# Controllability of giant connected components in a directed network

Xueming Liu,[1,2] Linqiang Pan,[1,3] H. Eugene Stanley,[2] and Jianxi Gao[4,*]

[1]*Key Laboratory of Image Information Processing and Intelligent Control of Education Ministry of China, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China*

[2]*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA*

[3]*School of Electric and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan, China*

[4]*Center for Complex Network Research and Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA*

(Received 10 March 2016; revised manuscript received 15 November 2016; published 24 April 2017)

When controlling a complex networked system it is not feasible to control the full network because many networks, including biological, technological, and social systems, are massive in size and complexity. But neither is it necessary to control the full network. In complex networks, the giant connected components provide the essential information about the entire system. How to control these giant connected components of a network remains an open question. We derive the mathematical expression of the degree distributions for four types of giant connected components and develop an analytic tool for studying the controllability of these giant connected components. We find that for both Erdős-Rényi (ER) networks and scale-free (SF) networks with $p$ fraction of remaining nodes, the minimum driver node density to control the giant component first increases and then decreases as $p$ increases from zero to one, showing a peak at a critical point $p = p_{\mathrm{m}}$. We find that, for ER networks, the peak value of the driver node density remains the same regardless of its average degree $\langle k \rangle$ and that it is determined by $p_{\mathrm{m}}\langle k \rangle$. In addition, we find that for SF networks the minimum driver node densities needed to control the giant components of networks decrease as the degree distribution exponents increase. Comparing the controllability of the giant components of ER networks and SF networks, we find that when the fraction of remaining nodes $p$ is low, the giant in-connected, out-connected, and strong-connected components in ER networks have lower controllability than those in SF networks.

DOI: 10.1103/PhysRevE.95.042318

## I. INTRODUCTION

Many real-world complex systems, including social systems [1], biological systems [2], and the Internet [3], can be modeled as complex networks [4]. In recent decades, studies of the structure and dynamics of complex networked systems [1–6] have enabled us to understand both natural and technological systems, but the ultimate proof that we understand these systems is indicated by how well we are able to control them [7]. According to control theory [8], a dynamic system is controllable if, using appropriate external inputs, the state of the system state can be driven from any initial state to any desired state within a finite time period. An analytic framework has been developed for studying the structural controllability of complex networks, which is denoted by the minimum driver node density [7], via a combination of tools from network science [9], control theory [10], and statistical physics [11], and this has triggered a hot spot of research activity [12–19]. These studies of the controllability of complex networks can help us understand complex systems in science and engineering and have the potential of generating technological breakthroughs in our ability to control them [20].

Most of the studies cited above have focused on complete control. Complete control is essential in some engineered systems, such as the fly-by-wire system for controlling the surfaces of an airplane and cruise control systems in automobiles [21], but because many biological, technological, and social systems are massive in size and complexity, it is not feasible to attempt complete control of the network [22]. But neither is it necessary. Gao *et al.* recently developed a approach for studying the targeted control of complex networks [22] and focused on two ways of choosing which subset of nodes to control: (1) a random scheme in which a fraction $f$ of nodes are chosen uniformly at random and (2) a local scheme in which the chosen nodes form a connected component that constitute a well-defined local neighborhood. Gao *et al.* [22] developed an alternative $k$-walk theory and a greedy algorithm for studying the target control of networks and found that the structural controllability approach used for full control overestimates the minimum number of driver nodes required for targeted control, and that many real-world networks are suitable for efficient targeted control.

We can also select the target node set to be the functional nodes belonging to the giant connected component of the network [23]. When researchers study the robustness of a networked system, only the nodes in giant connected components are considered functional [24–26]. The functioning of networked systems is crucially dependent on the connectivity between nodes that enables them to function cooperatively as a network, e.g., airline routes, electric power grids, and the Internet [25]. Here we choose the nodes of the giant connected components to be the target node sets. There are four types of giant connected components [23,27,28] in a directed network: (1) giant weakly connected components (GWCCs) in which each pair of nodes can connect via a path irrespective of the directionality of the link, (2) giant strongly connected components (GSCCs), in which each pair of nodes can connect by directed paths, (3) giant in-components (GINs), the set of nodes from which the GSCC are approachable by directed paths; and (4) the giant out-components (GOUTs), the set of nodes approachable from the GSCCs by directed paths. A theoretical

*jianxi.gao@gmail.com

 

tool for analyzing the controllability of the giant components of an arbitrary complex network remains an open problem.

Here we develop an analytic tool to study the controllability of the four giant components in a directed network with an arbitrary joint degree distribution. We first present a method to calculate the in-degree and out-degree distributions of the giant components. The mathematical expressions of the degree distributions of the giant components are interesting and fundamental in network science. If we knew the degree distributions of a network or a subnetwork, we could study the properties of the network or the subnetwork, such as controllability [7], robustness [26], synchronization [29], resilience [30], and many more. Though the degree distribution of the giant component of an undirected network [31] is known, we lack the analytical tools for degree distributions of the giant components of directed networks. We derived the in-degree and out-degree distributions of the four types of giant components of a network with arbitrary degree distributions. With our results of the giant components' degree distributions, one can use them to study the network robustness, synchronization, and resilience.

Then we use the structural controllability theory to compute the minimum driver node densities $g_D$ required to control the giant components. We also study how the minimum driver node densities $g_D$ change when a randomly chosen $1-p$ fraction of nodes are removed from the full network. When the fraction of the remaining nodes $p$ increases from zero to one, the minimum driver node density first increases and then decreases, showing a peak at a critical point $p = p_m$ for both Erdős-Rényi (ER) networks and scale-free (SF) networks. In particular, for each type of giant component in a ER network, the peak value $g_{Dm}$ is determined by the product of the fraction of the remaining nodes and the average degree $p\langle k\rangle$, indicating that for each giant component $p_m\langle k\rangle$ is a constant value. Thus, the critical point $p_m$ has a reciprocal relation with the average degree $\langle k\rangle$. For the SF networks, the in-degree and the out-degree follow power law distributions with $P(k_{in}) \propto k_{in}^{-\lambda_{in}}$ and $P(k_{out}) \propto k_{out}^{-\lambda_{out}}$. The minimum driver node densities required to control the giant components in SF networks decrease as the degree distribution exponents $\lambda_{in}$ and $\lambda_{out}$ increase. We find that the GWCCs in ER networks are easier to control than that in SF networks with the same average degree. For the GIN and GSCC, ER networks are not always easier to control than SF networks, which is similar to the findings in Ref. [22]. When the fraction of remaining nodes $p$ is low, ER networks have lower controllability than SF networks.

## II. THE DEGREE DISTRIBUTIONS OF THE GIANT COMPONENTS

Given a directed network with every node assigned with an in-degree $k_{in}$ and an out-degree $k_{out}$ from a joint probability distribution $P(k_{in},k_{out})$, we define the generating function of the degree distribution

$$\Phi(x,y) = \sum_{k_{in},k_{out}}^{\infty} P(k_{in},k_{out}) x^{k_{in}} y^{k_{out}}, \qquad (1)$$

where $x,y$ are arbitrary complex variables [27]. The Z transform of the out-degree distribution of the node, arrived by following a randomly chosen link when one moves along the link direction (called a branching process [23]), is $\Phi_1(1,y) = \partial_x \Phi(x,y)|_{x=1}/\partial_x \Phi(1,1)$, where $\partial_x \Phi(1,1) = \frac{\langle k\rangle}{2}$ with $\langle k\rangle$ being the average degree. Accordingly, the Z transform of the in-degree of the node can be found by considering motion in the direction opposite to that of the directed link (also called a branching process) that is $\Phi_1(x,1) = \partial_y \Phi(x,y)|_{y=1}/\partial_y \Phi(1,1)$. Once a fraction $1-p$ of nodes is randomly removed from a network, the generating functions of the degree distributions of the remaining network and the branching processes remain the same but must be computed from two new arguments: $z_{in} = px + 1 - p$ and $z_{out} = py + 1 - p$ [32,33].

For the four types of giant components, we use superscripts to the variables that represent the properties of the giant components: "(w)" for GWCCs, "(i)" for GINs, "(o)" for GOUTs, and "(s)" for GSCCs. For example, $P^{(w)}(k_{in},k_{out})$ represents the degree distribution of the GWCCs. We also use the superscript "(g)" to generalize all the superscripts: "(w)" for GWCCs, "(i)" for GINs, "(o)" for GOUTs, and "(s)" for GSCCs. For any of the giant components, the degree distribution is $P^{(g)}(k_{in},k_{out})$, and the generating function of the degree distribution of a giant component is $\Phi^{(g)}(x,y) = \sum_{k_{in},k_{out}}^{\infty} P^{(g)}(k_{in},k_{out}) x^{k_{in}} y^{k_{out}}$. The in-degree and out-degree distributions of the full network are assumed to be independent from each other, thus the in-degree and out-degree distributions of the giant component are also independent from each other: $P^{(g)}(k_{in},k_{out}) = P^{(g)}(k_{in})P^{(g)}(k_{out})$. The generating functions of the in-degree and out-degree distributions of a giant component are $\Phi^{(g)}(x,1) = \sum_{k_{in}}^{\infty} P^{(g)}(k_{in}) x^{k_{in}}$ and $\Phi^{(g)}(1,y) = \sum_{k_{out}}^{\infty} P^{(g)}(k_{out}) y^{k_{out}}$ respectively. In addition, the average degree of any giant component is denoted $\langle k\rangle_G$ and $\langle k\rangle_G = 2\partial_x \Phi^{(g)}(x,1)|_{x=1} = 2\partial_y \Phi^{(g)}(1,y)|_{y=1}$.

