1-a_{ji}) \right. \\ & \left. - \sum_{s_i+\sum_{j\neq i} s_{ij}=0}^{k-1} \frac{(1-a_i-b_i)^{s_i}}{s_i!} G_{ii}^{(s_i)}(b_i) \prod_{j\neq i}^m \frac{(1-a_{ji}-b_{ji})^{s_{ij}}}{s_{ij}!} G_{ij}^{(s_{ij})}(b_{ji})\right] \quad \forall i, \end{aligned} \tag{3.2}$$

$$l_{kc}^i = (1-a_i-b_i)^2 \frac{\langle q_i\rangle}{2} \quad \forall i,$$

$$l_{kc}^{ij} = (1-a_{ij}-b_{ij})(1-a_{ji}-b_{ji})\langle q_{ij}\rangle \quad \forall i<j,$$

and the probabilities $\{a_i\}_{i=1}^m, \{a_{ij}\}_{j\neq i}, \{b_i\}_{i=1}^m, \{b_{ij}\}_{j\neq i}$ are given by solving the $2m^2$ equations:

$$\begin{aligned} a_i = {} & \frac{1-r_i}{\langle q_i\rangle} \sum_{s_i=0}^{k-2} \frac{(1-a_i-b_i)^{s_i}}{s_i!} G_{ii}^{(s_i+1)}(b_i) \\ & + \frac{r_i}{\langle q_i\rangle} \sum_{s_i+\sum_{j\neq i} s_{ij}=0}^{k-2} \frac{(1-a_i-b_i)^{s_i}}{s_i!} G_{ii}^{(s_i+1)}(b_i) \\ & \cdot \prod_{j\neq i}^m \frac{(1-a_{ji}-b_{ji})^{s_{ij}}}{s_{ij}!} G_{ij}^{(s_{ij})}(b_{ji}) \quad \forall i, \end{aligned} \tag{3.3}$$

$$\begin{aligned} a_{ij} = {} & \frac{1}{\langle q_{ij}\rangle} \sum_{s_i+\sum_{j'\neq i} s_{ij'}=0}^{k-2} \frac{(1-a_i-b_i)^{s_i}}{s_i!} G_{ii}^{(s_i)}(b_i) \\ & \cdot \left(\prod_{j'\neq i}^m \frac{(1-a_{j'i}-b_{j'i})^{s_{ij'}}}{s_{ij'}!}\right) G_{ij}^{(s_{ij}+1)}(b_{ji}) \prod_{j'\neq i,j}^m G_{ij'}^{(s_{ij'})}(b_{j'i}) \quad \forall j\neq i, \end{aligned} \tag{3.4}$$

$$b_i = 1 - \frac{1}{\langle q_i\rangle} G_{ii}^{(1)}(1-a_i)\left(1-r_i+r_i\prod_{j\neq i}^m G_{ij}(1-a_{ji})\right) \quad \forall i, \tag{3.5}$$

$$b_{ij} = 1 - \frac{1}{\langle q_{ij}\rangle} G_{ii}(1-a_i) G_{ij}^{(1)}(1-a_{ji}) \prod_{j'\neq i,j}^m G_{ij'}(1-a_{j'i}) \quad \forall j\neq i. \tag{3.6}$$

Proof. Fix any $i \in [m]$. A node v in module i, which is reached by following a random intraedge in module i with the other end node u, is not a-removable or b-removable in $\mathcal{G}\backslash u$ if v links to $k-1$ neighbors which are not a-removable or b-removable in $\mathcal{G}\backslash v$ and v has other neighbors only of type b_j in $\mathcal{G}\backslash v$ for $j \in [m]$. Therefore, we have

$$\begin{aligned}1 - a_i - b_i = (1-r_i) \sum_{q_i=k}^{\infty} \frac{q_i P(q_i)}{\langle q_i \rangle} \sum_{s_i=k-1}^{q_i-1} \binom{q_i-1}{s_i} (1-a_i-b_i)^{s_i} b_i^{q_i-1-s_i} \\ + r_i \sum_{q_i+\sum_{j\neq i} q_{ij}=k}^{\infty} \frac{q_i P(q_i)}{\langle q_i \rangle} \prod_{j\neq i}^{m} P(q_{ij}) \\ \cdot \sum_{s_i+\sum_{j\neq i} s_{ij}=k-1}^{q_i-1+\sum_{j\neq i} q_{ij}} \Bigg[\binom{q_i-1}{s_i} \\ \cdot (1-a_i-b_i)^{s_i} b_i^{q_i-1-s_i} \prod_{j\neq i}^{m} \binom{q_{ij}}{s_{ij}} (1-a_{ji}-b_{ji})^{s_{ij}} b_{ji}^{q_{ij}-s_{ij}} \Bigg].\end{aligned} \tag{3.7}$$

There are two terms on the right-hand side of (3.7). The first term is responsible for the case that v is a noninterconnecting node, whose neighbors are all within module i. $\frac{q_i P(q_i)}{\langle q_i \rangle}$ is the probability that v has degree q_i [25]. Apart from the edge leading to v, it has $q_i - 1$ incident edges, s_i of whom must lead to those not of type a_j or b_j in $\mathcal{G}\backslash v$. Analogously, the second term is responsible for the case of v being an interconnecting node. In this term, q_i and q_{ij} represent the numbers of its neighbors in module i and module j ($j \neq i$), respectively. In this case, we have to consider the aggregation of neighbors of v in all m modules.

Next, fix any $j \neq i$ and $j \in [m]$. We consider a node v in module i, which is reached by following a random interedge between module i and module j with the other end node u in module j. Again, v is not a-removable or b-removable in $\mathcal{G}\backslash u$ if v links to $k-1$ neighbors which are not a-removable or b-removable in $\mathcal{G}\backslash v$ and v has other neighbors only of type b_j in $\mathcal{G}\backslash v$ for $j \in [m]$. We obtain

$$\begin{aligned}1 - a_{ij} - b_{ij} = \sum_{q_i+\sum_{j'\neq i} q_{ij'}=k}^{\infty} P(q_i) \frac{q_{ij} P(q_{ij})}{\langle q_{ij} \rangle} \prod_{j'\neq i,j}^{m} P(q_{ij'}) \sum_{s_i+\sum_{j'\neq i} s_{ij'}=k-1}^{q_i-1+\sum_{j'\neq i} q_{ij'}} \Bigg[\binom{q_i}{s_i} \\ \cdot (1-a_i-b_i)^{s_i} b_i^{q_i-s_i} \binom{q_{ij}-1}{s_{ij}} (1-a_{ji}-b_{ji})^{s_{ij}} b_{ji}^{q_{ij}-1-s_{ij}} \\ \cdot \prod_{j'\neq i,j}^{m} \binom{q_{ij'}}{s_{ij'}} (1-a_{j'i}-b_{j'i})^{s_{ij'}} b_{j'i}^{q_{ij'}-s_{ij'}} \Bigg],\end{aligned} \tag{3.8}$$

where q_i and $q_{ij'}$ represent the numbers of v's neighbors in module i and module j' ($j' \neq i$), respectively, and the expression can be similarly explained as in (3.7).

