Bridges in complex networks

Ang-Kun Wu,[1,2,3] Liang Tian,[1,4] and Yang-Yu Liu[1,5,*]

[1]*Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA*

[2]*Department of Physics, Chu Kochen Honors College, Zhejiang University, Hangzhou, Zhejiang 310027, China*

[3]*Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA*

[4]*College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*

[5]*Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA*

(Received 12 December 2016; revised manuscript received 22 April 2017; published 17 January 2018)

A bridge in a graph is an edge whose removal disconnects the graph and increases the number of connected components. We calculate the fraction of bridges in a wide range of real-world networks and their randomized counterparts. We find that real networks typically have more bridges than their completely randomized counterparts, but they have a fraction of bridges that is very similar to their degree-preserving randomizations. We define an edge centrality measure, called bridgeness, to quantify the importance of a bridge in damaging a network. We find that certain real networks have a very large average and variance of bridgeness compared to their degree-preserving randomizations and other real networks. Finally, we offer an analytical framework to calculate the bridge fraction and the average and variance of bridgeness for uncorrelated random networks with arbitrary degree distributions.

DOI: 10.1103/PhysRevE.97.012307

I. INTRODUCTION

A *bridge*, also known as a *cut-edge*, is an edge of a graph whose removal disconnects the graph, i.e., it increases the number of connected components (see Fig. 1, red or light gray edges) [1]. A dual concept is an *articulation point* or a *cut-vertex*, defined as a node in a graph whose removal disconnects the graph [2,3]. Both bridges and articulation points in a graph can be identified via a linear-time algorithm based on depth-first search [4], and they represent natural vulnerabilities of networked systems. Analysis of articulation points has recently provided us with a different angle to systematically investigate the structure and function of many real-world networks [5]. This prompts us to ask if a similar analysis can be applied to bridges.

Note that a bridge is similar to, but different from, the notion of a *red bond* introduced in percolation theory to characterize substructures of percolation clusters on lattices [6]. To define a red bond, we consider the percolation cluster as a network of wires carrying electrical current, and we impose a voltage drop between two nodes in the network. Then red bonds are those links that carry all current, whose removal stops the current. The definition of bridges does not require us to impose a voltage drop on the network. Instead, it just concerns the connectivity of the whole network.

Despite the fact that bridges play important roles in ensuring network connectivity, the notion of a bridge has never been systematically studied in complex networks. What is the typical number of bridges in a random graph with a prescribed degree distribution? Are the bridges in a real network overrepresented or underrepresented? How do we quantify network vulnerability in terms of a bridge attack? In this paper, we systematically address those questions in both real networks and random graphs.

II. EMPIRICAL RESULTS OF REAL NETWORKS

We first calculate the fraction of bridges ($f_b := L_b/L$) in a wide range of real-world networks, from infrastructure networks to food webs, neuronal networks, protein-protein interaction (PPI) networks, gene regulatory networks, and social graphs. Here L_b and L denote the number of bridges and total number of links in a network, respectively. We find that many real networks have a very small fraction of bridges, while a few of them (e.g., PPI networks) have a very large fraction of bridges [Fig. 2(a)]. To identify the topological characteristics that determine f_b in real networks, we compare f_b of a given network with that of its randomized counterpart. We first randomize each real network using a complete randomization procedure that turns the network into an Erdös-Rényi (ER) type of random graph with the number of nodes N and links L unchanged [11]. We find that most of the completely randomized networks possess very different f_b, compared to their corresponding real networks [Fig. 2(a)]. This indicates that complete randomization eliminates the topological characteristics that determine f_b. Moreover, we find that real networks typically display much higher f_b than their completely randomized counterparts [Fig. 2(a)]. By contrast, when we apply a degree-preserving randomization, which rewires the edges among nodes while keeping the degree k of each node unchanged, this procedure does not alter f_b significantly [Fig. 2(b)]. In other words, the characteristics of a real network in terms of f_b is largely encoded in its degree distribution $P(k)$.

To quantify the importance of an edge in damaging a network, we define an edge centrality measure B, called *bridgeness*, for each edge in a graph as the number of nodes

*Author to whom all correspondence should be addressed: yyl@channing.harvard.edu

 ©2018 American Physical Society

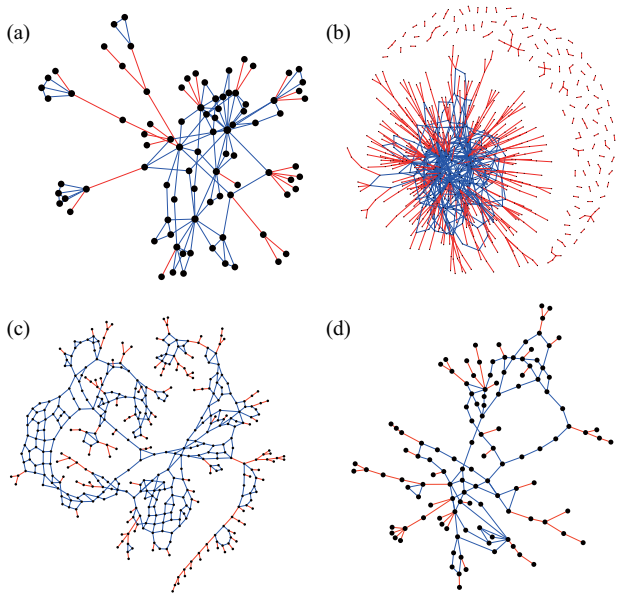

FIG. 1. Bridges in real-world networks. Bridges are edges (in red or light gray) whose removal will increase the number of connected components in a graph. (a) The *Grassland* food web [7]; (b) the protein-protein interaction network of *C. elegans* [8]; (c) a subgraph of the road network of California [9]; (d) a subgraph of the power grid in three western states of the United States [10].

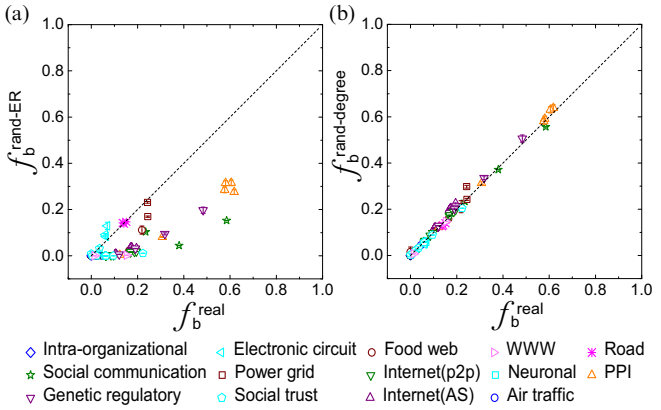

FIG. 2. Bridge fractions in real-world networks and their randomized counterparts. The dashed lines are $y = x$. The error bars represent the standard deviation, calculated from 100 randomizations. (a) Complete randomization of real networks. (b) Degree-preserving randomization.

disconnected from the giant connected component (GCC) [12] after the removal of this edge. Note that for finite networks, we can define bridgeness similarly as the number of nodes disconnected from the largest connected component (LCC). The algorithm to identify bridges and calculate bridgeness for an arbitrary finite network is detailed in Appendix A. By definition, if an edge is not a bridge or outside the GCC (or LCC), it has zero bridgeness. We notice that bridgeness has been defined differently in the literature. But none of the previous definitions of bridgeness is based on the notion of a bridge. Some of them are actually defined on nodes [13–16], rather than edges. Here we define bridgeness based on the notion of a bridge, and we focus on the damage to the GCC (or LCC), which is typically the main functional part of a network.