The directionality of a network is ignored in GWCCs, and the generating function can be reduced to be $G_0(x) = \Phi(x,x)$. Here the distribution of the degree minus one of nodes that are arrived at by following a randomly chosen link is generated by $G_1(x) = G_0'(x)/G_0'(1)$ [27]. The GWCC exists if $G_1'(x) > 1$ and the size of the GWCC $W$ can be obtained by

$$W = 1 - G_0(u_w), \quad u_w = G_1(u_w), \qquad (2)$$

where $u_w$ is the probability that one side of a randomly chosen link is connected to a finite weakly connected component. It follows that $u_w^2$ is the probability that a random link belongs to a finite weakly connected component, and $1 - u_w^2$ is the probability that a random link belongs to the GWCC [31]. We define $h_w = G_0(u_w)$ to be the probability that a randomly chosen node belongs to a finite connected component, and $1 - h_w$ is the probability that it belongs to the GWCC.

The degree distribution of the GWCC is examined in Ref. [31]. They find that $P(k) = (1-h_w)P^{(w)}(k) + h_w P^{(f)}(k)$, where $P(k)$, $P^{(w)}(k)$, and $P^{(f)}(k)$ are the degree distributions of the full network, the GWCCs, and the finite components, respectively. The computation of the in-degree and out-degree distributions of the GWCC is similar to the computation of its degree distribution, i.e., $P(k_{in}) = (1-h_w)P^{(w)}(k_{in}) + h_w P^{(f)}(k_{in})$ where $P(k_{in})$, $P^{(w)}(k_{in})$ and $P^{(f)}(k_{in})$ are the in-degree distributions of the full network, the GWCC, and the finite components, respectively. The in-degree distribution of the finite components $P^{(f)}(k_{in})$ is equivalent to the in-degree distribution of a randomly chosen part consisting of $h_{\mathrm{w}}$ fraction of nodes [34], and it follows that

$$P^{(f)}(k_{\mathrm{in}}) = \sum_{i \geqslant k_{\mathrm{in}}}^{\infty} P(k_{\mathrm{in}}) \binom{i}{k} h_{\mathrm{w}}^{k_{\mathrm{in}}} (1-h_{\mathrm{w}})^{i-k_{\mathrm{in}}}, \qquad (3)$$

where $\binom{i}{k} = \frac{i!}{k!(i-k)!}$ is a combination, and the generating function of $P^{(f)}(k_{\mathrm{in}})$ is $\Phi^{(f)}(x,1) = \Phi(1-h_{\mathrm{w}}(1-x),1)$. We obtain the in-degree distribution of the GWCC:

$$P^{(\mathrm{w})}(k_{\mathrm{in}}) = \frac{P(k_{\mathrm{in}}) - h_{\mathrm{w}} P^{(f)}(k_{\mathrm{in}})}{1-h_{\mathrm{w}}}. \qquad (4)$$

Substituting Eqs. (3) and (4) into Eq. (1), the generating function of the in-degree distribution of the GWCC is

$$\Phi^{(\mathrm{w})}(x,1) = \frac{\Phi(x,1) - h_{\mathrm{w}} \Phi(1-h_{\mathrm{w}}(1-x),1)}{1-h_{\mathrm{w}}}. \qquad (5)$$

Accordingly, we can get the out-degree distribution of the GWCC:

$$P^{(\mathrm{w})}(k_{\mathrm{out}}) = \frac{P(k_{\mathrm{out}}) - h_{\mathrm{w}} P^{(f)}(k_{\mathrm{out}})}{1-h_{\mathrm{w}}}, \qquad (6)$$

where $P^{(f)}(k_{\mathrm{out}})$ satisfies Eq. (3) by replacing $k_{\mathrm{in}}$ as $k_{\mathrm{out}}$. The generating function of the out-degree distribution of the GWCC is

$$\Phi^{(\mathrm{w})}(1,y) = \frac{\Phi(1,y) - h_{\mathrm{w}} \Phi(1,1-h_{\mathrm{w}}(1-y))}{1-h_{\mathrm{w}}}. \qquad (7)$$

Figures 1(a) and 1(b) show the analytic results (solid lines) of the in-degree distribution [Eq. (4)] and out-degree distribution [Eq. (6)] of the GWCC in a network whose in-degrees are Possion-distributed and out-degrees are power-law distributed, respectively. They agree well with the corresponding quantities calculated from large synthetic random networks (symbols).

In a directed network with an arbitrary in-degree distribution $P(k_{\mathrm{in}})$ and an arbitrary out-degree distribution $P(k_{\mathrm{out}})$, we find the following four important properties for the in-degree and out-degree distributions of the GIN, GOUT, and GSCC (shown in Table I): (1) the in-degree distribution of the GIN is the same as that of the full network $P^{(\mathrm{in})}(k_{\mathrm{in}}) = P(k_{\mathrm{in}})$; (2) the out-degree distribution of the GOUT is the same as that of the full network $P^{(\mathrm{out})}(k_{\mathrm{out}}) = P(k_{\mathrm{out}})$; (3) the in-degree distribution of the GSCC is the same as the in-degree distribution of the GOUT $P^{(\mathrm{s})}(k_{\mathrm{in}}) = P^{(\mathrm{out})}(k_{\mathrm{in}})$; and (4) the out-degree distribution of GSCC is the same as the out-degree distribution of the GIN $P^{(\mathrm{s})}(k_{\mathrm{out}}) = P^{(\mathrm{in})}(k_{\mathrm{out}})$. Thus computing the in-degree distribution of the GOUT and the out-degree distribution of the GIN gives us all of the in-degree and out-degree distributions of the GIN, GOUT, and GSCC.

The GIN is present when $\Phi_1'(x,1)|_{x=1} > 1$, and its size is $I = 1 - \Phi(u_{\mathrm{in}},1)$, where $u_{\mathrm{in}} = \Phi_1(u_{\mathrm{in}},1)$ is the probability that one side of a randomly chosen link connects to a finite in-component [23]. It follows that $u_{\mathrm{in}}^2$ is the probability that a randomly chosen link belongs to a finite in-component. The GOUT is present when $\Phi_1'(1,y)|_{y=1} > 1$, and its size is $O = 1 - \Phi(1,u_{\mathrm{out}})$, where $u_{\mathrm{out}} = \Phi_1(1,u_{\mathrm{out}})$ is the probability that one side of a randomly chosen link connects to a finite out-component. It follows that $u_{\mathrm{out}}^2$ is the probability that a randomly chosen link belongs to a finite out-component. We define $h_{\mathrm{in}} = \Phi(u_{\mathrm{in}},1)$ and $h_{\mathrm{out}} = \Phi(1,u_{\mathrm{out}})$ as the fraction

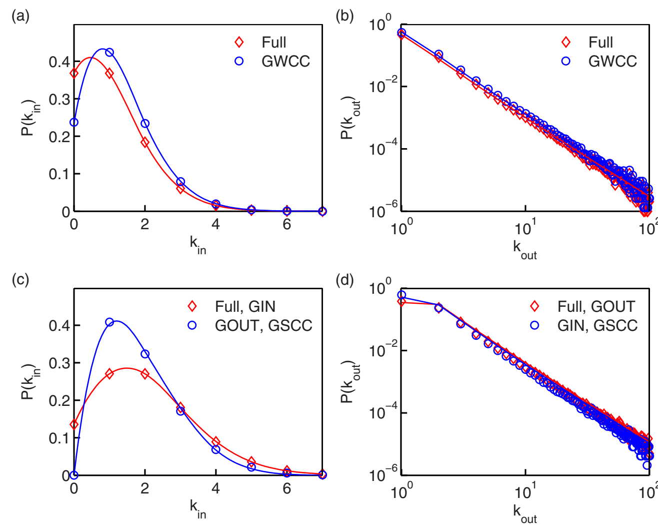


FIG. 1. The in- and out-degree distributions of the full network, and the four giant components of a network with the in-degrees Poisson distributed and the out-degrees power law distributed, after removing $p = 0.5$ fraction nodes from the original network. (a), The in-degree distribution and (b) the out-degree distribution the remaining part after nodes removal (full) and the GWCC of a network with the average degree $\langle k \rangle = 4$. The average degree of the original network of (c) and (d) is $\langle k \rangle = 4$. From (c) and (d), we find that (1) the in-degree distribution of the GIN is the same as that of the full network $P^{(\mathrm{in})}(k_{\mathrm{in}}) = P(k_{\mathrm{in}})$; (2) the out-degree distribution of the GOUT is the same as that of the full network $P^{(\mathrm{out})}(k_{\mathrm{out}}) = P(k_{\mathrm{out}})$; (3) the in-degree distribution of the GSCC is the same as the in-degree distribution of the GOUT $P^{(\mathrm{s})}(k_{\mathrm{in}}) = P^{(\mathrm{out})}(k_{\mathrm{in}})$; and (4) the out-degree distribution of GSCC is the same as the out-degree distribution of the GIN $P^{(\mathrm{s})}(k_{\mathrm{out}}) = P^{(\mathrm{in})}(k_{\mathrm{out}})$. The out-degree distribution exponent of the network in (b) and (d) is $\lambda = 2.5$. The lines represent theoretic results and they agree with the simulation results (denoted by symbols and the size of each network is $N = 10^6$).

of nodes that do not belong to the GIN and the GOUT, respectively. The derivation of the out-degree distribution of the GIN is similar to that of the above in-degree distribution of the GOUT. We next compute the in-degree distribution of the GOUT.