We have obtained m^2 equations from (3.7) and (3.8), and the rest of equations can be found by examining b-removable nodes in the network. A node v in module i, which is reached by following a random intraedge in module i with the other end node u, is b-removable if it has a neighbor of type a_j for some $j \in [m]$ in $\mathcal{G}\backslash v$. We derive that

$$\begin{aligned}1 - b_i = (1-r_i) \sum_{q_i=1}^{\infty} \frac{q_i P(q_i)}{\langle q_i \rangle} (1-a_i)^{q_i-1} \\ + r_i \sum_{q_i+\sum_{j\neq i} q_{ij}=1}^{\infty} \frac{q_i P(q_i)}{\langle q_i \rangle} \prod_{j\neq i}^{m} P(q_{ij}) (1-a_i)^{q_i-1} \prod_{j\neq i}^{m} (1-a_{ji})^{q_{ij}},\end{aligned} \tag{3.9}$$

where the first term on the right-hand side above accounts for the case that v is a noninterconnecting node whose outgoing neighbors are all of type a_i, and the second term corresponds to the case where v is an interconnecting node having q_i-1 outgoing neighbors in module i and q_{ij} outgoing neighbors in module j for $j\neq i$. Similarly, if we follow a random interedge between module i and module j to a node v in module i, it is b-removable when it has a neighbor of type a_j in $\mathcal{G}\backslash v$ for some $j\in[m]$. For any fixed $j\neq i$, we have

$$
\begin{aligned}
1-b_{ij} = & \sum_{q_i+\sum_{j'\neq i}^m q_{ij'}=1}^{\infty} P(q_i)\frac{q_{ij}P(q_{ij})}{\langle q_{ij}\rangle} \\
& \cdot\prod_{j'\neq i,j}^m P(q_{ij'})(1-a_i)^{q_i}(1-a_{ji})^{q_{ij}-1}\prod_{j'\neq i,j}^m(1-a_{j'i})^{q_{ij'}},
\end{aligned} \tag{3.10}
$$

where q_i and $q_{ij'}$ represent the numbers of v's neighbors in module i and module j' ($j'\neq i$), respectively, and the expression can be similarly explained as in (3.9).

In (3.9), using the binomial expansion, we can write

$$
(1-a_i)^{q_i-1}=\sum_{s_i=0}^{q_i-1}\binom{q_i-1}{s_i}(1-a_i-b_i)^{s_i}b_i^{q_i-1-s_i}
$$

and

$$
\begin{aligned}
(1-a_i)^{q_i-1}\prod_{j\neq i}^m(1-a_{ji})^{q_{ij}} = & \sum_{s_i+\sum_{j\neq i}s_{ij}=0}^{q_i-1+\sum_{j\neq i}q_{ij}}\Bigg[\binom{q_i-1}{s_i}(1-a_i-b_i)^{s_i}b_i^{q_i-1-s_i} \\
& \cdot\prod_{j\neq i}^m\binom{q_{ij}}{s_{ij}}(1-a_{ji}-b_{ji})^{s_{ij}}b_{ji}^{q_{ij}-s_{ij}}\Bigg].
\end{aligned}
$$

Therefore, (3.3) readily follows from (3.7), (3.9), and some standard algebra. In a similar way, we can derive (3.4) by drawing on (3.8) and (3.10). Moreover, it is clear that (3.5) and (3.6) are direct results of (3.9) and (3.10), respectively.

For any $i\in[m]$, let n_{kc}^i be the probability that a randomly chosen node v in module i belongs to Gk-core. The node v belongs to Gk-core if it has k neighbors which are not a-removable or b-removable and has other neighbors only of type b_j in $\mathcal{G}\backslash v$ for $j\in[m]$. Hence,

$$
\begin{aligned}
n_{kc}^i = & (1-r_i)\sum_{q_i=k}^{\infty}P(q_i)\sum_{s_i=k}^{q_i}\binom{q_i}{s_i}(1-a_i-b_i)^{s_i}b_i^{q_i-s_i} \\
& +r_i\sum_{q_i+\sum_{j\neq i}q_{ij}=k}^{\infty}P(q_i)\prod_{j\neq i}^m P(q_{ij}) \\
& \cdot\sum_{s_i+\sum_{j\neq i}s_{ij}=k}^{q_i+\sum_{j\neq i}q_{ij}}\Bigg[\binom{q_i}{s_i}(1-a_i-b_i)^{s_i}b_i^{q_i-s_i}\prod_{j\neq i}^m\binom{q_{ij}}{s_{ij}}(1-a_{ji}-b_{ji})^{s_{ij}}b_{ji}^{q_{ij}-s_{ij}}\Bigg],
\end{aligned} \tag{3.11}
$$

where the two terms on the right-hand side are analogous to those in (3.7) except we here choose v_0 uniformly at random (rather than follow a random edge). The compact expression (3.2) follows directly from (3.11), the binomial expansion, and some standard algebra. Recall that n_{kc} is the expected fraction of nodes in the whole network. Hence, the first part of (3.1) follows.

Next, we consider the expected normalized number of edges l_{kc} (with respect to the number of nodes) in the network. For any $i, j \in [m]$, and $i < j$, denote by l^i_{kc} the expected normalized number of edges in the Gk-core with both end nodes within module i and similarly, l^{ij}_{kc} the expected normalized number of edges in the Gk-core with one end node in module i and the other in module j. Let E_i and E_{ij} be the expected numbers of edges in module i and those joining modules i and j, respectively. Since an edge is in the Gk-core if both of its end nodes are in the Gk-core, we have

$$l^i_{kc} = (1 - a_i - b_i)^2 \frac{E_i}{N_i} = (1 - a_i - b_i)^2 \frac{\langle q_i \rangle}{2}$$

and

$$l^{ij}_{kc} = (1 - a_{ij} - b_{ij})(1 - a_{ji} - b_{ji}) \frac{E_{ij}}{r_i N_i} = (1 - a_{ij} - b_{ij})(1 - a_{ji} - b_{ji}) \langle q_{ij} \rangle.$$

Finally, the expected number of edges in the Gk-core can be calculated as $\sum_{i=1}^m N_i l^i_{kc} + \sum_{1 \le i < j \le m} N_i r_i l^{ij}_{kc}$, which proves the second part of (3.1) and hence concludes the proof of Theorem 1. ∎

It can be seen that when $r_i = 0$ for $i \in [m]$ and $m = 1$ (namely, the network under consideration has a single degree distribution $P(q_i)$), the main result in [1] can be recovered from Theorem 1.

3.2. Rate equation method. Rate function method has been proved to be instrumental in modeling the evolution of degree distribution of the remaining network in the process of percolation [24, 1, 14]. In our network with community structure, recall that each module i has N_i nodes for $i \in [m]$. For simplicity, we assume the modules have the same size, i.e., the number of nodes in the network can be denoted by $N := mN_1$. The following method can be generalized for the general scenario of N_i.

To obtain the evolution equation for the expected number of nodes, we consider a modified k-leaf removal algorithm, where at each time step all incident edges of a randomly chosen k-leaf and those of its nearest neighbors are removed. The process continues until there is no k-leaves except isolated nodes in the remaining network. In other words, we only delete edges as compared to the original algorithm described in section 2, and the number of nodes in the network, N, is kept constant.