We emphasize that the bridgeness centrality defined here is fundamentally different from the edge betweenness centrality [17]. For an edge e in a graph, its betweenness centrality is defined as follows: $C_B(e) = \sum_{i \neq j} \sigma_{ij}(e)/\sigma_{ij}$, where σ_{ij} is the number of shortest paths between any two nodes i and j in the graph, and $\sigma_{ij}(e)$ is the number of shortest paths between i and j that pass through edge e. By definition, each edge in a graph can be associated with a nonzero betweenness centrality. And an edge with very high betweenness centrality can be considered as a "bottleneck" of the graph, whose removal may block the communications between many pairs of nodes with the shortest paths between them passing through the bottleneck. Yet, such a bottleneck edge, e.g., (5,7) in Fig. 3(a), might not be a bridge. Similarly, the fact that an edge is a bridge does not mean its betweenness centrality is always very high. For example, in Fig. 3(b), node 1 is a leaf (with degree 1), and it is connected to the rest of the network through a bridge, i.e., edge (1,3) shown in red. But the betweenness of this bridge (1,3) is much lower than that of a nonbridge edge (4,5) for large n.

Bridgeness differentiates edges based on their structural importance. Consider all bridges that have nontrivial bridgeness, i.e., $B > 0$. Denote their average and variance as $\langle B \rangle$ and var(B), respectively. We calculate $\langle B \rangle$ and var(B) for the same set of real-world networks analyzed in Fig. 2, and we compare the values with those of their degree-preserving randomizations (see Fig. 4). We find that road networks and certain domains (stanford.edu and nd.edu) of the World Wide Web (WWW) have much larger var(B) than their randomized counterparts and other real networks [Fig. 3(b)]. Structural analysis shows that those networks have many large bicon-

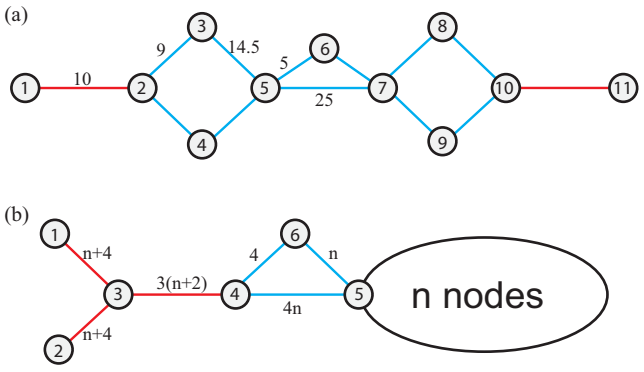

FIG. 3. Bridgeness and edge betweenness. Bridges are shown in red (or light gray), nonbridge edges are shown in blue (or black). The values next to the edges represents their edge betweenness. Only bridges have nonzero bridgeness. In those two examples, all the bridges shown in red have bridgeness 1 except for the one (3,4) with bridgeness 3. (a) Edges with very high betweenness centrality are not necessarily bridges. For example, the central edge (5,7) has the largest betweenness centrality 25 among all the edges, but it is not a bridge. (b) A bridge does not necessarily have very high betweenness centrality. Here, the bridge (1,3) has betweenness centrality $n + 4$, which is much lower than the betweenness centrality $4n$ of a nonbridge edge (4,5), when n is large. Note that here node-5 is part of the n nodes.

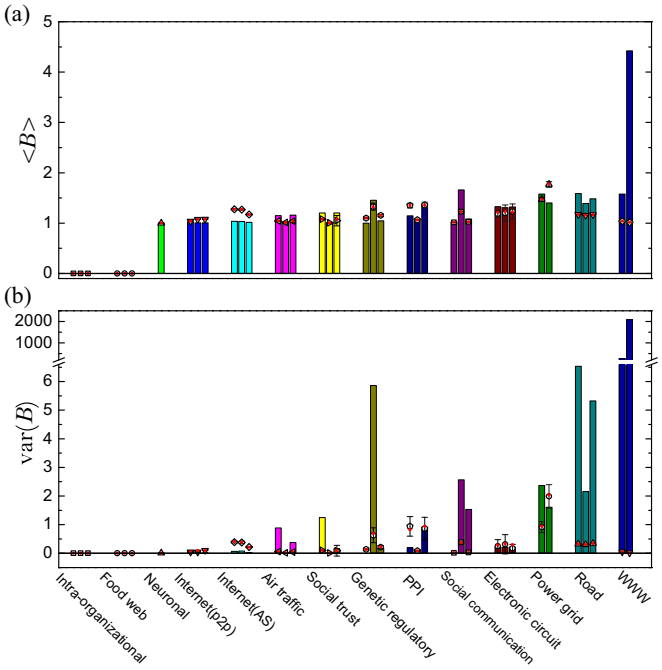

FIG. 4. Average and variance of bridgeness in real-world networks and their randomized counterparts. The bars represent the values of real networks, empty symbols represent the average values of their degree-preserving randomizations, and solid red points are theoretical results derived from the analytical framework presented in Sec. III F, using the degree distribution of the real networks as the only inputs. The error bars represent the standard deviation, calculated from 10 randomizations. (a) Average bridgeness. (b) Variance of bridgeness.

nected components (BCCs) (see Table I). Here, a BCC is a connected subgraph where for any two nodes there are at least two paths connecting them that have no nodes in common other than these two nodes [19]. Note that, by definition, no bridge exists in a BCC. But bridges can connect BCCs in the GCC. In those two domains of the WWW, certain bridges connect very large BCCs in the GCC, yielding very large bridgeness and var(B).

Since the bridge fractions in real networks are almost the same as their degree-preserving randomized counterparts, the difference of average bridgeness between real networks and their degree-preserving randomizations indicates variations of vulnerability of those networks in terms of bridge attack.

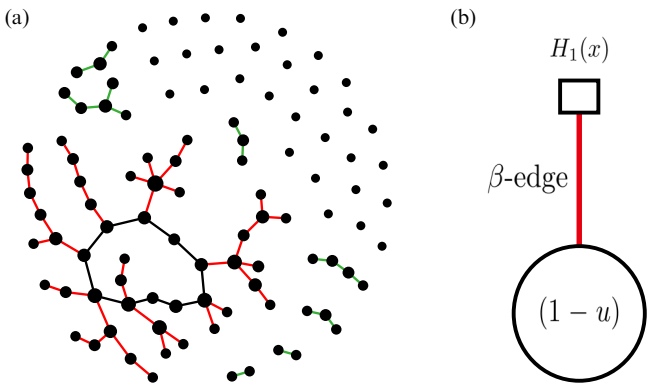

FIG. 5. Demonstration of different types of edges. (a) The green (or light gray), red (or dark gray), and black lines represent α-edges, β-edges, and γ-edges, respectively. (b) Neighborhood of a β-edge. The red edge is a β-edge and the black square and ellipse represent an FCC and the GCC, respectively, after the removal of the β-edge.

Figure 4(a) shows that certain types of networks, such as road networks and WWW, are more vulnerable, displaying much larger $\langle B\rangle$ than their randomizations. By contrast, the Internet at the autonomous system (AS) level and the Internet peer-to-peer (p2p) file sharing networks have smaller $\langle B\rangle$ than their randomized counterparts, indicating that those networks are robust from the bridge attack perspective.