Between the GOUT and the rest of the nodes that do not belong to the GOUT (non-GOUT), no link points from the GOUT to the non-GOUT, but there are links that point from the non-GOUT to the GOUT, which are ignored when computing the in-degree distribution of the GOUT. If we take these links connecting the GOUT and the non-GOUT into consideration, the formula of the in-degree distribution of the GOUT is analogous to the degree distribution of the

TABLE I. The in-degree and out-degree distribution of the GIN, GOUT, and GSCC

| | In-degree distribution | Out-degree distribution |
|---|---|---|
| Full network | $P(k_{\mathrm{in}})$ | $P(k_{\mathrm{out}})$ |
| GIN | $P^{(\mathrm{in})}(k_{\mathrm{in}}) = P(k_{\mathrm{in}})$ | $P^{(\mathrm{in})}(k_{\mathrm{out}})$ |
| GOUT | $P^{(\mathrm{out})}(k_{\mathrm{in}})$ | $P^{(\mathrm{out})}(k_{\mathrm{out}}) = P(k_{\mathrm{out}})$ |
| GSCC | $P^{(\mathrm{s})}(k_{\mathrm{in}}) = P^{(\mathrm{out})}(k_{\mathrm{in}})$ | $P^{(\mathrm{s})}(k_{\mathrm{out}}) = P^{(\mathrm{in})}(k_{\mathrm{out}})$ |

GWCC [31], which is not the real in-degree distribution of the GOUT, but it will help us find the real in-degree distribution. We use $P_0^{(\mathrm{out})}(k_{\mathrm{in}})$ to denote such in-degree distribution:

$$P_0^{(\mathrm{out})}(k_{\mathrm{in}}) = \frac{P(k_{\mathrm{in}})\left(1-u_{\mathrm{out}}^{k_{\mathrm{in}}}\right)}{(1-h_{\mathrm{out}})}. \tag{8}$$

In addition, the number of links in the GOUT and in the non-GOUT can be computed explicitly, and they are $1-h_{\mathrm{out}}$ and $u_{\mathrm{out}}^2$, respectively. Because the out-degree distribution of the GOUT is the same as the full network, the fraction of links belonging to the GOUT is the same as the fraction of nodes belonging to the GOUT, which is $1-h_{\mathrm{out}}$. A link belongs to the non-GOUT, meaning that this link belongs to a finite out-component. The probability that a randomly chosen link belongs to a finite out-component is $u_{\mathrm{out}}^2$. Thus the fraction of links pointing from the non-GOUT to the GOUT is $f_{\mathrm{out}} = h_{\mathrm{out}} - u_{\mathrm{out}}^2$.

To determine the in-degree distribution of the GOUT, we must eliminate the influence of the links pointing from the non-GOUT to the GOUT. Before removing the links pointing from the non-GOUT to the GOUT, the probability that a node with in-degree $k_{\mathrm{in}}$ is $P_0^{(\mathrm{out})}(k_{\mathrm{in}})$, as Eq. (8) shows. The probability that one link pointing from the non-GOUT to this node being removed is $p_r(k_{\mathrm{in}}) = k_{\mathrm{in}} f_{\mathrm{out}}/h_{\mathrm{out}}$. In the GOUT, the probability that the degree of a node decreases from $k$ to $k-1$ is $P_0^{(\mathrm{out})}(k_{\mathrm{in}})p_r(k_{\mathrm{in}})$. After removing the links pointing from the non-GOUT to the GOUT, the probability that a node with in-degree $k_{\mathrm{in}}$ is $P^{(\mathrm{out})}(k_{\mathrm{in}}) = P_0^{(\mathrm{out})}(k_{\mathrm{in}}) - P_0^{(\mathrm{out})}(k_{\mathrm{in}})p_r(k_{\mathrm{in}}) + P_0^{(\mathrm{out})}(k_{\mathrm{in}}+1)p_r(k_{\mathrm{in}}+1)$. In fact, there may be the case that two or more than two links of a node have been removed. However, in a large network, the fraction of links pointing from the non-GOUT to the GOUT is too small, and the probability that two or more than two such links pointing to a same node is even smaller, such cases can be ignored. Thus we obtain the probability of a node with in-degree $k_{\mathrm{in}}$ in the GOUT:

$$P^{(\mathrm{out})}(k_{\mathrm{in}}) = \begin{cases} P_0^{(\mathrm{out})}(k_{\mathrm{in}}) + P_0^{(\mathrm{out})}(k_{\mathrm{in}}+1)p_r(k_{\mathrm{in}}+1) & k_{\mathrm{in}} = 1, \\ P_0^{(\mathrm{out})}(k_{\mathrm{in}})[1-p_r(k_{\mathrm{in}})] + P_0^{(\mathrm{out})}(k_{\mathrm{in}}+1)p_r(k_{\mathrm{in}}+1) & 1 < k_{\mathrm{in}} < k_{\max}^{(\mathrm{out})}, \\ P_0^{(\mathrm{out})}(k_{\mathrm{in}})[1-p_r(k_{\mathrm{in}})] & k_{\mathrm{in}} = k_{\max}^{(\mathrm{out})}, \end{cases} \tag{9}$$

where $k_{\max}^{(\mathrm{out})}$ is the maximum in-degree of the GOUT. Accordingly, we get the out-degree distribution of the GIN $P^{(\mathrm{in})}(k_{\mathrm{out}})$,

$$P^{(\mathrm{in})}(k_{\mathrm{out}}) = \begin{cases} P_0^{(\mathrm{in})}(k_{\mathrm{out}}) + P_0^{(\mathrm{in})}(k_{\mathrm{out}}+1)p_r(k_{\mathrm{out}}+1) & k_{\mathrm{out}} = 1, \\ P_0^{(\mathrm{in})}(k_{\mathrm{out}})[1-p_r(k_{\mathrm{out}})] + P_0^{(\mathrm{in})}(k_{\mathrm{out}}+1)p_r(k_{\mathrm{out}}+1) & 1 < k_{\mathrm{out}} < k_{\max}^{(\mathrm{in})}, \\ P_0^{(\mathrm{in})}(k_{\mathrm{out}})[1-p_r(k_{\mathrm{out}})] & k_{\mathrm{out}} = k_{\max}^{(\mathrm{in})}, \end{cases} \tag{10}$$

where $k_{\max}^{(\mathrm{in})}$ is the maximum out-degree of the GIN, $P_0^{(\mathrm{in})}(k_{\mathrm{out}}) = P(k_{\mathrm{out}})(1-u_{\mathrm{in}}^{k})/(1-h_{\mathrm{in}})$ with $P(k_{\mathrm{out}})$ being the out-degree distribution of the full network, and $p_r(k_{\mathrm{out}}) = k_{\mathrm{out}}(h_{\mathrm{in}} - u_{\mathrm{in}}^2)/h_{\mathrm{in}}$.

The GSCC appears at the intersection of the GIN and GOUT, and its relative size takes the form $S = 1 - \Phi^{(s)}(u_{\mathrm{in}},u_{\mathrm{out}}) = 1 - \Phi(u_{\mathrm{in}},1) - \Phi(1,u_{\mathrm{out}}) + \Phi(u_{\mathrm{in}},u_{\mathrm{out}})$ [23], where $u_{\mathrm{in}} = \Phi_1(u_{\mathrm{in}},1)$ and $u_{\mathrm{out}} = \Phi_1(1,u_{\mathrm{out}})$. The in-degree and out-degree distributions of the GSCC are $P^{(s)}(k_{\mathrm{in}}) = P^{(\mathrm{out})}(k_{\mathrm{in}})$ and $P^{(s)}(k_{\mathrm{out}}) = P^{(\mathrm{in})}(k_{\mathrm{out}})$ respectively. Figures 1(c) and 1(d) show the analytic solutions of the in-degree and out-degree distribution of the GSCC (solid lines) agree well with the simulation results (symbols).

## III. CONTROLLABILITY OF THE GIANT COMPONENTS

The controllability of nonlinear systems is structurally similar to that of linear systems [7,8]. We study a system with canonical linear, time-invariant dynamics formulated by [35]

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t), \tag{11}$$

where the vector $x(t) = (x_1(t),x_2(t),\dots,x_N(t))^T$ describes the states of the $N$ nodes of the networked system at time $t$. The $N \times N$ matrix $A$ is the transposition of the adjacency matrix and captures the wiring diagram of the system and the interaction strengths between nodes. The $N \times M$ matrix $B$ is the input matrix ($N \geqslant M$) that identifies the nodes into which the input signals are injected, $M$ is the number of input signals, and $u(t) = (u_1(t),u_2(t),\dots,u_m(t))^T$ is the input vector.

In control theory, a system is controllable if it can be driven from any initial state to any desired final state during a finite time period [10]. According to Kalman's controllability rank condition [10], the system represented by Eq. (11) is controllable if and only if the $N \times NM$ controllability matrix $C$ has full rank:

$$\mathrm{rank}(C) = \mathrm{rank}[B,AB,A^2B,\dots,A^{N-1}B] = N. \tag{12}$$

This controllability rank condition indicates that to control the full network we must identify the number of signals and the nodes into which the signals are injected, called driver nodes. Liu *et al.* [7] recently showed that a system can be structurally controllable by a minimal set of driver nodes. A system is structurally controllable if it is possible to choose nonzero weights in $A$ and $B$ such that Eq. (12) holds [7]. The minimum number of driver nodes for controlling a full network is denoted $N_D$ and the minimum driver node density is $n_D = N_D/N$. The minimum driver node density required to control the full complex network quantifies its structural controllability.

To study the controllability of the giant components of a network, we define the minimum number of driver node density to control a giant component as $G_D$ and the corresponding minimum driver node density as $g_D = G_D/N$, where $N$ is the number of nodes in the full network. The minimum driver node density required to control the GWCC is $g_D^{(w)}$, and the minimum driver node densities to control the GIN, GOUT, and GSCC are denoted as $g_D^{(in)}$, $g_D^{(out)}$, and $g_D^{(s)}$ respectively. The minimum driver node density $g_D$ is determined by the in-degree and out-degree distributions of a giant component and the size of the giant component. We next compute the minimum driver node densities of the giant components in a directed network.