Set $\Delta t := N^{-1}$ as the scaled time step of the modified removal process. Fix any $i, j \in [m]$ and $j \neq i$. Denote by $N_i(q_i, t)$ the expected number of nodes in module i, which has q_i neighbors within module i at time t. Similarly, denote by $N_{ij}(q_{ij}, t)$ the expected number of interconnecting nodes in module i, which has q_{ij} neighbors in module j at time t. Recalling the degree distributions defined in section 2, we have $N_i(q_i, t) = N_i P_i(q_i, t) = \frac{N}{m} P(q_i, t)$ and $N_{ij}(q_{ij}, t) = r_i N_i P_{ij}(q_{ij}, t) = \frac{r_i N}{m} P(q_{ij}, t)$. Hence, the change of number of nodes after one iteration can be delineated by

$$N_i(q_i, t + \Delta t) - N_i(q_i, t) = \frac{1}{m\Delta t}(P(q_i, t + \Delta t) - P(q_i, t)) = \frac{1}{m} \frac{\partial P(q_i, t)}{\partial t} \tag{3.12}$$

and

$$N_{ij}(q_{ij}, t + \Delta t) - N_{ij}(q_{ij}, t) = \frac{r_i}{m\Delta t}(P(q_{ij}, t + \Delta t) - P(q_{ij}, t)) = \frac{r_i}{m} \frac{\partial P(q_{ij}, t)}{\partial t} \tag{3.13}$$

for all $i, j \in [m]$ and $i \neq j$. Let $L(t)$ be expected number of edges in the network at time t. We have

$$L(t+\Delta t) - L(t) = \frac{1}{N\Delta t}(L(t+\Delta t) - L(t)) = \frac{1}{N}\frac{\mathrm{d}L(t)}{\mathrm{d}t}. \tag{3.14}$$

The evolution equations for (3.12)–(3.14) will be given in Theorem 2 below. We can obtain the degree distributions $P_i(q_i, t)$, $P_{ij}(q_{ij}, t)$ and the expected normalized number of edges $\frac{L(t)}{N}$ at each time step t by solving these equations. Moreover, define $\hat{P}(q, t)$ be the (overall) degree distribution of a random node in our network model at time t. By using the total probability rule, we can calculate the degree distribution as

$$\hat{P}(q,t) = \sum_{i=1}^{m} \frac{1-r_i}{m} P_i(q,t) + \sum_{i=1}^{m} \frac{r_i}{m} \sum_{q_i + \sum_{j\neq i} q_{ij} = q} P_i(q_i, t) \prod_{j\neq i}^{m} P_{ij}(q_{ij}, t). \tag{3.15}$$

Suppose that the initial network has intradegree distribution $P_i(q_i, t = 0)$ and interdegree distribution $P_{ij}(q_{ij}, t = 0)$ for $i, j \in [m]$ and $i \neq j$. For a given k, the algorithm is iterated until some time $\hat{t}_k$ at which $\hat{P}(q, \hat{t}_k) = 0$ for all $1 \leq q \leq k - 1$. Once $\hat{t}_k$ is obtained, we can compute n_{kc} and l_{kc} as follows:

$$n_{kc} = \frac{\sum_{i=1}^{m} N_i n_{kc}^i}{N} = \frac{\sum_{i=1}^{m} n_{kc}^i}{m} \quad \text{and} \quad l_{kc} = \frac{L(\hat{t}_k)}{N}, \tag{3.16}$$

where $n_{kc}^i = 1 - P_i(0, \hat{t}_k)$.

In addition to the average degrees introduced in section 3.1, we further define $\langle q_{ij}^2 \rangle := \sum_{q_{ij}=0}^{\infty} q_{ij}^2 P(q_{ij})$ and a time-dependent version $\langle q_{ij}^2 \rangle_t := \sum_{q_{ij}=0}^{\infty} q_{ij}^2 P(q_{ij}, t)$. To complete this section, we present our rate equations as follows.

THEOREM 2. *Given $k \geq 2$, the rate equations for Gk-core percolation over a network $\mathcal{G}$ with m communities are given by*

$$\frac{1}{m}\frac{\partial P(q_i, t)}{\partial t} = R1_i + R2_i + R3_i + R4_i \quad \forall i \in [m], \tag{3.17}$$

where

$$R1_i = -\frac{1}{m}\left((1-r_i)\frac{\theta(k-q_i)P(q_i,t)}{\sum_{s_i=0}^{\infty}\theta(k-s_i)P(s_i,t)} \right.$$
$$\left. + r_i \frac{\theta(k-q_i) \sum_{s_i + \sum_{j\neq i} s_{ij} = q_i} P(s_i,t) \prod_{j\neq i}^{m} P(s_{ij},t)}{\sum_{\substack{s_i=0, s_{ij}=0 \\ (j\neq i)}}^{\infty} \theta(k - s_i - \sum_{j\neq i} s_{ij}) P(s_i,t) \prod_{j\neq i}^{m} P(s_{ij},t)} \right),$$

$$R2_i = \delta_{q_i,0}\left(\frac{1}{m}(1+\Phi_i(t)) + \sum_{j\neq i}^{m} \frac{r_j}{m}\Theta_{ij}(t) \right),$$

$$R3_i = -\frac{1}{m}\Phi_i(t)\left(r_i \sum_{s_i + \sum_{j\neq i} s_{ij} = q_i} \frac{s_i P(s_i,t)}{\langle q_i \rangle_t} \prod_{j\neq i}^{m} P(s_{ij},t) + (1-r_i)\frac{q_i P(q_i,t)}{\langle q_i \rangle_t} \right)$$
$$- \sum_{j\neq i}^{m} \frac{r_j}{m}\Theta_{ij}(t)\left(\sum_{s_i + \sum_{j\neq i} s_{ij} = q_i} P(s_i,t)\frac{s_{ij}P(s_{ij},t)}{\langle q_{ij} \rangle_t} \prod_{j'\neq i,j}^{m} P(s_{ij'},t) \right),$$

$$R4_i = \frac{1}{m}\Phi_i(t)\frac{\sum_{s_i=0}^{\infty} s_i(s_i-1)P(s_i,t)}{\langle q_i\rangle_t}\left(\frac{(q_i+1)P(q_i+1,t)-q_iP(q_i,t)}{\langle q_i\rangle_t}\right)$$
$$+\frac{r_i}{m}\sum_{j\neq i}^{m}\left(\Theta_{ji}(t)\frac{\sum_{s_{ji}=0}^{\infty} s_{ji}(s_{ji}-1)P(s_{ji},t)}{\langle q_{ji}\rangle_t}(P(q_i+1,t)-P(q_i,t))\right)$$
$$+\sum_{j\neq i}^{m}\frac{r_j}{m}\Theta_{ij}(t)\left(\sum_{s_i=0}^{\infty}(s_i-1)P(s_i,t)\right)\left(\frac{(q_i+1)P(q_i+1,t)-q_iP(q_i,t)}{\langle q_i\rangle_t}\right),$$

$$\Phi_i(t) := \frac{\sum_{s_i=0}^{\infty} s_i\theta(k-s_i)P(s_i,t)}{\sum_{s_i=0}^{\infty}\theta(k-s_i)P(s_i,t)},$$

$$\Theta_{ij}(t) := \frac{\sum_{s_{ji}=0}^{\infty} s_{ji}\sum_{\substack{s_j=0,s_{ji'}=0\\(i'\neq i,j)}}^{\infty}\theta(k-s_j-\sum_{i'\neq j}s_{ji'})P(s_j,t)\prod_{i'\neq j}^{m}P(s_{ji'},t)}{\sum_{\substack{s_j=0,s_{ji'}=0\\(i'\neq j)}}^{\infty}\theta(k-s_j-\sum_{i'\neq j}s_{ji'})P(s_j,t)\prod_{i'\neq j}^{m}P(s_{ji'},t)};$$