III. ANALYTICAL FRAMEWORK

The results of real-world networks, especially the interesting results on f_b in real networks and degree-preserving randomizations [Fig. 2(b)], prompt us to analytically decipher bridge structure for large uncorrelated random networks with prescribed degree distributions. To begin with, we adopt the local tree approximation (LTA), which assumes the absence of finite loops in the thermodynamic limit (i.e., as the network size $N \to \infty$) and allows only infinite loops [5,20–23]. The LTA leads to three important properties: (i) all finite connected components (FCCs) are trees, and hence all edges inside them are bridges; (ii) there exists only one GCC [22], only one BCC (which has no bridges), and the BCC is a subgraph of the GCC; (iii) subgraphs inside the GCC but outside the BCC are trees, and all edges in those subgraphs are bridges [5,21–25]. Those properties enable us to categorize all the edges in a graph into three types [Fig. 5(a)]: (i) α-edge: edges in FCCs, which

TABLE I. Detailed information of the road networks and the World Wide Webs analyzed in this paper. Here, N and L represent the number of nodes and edges in the networks, respectively. s_{GCC} is the relative size of the GCC. N_{BCC} and L_b are the numbers of BCCs and bridges, respectively. $\langle s_{BCC}\rangle$ is the average size of all BCCs in the network except the largest one. B_{max} represents the largest bridgeness.

Category	name	N	L	s_{GCC}	N_{BCC}	$\langle s_{BCC}\rangle$	L_b	$\langle B\rangle$	var(B)	B_{max}	Description
Road [9]	RoadNet-CA	1965206	2766607	1957027	4042	5.94	376517	1.598	6.53	162	California road network
networks	RoadNet-PA	1088092	1541898	1087562	1815	4.67	216994	1.39	2.16	94	Pennsylvania road network
	RoadNet-TX	1379917	1921660	1351137	3054	13.35	290333	1.48	5.32	209	Texas road network
World	stanford.edu [9]	281903	1992636	255265	1073	31.77	27344	4.42	2081.1	4907	WWW from the stanford
Wide											.edu domain
Web	nd.edu [18]	325729	1090108	325729	308	45.77	166376	1.58	264.29	2660	WWW from the nd.edu domain

are bridges; (ii) β-edge: edges inside the GCC but outside the BCC, which are also bridges; (iii) γ-edge: edges inside the BCC, which are not bridges. Hereafter, we also use α, β, or γ to denote the probability that a randomly chosen edge is a α-edge, β-edge, or γ-edge, respectively. By definitions, we have $\alpha+\beta+\gamma=1$ and $f_{\rm b}=\alpha+\beta$. Note that according to our definition of bridgeness, only β-edges have nontrivial bridgenesses, i.e., $B>0$.

Generating functions

$$G_0(x)=\sum_{k=0}^{\infty}P(k)x^k \tag{1}$$

and

$$G_1(x)=\sum_{k=1}^{\infty}Q(k)x^{k-1} \tag{2}$$

are very useful in calculating key quantities of random graphs, such as the mean component size and the size of the GCC [22]. Here $Q(k)=kP(k)/c$, and $c=\sum_{k=0}^{\infty}kP(k)$ is the mean degree. To calculate α, β, and γ, we introduce the generating function $H_1(x)$ for the size distribution of the components that are reachable by choosing a random edge and following one of its ends. [The notation $H_0(x)$ is reserved for the generating function of the size distribution of the components in which a randomly chosen node is located [22].]

According to the LTA, $H_1(x)$ satisfies the following self-consistency equation [22]:

$$H_1(x)=\sum_{k=1}^{\infty}xQ(k)[H_1(x)]^{k-1}. \tag{3}$$

Note that we only include the finite components in calculating $H_1(x)$, which means that the chosen edge must be a bridge due to the LTA, namely either α- or β-edge. Equation (3) implies that following a bridge, the excess edges of its end to finite subcomponents should also be bridges. We can rewrite Eq. (3) using the generation function of $Q(k)$, i.e., $G_1(x)$, yielding

$$H_1(x)=xG_1(H_1(x)). \tag{4}$$

Define $u:=H_1(1)$, representing the probability that following a randomly chosen edge to one of its end nodes, the node belongs to an FCC after removing this edge. Then the probability that a randomly chosen edge is an α-edge or belongs to an FCC is simply

$$\alpha=u^2. \tag{5}$$

For a β-edge, one of its end nodes belongs to an FCC and the other one belongs to the GCC after removing this edge. Hence we have

$$\beta=2u(1-u). \tag{6}$$

For a γ-edge, both of its end nodes belong to the GCC after its removal, and hence

$$\gamma=(1-u)^2. \tag{7}$$

Note that the normalization condition $\alpha+\beta+\gamma=1$ is naturally satisfied. The fraction of bridges is simply given by

$$f_{\rm b}=\alpha+\beta=1-(1-u)^2. \tag{8}$$

Besides $f_{\rm b}$, we can also calculate the bridgeness distribution $P(B)$ from $H_1(x)$. For nontrivial bridgeness ($B>0$) we only consider the bridges in the GCC. In other words, we calculate $P(B)$ for β-edges in random graphs. Define the generating function of $P(B)$ as

$$F(x)=\sum_{B=1}^{\infty}P(B)x^B, \tag{9}$$

which leads to $P(B)=\frac{1}{B!}\frac{d^BF(x)}{dx^B}|_{x=0}$. Since one end node of a β-edge is located in the GCC after the removal of this edge [Fig. 5(b)], we have

$$F(x)=\frac{2(1-u)H_1(x)}{\beta}, \tag{10}$$

where the numerator represents the generating function for the bridgeness distribution of a randomly chosen β-edge, and the denominator originates from the fact that we focus on β-edges. The moments of $P(B)$ are then given by

$$\langle B^k\rangle=\sum_{B=1}^{\infty}B^kP(B)=\left[\left(x\frac{d}{dx}\right)^kF(x)\right]_{x=1}. \tag{11}$$

For example, the average bridgeness is given by

$$\langle B\rangle=F'(1)=\frac{H_1'(1)}{u}=\frac{H_1'(1)}{H_1(1)} \tag{12}$$

and

$$\langle B^2\rangle=F'(1)+F''(1)=\frac{H_1'(1)+H_1''(1)}{u}. \tag{13}$$

Consequently, the variance of bridgeness is

$$\mathrm{var}(B)=\frac{H_1'(1)[u-H_1'(1)]+uH_1''(1)}{u^2}. \tag{14}$$

In Fig. 6(a), we show the bridge fraction $f_{\rm b}$ calculated from Eqs. (4) and (8), the relative size of BCC ($s_{\rm BCC}$) [19], and the relative size of GCC ($s_{\rm GCC}$) [22] as functions of mean degree c in ER random graphs with Poisson degree distribution $P(k)=e^{-c}c^k/k!$ [11]. (See Appendix B for the generating function formalism in calculating $s_{\rm GCC}$ and $s_{\rm BCC}$.) We find that before the GCC and BCC emerge at the percolation threshold $c^*=1$, all components are FCCs and all edges are α-edges, rendering $f_{\rm b}=1$. After the emergence of the GCC and BCC at $c^*=1$, f_b begins to deviate from 1, and the fraction of β-edges displays a nonmonotonic behavior (because the difference between $s_{\rm GCC}$ and $s_{\rm BCC}$ increases first and then decreases). We also calculate $f_{\rm b}$ for scale-free (SF) networks with power-law degree distribution $P(k)\sim k^{-\lambda}$ generated by the static model [26–28]. For SF networks, the smaller the degree exponent λ, the smaller is the percolation threshold c^* [23], causing $f_{\rm b}$ to deviate from 1 at smaller c^* [Fig. 6(b)].