We compute the minimum driver node density required to control a giant component by substituting its degree distributions (see Sec. II) for the degree distributions of the original networks into the equation for computing the minimum driver node density for controlling the full network [7], which is

$$g_D = (1-h_g)\frac{1}{2}\{[\Phi^{(g)}(\widehat{\omega}_2,1)+\Phi^{(g)}(1-\widehat{\omega}_1,1)-1] + [\Phi^{(g)}(1,\omega_2)+\Phi^{(g)}(1,1-\omega_1)-1] + \frac{\langle k\rangle_G}{2}[\widehat{\omega}_1(1-\omega_2)+\omega_1(1-\widehat{\omega}_2)]\}, \tag{13}$$

where the variables $\omega_1$, $\omega_2$, $\widehat{\omega}_1$ and $\widehat{\omega}_2$ satisfy

$$\begin{aligned} \omega_1 &= \Phi_1^{(g)}(\widehat{\omega}_2,1) \quad \omega_2 = 1-\Phi_1^{(g)}(1-\widehat{\omega}_1,1),\\ \widehat{\omega}_1 &= \Phi_1^{(g)}(1,\omega_2) \quad \widehat{\omega}_2 = 1-\Phi_1^{(g)}(1,1-\omega_1), \end{aligned} \tag{14}$$

with $\Phi_1^{(g)}(x,1) = \partial_x \Phi^{(g)}(x,1)|_{x=1}/\frac{\langle k\rangle_G}{2}$ and $\Phi_1^{(g)}(1,y) = \partial_y \Phi^{(g)}(1,y)|_{y=1}/\frac{\langle k\rangle_G}{2}$.

To study the controllability of the giant components when there are node or link failures, we randomly remove a fraction $1-p$ of nodes and all the links connecting to these nodes. Given a directed network with a degree distribution $P(k_{in},k_{out})$, whose generating function is $\Phi(x,y) = \sum_{k_{in},k_{out}}^{\infty} P(k_{in},k_{out})x^{k_{in}}y^{k_{out}}$. After the nodes removal, the in-degree and out-degree distributions of the network [34] are

$$\begin{aligned} P^{(p)}(k_{in}) &= \sum_{i\geqslant k_{in}}^{\infty} P(k_{in})\binom{i}{k} p^{k_{in}}(1-p)^{i-k_{in}},\\ P^{(p)}(k_{out}) &= \sum_{i\geqslant k_{out}}^{\infty} P(k_{out})\binom{i}{k} p^{k_{out}}(1-p)^{i-k_{out}}. \end{aligned} \tag{15}$$

The generating function of the degree distribution of the remaining part is $\Phi(z_{in},z_{out})$ with $z_{in} = px+1-p$ and $z_{out} = py+1-p$ [32,33]. Substituting $\Phi(z_{in},z_{out})$ to Eqs. (5) and (7), we get the generating functions of the in-degree distribution and the out-degree distribution of the GWCC of a network after $1-p$ fraction of nodes removal. Similarly, the in-degree distributions and out-degree distributions of the GIN(GOUT) and the GSCC can be computed by substituting $\Phi(z_{in},z_{out})$ into Eqs. (9) and (10). With the in-degree and out-degree distributions of a giant component and its relative size among the full network, the minimum driver node density $g_D$ of the giant component can be obtained by substituting them to Eq. (13). We next compute the driver node density $g_D$ for controlling the four giant connected components (GWCCs, GINs, GOUTs, and GSCCs) in ER networks and SF networks.

## IV. CONTROLLABILITY OF THE GIANT COMPONENTS OF ER NETWORKS

In a directed ER network with an average degree $\langle k\rangle$, the in-degree and out-degree follow a Poisson distribution, and there is no correlation between the in-degree distribution and the out-degree distribution. After removing a fraction $1-p$ of nodes from the ER network, the remaining subnetwork continues to follow a Poisson distribution with an average degree $\langle k\rangle p$. The joint degree distribution of the ER network is

$$P(k_{in},k_{out}) = P(k_{in})P(k_{out}) = \frac{e^{-\frac{\langle k\rangle}{2}}\left(\frac{\langle k\rangle}{2}\right)^{k_{in}}}{k_{in}!}\frac{e^{\frac{\langle k\rangle}{2}}\left(\frac{\langle k\rangle}{2}\right)^{k_{out}}}{k_{out}!}, \tag{16}$$

and its generating function is

$$\Phi(x,y) = e^{\frac{\langle k\rangle p}{2}(x+y-2)}. \tag{17}$$

### A. Controllability of the GWCC of ER networks

When we compute the size of the GWCC we ignore the directionality of the links, and thus the generating functions of the degree distribution and the branching process of the full network after removing $1-p$ nodes is

$$G_0(x) = G_1(x) = e^{\langle k\rangle p(x-1)}. \tag{18}$$

Because the generating functions of the degree distribution and the branching process are the same, thus $h_w = u_w$. The size of the GWCC is $1-h_w$, where $h_w \in [0,1]$ is a solution of the equation $h_w = e^{\langle k\rangle p(h_w-1)}$. If the equation $h_w = e^{\langle k\rangle p(h_w-1)}$ has multiple solutions, then $h_w$ is the one that is closest to 1 [23]. Substituting Eq. (18) and $h_w$ to Eqs. (5) and (7), the generating functions of the in-degree and out-degree distributions of the GWCC are

$$\begin{aligned} \Phi^{(w)}(x,1) &= \Phi^{(w)}(1,x) = \frac{1}{1-h_w}\left(e^{\frac{\langle k\rangle p}{2}(x-1)} - h_w e^{h_w\frac{\langle k\rangle p}{2}(x-1)}\right),\\ \Phi_1^{(w)}(x,1) &= \Phi_1^{(w)}(1,x) = \frac{\Phi^{(w)'}(x,1)}{\Phi^{(w)'}(1,1)}\\ &= \frac{1}{1-h_w^2}\left(e^{\frac{\langle k\rangle p}{2}(x-1)} - h_w^2 e^{h_w\frac{\langle k\rangle p}{2}(x-1)}\right). \end{aligned} \tag{19}$$

The average degree in the GWCC is $\langle k\rangle p(1+h_w)$. Substituting Eq. (19) into Eq. (13), the minimum density of driver nodes of the GWCC after removing a fraction $1-p$ of nodes is

$$g_D^{(w)} = \left[(1+h_w)(\omega_1+1-\omega_2) + \frac{\langle k\rangle}{2}p(1+h_w)\omega_1(1-\omega_2) - h_w\left(e^{-\frac{\langle k\rangle}{2}p\omega_1} + e^{\frac{\langle k\rangle}{2}ph_w(1-\omega_2)}\right) - 1\right](1-h_w), \tag{20}$$

where $\omega_1$ and $\omega_2$ are determined by

$$\begin{aligned} \omega_1 &= \frac{1}{1-h_w^2}\left(e^{-\frac{\langle k\rangle}{2}p(1-\omega_2)} - h_w^2 e^{-\frac{\langle k\rangle}{2}ph_w(1-\omega_2)}\right),\\ 1-\omega_2 &= \frac{1}{1-h_w^2}\left(e^{-\frac{\langle k\rangle}{2}p\omega_1} - h_w^2 e^{-\frac{\langle k\rangle}{2}ph_w\omega_1}\right). \end{aligned} \tag{21}$$

### B. Controllability of the GIN and GOUT of ER networks

Because of the symmetries between component sizes and degree distributions between GIN and GOUT, the driver node densities of the GIN and GOUT are also the same, i.e., $g_{\mathrm{D}}^{(\mathrm{in})} = g_{\mathrm{D}}^{(\mathrm{out})}$. We next we show the analysis of the driver node density $g_{\mathrm{D}}^{(\mathrm{out})}$ for controlling the GOUT among the full network after remove a fraction $1-p$ nodes.

After randomly removing $1-p$ fraction nodes from a ER network with the average degree $\langle k\rangle$, its in-degree and out-degree distributions are $P(k_{\mathrm{in}}) = e^{-\frac{\langle k\rangle p}{2}}(\frac{\langle k\rangle p}{2})^{k_{\mathrm{in}}}/k_{\mathrm{in}}!$ and $P(k_{\mathrm{out}}) = e^{-\frac{\langle k\rangle p}{2}}(\frac{\langle k\rangle p}{2})^{k_{\mathrm{out}}}/k_{\mathrm{out}}!$. The generating functions of the out-degree distribution $\Phi(1,y)$ and the branching process $\Phi_1(1,y)$ in the full network after remove $1-p$ fraction of nodes are $\Phi(1,y) = \Phi_1(1,y) = e^{\frac{\langle k\rangle p}{2}(y-1)}$. The GOUT emerges when there are nontrivial solutions of the equation $u_{\mathrm{out}} = \Phi_1(1,u_{\mathrm{out}}) = e^{\frac{\langle k\rangle p}{2}(u_{\mathrm{out}}-1)}$, and the size of GOUT is $I = 1 - h_{\mathrm{out}}$ where $h_{\mathrm{out}} = e^{\frac{\langle k\rangle}{2}(h_{\mathrm{out}}-1)}$. Thus for the GOUT of a ER network, $h_{\mathrm{out}} = u_{\mathrm{out}}$. In addition, the generating functions of the out-degree distribution and the branching process of the GOUT are

$$\Phi^{(\mathrm{out})}(1,y) = \Phi_1^{(\mathrm{out})}(1,y) = e^{\frac{\langle k\rangle p}{2}(y-1)}. \tag{22}$$