and

$$\frac{r_i}{m}\frac{\partial P(q_{ij},t)}{\partial t} = R5_{ij}+R6_{ij}+R7_{ij}+R8_{ij}\quad \forall i,j\in[m], i\neq j, \tag{3.18}$$

where

$$R5_{ij} = -\frac{r_i}{m}\frac{\theta(k-q_{ij})P(q_{ij},t)}{\sum_{s_{ij}=0}^{\infty}\theta(k-s_{ij})P(s_{ij},t)},\quad R6_{ij} = \delta_{q_{ij},0}\left(\frac{r_i}{m}(1+\Phi_i(t))+\frac{r_j}{m}\Phi_{ij}(t)\right),$$

$$R7_{ij} = -\frac{r_i}{m}\Phi_i(t)P(q_{ij},t)-\frac{r_j}{m}\Theta_{ij}(t)\frac{q_{ij}P(q_{ij},t)}{\langle q_{ij}\rangle_t},$$

$$R8_{ij} = \left[-\frac{r_i}{m}\Theta_{ji}(t)+\left(\sum_{i'\neq j}^{m}\frac{r_{i'}}{m}\Theta_{ji'}(t)+\frac{r_j}{m}\Phi_j(t)\right)\frac{\langle q_{ji}^2\rangle_t}{\langle q_{ji}\rangle_t}\right]$$
$$\cdot\left(\frac{(q_{ij}+1)P(q_{ij}+1,t)-q_{ij}P(q_{ij},t)}{\langle q_{ij}\rangle_t}\right);$$

and

$$\frac{1}{N}\frac{\mathrm{d}L(t)}{\mathrm{d}t} = R9, \tag{3.19}$$

where

$$R9 = -\frac{1}{m}\sum_{i=1}^{m}\Phi_i(t)\left(\frac{\langle q_i^2\rangle_t}{\langle q_i\rangle_t}+r_i\sum_{j\neq i}^{m}\langle q_{ij}\rangle_t\right)$$
$$-\sum_{i=1}^{m}\frac{r_i}{m}\sum_{j\neq i}^{m}\Theta_{ji}(t)\left(\frac{\langle q_j^2\rangle_t}{\langle q_j\rangle_t}+r_j\sum_{j'\neq j}^{m}\langle q_{jj'}\rangle_t\right).$$

Proof. Define the step function $\theta(q)=1$ if $q>0$ and $\theta(q)=0$ if $q\leq 0$. We prove the three equations (3.17)–(3.19), respectively.

(i) Equation (3.17): We first choose randomly a k-leaf in the network and remove all its incident edges. If the chosen node resides in module i and is a noninterconnecting node, then with probability $\frac{\theta(k-q_i)P(q_i,t)}{\sum_{s_i=0}^{\infty}\theta(k-s_i)P(s_i,t)}$ the number of nodes in module i having degree $q_i < k$ within module i decreases by 1. On the other hand, if the chosen node is in module i but is interconnecting, then with probability

$$\frac{\theta(k-q_i)\sum_{s_i+\sum_{j\neq i}s_{ij}=q_i} P(s_i,t)\prod_{j\neq i}^{m}P(s_{ij},t)}{\sum_{p_i=0}^{\infty}\theta(k-p_i)\sum_{s_i+\sum_{j\neq i}s_{ij}=p_i} P(s_i,t)\prod_{j\neq i}^{m}P(s_{ij},t)}$$

the number of nodes in module i having degree $q_i < k$ within module i decreases by 1. By the total probability formula, the $R1_i$ term represents the contribution of the chosen node to the left-hand side of (3.17) or equivalently (3.12).

When all edges incident to the chosen k-leaf and its neighbors are removed, the k-leaf and all the neighbors become isolated nodes. If the chosen node lies in module i, then on average $1+\Phi_i(t)$ nodes in module i become isolated. Here, $\Phi_i(t) = \frac{\sum_{s_i=0}^{\infty}s_i\theta(k-s_i)P(s_i,t)}{\sum_{s_i=0}^{\infty}\theta(k-s_i)P(s_i,t)}$ counts the average number of neighbors of the k-leaf node in module i. If the chosen k-leaf lies in module j ($j\neq i$) and is an interconnecting node, then the average number of nodes in module i that become isolated turns out to be

$$\frac{\sum_{p_j=0}^{\infty}s_{ji}\theta(k-p_j)\sum_{s_j+\sum_{i'\neq j}s_{ji'}=p_j} P(s_j,t)\prod_{i'\neq j}^{m}P(s_{ji'},t)}{\sum_{p_j=0}^{\infty}\theta(k-p_j)\sum_{s_j+\sum_{i'\neq j}s_{ji'}=p_j} P(s_j,t)\prod_{i'\neq j}^{m}P(s_{ji'},t)} = \Theta_{ij}(t).$$

Therefore, the term in the square brackets of $R2_i$ counts the average number of nodes that become isolated in module i. The quantity $\delta_{q,0}$ is the Kronecker delta, where $\delta_{q,0} = 1$ if $q = 0$ and vanishes otherwise. Hence, the $R2_i$ term summarizes the contribution of isolated nodes to the left-hand side of (3.17).

On the other hand, when edges incident to the nearest neighbors of the chosen k-leaf are removed, the contribution of these neighbors of degree q_i is represented by the term $R3_i$. In fact, when the chosen k-leaf node resides in module i, the number of neighbors of degree q_i within module i decreases by the average number of neighbors of the chosen node in module i, i.e., $\Phi_i(t)$, with probability

$$r_i \sum_{s_i+\sum_{j\neq i}s_{ij}=q_i} \frac{s_iP(s_i,t)}{\langle q_i\rangle_t}\prod_{j\neq i}^{m}P(s_{ij},t) + (1-r_i)\frac{q_iP(q_i,t)}{\langle q_i\rangle_t}.$$

Here, the probability is calculated by considering whether the chosen node is interconnecting (the first term) or noninterconnecting (the second term), and we recall that $\langle q_i\rangle_t := \sum_{s_i=0}^{\infty}s_iP(s_i,t)$. Next, when the chosen k-leaf node resides in module j ($j\neq i$), the number of neighbors of degree q_i within module i decreases by the average number of neighbors of the chosen node in module i (which is $\Theta_{ij}(t)$ in this case) with probability

$$\sum_{s_i+\sum_{j\neq i}s_{ij}=q_i} P(s_i,t)\frac{s_{ij}P(s_{ij},t)}{\langle q_{ij}\rangle_t}\prod_{j'\neq i,j}^{m}P(s_{ij'},t),$$

where we recall the definition $\langle q_{ij}\rangle_t := \sum_{s_{ij}=0}^{\infty}s_{ij}P(s_{ij},t)$.