We also calculate the average bridgeness $\langle B\rangle$ and the variance of bridgeness $\mathrm{var}(B)(:=\langle B^2\rangle-\langle B\rangle^2)$ in ER and SF random networks [Figs. 6(c) and 6(d)]. We find that for both ER and SF networks, $\langle B\rangle$ and $\mathrm{var}(B)$ decrease monotonically as c increases. Note that $\langle B\rangle$ and $\mathrm{var}(B)$ of SF networks are typically lower than those of ER networks for small c, and

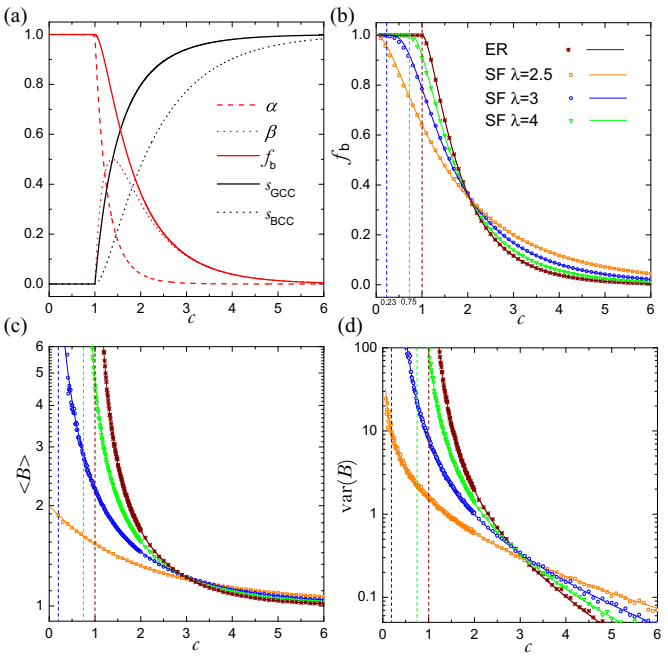

FIG. 6. Bridge and bridgeness in ER and SF random networks. The size of all these random network is 10^6. Curves in (a)–(d) are analytical predication about bridges, and symbols are simulation results. (a) The behavior of the bridge fraction f_b, relative sizes of the BCC and GCC (denoted as s_{BCC} and s_{GCC}, respectively) in ER random graphs. (b) Bridge fraction in different random networks with dashed vertical lines representing the corresponding percolation threshold c^* where the GCC and BCC emerge. Note that $c^*_{SF}(\lambda = 2.5) = 0$, $c^*_{SF}(\lambda = 3.0) \approx 0.23$, $c^*_{SF}(\lambda = 4.0) \approx 0.75$, and $c^*_{ER} = 1.0$. (c) Average bridgeness in random networks. (d) The variance of bridgeness in random networks. Detailed calculation and more distributions can be found below.

higher for large c. This is because SF networks tend to first form densely connected components of hub nodes when c is small, and then gradually stretch out as c increases. The divergent behavior of bridgeness around the percolation threshold c^* is due to the emergence of the GCC, which initially is treelike and therefore contains bridges with a huge range of bridgeness.

For the sake of completeness, in the following subsections we demonstrate the analytical calculation of f_b, $\langle B\rangle$, var(B), as well as s_{GCC} and s_{BCC} for uncorrelated random graphs with various canonical degree distributions. For certain random graphs with simple degree distributions and generating functions, e.g., the ER random graphs, those quantities have simple expressions.

A. Poisson-distributed graphs

The degree distribution $P(k)$ for Erdös-Rényi random graphs follows Poisson distribution [11]:

$$P(k) = \frac{e^{-c}c^k}{k!},$$

where c is the mean degree. Then the generating functions of $P(k)$ and $Q(k)$ are

$$G_0(x,c) = G_1(x,c) = e^{c(x-1)}$$

with derivatives

$$G_1'(x) = ce^{c(x-1)},$$
$$G_1''(x) = c^2e^{c(x-1)}.$$

With

$$u = H_1(1) = G_1(H_1(1)) = e^{c(u-1)}$$

and

$$H_1'(1) = \frac{G_1(u)}{1 - G_1'(u)} = \frac{u}{1 - cu},$$

we have

$$f_b = [1 - (1 - u)^2]$$

and

$$\langle B\rangle = \frac{1}{1 - cu}.$$

Substituting the above equations into Eqs. (12)–(14), we can get

$$\text{var}(B) = \frac{cu}{(1 - cu)^3}.$$

Moreover, according to Eqs. (B2)–(B4), we have

$$s_{GCC} = 1 - e^{c(u-1)}$$

and

$$s_{BCC} = 1 - e^{c(u-1)} - u(1 - u)c.$$

Results are shown in Fig. 6.

B. Exponentially distributed graphs

The degree distribution for exponentially distributed graphs is [22,29]

$$P(k) = (1 - e^{-1/\kappa})e^{-k/\kappa},$$

and the mean degree is

$$c = \frac{e^{-1/\kappa}}{1 - e^{-1/\kappa}}.$$

The generating functions of $P(k)$ and $Q(k)$ are

$$G_0(x) = \frac{1 - e^{-1/\kappa}}{1 - xe^{-1/\kappa}},$$
$$G_1(x) = \left[\frac{1 - e^{-1/\kappa}}{1 - xe^{-1/\kappa}}\right]^2,$$

respectively. Their derivatives are

$$G_1'(x) = \frac{2e^{-1/\kappa}(1 - e^{-1/\kappa})^2}{(1 - xe^{-1/\kappa})^3},$$
$$G_1''(x) = \frac{6e^{-2/\kappa}(1 - e^{-1/\kappa})^2}{(1 - xe^{-1/\kappa})^4}.$$

Substituting the above equations into Eqs. (12)–(14) and (B2)–(B4), we can get f_b, s_{GCC}, s_{BCC}, $\langle B\rangle$, and var(B).

Unfortunately, their expressions are too complicated to show. The results of these quantities can be found in Fig. 7.

C. Purely power-law distributed graphs

The degree distribution for purely power-law distributed graphs is [22,29]

$$P(k) = \frac{k^{-\lambda}}{\zeta(\lambda)} \quad \text{for} \quad k \geqslant 1,$$

where $\zeta(\lambda) = \sum_{k=1}^{\infty} k^{-\lambda}$ is the Riemann zeta function. Note that $P(k)$ can be normalized only for $\lambda \geqslant 2$.

The generating functions of $P(k)$ and $Q(k)$ are

$$G_0(x) = \frac{\mathrm{Li}_\lambda(x)}{\zeta(\lambda)},$$
$$G_1(x) = \frac{\mathrm{Li}_{\lambda-1}(x)}{x\zeta(\lambda-1)},$$

respectively. Here, $\mathrm{Li}_n(x) = \sum_{k=1}^{\infty} x^k/k^n$ is the nth polylogarithm of x, whose derivative is $\frac{d\mathrm{Li}_n(x)}{dx} = \frac{\mathrm{Li}_{\lambda-1}(x)}{x}$. The derivatives of the generating functions are

$$G_1'(x) = \frac{\mathrm{Li}_{\lambda-2}(x) - \mathrm{Li}_{\lambda-1}(x)}{x^2\zeta(\lambda-1)},$$
$$G_1''(x) = \frac{\mathrm{Li}_{\lambda-3}(x) - 3\mathrm{Li}_{\lambda-2}(x) + 2\mathrm{Li}_{\lambda-1}(x)}{x^3\zeta(\lambda-1)}.$$

Substituting the above equations into Eqs. (12)–(14) and (B2)–(B4), we can get f_b, s_{GCC}, s_{BCC}, $\langle B\rangle$, and var(B). The results of these quantities can be found in Fig. 7.