The average degree of the GOUT is the same as the full network, i.e., $\langle k\rangle_I = \langle k\rangle p$. Between the GOUT and the part of nodes that do not belong to the GOUT (non-GOUT), there is a fraction $f_{\mathrm{out}} = h_{\mathrm{out}} - h_{\mathrm{out}}^2$ of links that point from the non-GOUT to the GOUT. If we take the links pointing from the non-GOUT to the nodes in GOUT into consideration, the in-degree distribution of the GOUT is $P_0^{(\mathrm{out})}(k_{\mathrm{in}}) = P(k_{\mathrm{in}})(1 - h_{\mathrm{out}}^{k_{\mathrm{in}}})/(1 - h_{\mathrm{out}})$. We remove the links connecting the GOUT and the non-GOUT to eliminate their influence on the in-degree distribution of the GOUT. Among all the links pointing into the nodes of the GOUT, the probability that a link has been removed is $f_{\mathrm{out}}/h_{\mathrm{out}} = 1 - h_{\mathrm{out}}$, and the probability of a link of a node with in-degree $k_{\mathrm{in}}$ has been removed is $p_r(k_{\mathrm{in}}) = k_{\mathrm{in}}(1 - h_{\mathrm{out}})$. Thus the in-degree distribution of the GOUT is

$$P^{(\mathrm{out})}(k_{\mathrm{in}}) = \begin{cases} e^{-\frac{\langle k\rangle p}{2}}\left(\frac{\langle k\rangle p}{2}\right)\left[1 + \frac{\langle k\rangle p}{2}\left(1 - h_{\mathrm{out}}^2\right)\right] & k_{\mathrm{in}} = 1, \\ \frac{e^{-\frac{\langle k\rangle p}{2}}\left(\frac{\langle k\rangle p}{2}\right)^{k_{\mathrm{in}}}}{k_{\mathrm{in}}!}\left\{\frac{(1-h_{\mathrm{out}}^{k})([-k_{\mathrm{in}}(1-h_{\mathrm{out}})]}{1-h_{\mathrm{out}}} + \frac{\langle k\rangle p}{2}\left(1 - h_{\mathrm{out}}^{k_{\mathrm{in}}+1}\right)\right\} & 1 < k_{\mathrm{in}} < k_{\mathrm{max}}^{(\mathrm{out})}, \\ \frac{e^{-\frac{\langle k\rangle p}{2}}\left(\frac{\langle k\rangle p}{2}\right)^{k_{\mathrm{in}}}}{k_{\mathrm{in}}!}\left\{\frac{(1-h_{\mathrm{out}}^{k})[1-k_{\mathrm{in}}(1-h_{\mathrm{out}})]}{1-h_{\mathrm{out}}}\right\} & k_{\mathrm{in}} = k_{\mathrm{max}}^{(\mathrm{out})}, \end{cases} \tag{23}$$

where $k_{\mathrm{max}}^{(\mathrm{out})}$ is the maximum in-degree of the GOUT of a ER network. Thus the generating functions of the in-degree distribution of the GOUT and the branching process in GOUT are

$$\Phi^{(\mathrm{out})}(x,1) = \sum_{k_{\mathrm{in}}=1}^{k_{\mathrm{max}}^{(\mathrm{out})}} P^{(\mathrm{out})}(k_{\mathrm{in}})x^{k_{\mathrm{in}}}, \qquad \Phi_1^{(\mathrm{out})}(x,1) = \frac{\Phi^{(I)'}(x,1)}{\Phi^{(I)'}(x,1)|_{x=1}}. \tag{24}$$

Substituting Eq. (24) into Eqs. (13) and (14), we can get the minimum fraction driver nodes for controlling the GOUT among the full network $g_{\mathrm{D}}^{(\mathrm{out})}$.

### C. Controllability of the GSCC of ER networks

For the GSCC of a ER network with the average degree being $\langle k\rangle$, its in-degree distribution and the out-degree distribution are the same with each other, and they all equal to the out-degree distribution of the GIN $P^{(\mathrm{in})}(k_{\mathrm{out}})$. Thus the generating functions of the in-degree and out-degree distributions and the branching processes of the GSCC are

$$\Phi^{(s)}(x,1) = \Phi^{(s)}(1,x) = \sum_{k_{\mathrm{out}}=1}^{k_{\mathrm{max}}^{(\mathrm{in})}} P^{(\mathrm{in})}(k_{\mathrm{out}})x^{k_{\mathrm{out}}},$$
$$\Phi_1^{(s)}(x,1) = \Phi_1^{(s)}(1,x) = \frac{\Phi^{(I)'}(1,x)}{\Phi^{(I)'}(1,x)|_{x=1}}. \tag{25}$$

In addition, the average degree of the GSCC is the same as the full network after removing $1-p$ fraction of nodes $\langle k\rangle_{\mathrm{s}} = \langle k\rangle p$, and the size of the GSCC is $1 - h_{\mathrm{s}} = (1 - h_{\mathrm{in}})^2$. The minimum driver node density of the GSCC $g_{\mathrm{D}}^{(s)}$ can be computed by substituting the generating functions of the in-degree and out-degree distributions of the GSCC to Eqs. (13) and (14).

In any of the four giant components (GWCCs, GINs, GOUTs, and GSCCs), the driver node density $g_{\mathrm{D}}$ first increases and then decreases as the fraction of the remaining nodes $p$ or the average degree $\langle k\rangle$ increasing and shows a peak. Figure 2

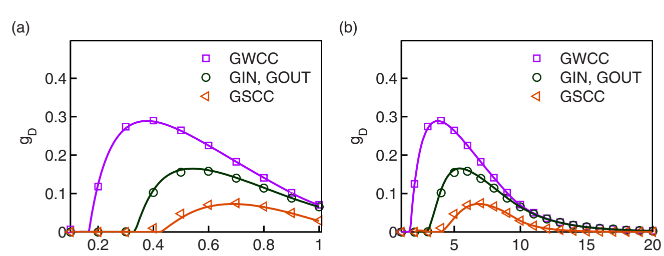


FIG. 2. The driver node densities of the GWCC, GIN(GOUT) and GSCC of ER networks. The driver node density $g_{\mathrm{D}}$ for each giant component first increases and then decreases, as (a) the fraction of the remaining nodes $p$ is increasing and (b) the average degree $\langle k\rangle$ is increasing, showing a peak. The solid lines represent the theoretic results, when $\langle k\rangle = 6$ in (a) and $p = 0.6$ in (b), and the symbols are simulation results where each network with the number of nodes $N = 10^6$. They agree well with each other.

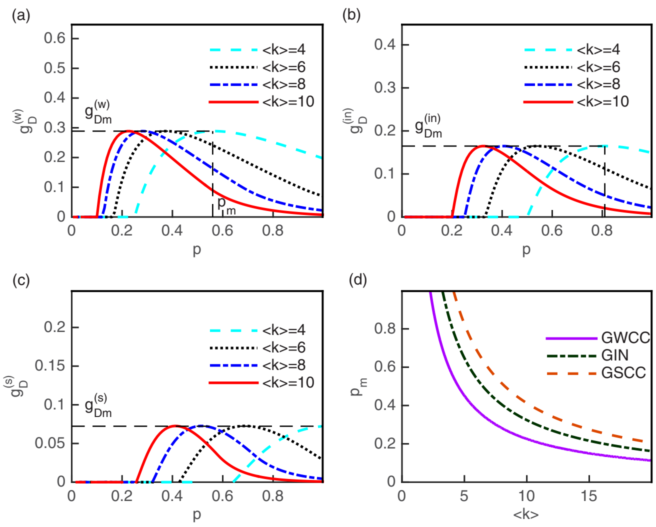


FIG. 3. Analysis of the peak value of driver node density for controlling the giant components of ER networks. The peak value $g_{\mathrm{Dm}}$ remains the same under different fraction of remaining nodes $p$ and different average $\langle k\rangle$ for (a) the GWCC, (b) the GIN/GOUT, and (c) the GSCC. (d) The critical point $p_{\mathrm{m}}$ in an ER network with average degree $\langle k\rangle$ where the driver node densities of the GWCC, GIN(GOUT) and GSCC reach their maximums: for the GWCC (solid line), the minimum driver node density reaches its maximal value $n_{pD}^{(\mathrm{w})} = 0.2888$ when $p_{\mathrm{m}} = \frac{2.2567}{\langle k\rangle}$; for the GIN(GOUT) (dot dash line), the minimum driver node density reaches its maximal value $n_{pD}^{(\mathrm{in})} = 0.1648$ when $p_{\mathrm{m}} = \frac{3.2385}{\langle k\rangle}$; for the GSCC (dash line), the minimum driver node density reaches its maximal value $n_{pD}^{(\mathrm{s})} = 0.0726$ when $p_{\mathrm{m}} = \frac{4.1234}{\langle k\rangle}$.

shows the analytic results of the minimum driver node densities of the GWCC, GIN (GOUT), and GSCC (solid lines), and they agree with the simulation results (symbols).

## V. THE MAXIMAL DRIVER NODE DENSITY

The minimum driver node density $g_{\mathrm{D}}$ required to control the giant component first increases and then decreases as the fraction of the remaining nodes $p$ increases, and it shows a peak at a critical point $p = p_{\mathrm{m}}$ (see Fig. 3). As $1 - p$ increases, the number of nodes and links in the giant component decrease simultaneously. When $p = 1$, there is no node removal. As we remove nodes, the network fragments into smaller components, the sparseness level increases, and the result is that the sparseness level of the giant component increases and its size decreases. As the sparseness level of the giant component increases, the number of driver nodes required to maintain control increases, but as the size the giant component decreases, the number of driver nodes required to maintain control decreases. At the beginning of the node removal process, the size of the giant component decreases slowly but quickly becomes more sparse. Thus the minimum driver node density $g_{\mathrm{D}}$ required to control the giant component at first increases. After the critical point, the size of the giant component decreases rapidly and the sparseness level slowly increases, so the minimum driver node density $g_{\mathrm{D}}$ decreases and continues to decrease until there is no giant component in which $g_{\mathrm{D}} = 0$.

An ER network with an average degree $\langle k\rangle$ continues to be a ER network after a $1 - p$ fraction of nodes is randomly removed, and the average degree becomes $p\langle k\rangle$ [32]. The peak value of the minimum driver node density is denoted $g_{\mathrm{Dm}}$. The value of $g_{\mathrm{Dm}}$ is determined by the average degree $p\langle k\rangle$, rather than by $p$ or $\langle k\rangle$ separately. Thus for different $p$ and $\langle k\rangle$, $g_{\mathrm{Dm}}$ remains the same [see Figs. 3(a), 3(b), and 3(c)]. At the point where the peak appears, $p_m\langle k\rangle = C$ where $C$ is a constant. Thus the critical point $p_m$ has a reciprocal relation with $\langle k\rangle$.