The final term $R4_i$ is responsible for the contribution of the second nearest neighbors due to modification of their degrees. When all edges incident to the chosen k-leaf and its nearest neighbors are removed, the degree of each of the second nearest neighbors of the k-leaf node decreases by 1. Therefore, any second nearest neighbor will contribute 1 to the left-hand side of (3.17) if its degree in module i at time t is $q_i + 1$ while contribute -1 if its degree in module i at time t is q_i. In the first term of $R4_i$, when the chosen k-leaf is in module i, it has $\Phi_i(t)$ neighbors in module i and each of them has $\sum_{s_i=0}^{\infty} s_i(s_i - 1)P(s_i, t)/\langle q_i \rangle_t$ outgoing neighbors on average within module i. In the second term of $R4_i$, when the chosen node is an interconnecting node in module i, it has $\Theta_{ji}(t)$ neighbors in module j for $j \neq i$ and each of them leads to $\sum_{s_{ji}=0}^{\infty} s_{ji}(s_{ji} - 1)P(s_{ji}, t)/\langle q_{ji} \rangle_t$ nodes on average back to module i. In the third term of $R4_i$ we consider the situation when the chosen node is an interconnecting node in module j ($j \neq i$). In this case, it has $\Theta_{ij}(t)$ neighbors in module i on average, and each of them has $\sum_{s_i=0}^{\infty}(s_i - 1)P(s_i, t)$ outgoing neighbors on average within module i.

(ii) Equation (3.18): First, we choose randomly a k-leaf in the network and remove all its incident edges. If the chosen node resides in module i and is an interconnecting node, then with probability $\theta(k - q_{ij})P(q_{ij}, t)/\sum_{s_{ij}=0}^{\infty} \theta(k - s_{ij})P(s_{ij}, t)$ the number of interconnecting nodes in module i having $q_{ij} < k$ neighbors in module j decreases by 1. This gives the $R5_{ij}$ term, which represents the contribution of the chosen k-leaf to the left-hand side of (3.18) or equivalently (3.13).

Next, when all edges incident to the chosen k-leaf and its neighbors are removed, the k-leaf and all the neighbors become isolated nodes (hence no edge running between modules i and j). If the chosen node is a noninterconnecting node in module i, then $\Phi_i(t)$ nodes on average in module i become isolated. As edges are randomly connected, $\Phi_i(t)r_i$ of these nodes would be interconnecting. Similarly, if the chosen node is an interconnecting node in module i, on average $1 + \Phi_i(t)r_i$ nodes become isolated. Adding these two parts gives

$$\frac{1 - r_i}{m}\Phi_i(t)r_i + \frac{r_i}{m}(1 + \Phi_i(t)r_i) = \frac{r_i}{m}(1 + \Phi_i(t)),$$

which leads to the first half of $R6_{ij}$. If the chosen node is an interconnecting node in module j, the number of nodes becoming isolated is $\Theta_{ij}(t)$ on average, which leads to the second half of $R6_{ij}$.

On the other hand, when edges incident to the nearest neighbors of the chosen k-leaf are removed, the contribution of these neighbors in module i running q_{ij} edges between modules i and j is represented by the term $R7_{ij}$ (each of these neighbors will contribute -1 to the the left-hand side of (3.18)). In fact, when the chosen k-leaf node is a noninterconnecting node in module i, there are on average $\Phi_i(t)r_i$ interconnecting neighbors in module i, each of which has q_{ij} outgoing neighbors in module j with probability $P(q_{ij}, t)$. When the chosen k-leaf node is an interconnecting node in module i, there are again $\Phi_i(t)r_i$ interconnecting neighbors in module i, each of which has q_{ij} outgoing neighbors in module j with probability $P(q_{ij}, t)$. Combining these two parts yields

$$\frac{1 - r_i}{m}\Phi_i(t)r_iP(q_{ij}, t) + \frac{r_i}{m}\Phi_i(t)r_iP(q_{ij}, t) = \frac{r_i}{m}\Phi_i(t)P(q_{ij}, t).$$

This gives the first half of $R7_{ij}$. If the chosen node is an interconnecting node in module j, it has $\Theta_{ij}(t)$ interconnecting neighbors in module i, each of which has outgoing degree q_{ij} (including the edge leading to it) to module j with probability $q_{ij}P(q_{ij},t)/\langle q_{ij}\rangle_t$. This provides the second half of $R7_{ij}$.

The last term $R8_{ij}$ is responsible for the contribution of the second nearest neighbors due to modification of their degrees. When all edges incident to the chosen k-leaf and its nearest neighbors are removed, the degree of each of the second nearest neighbors of the k-leaf node decreases by 1. Hence, any second nearest neighbor will contribute 1 to the left-hand side of (3.18) if it has $q_{ij}+1$ neighbors in module j at time t while contribute -1 if it has q_{ij} neighbors in module j at time t. When the chosen k-leaf is an interconnecting node in module i, it has $\Theta_{ji}(t)$ neighbors in module j and each of them has $\sum_{s_{ji}=0}^{\infty} s_{ji}(s_{ji}-1)P(s_{ji},t)/\langle q_{ji}\rangle_t$ outgoing neighbors on average back to module i. This yields the following contribution to the left-hand side of (3.18):

$$\frac{r_i}{m}\Theta_{ji}(t)\frac{\sum_{s_{ji}=0}^{\infty} s_{ji}(s_{ji}-1)P(s_{ji},t)}{\langle q_{ji}\rangle_t}\left(\frac{(q_{ij}+1)P(q_{ij}+1,t)-q_{ij}P(q_{ij},t)}{\langle q_{ij}\rangle_t}\right). \tag{3.20}$$

When the chosen k-leaf is an interconnecting node in module i' for $i'\neq i,j$, it has $\Theta_{ji'}(t)$ neighbors in module j and each of them has $\langle q_{ji}^2\rangle_t/\langle q_{ji}\rangle_t$ outgoing neighbors on average to module i. This yields the contribution:

$$\sum_{i'\neq i,j}^{m}\frac{r_{i'}}{m}\Theta_{ji'}(t)\frac{\langle q_{ji}^2\rangle_t}{\langle q_{ji}\rangle_t}\left(\frac{(q_{ij}+1)P(q_{ij}+1,t)-q_{ij}P(q_{ij},t)}{\langle q_{ij}\rangle_t}\right). \tag{3.21}$$

When the chosen k-leaf is in module j, it has $\Phi_j(t)r_j$ interconnecting neighbors in module j on average. Each of them has $\langle q_{ji}^2\rangle_t/\langle q_{ji}\rangle_t$ outgoing neighbors on average to module i. Therefore, this gives the following contribution:

$$\frac{1}{m}\Phi_j(t)r_j\frac{\langle q_{ji}^2\rangle_t}{\langle q_{ji}\rangle_t}\left(\frac{(q_{ij}+1)P(q_{ij}+1,t)-q_{ij}P(q_{ij},t)}{\langle q_{ij}\rangle_t}\right). \tag{3.22}$$

Adding (3.20), (3.21), and (3.22) up, we derive the last contribution $R8_{ij}$.