D. Power-law distribution with exponential cutoff

The degree distribution for a purely power-law distribution with exponent λ and exponential cutoff is [22,29]

$$P(k) = \frac{k^{-\lambda}e^{-k/\kappa}}{\mathrm{Li}_\lambda(e^{-1/\kappa})} \quad \text{for} \quad k \geqslant 1.$$

This distribution can be normalized for any λ.

The generating functions of $P(k)$ and $Q(k)$ are

$$G_0(x) = \frac{\mathrm{Li}_\lambda(xe^{-1/\kappa})}{\mathrm{Li}_\lambda(e^{-1/\kappa})},$$
$$G_1(x) = \frac{\mathrm{Li}_{\lambda-1}(xe^{-1/\kappa})}{x\mathrm{Li}_{\lambda-1}(e^{-1/\kappa})},$$

respectively, and their derivatives are

$$G_1'(x) = \frac{\mathrm{Li}_{\lambda-2}(xe^{-1/\kappa}) - \mathrm{Li}_{\lambda-1}(xe^{-1/\kappa})}{x^2\mathrm{Li}_{\lambda-1}(e^{-1/\kappa})},$$
$$G_1''(x) = \frac{\mathrm{Li}_{\lambda-3}(xe^{-1/\kappa}) - 3\mathrm{Li}_{\lambda-2}(xe^{-1/\kappa}) + 2\mathrm{Li}_{\lambda-1}(xe^{-1/\kappa})}{x^3\mathrm{Li}_{\lambda-1}(e^{-1/\kappa})}.$$

Substituting the above equations into Eqs. (12)–(14) and (B2)–(B4), we can get f_b, s_{GCC}, s_{BCC}, $\langle B\rangle$, and var(B). The results of these quantities can be found in Fig. 7.

E. Static model

In the main text, we use the static model to generate scale-free (SF) random graphs [26]. This model consists of the following steps [29]: (i) Given N isolated nodes, we label them from 1 to N. For each node i, we assign a weight $p_i \propto i^{-a}$, where $a = \frac{1}{\lambda-1}$, and λ is the characteristic parameter of the SF graphs. (ii) Then we randomly choose two nodes according to their weight and connect them if they are not connected. Self-links and multilinks are forbidden here. We repeat this step until $M = cN/2$ links are added.

The degree distribution of the static mode can be analytically derived as [27,28]

$$P(k) = \frac{[c(1-a)/2]^{1/a}}{a}\frac{\Gamma(k-1/a, c(1-a)/2)}{\Gamma(k+1)},$$

with $\Gamma(s)$ the gamma function and $\Gamma(s.x)$ the upper incomplete gamma function. When $k \to \infty$, $P(k) \sim k^{-(1+1/a)} = k^{-\lambda}$. Therefore, we can build different SF random graphs by tuning a. The generating functions are

$$G_0(x) = \frac{1}{a}E_{1+\frac{1}{a}}[c(1-a)(1-x)],$$
$$G_1(x) = \frac{1-a}{a}E_{\frac{1}{a}}[c(1-a)(1-x)],$$

where $E_n = \int_1^\infty e^{-xy}y^{-n}dy$ is the exponential integral. Note that the derivative of E_n follows $E_n' = -E_{n-1}(x)$. From the generating functions, we can derive f_b, s_{GCC}, s_{BCC}, $\langle B\rangle$, and var(B). Results are shown in Fig. 6.

F. Degree distributions of real networks

Our analytical framework can also be applied to real-world networks with degree distributions as the only input. For these networks, we count the number of k-degree nodes, n_k, and the degree distribution is $P(k) = n_k/N$. The generating functions of $P(k)$ and $Q(k)$ are finite polynomials,

$$G_0(x) = \frac{\sum_{k=0}^{k_{\max}} n_k x^k}{N}, \tag{15}$$
$$G_1(x) = \frac{\sum_{k=0}^{k_{\max}} (k+1)n_{k+1}x^k}{N}, \tag{16}$$

respectively. Substituting the above equations and their derivatives into Eqs. (12)–(14), we can get f_b, $\langle B\rangle$, and var(B).

In this case, the theoretical results are considered as the ensemble average of all the networks with the prescribed degree distribution in the thermodynamic limit. As expected, we find that the theoretical results (red points) agree well with the numerical results (empty symbols) calculated from degree-preserving randomized counterparts.

As shown in Fig. 2(b), we find that the fraction of bridges is largely determined by the degree distribution. However, higher-order correlation of real networks, such as assortativity, clustering, and modularity, may affect the structure of

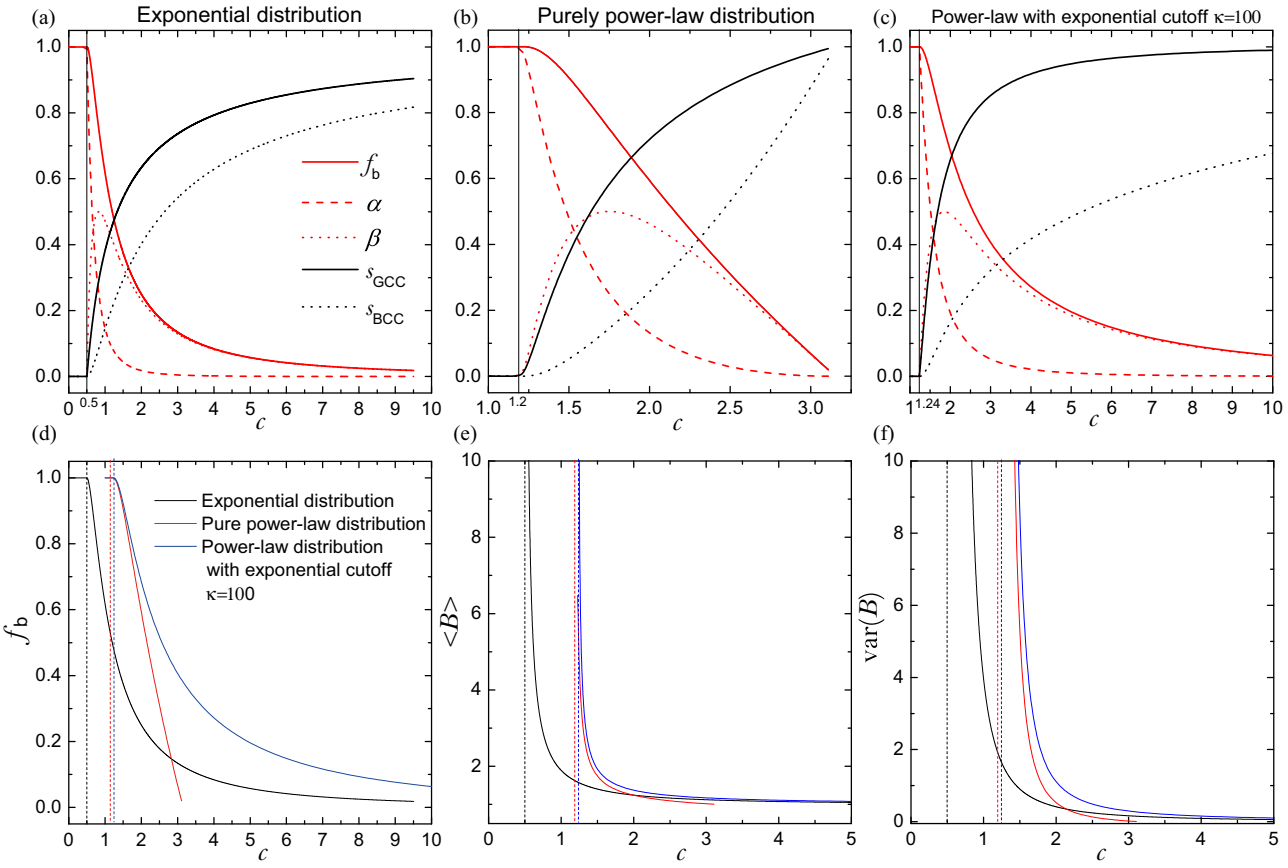

FIG. 7. The analytical results of the bridge fraction f_b, s_{GCC}, s_{BCC}, $\langle B\rangle$, and var(B) in different random graphs. All the results are analytical, and the dashed lines mark their percolation positions. The corresponding degree distribution is denoted in the figures.

the bridges (bridgeness). Therefore, our theoretical approach serves as a starting or reference point to study bridge structure in real-world networks.