We next calculate the maximum value of the minimum driver node density for each giant component. By solving Eqs. (20) and (21) where $p\langle k\rangle$ is the variable, we find that the minimum driver node density of the GWCC reaches its maximum value of $g_{\mathrm{Dm}}^{(\mathrm{w})} = 0.2888$. Thus the critical point $p_{\mathrm{m}}^{(\mathrm{w})}$ is

$$p_{\mathrm{m}}^{(\mathrm{w})} = \frac{2.2567}{\langle k\rangle}. \tag{26}$$

Accordingly, for the GIN(GOUT), by substituting the generating functions of their in-degree distribution, out-degree distribution, and the corresponding branching processes, Eqs. (22) and (24), into Eqs. (13) and (14), we can calculate the minimum driver node density for controlling the GIN(GOUT). By doing the numerical computing with $p\langle k\rangle$ changes, we find the minimum driver node density for controlling the GIN(GOUT) reaches its maximal value of $g_{\mathrm{Dm}}^{(\mathrm{in})} = g_{\mathrm{Dm}}^{(\mathrm{out})} = 0.1648$. Thus the critical points $p_{\mathrm{m}}^{(\mathrm{in})}$ and $p_{\mathrm{m}}^{(\mathrm{out})}$ are

$$p_{\mathrm{m}}^{(\mathrm{in})} = p_{\mathrm{m}}^{(\mathrm{out})} = \frac{3.2385}{\langle k\rangle}. \tag{27}$$

Similarly, the minimum driver node density for controlling the GSCC can be calculated by substituting the Eq. (25) into Eqs. (13) and (14). The minimum driver node density for controlling the GSCC reaches its maximal value of $g_{\mathrm{Dm}}^{(\mathrm{s})} = 0.0726$. Thus the critical point $p_{\mathrm{m}}^{(\mathrm{s})}$ is

$$p_{\mathrm{m}}^{(\mathrm{s})} = \frac{4.1234}{\langle k\rangle}. \tag{28}$$

In each giant component, the critical point $p_{\mathrm{m}}$ shows a reciprocal relation with $\langle k\rangle$. As shown in Fig. 3(d), $p_{\mathrm{m}}$ decreases gradually as the average degree $\langle k\rangle$ increases.

The fact that the critical fraction of remaining nodes $p_{\mathrm{m}}$ can be regarded as a reciprocal function of $\langle k\rangle$ reminds us the critical threshold $p_c = 1/\langle k\rangle$ after which the size of the giant connected component of a undirected ER network changes from zero to nonzero in a percolation attacking process (as the remaining fraction of nodes $p$ increasing from zero to one) [36]. In addition, in a system of two interdependent ER networks with the average degree $\langle k\rangle$ [24], there is also a critical threshold $p_c = 2.4554/\langle k\rangle$ where the size of the mutual giant component jumps from a nonzero to zero as $p$ decreasing in a percolation process. The critical threshold $p_c$ characterizing the system robustness has captured plenty of attention from scientists [23,27,33,37,38]. The critical fraction $p_{\mathrm{m}}$ at which the minimum driver node density required to control the giant component reaches its maximum value also characterizes the robustness of the controllability of the giant component. There are additional interesting properties yet to be uncovered.

## VI. CONTROLLABILITY OF THE GIANT COMPONENTS IN SF NETWORKS

SF networks approximate real networks such as protein-protein interaction network [39], Internet [40], and social networks [41]. The degrees of the nodes of SF networks follow a power law distribution, and there are different models of SF networks, such as the Barabáse-Albert model [1], the static model [7], and SF networks with a structural cutoff [26,33]. We study a directed SF network [26] which are characterized by a power law in-degree distribution, $P(k_{\rm in}) \sim k^{-\lambda_{\rm in}}$ with $m_{\rm in} \leqslant k \leqslant M_{\rm in}$, and a power law out-degree distribution, $P(k_{\rm out}) \sim k^{-\lambda_{\rm out}}$ with $m_{\rm out} \leqslant k \leqslant M_{\rm out}$, where $m_{\rm in}$ and $m_{\rm out}$ are the minimum in-degree and out-degree, respectively, and $M_{\rm in}$ and $M_{\rm out}$ are the maximum in-degree and out-degree respectively. The average degree of a SF network $\langle k\rangle$ changes when the minimum in-degree $m_{\rm in}$ and minimum out-degree $m_{\rm out}$ changes. Even if two networks have the same parameter $\lambda$, their average degrees are different with different minimum degree, so in this case $\lambda$ and $\langle k\rangle$ are relatively independent [42]. In addition, the analytical framework in our paper is general for all networks with arbitrary degree distributions. Here we use ER networks and SF networks [26] as examples to show the results of the framework. Like many previous work [25,26,43], this kind of SF network works well for our theory framework. Our analytical framework can also be applied to other models, for example, a static model, by substituting the degree distribution of our SF network [Eq. (29)] with the equation of the degree distribution of static model.

The in-degree and out-degree distributions of a SF network [33] are

$$P(k_{\rm in}) = \frac{(k_{\rm in}+1)^{1-\lambda_{\rm in}} - k_{\rm in}^{1-\lambda_{\rm in}}}{\left[(M_{\rm in}+1)^{1-\lambda_{\rm in}} - m_{\rm in}^{1-\lambda_{\rm in}}\right]},$$
$$P(k_{\rm out}) = \frac{(k_{\rm out}+1)^{1-\lambda_{\rm out}} - k_{\rm out}^{1-\lambda_{\rm out}}}{\left[(M_{\rm out}+1)^{1-\lambda_{\rm out}} - m_{\rm out}^{1-\lambda_{\rm out}}\right]}. \quad (29)$$

For a directed SF network where no correlation between the in- and out-degree of a given node exists, the degree distribution is defined by the generating function

$$\Phi(x,y) = \frac{\sum_{m_{\rm in}}^{M_{\rm in}}\left[(k_{\rm in}+1)^{1-\lambda_{\rm in}} - k_{\rm in}^{1-\lambda_{\rm in}}\right]x^{k_{\rm in}}}{\left[(M_{\rm in}+1)^{1-\lambda_{\rm in}} - m_{\rm in}^{1-\lambda_{\rm in}}\right]} \times \frac{\sum_{m_{\rm out}}^{M_{\rm out}}\left[(k_{\rm out}+1)^{1-\lambda_{\rm out}} - k_{\rm out}^{1-\lambda_{\rm out}}\right]y^{k_{\rm out}}}{\left[(M_{\rm out}+1)^{1-\lambda_{\rm out}} - m_{\rm out}^{1-\lambda_{\rm out}}\right]}. \quad (30)$$

After randomly removing a fraction $1-p$ nodes from the SF network, the in-degree distribution of the remaining network [34] is

$$P^{(p)}(k_{\rm in}) = \sum_{i\geqslant k_{\rm in}}^{\infty} P(k_{\rm in})\binom{i}{k} p^{k_{\rm in}}(1-p)^{i-k_{\rm in}}, \quad (31)$$

and the out-degree distribution is

$$P^{(p)}(k_{\rm out}) = \sum_{i\geqslant k_{\rm out}}^{\infty} P(k_{\rm out})\binom{i}{k} p^{k_{\rm out}}(1-p)^{i-k_{\rm out}}. \quad (32)$$

The generating function of the degree distribution of the remaining network is

$$\Phi(z_{\rm in},z_{\rm out}) = \frac{\sum_{m_{\rm in}}^{M_{\rm in}}\left[(k_{\rm in}+1)^{1-\lambda_{\rm in}} - k_{\rm in}^{1-\lambda_{\rm in}}\right]z_{\rm in}^{k_{\rm in}}}{\left[(M_{\rm in}+1)^{1-\lambda_{\rm in}} - m_{\rm in}^{1-\lambda_{\rm in}}\right]} \times \frac{\sum_{m_{\rm out}}^{M_{\rm out}}\left[(k_{\rm out}+1)^{1-\lambda_{\rm out}} - k_{\rm out}^{1-\lambda_{\rm out}}\right]z_{\rm out}^{k_{\rm out}}}{\left[(M_{\rm out}+1)^{1-\lambda_{\rm out}} - m_{\rm out}^{1-\lambda_{\rm out}}\right]}, \quad (33)$$

where $z_{\rm in} = px+1-p$ and $z_{\rm out} = py+1-p$ [32]. For simplicity and without lose of generalization, we assume that the SF network has the same in-degree distribution and out-degree distribution, i.e., $m_{\rm in} = m_{\rm out} = m$, $M_{\rm in} = M_{\rm out} = M$ and $\lambda_{\rm in} = \lambda_{\rm out} = \lambda$.

The procedure for calculating the minimum driver node densities required to control the giant connected components in SF networks is similar to that used in ER networks. The generating functions for computing the size of GWCCs are

$$\Phi^{(\rm w)}(x) = \left\{\frac{\sum_m^M[(k+1)^{1-\lambda} - k^{1-\lambda}](px+1-p)^k}{[(M+1)^{1-\lambda} - m^{1-\lambda}]}\right\}^2, \quad (34)$$

and the generating function of the branching process is

$$\Phi_1^{(\rm w)}(x) = \Phi^{(\rm w)'}(x)/\Phi^{(\rm w)'}(1). \quad (35)$$

Substituting Eqs. (34) and (35) into Eq. (2) gives us the size of the GWCC $1-h_w$. Then substituting the in-degree and out-degree distribution of the remaining part of network after an initial failure, Eqs. (31) and (32), respectively into Eq. (4) and Eq. (6) gives us the in-degree and out-degree distributions of the GWCC, which are $P^{(\rm w)}(k_{\rm in})$ and $P^{(\rm w)}(k_{\rm out})$. Substituting $P^{(\rm w)}(k_{\rm in})$ and $P^{(\rm w)}(k_{\rm out})$ into Eqs. (5) and (7) produces the generating functions of the in-degree and out-degree distributions of the GWCC of the SF network. Then the Eqs. (13) and (14) gives us the minimum driver node density to control the GWCC of a SF network $g_D^{(\rm w)}$.