(iii) Equation (3.19): We consider the Gk-core percolation process. At each step, the average number of edges removed is equal to the product of the average number of nearest neighbors of the k-leaf and the average degree of a neighbor. When the chosen k-leaf node is a noninterconnecting node in some module $i\in[m]$, the average number of neighbors is given by $\Phi_i(t)$. If a neighbor is noninterconnecting, its mean degree is $\langle q_i^2\rangle_t/\langle q_i\rangle_t$; if it is interconnecting, the mean degree becomes $\langle q_i^2\rangle_t/\langle q_i\rangle_t+\sum_{j\neq i}^{m}\langle q_{ij}\rangle_t$. Hence, this yields the following contribution to the left-hand side of (3.19) or equivalently (3.14):

$$-\sum_{i=1}^{m}\frac{1-r_i}{m}\Phi_i(t)\left[(1-r_i)\frac{\langle q_i^2\rangle_t}{\langle q_i\rangle_t}+r_i\left(\frac{\langle q_i^2\rangle_t}{\langle q_i\rangle_t}+\sum_{j\neq i}^{m}\langle q_{ij}\rangle_t\right)\right]. \tag{3.23}$$

When the chosen k-leaf node is an interconnecting node in some module $i\in[m]$, the average number of its neighbors within module i is again given by $\Phi_i(t)$. By analyzing a neighbor in module i according to whether it is noninterconnecting or interconnecting as above, we obtain a similar contribution as in (3.23). Moreover, the average number of the chosen k-leaf's neighbors in module j ($j\neq i$) is given by $\Theta_{ji}(t)$. If such a neighbor is noninterconnecting, its mean degree is $\langle q_j^2\rangle_t/\langle q_j\rangle_t$; if it is interconnecting, the mean degree is $\langle q_j^2\rangle_t/\langle q_j\rangle_t+\sum_{j'\neq j}^{m}\langle q_{jj'}\rangle_t$. Therefore, the above discussion yields the following contribution to the left-hand side of (3.19):

$$- \sum_{i=1}^{m} \frac{r_i}{m} \left[\Phi_i(t) \left((1-r_i) \frac{\langle q_i^2 \rangle_t}{\langle q_i \rangle_t} + r_i \left(\frac{\langle q_i^2 \rangle_t}{\langle q_i \rangle_t} + \sum_{j \neq i}^{m} \langle q_{ij} \rangle_t \right) \right) \right.$$

$$\left. + \sum_{j \neq i}^{m} \Theta_{ji}(t) \left((1-r_j) \frac{\langle q_j^2 \rangle_t}{\langle q_j \rangle_t} + r_j \left(\frac{\langle q_j^2 \rangle_t}{\langle q_j \rangle_t} + \sum_{j' \neq j}^{m} \langle q_{jj'} \rangle_t \right) \right) \right]. \tag{3.24}$$

Adding (3.23) and (3.24) up, we obtain $R9$. This completes the proof of Theorem 2.□

4. Numerical applications in networks with Erdős–Rényi modules. To verify the theoretical results and explore the influence of communities on the Gk-core percolation process, we consider a modular network of $N = 10^8$ nodes and $m = 2$ modules. We set $N_1 = N_2 = \frac{N}{2}$ and both modules are modeled by Erdős–Rényi (ER) random networks following degree distribution $P_1(q) = P_2(q) := P(q) = e^{-\lambda} \frac{\lambda^q}{q!}$ for $q \geq 0$. Hence, the average degree of each module is given by $\langle q \rangle = \lambda$. Set $r_1 = r_2 = r$ and edges joining the interconnecting nodes in the two modules are characterized by a random matching. In other words, $P_{12}(q) = P_{21}(q) = \delta_{1,0}$.

We show the expected relative fractions of nodes and edges in Gk-core with $k = 2$ in Figure 3 and with $k = 3$ in Figure 4. In particular, we use the generating function formalism, i.e., Theorem 1, for calculating analytical $G2$-core and apply the rate equation method developed in section 3.2 for calculating analytical $G3$-core. As the Poisson degree distribution decays rapidly, we set the maximum degree as $q_{\max} = 40$ and only solve the set of first $41 \times 3 = 123$ equations in (3.17) and (3.18). In both cases we observe good agreement between numerical and simulation results.

Several interesting observations can be drawn as follows. First, $G2$-core emerges continuously in terms of both n_{kc} and l_{kc} for any fraction r of interconnecting nodes; see Figure 3. In contrast, $G3$-core displays a first-order percolation transition for both n_{kc} and l_{kc} and any fraction r; see Figure 4. We confirm this discontinuous phase transition for Gk-core with $k \geq 3$ by performing extensive calculations. In other words, the type of phase transition in Gk-core percolation is determined by k and does not change with the intensity of interconnections between modules. Note

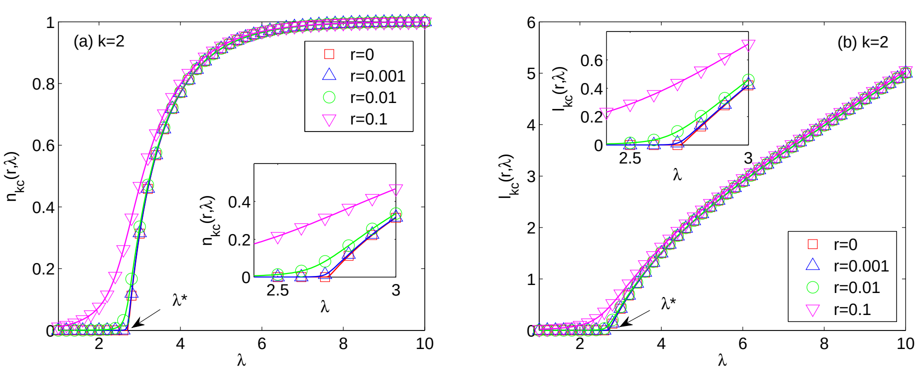

FIG. 3. (a) *Expected relative fraction of nodes in Gk-core, n_{kc}, as a function of λ for modular networks with two ER modules having size $N = 10^8$ and $k = 2$. Corresponding expected relative fraction of edges in Gk-core, l_{kc}, as a function of λ is shown in* (b). *Insets show the magnified views around the percolation threshold λ^*. Analytical results (solid lines) are based on the generating function formalism in Theorem* 1, *and simulations (red squares for $r = 0$, blue upper triangles for $r = 0.001$, green circles for $r = 0.01$, and magenta lower triangles for $r = 0.1$) are averaged over* 300 *realizations.*

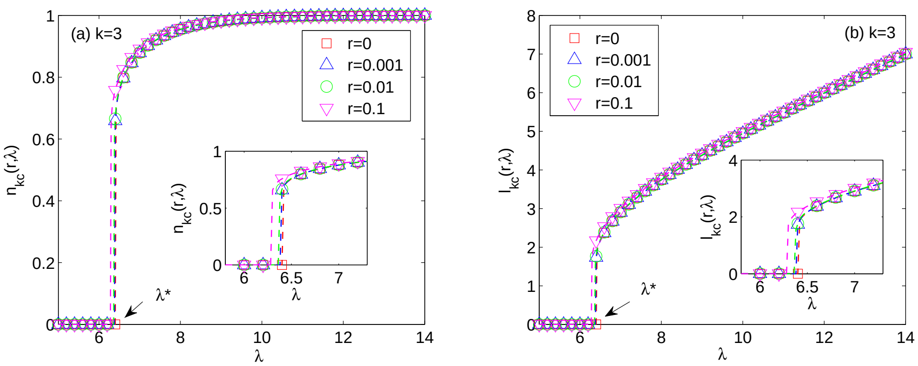

FIG. 4. (a) *Expected relative fraction of nodes in Gk-core, n_{kc}, as a function of λ for modular networks with two ER modules having size $N = 10^8$ and $k = 3$. Corresponding expected relative fraction of edges in Gk-core, l_{kc}, as a function of λ is shown in* (b). *Insets show the magnified views around the percolation threshold λ^*. Analytical results (dashed lines) are based on the rat equation method in section* 3.2*, and simulations (red squares for $r = 0$, blue upper triangles for $r = 0.001$, green circles for $r = 0.01$, and magenta lower triangles for $r = 0.1$) are averaged over* 300 *realizations.*