IV. NUMERICAL STUDY ON NETWORKS WITH MESOSCOPIC STRUCTURE

The above analytical framework does not apply to networks with certain mesoscopic structure, e.g., the presence of communities [30] or core-periphery structure [31]. To explore the impact of mesoscopic structure on the bridge fraction and bridgeness distribution, here we numerically investigate those quantities using the stochastic block model (SBM) [32].

Consider a partition $\boldsymbol{g}$ of total N vertices into K groups, where $\boldsymbol{g}$ is a vector of N elements and g_i represents the group to which vertex i belongs. We can define a symmetric $K \times K$ matrix ψ, whose element $\psi_{g_i g_j}$ is the probability of an edge existing between vertices i and j.

To study the effect of mesoscopic structure on bridges in random graphs, we mix the above SBM with an ER random graph of the same mean degree. The elements of the probability matrix ψ can then be written as

$$\lambda\psi_{g_i g_j} + (1-\lambda)p,$$

where p is the connection probability for the ER random graph depending on mean degree c, and λ is a tuning parameter ($0 \leqslant \lambda \leqslant 1$) [33].

We study two types of mesoscopic structure here. The first is community structure with

$$\psi = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix},$$

where three groups (communities) exist and each group has the same number of vertices and connecting probability ($\psi_{g_i g_i} = a$), and a is a variable depending on mean degree c. There is no edge between groups ($\psi_{g_i g_j} = 0$). The second is the core-periphery structure [31] with

$$\psi = \begin{pmatrix} b & 0.1b \\ 0.1b & 0 \end{pmatrix},$$

where two groups exist and each group has the same number of vertices, but one group has zero inner connection. b is a variable depending on mean degree c.

Figure 8 shows how bridge characteristics [f_b,$\langle B\rangle$,var(B)] change from pure ER random networks to networks with the above two types of mesoscopic structures, under three different mean degrees. In the first case [Figs. 8(a)–8(c)], namely networks with three communities, bridge characteristics do not change significantly at different mean degrees. Indeed, it is very unlikely that bridges will form to connect two communities. The transition from a single ER network into three smaller ER networks has little influence on the bridge characteristics [see Figs. 9(a)–9(c)].

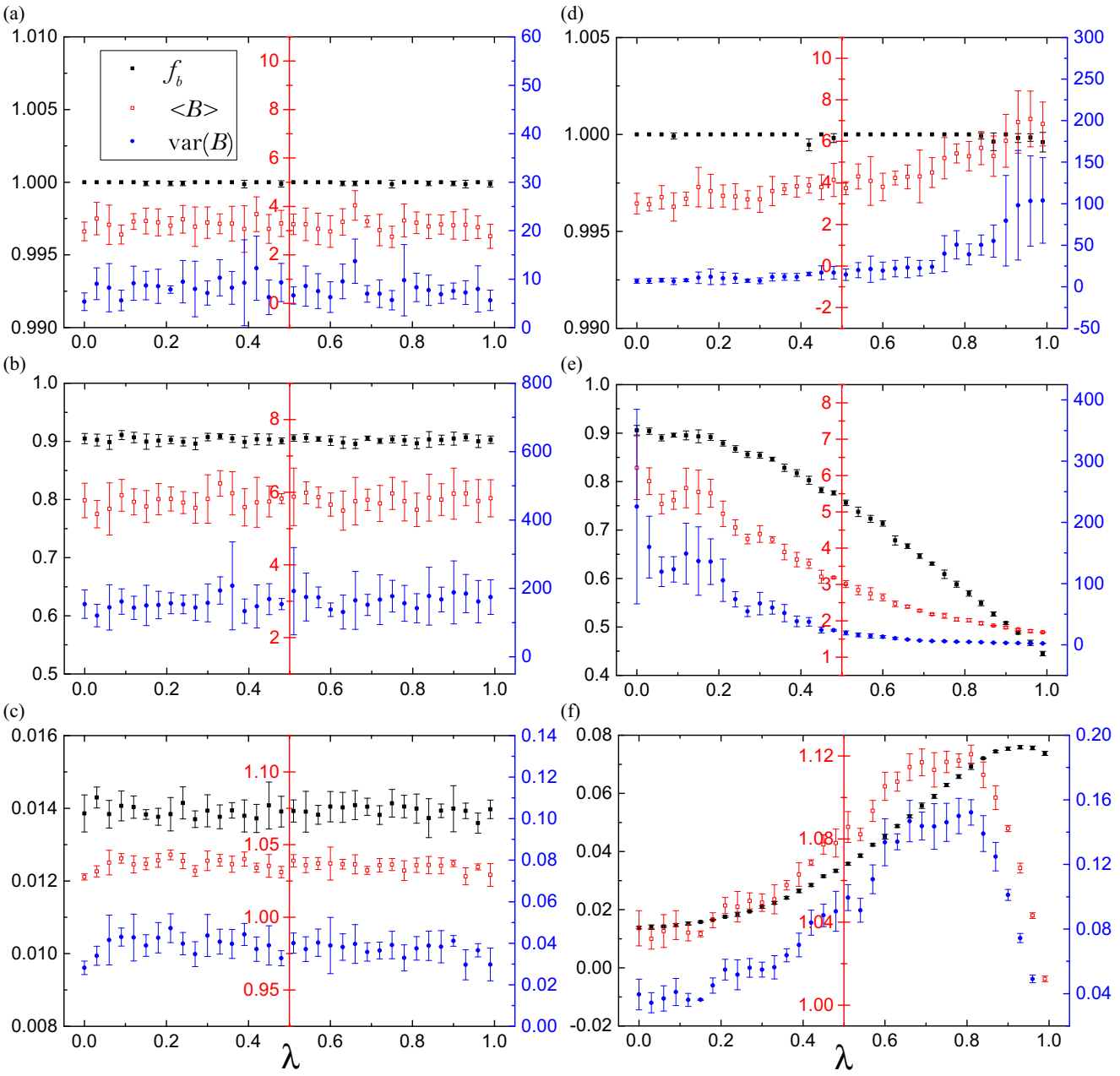

FIG. 8. Bridge structure in networks with a mesoscopic structure. The x axis is λ, the extent to which networks are constructed close to a specific mesoscopic structure from ER networks. The black (or solid square), red (or empty square) and blue (or solid circle) symbols represent the bridge fraction, average, and variance of bridgeness, respectively, in different networks. All networks have 30 000 nodes. The error bar represents standard deviation from five repeated simulations. (a)–(c) Networks with three equal-sized communities. For (a), (b), and (c), they have mean degree 0.5, 1.2, and 5.0, respectively. (d)–(f) Networks with core-periphery structure. For (d), (e), and (f), they have mean degree 0.5, 1.2, and 5, respectively.