Substituting the in-degree and out-degree distributions of the remaining SF network, Eqs. (31) and (32), into Eq. (9) gives us the in-degree distribution of the GOUT $P^{\rm out}(k_{\rm in})$, and substituting them into Eq. (10) gives us the out-degree distribution of the GIN $P^{\rm in}(k_{\rm out})$. Using the properties of the in-degree and out-degree distributions of the GIN, GOUT, and GSCC (as shown in Table I), with $P^{\rm out}(k_{\rm in})$, $P^{\rm in}(k_{\rm out})$, and Eqs. (31) and (32), we obtain all the in-degree and out-degree distributions of the GIN, GOUT and GSCC of a SF network. Using Eq. (1) and these in-degree and out-degree distributions, we can get their generating functions. Substituting their generating functions into Eqs. (13) and (14) gives us all the minimum driver node densities required to control the GIN, GOUT, and GSCC.

As the fraction of remaining nodes $p$ after initial node failure increases, all the minimum driver node densities required to control the GWCC, GIN, GOUT, and GSCC of a SF network first increase and then decrease, showing peaks [see Fig. 4(a)]. The mechanism that forms these peaks is the same as that found in ER networks. It is produced by the competition between the decrease of the minimum driver node density caused by the decrease of the giant component size, and the increase of the minimum driver nodes number caused

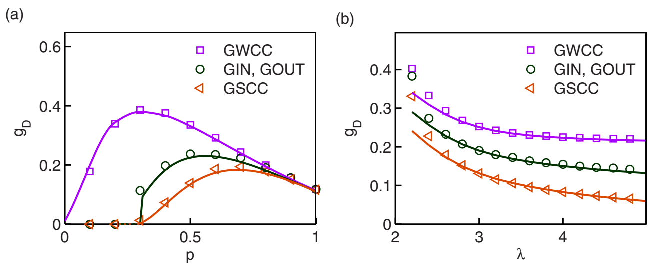


FIG. 4. The minimum driver node densities of the giant components of SF networks. (a) As the fraction of remaining nodes $p$ changes from zero to one, for each giant component, the driver node density $g_D$ increases and then decreases, displaying a peak. (b) The driver node density $g_D$ decreases constantly as the degree distribution exponent $\lambda$ increases. The solid lines are the theoretic results when $\lambda = 2.6$ for (a) and $p = 0.6$ for (b). The symbols are simulation results where each network with the number of nodes $N = 10^6$ and agree well with each other.

by the increase in giant component sparseness. We also find that the minimum driver node density of the giant component $g_D$ decreases as the degree distribution exponent $\lambda$ increases [see Fig. 4(b) and Fig. 5]. Comparing the controllability of the giant components in ER networks and SF networks, the giant components in ER networks are not always easier to control than the giant components of SF networks. For the GINs, GOUTs, and GSCCs, when the fraction of remaining nodes $p$ is low, ER networks have lower controllability than SF networks, which is similar to the findings in Ref. [22].

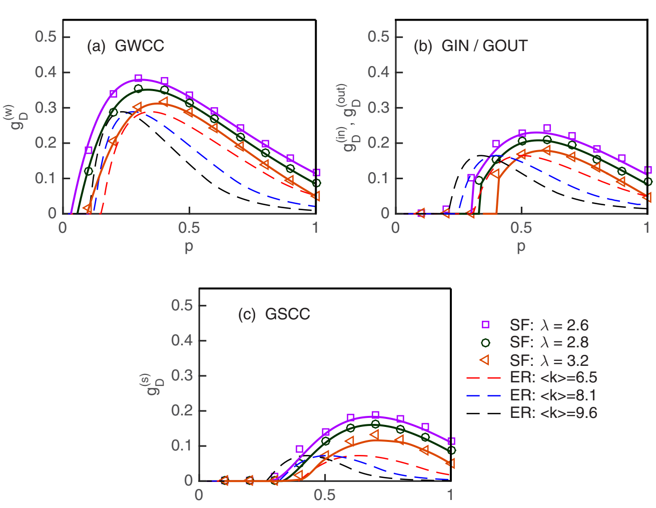


FIG. 5. The minimum driver node densities of the (a) GWCC, (b) GIN, and (c) GSCC of SF networks with different values of $\lambda$ and ER networks. For each giant component, the driver node densities first increase and then decrease as the remaining node fraction $p$ increases, showing a peak value at the critical point $p_m$. The peak values are different for different values of $\lambda$. For SF networks with different $\lambda$, the driver node densities $g_D$ for decrease as $\lambda$ increases. The average degrees of SF networks with $\lambda = 2.6$, $\lambda = 2.8$, and $\lambda = 3.2$ are respectively equal to the average degrees of ER network with $\langle k \rangle = 9.6$, $\langle k \rangle = 8.1$, and $\langle k \rangle = 6.5$. The GWCC in ER networks is easier to control than that in SF networks with the same average degree. For the GIN and GSCC, ER networks are not always easier to control than SF networks. When the fraction of remaining nodes $p$ is low, ER networks have lower controllability than SF networks.

## VII. CONTROLLABILITY OF GIANT COMPONENTS UNDER NODE CLASSIFICATION BASED ATTACK

In a network, nodes are different based on their role in maintaining controllability [7], and nodes can be classified into three different groups: critical, ordinary, and redundant. A critical, ordinary, and redundant node respectively acts as a driver node in all, some, or none of the control configurations [14]. We study how the minimum driver node density $g_D$ for controlling the giant components behaves if a fraction of only critical, ordinary, or redundant nodes being removed.

We first use the random sampling method in Ref. [44] to classify the nodes into critical, ordinary, and redundant nodes. Then in the set of critical nodes, we randomly remove $1 - p_{cr}$ fraction of nodes and compute the minimum driver node density $g_D$ that needs to control the giant components in the remaining network. As shown in Fig. 6, as the fraction of remaining critical nodes $p_{cr}$ increase, $g_D$ shows different behavior for different networks: (1) when $\langle k \rangle$ is small (for example, ER network with $\langle k \rangle = 3$), it continuously increases; (2) when $\langle k \rangle$ is high (for example, ER network with $\langle k \rangle = 6$), it continuously decreases. The behavior of $g_D$ is related to the density of the critical nodes among the whole network. When the density of critical nodes is high where $\langle k \rangle$ is small, the network is relatively sparse and the removal of critical nodes could cause a giant component being even sparser, which requires more driver node density to control a giant component. When the density of critical nodes is low where $\langle k \rangle$ is high, the removal of the critical nodes has little impact on the density of a giant component. For more critical nodes in a giant component it requires more driver nodes, so the removal of critical nodes could result in the decrease of the driver node density.

When we remove $1 - p_{re}$ fraction of redundant nodes, $g_D$ also shows different behavior when $p_{re}$ changes: (1) when the giant component is sparse (for example, ER network with $\langle k \rangle = 3$), continuously increases; (2) when the giant component is dense (for example, ER networks with $\langle k \rangle = 6$ and SF networks with $\lambda = 2.6$ whose $\langle k \rangle$ is nearly 10), first increases and then decreases; (3) when the giant component is even dense (for example, the GWCC in SF networks with $\lambda = 2.2$ and $\langle k \rangle = 17$), continuously decreases. When the density is relatively low as in the cases of the GIN and GSCC of ER networks with $\langle k \rangle = 3$, the whole network is relatively sparse and the giant components are also sparse. The removal of redundant nodes could cause the giant components to be ever sparser, which results in the increasing of the $g_D$. When the density of the redundant nodes is relatively high as the case of ER network with $\langle k \rangle = 6$. As we remove less redundant nodes, i.e. $p_{re}$ increases, the giant components become larger and $g_D$ increases. As $p_{re}$ becomes larger than a critical value $p_m$, the giant components become denser, so that $g_D$ decreases. For some highly dense networks such as the cases of SF networks with $\lambda = 2.2$ and $\lambda = 2.6$, their GWCC always exists and also dense. More redundant nodes remain, and the GWCC is even denser, so that less $g_D$ is needed. That is, $g_D$ continuously decreases. For the case of removing $1 - p_{or}$ ordinary nodes,