that for $r = 0$, n_{kc} and l_{kc} for our example are equivalent to those for a single ER network by Theorem 1. Our observation is in line with the known results found in [1, 31] in the case of $r = 0$. When $r = 0$, the percolation threshold $\lambda^*(k = 2)$ is estimated as 2.709 (c.f. Figure 3) and $\lambda^*(k = 3)$ is estimated as 6.415 (c.f. Figure 4). Second, we find that $n_{kc}(r, \lambda) > n_{kc}(0, \lambda) = 0$ and $l_{kc}(r, \lambda) > l_{kc}(0, \lambda) = 0$ for $r > 0$, showing a nontrivial Gk-core at the percolation threshold λ^*. The small fraction of interconnecting nodes boosts n_{kc} and l_{kc} prominently only around the critical point λ^*. This contrasts with the recent observation in [31], where increase of average degree through random addition, hub-targeted addition, and localized addition invariably yields uniform growth of n_{kc} and l_{kc} in ER networks with any λ. This phenomenon reveals that the network with community structure is essentially different from the no-community case. The interconnecting nodes play a unique role in enhancing the network robustness in terms of Gk-core.

Next, we further examine the field-type physical effect of the fraction r of interconnecting nodes on the percolation transitions by investigating the scaling relationship between $n_{kc}(r, \lambda)$, r, and λ, possibly serving as order parameter (magnetization), external field and temperature, respectively [26]. The effect of a magnetic field in a spin system can be described by three critical exponents β, δ, and γ near the criticality following the universal Widom's identity $\beta(\delta - 1) = \gamma$ [4]. In our Gk-core percolation processes, the exponent β characterizes the behavior of order parameter n_{kc} with zero magnitude of the field near the criticality λ^* and is given by

$$n_{kc}(0, \lambda) - n_{kc}(0, \lambda^*) \sim (\lambda - \lambda^*)^{\beta}.$$

The exponent δ relates the order parameter with the magnitude of the magnetic field, r, at the critical point following

$$n_{kc}(r, \lambda^*) \sim r^{\frac{1}{\delta}}.$$

The exponent γ describes the susceptibility of the spin system in the critical region as

$$\Big(\frac{\partial n_{kc}(r, \lambda)}{\partial r}\Big)_{r \to 0} \sim |\lambda - \lambda^*|^{-\gamma}.$$

Figure 5 presents our results for $G2$-core (or simply core) percolation over modular networks with two ER modules as above. We obtain $\beta = 2.6$, $\delta = 1.292$, and $\gamma = 0.759$ based on numerical calculations using the generating function formalism. These values are consistent with the Widom identity, indicating that the interconnecting nodes act analogously as an external field on the $G2$-core percolation. Hence, the system of three exponents β, δ and γ has two degrees of freedom in $G2$-core percolation. This phenomenon echoes a recent discovery for giant component based metric in modular network resilience [8, 10] for continuous phase transitions.

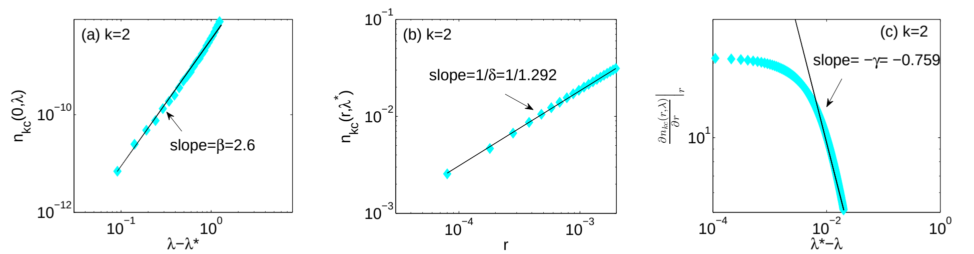

FIG. 5. *Estimation of critical components for $G2$-core of the modular networks with two ER modules.* (a) $n_{kc}(0,\lambda)$ *as a function of* $\lambda - \lambda^*$. (b) $n_{kc}(r,\lambda^*)$ *as a function of* r. (c) $\frac{\partial n_{kc}(r,\lambda)}{\partial r}\Big|_r$ *as a function of* $\lambda^* - \lambda$ *with* $r = 0.0001$. *Numerical results (cyan diamonds) are plotted based on Theorem* 1 *with added lines as an aid to eye. All parameters are the same as in Figure* 3.

As a comparison, we show in Figure 6 the estimates of $\beta = 0.07$, $\delta = 87$, and $\gamma = 0.08$ for $G3$-core percolation, which apparently do not satisfy the Widom identity. Very recently, it is discovered remarkably in [13] that an interdependent ER network with community structure undergoing first-order percolation transitions fulfills the Widom identity with nontrivial exponents ($\delta = 2$, $\gamma = \beta = 0.5$). Therefore, our observation from Figure 6 suggests that the violation is presumably not a result from abrupt transitions but uncovers a radical physical discrepancy between $G2$-core percolation and Gk-core percolation for $k \geq 3$ over networks with community structure.

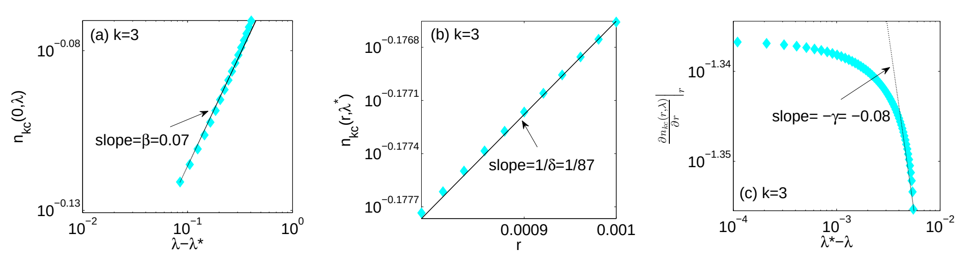

FIG. 6. *Estimation of critical components for $G3$-core of the modular networks with two ER modules.* (a) $n_{kc}(0,\lambda)$ *as a function of* $\lambda - \lambda^*$. (b) $n_{kc}(r,\lambda^*)$ *as a function of* r. (c) $\frac{\partial n_{kc}(r,\lambda)}{\partial r}\Big|_r$ *as a function of* $\lambda^* - \lambda$ *with* $r = 0.0001$. *Numerical results (cyan diamonds) are plotted based on Theorem* 1 *with added lines as an aid to eye. All parameters are the same as in Figure* 4.