Interestingly, the presence of core-periphery structure has a big impact on bridges [Figs. 8(d)–8(f)]. When mean degree $c = 0.5$ (almost all edges in ER networks are bridges), the transition from ER networks to core-periphery networks increases, $\langle B \rangle$ and var(B), while f_b decreases only slightly in the end, which can be viewed as the process of turning edges (or bridges) in the periphery into edges in the core, and thus the size of trees increases. Finally, some BCCs are created. When $c = 1.2$, the GCC and the BCC just emerge and the transition process turns more bridges into BCCs, leading to lower f_b, $\langle B \rangle$, and var(B). However, when $c = 5.0$, since there are few bridges in the very beginning, the process disassembles the BCC in the periphery and creates more bridges and larger trees (larger bridgeness). When this process continues, bridges between peripheral vertices are also transformed into the core, leading to the final decrease. We can also see the transition of the adjacency matrix in Figs. 9(d)–9(f). When λ is large [$\lambda = 0.9$, Fig. 9(e)], most of the bridges are between the core and the periphery (the off-diagonal parts). Larger λ [$\lambda = 1.0$, Fig. 9(f)] leads to fewer bridges in the off-diagonal parts. Note that in

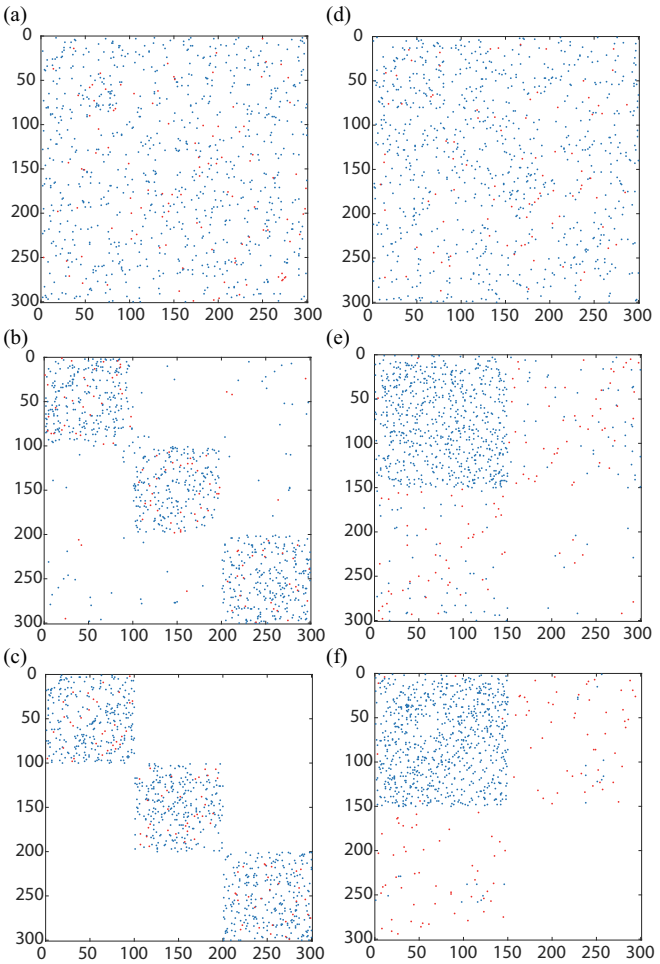

FIG. 9. Adjacency matrices of networks with mesoscopic structure. Each point represents an edge here, while the red (or the light gray) are bridges and the blue (or the dark gray) are not. All networks have 300 nodes and mean degree $c = 3.0$. (a)–(c) Networks with three equivalent communities. From (a), (b), and (c), they have $\lambda = 0.1$, 0.9, and 1.0, respectively. (d)–(f) Networks with a core-periphery structure. From (d), (e), and (f), they have $\lambda = 0.5$, 1.2, and 5.0, respectively.

such a core-periphery structure, not all edges in the periphery part are bridges. Actually, the definition or detection of the periphery is much more rigorous and may be different in various real networks [34], and in this case identifying bridges can also be a new way to define the periphery of a network.

V. CONCLUSION

In conclusion, we systematically investigate the bridge structure in complex networks. We demonstrate bridges in real-world networks, calculate the fraction of bridges in different networks, and define an edge centrality measure, called bridgeness, to quantify the importance of bridges in damaging a network. Finally, we analytically calculate bridge structure in random graphs with prescribed degree distributions. The presented results help us understand the complex architecture of real-world networks, and they may shed light on the design of more robust networks against edge attack.

ACKNOWLEDGMENTS

We thank W. Chen for valuable discussions. This work is supported by the John Templeton Foundation (Award No. 51977), National Natural Science Foundation of China (Grants No. 11505095 and No. J1210046), Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20133218120033), and the Fundamental Research Funds for the Central Universities of China (Grants No. NS2014072 and No. NZ2015110).

APPENDIX A: BRIDGE IDENTIFICATION AND BRIDGENESS CALCULATION

The algorithm for identifying bridges in a network is based on depth-first search (DFS), which has linear time complexity [4]. Randomly choosing a node from the network, we start DFS and track two indices for each node i: its DFS visited time stamp (DFS[i]) and the lowest DFS reachable ancestor (low[i]). DFS[i] is defined as the number of other visited nodes until the current one in DFS, and low[i] represents the lowest DFS[j] of a previously visited node j that can be reached again by current node i in the later DFS. Note that, for two successively visited nodes i and j in the DFS, the index low[i] is updated by min(low[i], low[j]) after j is visited.

Note that low[i] marks the node's topological position in the network. For two nodes i and j in the same biconnected component (BCC), low[i] = low[j]. For nodes in a tree structure, low[i] = DFS[i], which is different for each node. A bridge between two nearest-neighboring nodes (i and j) is identified whenever the later visited node, say node i, has larger low[i] than that of the previously visited node j.

To calculate the size (b) of the subgraph that will be cut from the network due to the removal of a bridge, we can simply use the current time step (T), i.e., the number of visited nodes, to subtract the DFS visited time stamp of the end of the bridge (which is inside the BCC), and add 1. For instance, in Fig. 10(c), $b = T - \mathrm{DFS}[3] + 1 = 6 - 3 + 1 = 4$. Naturally each bridge has two components to be cut from the network, and we define bridgeness to be the smaller size of the two components. Thus, to calculate the bridgeness, we need to go through the LCC again, and B is calculated as $\min\{b, S - b\}$, where S represents the size of the whole connected component; see Fig. 10(e).

To summarize, we first conduct DFS in each connected component of a graph to identify bridges with one of the separating parts (b) after their removal and get the size of each connected component. Then we go through the LCC again to get the bridgeness (B) of each bridge in the LCC.