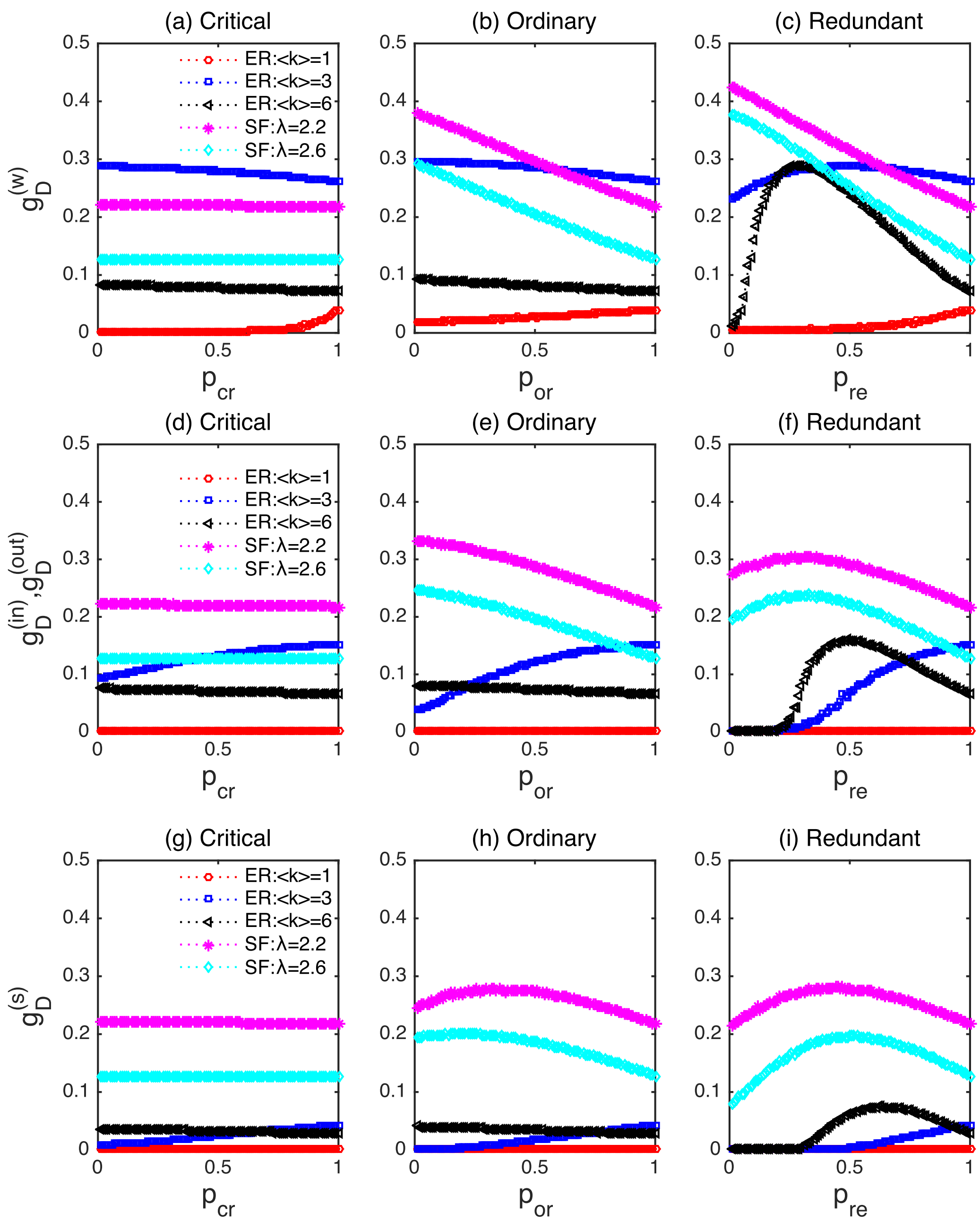


FIG. 6. Critical, redundant, and ordinary nodes based attack. Panels (a), (b), and (c) respectively show how the minimum driver node density $g_D^{(w)}$ required to control the GWCC changes when $1 - p_{cr}$ fraction of critical nodes, $1 - p_{or}$ fraction of ordinary nodes, and $1 - p_{re}$ redundant nodes are removed from the network. Panels (d), (e), and (f) show the behavior of the minimum driver node density $g_D^{(\text{in})}$ or $g_D^{(\text{out})}$ required to control the GIN or GOUT. Panels (g), (h), and (i) show the behavior of the minimum driver node density $g_D^{(s)}$ required to control the GSCC. When we only remove critical nodes, $g_D$ may continuously increase, or continuously decrease for different densities of critical nodes among the network. When only redundant or ordinary nodes removed, $g_D$ may continuously increase, continuously decrease, or first increase and then decrease for different densities of redundant and ordinary nodes among the whole network.

$g_D$ also shows different behaviors that are similar to the case of redundant nodes removal, and different behaviors are also related to the density of the ordinary nodes.

For the critical, redundant, and ordinary nodes base attack, the minimum driver node density $g_D$ shows different behaviors in different networks. The appearances of different behaviors are related to different densities of critical, redundant, and ordinary nodes among networks. In the future, our work can be extended to other modes of attack, for example, different centralities (degree, betweenness, closeness, etc.) -based attack.

## VIII. CONCLUSION

We have studied the controllability of four types of giant components in directed networks with arbitrary joint degree distributions. We first present our method for deriving the in-degree and the out-degree distributions of the four giant components: GWCCs, GINs, GOUTs, and GSCCs, and we then use structural controllability theory to calculate the minimum driver node density $g_{\mathrm{D}}$ required to control these giant components.

We find that for both ER networks and SF networks with $p$ fraction of remaining nodes, the minimum driver node density to control the giant component first increases and then decreases as $p$ increases from zero to one, showing a peak at a critical point $p = p_m$. Especially, for ER networks, the peak values of the driver node density remain the same regardless of its average degree $\langle k \rangle$, and the products of the critical point and average degrees are the newly found constant values: $p_m^{(w)}\langle k \rangle = 2.2567$ for the GWCC, $p_m^{(\mathrm{in})}\langle k \rangle = p_m^{(\mathrm{out})}\langle k \rangle = 3.2385$ for the GIN and GOUT, and $p_m^{(s)}\langle k \rangle = 4.1234$ for the GSCC. These constant values are important for understanding the controllability of the giant components under node failures. Similar to two important constant values in the percolation of complex networks: (1) in a single ER network, the critical fraction of remaining nodes $p_c$ satisfies that $\langle k \rangle p_c = 1$ [27]; (2) in two fully interdependent networks, it satisfies $\langle k \rangle p_c = 2.4554$ [24]. These two constant values are very important in understanding the robustness of complex networks. In addition, we find that the giant components of ER networks are not always easier to control than the giant components of SF networks. For the GIN and GSCC, when the fraction of remaining nodes $p$ is low, ER networks have lower controllability than SF networks.

Our results suggest questions for further studies (1) How can we interpret the physical properties of the critical point $p_{\mathrm{m}}$ at which the minimum driver node density required to control the giant component reaches its maximum value? (2) What properties do the driver nodes for controlling giant components have in such real-world systems as gene regulatory networks and social networks? (3) How can we quantify the controllability of a giant component with nonlinear dynamics? Understanding these questions would significantly improve our understanding of the control principles in complex systems.

## ACKNOWLEDGMENTS

We acknowledge the support of NSF Grants PHY-1505000 and CMMI 1125290; NIH Grant 5R01AG021133; DOD Grant HDTRA1-14-1-0017; ONR Grant N00014-14-1-0738; DOE Grant DE-AC07-051D14517; Army Research Laboratory Cooperative Agreement No. W911NF-09-2-0053 (ARL Network Science CTA); European Commission Grants FP7317532 (MULTIPLEX) and 641191 (CIMPLEX); National Natural Science Foundation of China Grants 61320106005, 91530320, and 61374160; and the Innovation Scientists and Technicians Troop Construction Projects of Henan Province (154200510012). We thank Yang-Yu Liu for discussions.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **407**, 651 (2000).
[3] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[4] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[5] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, Cambridge, 2010).
[6] S. N. Dorogovtsev and J. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2013).
[7] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, Nature (London) **473**, 167 (2011).
[8] J.-J. E. Slotine, W. Li *et al.*, *Applied Nonlinear Control*, Vol. 199 (Prentice-Hall, Englewood Cliffs, NJ, 1991 ).
[9] C. T. Lin, IEEE Trans. Automat. Control **19**, 201 (1974).
[10] R. E. Kalman, J. Soc. Indus. Appl. Math. A **1**, 152 (1963).
[11] M. Mézard and G. Parisi, Eur. Phys. J. B **20**, 217 (2001).
[12] Y. Liu, J. Slotine, and A. Barabási, PloS One **7**, e44459 (2011).
[13] G. Yan, J. Ren, Y.-C. Lai, C.-H. Lai, and B. Li, Phys. Rev. Lett. **108**, 218703 (2012).
[14] T. Jia, Y.-Y. Liu, E. Csóka, M. Pósfai, J.-J. Slotine, and A.-L. Barabási, Nat. Commun. **4**, 2002 (2013).
[15] S. P. Cornelius, W. L. Kath, and A. E. Motter, Nat. Commun. **4**, 1942 (2013).
[16] J. Ruths and D. Ruths, Science **343**, 1373 (2014).
[17] X. Liu and L. Pan, BMC Syst. Biol. **8**, 51 (2014).
[18] G. Yan, G. Tsekenis, B. Barzel, J.-J. Slotine, Y.-Y. Liu, and A.-L. Barabási, Nat. Phys. **11**, 779 (2015).
[19] R. Badhwar and G. Bagler, PloS One **10**, e0139204 (2015).
[20] X. Liu, H. Peng, and J. Gao, Chaos Solitions Fractals **80**, 125 (2015).
[21] J. A. Stankovic, M. Spuri, K. Ramamritham, and G. C. Buttazzo, in *Deadline Scheduling for Real-Time Systems* (Springer, New York, 1998), pp. 1–11.
[22] J. Gao, Y.-Y. Liu, R. M. D'Souza, and A.-L. Barabási, Nat. Commun. **5**, 5415 (2014).
[23] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **64**, 025101 (2001).
[24] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, Nature (London) **464**, 1025 (2010).
[25] J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, Phys. Rev. Lett. **107**, 195701 (2011).
[26] X. Liu, H. E. Stanley, and J. Gao, Proc. Natl. Acad. Sci. USA **113**, 1138 (2016).
[27] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[28] N. Schwartz, R. Cohen, D. ben-Avraham, A.-L. Barabási, and S. Havlin, Phys. Rev. E **66**, 015104 (2002).
[29] H. Peng, D. Zhao, X. Liu, and J. Gao, PloS One **10**, e0144153 (2015).
[30] J. Gao, B. Barzel, and A.-L. Barabási, Nature (London) **530**, 307 (2016).
[31] P. Bialas and A. K. Oleś, Phys. Rev. E **77**, 036124 (2008).
[32] M. P. Stumpf and C. Wiuf, Phys. Rev. E **72**, 036118 (2005).
[33] J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, Phys. Rev. E **85**, 066134 (2012).
[34] M. P. Stumpf, C. Wiuf, and R. M. May, Proc. Natl. Acad. Sci. USA **102**, 4221 (2005).

[35] A. Lombardi and M. Hörnquist, Phys. Rev. E **75**, 056110 (2007).
[36] M. Molloy and B. A. Reed, Random Struct. Algor. **6**, 161 (1995).
[37] M. Molloy and B. Reed, Combin. Probab. Comput. **7**, 295 (1998).
[38] R. Parshani, S. V. Buldyrev, and S. Havlin, Phys. Rev. Lett. **105**, 048701 (2010).
[39] J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou *et al.*, Nature (London) **437**, 1173 (2005).
[40] A.-L. Barabási *et al.*, Science **325**, 412 (2009).
[41] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, Science **323**, 892 (2009).
[42] D. Zhou, J. Gao, H. E. Stanley, and S. Havlin, Phys. Rev. E **87**, 052812 (2013).
[43] X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, Phys. Rev. E **83**, 065101 (2011).
[44] T. Jia and A.-L. Barabási, Sci. Rep. **3**, 2354 (2013).