5. Concluding remarks. In this paper, we have studied a complex network model with community structure mediated by a small fraction of interconnecting nodes. We have developed two complementary theoretical frameworks, namely, the generating function formalism and the rate equation method, to tackle the Gk-core percolation over the random networks with community structure. Different from traditional giant component based metrics, the Gk-core structure is ideal for reflecting network resilience affected by spreading virus or epidemics when weak nodes (k-leaves) and their nearest neighbors are knocked out. It is found numerically and by simulations that $G2$-core displays a continuous phase transition while Gk-core for $k \geq 3$ undergoes a discontinuous percolation transition for any fraction of interconnecting nodes. The influence of interconnecting nodes is salient around the critical value. Moreover, our numerical results reveal that the effect of interconnections on $G2$-core percolation behaves like a magnetic field in a ferromagnetic-paramagnetic spin system, following Widom's identity. Interestingly, this relationship in general does not hold for $G3$-core experiencing discontinuous percolation transitions. However, given our simulations are performed on a finite system, it might also be likely that the critical exponents observed in Figure 6 are numerical approximates to the true values $\beta = 0$, $\delta = \infty$, and $\gamma = 0$, which are consistent with Widom's identity $\gamma = \beta(\delta - 1)$. It is hoped that the methodology developed in this paper could be applied to deepen our understanding of structure characteristics of modular networks and facilitate the design of resilient networked systems featuring community structures.

Acknowledgement. The author is very grateful to the reviewers for their detailed and constructive comments that have helped to improve the quality of the paper.

REFERENCES

[1] N. Azimi-Tafreshi, S. Osat, and S. N. Dorogovtsev, *Generalization of core percolation on complex networks*, Phys. Rev. E, 99 (2019), p. 022312.
[2] M. Bauer and O. Golinelli, *Core percolation in random graphs: A critical phenomena analysis*, Eur. Phys. J. B, 24 (2001), pp. 339–352.
[3] G. J. Baxter, S. N. Dorogovtsev, K.-E. Lee, J. F. F. Mendes, and A. V. Goltsev, *Critical dynamics of the k-core pruning process*, Phys. Rev. X, 5 (2015), 031017.
[4] A. Bunde and S. Havlin, *Fractals and Disordered Systems*, Springer, New York, 2012.
[5] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function*, Cambridge University Press, Cambridge, 2010.
[6] A. Coja-Oghlan, O. Cooley, M. Kang, and K. Skubch, *Core forging and local limit theorems for the k-core of random graphs*, J. Combin. Theory, Ser. B, 137 (2019), pp. 178–231.
[7] N. Dehmamy, S. Milanlouei, and A.-L. Barabási, *A structural transition in physical networks*, Nature, 563 (2018), pp. 676–680.
[8] G. Dong, J. Fan, L. M. Shekhtman, S. Shai, R. Du, L. Tian, X. Chen, H. E. Stanley, and S. Havlin, *Resilience of networks with community structure behaves as if under an external field*, Proc. Natl. Acad. Sci. USA, 115 (2018), pp. 6911–6915.
[9] E. Estrada and P. A. Knight, *A First Course in Network Theory*, Oxford University Press, Oxford, 2015.
[10] J. Fan, G. Dong, L. M. Shekhtman, D. Zhou, J. Meng, X. Chen, and S. Havlin, *Structural resilience of spatial networks with inter-links behaving as an external field*, New J. Phys., 20 (2018), 093003.
[11] S. Fortunato and D. Hric, *Community detection in networks: A user guide*, Phys. Rep., 659 (2016), pp. 1–44.
[12] M. Girvan and M. E. J. Newman, *Community structure in social and biological networks*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 7821–7826.
[13] B. Gross, H. Sanhedrai, L. Shekhtman, and S. Havlin, *Interconnections between networks acting like an external field in a first-order percolation transition*, Phys. Rev. E, 101 (2020), 022316.
[14] A. K. Hartmann and M. Weigt, *Phase Transitions in Combinatorial Optimization Problems*, Wiley-VCH, Weinheim, 2005.
[15] D. Helbing, D. Brockmann, T. Chadefaux, K. Donnay, U. Blanke, O. Woolley-Meza, M. Moussaid, A. Johansson, J. Krause, S. Schuttle, and M. Perc, *Saving human*

lives: What complexity science and information systems can contribute, J. Stat. Phys., 158 (2015), pp. 735–781.

[16] P. Hell and J. Nešetřil, *The core of a graph*, Discrete Math., 109 (1992), pp. 117–126.

[17] M. Jalili and M. Perc, *Information cascades in complex networks*, J. Complex Netw., 5 (2017), pp. 665–693.

[18] S. Janson and M. J. Luczak, *Asymptotic normality of the k-core in random graphs*, Ann. Appl. Probab., 18 (2008), pp. 1085–1137.

[19] T. Jia and M. Pósfai, *Connecting core percolation and controllability of complex networks*, Sci. Rep., 4 (2014), 5379.

[20] Y.-Y. Liu, E. Csóka, H. Zhou, and M. Pósfai, *Core percolation on complex networks*, Phys. Rev. Lett., 109 (2012), 205703.

[21] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, *Controllability of complex networks*, Nature, 473 (2011), pp. 167–173.

[22] F. Morone, G. D. Ferraro, and H. A. Makse, *The k-core as a predictor of structural collapse in mutualistic ecosystems*, Nat. Phys., 15 (2019), pp. 95–102.

[23] M. E. J. Newman, *The structure and function of complex networks*, SIAM Rev., 45 (2003), pp. 167–256.

[24] M. E. J. Newman, *Networks*, 2nd ed., Oxford University Press, Oxford, 2018.

[25] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Random graphs with arbitrary degree distributions and their applications*, Phys. Rev. E, 64 (2001), 026118.

[26] H. Nishimori and G. Ortiz, *Elements of Phase Transitions and Critical Phenomena*, Oxford University Press, New York, 2011.

[27] L. Peel, D. B. Larremore, and A. Clauset, *The ground truth about metadata and community detection in networks*, Sci. Adv., 3 (2017), e1602548.

[28] P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha, *Core-periphery structure in networks*, SIAM J. Appl. Math., 74 (2014), pp. 167–190.

[29] P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha, *Core-periphery structure in networks (revisited)*, SIAM Rev., 59 (2017), pp. 619–646.

[30] S. Shai, D. Y. Kenett, Y. N. Kenett, M. Faust, S. Dobson, and S. Havlin, *Critical tipping point distinguishing two types of transitions in modular network structures*, Phys. Rev. E, 92 (2015), 062805.

[31] Y. Shang, *Attack robustness and stability of generalized k-cores*, New J. Phys., 21 (2019), 093013.

[32] Y. Shang, *Multi-hop generalized core percolation on complex networks*, Adv. Complex Syst., 23 (2020), 2050001.

[33] L. M. Shekhtman, M. M. Danziger, I. Bonamassa, S. V. Buldyrev, G. Caldarelli, V. Zlatić, and S. Havlin, *Critical field-exponents for secure message-passing in modular networks*, New J. Phys., 20 (2018), 053001.

[34] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, *Comparing community structure to characteristics in online collegiate social networks*, SIAM Rev., 53 (2011), pp. 526–543.

[35] F. Tudisco, P. Mercado, and M. Hein, *Community detection in networks via nonlinear modularity eigenvectors*, SIAM J. Appl. Math., 78 (2018), pp. 2393–2419.

[36] C. Witt, *Analysis of an iterated local search algorithm for vertex cover in sparse random graphs*, Theor. Comput. Sci., 425 (2012), pp. 117–125.

[37] J. Zhang, L. Fu, S. Li, Y. Yao, and X. Wang, *Core percolation in interdependent networks*, IEEE Trans. Netw. Sci. Engin., (2018), https://doi.org/10.1109/TNSE.2018.2884774.

[38] J.-H. Zhao and H.-J. Zhou, *Controllability and maximum matchings of complex networks*, Phys. Rev. E, 99 (2019), 012317.