APPENDIX B: CALCULATION OF s_{FCC}, s_{GCC}, AND s_{BCC}

The generating function formalism allows us to easily calculate the relative size of FCC, GCC, and BCC. We let s_{FCC} be the fraction of vertices in the graph that do not belong to the giant component. Hence we have [22]

$$s_{\mathrm{FCC}} = H_0(1) = G_0(u). \tag{B1}$$

Then the relative size of the GCC is given by [22]

$$s_{\mathrm{GCC}} = 1 - s_{\mathrm{FCC}} = 1 - G_0(u). \tag{B2}$$

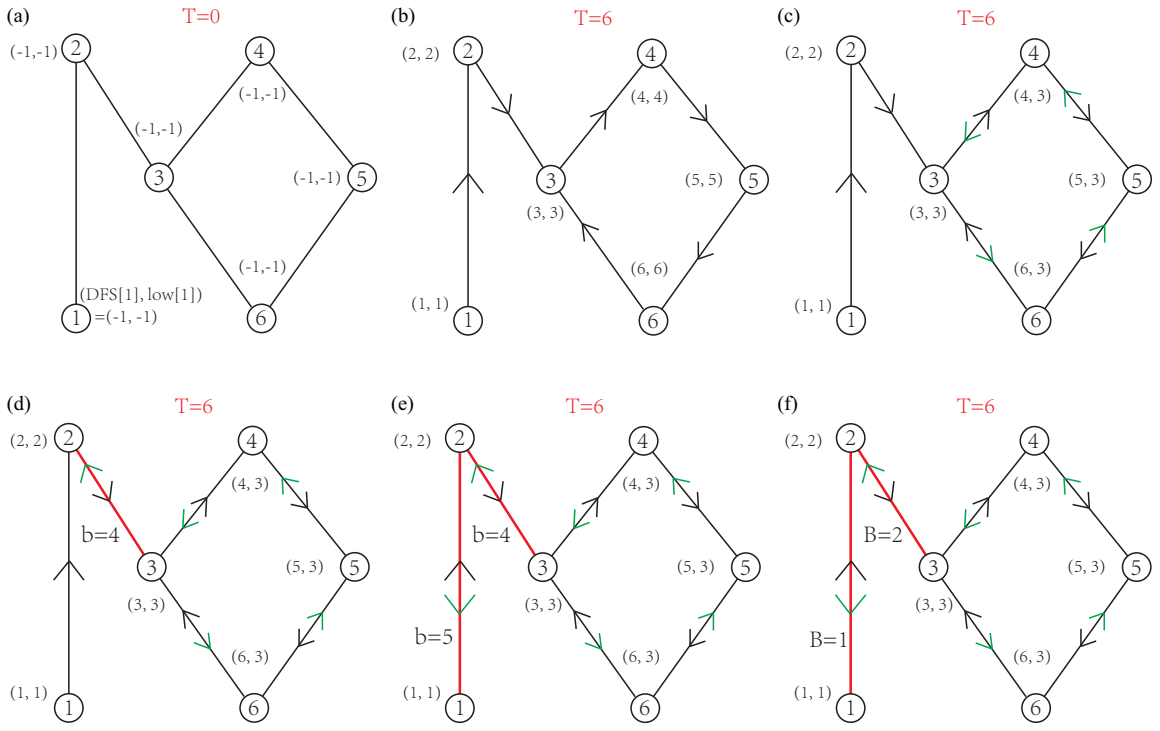

FIG. 10. Using depth-first search to identify bridges and calculate bridgeness. The labels in nodes and arrows on edges represent the sequence of DFS, red (or thick) edges are bridges, and for each node its two indices are presented as a coordinate (DFS[i], low[i]). (a) Initial state of indices. (b) The moment when DFS just finishes visiting all the nodes in the connected component. The size S of this connected component is given by the current time step T. (c)–(e) Updates of indices when the search goes back. (f) Checking through all nodes in the component, and let bridgeness be the size of the smaller part separated from that component.

As for the relative size of the BCC, we consider that if a node is outside the BCC, then its surroundings should have at most one edge along with a node that is not u [19], hence we have

$$s_{\mathrm{BCC}} = 1 - \sum_{k=1}^{\infty} P(k)u^k - \sum_{k=1}^{\infty} kP(k)(1-u)u^{k-1} = 1 - G_0(u) - (1-u)G_0'(u). \tag{B3}$$

Here we propose a method to calculate s_{BCC}, which relies on the result of s_{GCC} and β. Consider the β-edges, which are inside the GCC but outside of the BCC. Note that each β-edge can be assigned to one node that is inside the GCC but outside the BCC. Hence s_{BCC} can be calculated as

$$s_{\mathrm{BCC}} = s_{\mathrm{GCC}} - \beta c/2 = s_{\mathrm{GCC}} - u(1-u)c, \tag{B4}$$

where $\beta c/2 = u(1-u)c$ represents the fraction of β-edges normalized by the total number of nodes. Note that the above two equations are equivalent, because $G_0'(u) = cG_1(u) = cu$.

[1] B. Bollobás, *Modern Graph Theory,* Graduate Texts in Mathematics Vol. 184 (Springer-Verlag, New York, 1998).

[2] M. Behzad and G. Chartrand, *Introduction to the Theory of Graphs* (Allyn and Bacon, Boston, 1972).

[3] F. Harary *et al.*, *Graph Theory* (Addison-Wesley, Reading, MA, 1969).

[4] R. Tarjan, SIAM J. Comput. **1**, 146 (1972).

[5] L. Tian, A. Bashan, D.-N. Shi, and Y.-Y. Liu, Nat. Commun. **8**, 14223 (2017).

[6] R. Pike and H. E. Stanley, J. Phys. A **14**, L169 (1981).

[7] J. A. Dunne, R. J. Williams, and N. D. Martinez, Proc. Natl. Acad. Sci. (USA) **99**, 12917 (2002).

[8] N. Simonis, J.-F. Rual, A.-R. Carvunis, M. Tasan, I. Lemmens, T. Hirozane-Kishikawa, T. Hao, J. M. Sahalie, K. Venkatesan, F. Gebreab *et al.*, Nat. Methods **6**, 47 (2009).

[9] J. Leskovec and A. Krevl, SNAP Datasets: Stanford Large Network Dataset Collection (https://snap.stanford.edu/data/) (2014).

[10] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[11] P. Erdös and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci **5**, 17 (1960).

[12] B. Bollobás, W. Fulton, A. Katok, F. Kirwan, and P. Sarnak, in *Random Graphs* (Cambridge University Press, New York, 2001).

[13] T. Nepusz, A. Petróczi, L. Négyessy, and F. Bazsó, Phys. Rev. E **77**, 016107 (2008).

[14] K. Gong, M. Tang, H. Yang, and M. Shang, Chaos **21**, 043130 (2011).

[15] P. Jensen, M. Morini, M. Karsai, T. Venturini, A. Vespignani, M. Jacomy, J.-P. Cointet, P. Mercklé, and E. Fleury, J. Complex Netw. **4**, 319 (2015).

[16] X.-Q. Cheng, F.-X. Ren, H.-W. Shen, Z.-K. Zhang, and T. Zhou, J. Stat. Mech. (2010) P10011.
[17] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, Phys. Rev. E **65**, 056109 (2002).
[18] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[19] M. E. J. Newman and G. Ghoshal, Phys. Rev. Lett. **100**, 138701 (2008).
[20] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).
[21] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[22] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[23] S. N. Dorogovtsev, A. V. Goltsev, and J. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008).
[24] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[25] M. Mezard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009).
[26] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
[27] M. Catanzaro and R. Pastor-Satorras, Eur. Phys. J. B **44**, 241 (2005).
[28] J.-S. Lee, K.-I. Goh, B. Kahng, and D. Kim, Eur. Phys. J. B **49**, 231 (2006).
[29] Y.-Y. Liu, E. Csóka, H. Zhou, and M. Pósfai, Phys. Rev. Lett. **109**, 205703 (2012).
[30] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. (USA) **99**, 7821 (2002).
[31] S. P. Borgatti and M. G. Everett, Social Netw. **21**, 375 (2000).
[32] P. W. Holland, K. B. Laskey, and S. Leinhardt, Social Netw. **5**, 109 (1983).
[33] B. Karrer and M. E. J. Newman, Phys. Rev. E **83**, 016107 (2011).
[34] S. H. Lee, M. Cucuringu, and M. A. Porter, Phys. Rev. E **89**, 032810 (2014).