Betweenness centrality in a weighted network

Huijuan Wang, Javier Martin Hernandez, and Piet Van Mieghem

Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands

When transport in networks follows the shortest paths, the union of all shortest path trees $G_{\cup SPT}$ can be regarded as the "transport overlay network." Overlay networks such as peer-to-peer networks or virtual private networks can be considered as a subgraph of $G_{\cup SPT}$. The traffic through the network is examined by the betweenness B_l of links in the overlay $G_{\cup SPT}$. The strength of disorder can be controlled by, e.g., tuning the extreme value index α of the independent and identically distributed polynomial link weights. In the strong disorder limit ($\alpha \to 0$), all transport flows over a critical backbone, the minimum spanning tree (MST). We investigate the betweenness distributions of wide classes of trees, such as the MST of those well-known network models and of various real-world complex networks. All these trees with different degree distributions (e.g., uniform, exponential, or power law) are found to possess a power law betweenness distribution $\Pr[B_l = j] \sim j^{-c}$. The exponent c seems to be positively correlated with the degree variance of the tree and to be insensitive of the size N of a network. In the weak disorder regime, transport in the network traverses many links. We show that a link with smaller link weight tends to carry more traffic. This negative correlation between link weight and betweenness depends on α and the structure of the underlying topology.

I. INTRODUCTION

Routing in communication networks is based on shortest paths [or the best approximation due to, e.g., the distracting influence of border gateway protocol (BGP)] between any two nodes of the network. The resources of a network are most efficiently used when traffic follows shortest paths [1]. In large complex networks, not all links have equal importance. For example, if two clusters are connected by one link, the removal of this link will disable all the traffic flowing between these two clusters. In contrast, the removal of a link connecting to a dead end whose degree is one, will have no effect on the other parts of the network. The importance of links is of primary interest for network resilience to attacks [2,3] and immunization against epidemics [4]. A good measure for "link or node importance" is the betweenness $B_l(B_n)$ of a link (node), which is defined as the number of shortest paths between all possible pairs of nodes in the network that traverse the link (node). The betweenness $B_l(B_n)$ which incorporates global information is a simplified quantity to assess the maximum possible traffic. Assuming that a unit packet is transmitted between each node pair, the betweenness B_l is the total amount of packets passing through a link.

The overlay $G_{\cup SPT}$, as shown in Fig. 1, is the union of the shortest paths between all possible node pairs, and it can be regarded as the "transport overlay network" on top of the underlying network topology or substrate. The overlay $G_{\cup SPT}$, which is a subgraph of the substrate in a weighted graph, determines the network's performance: any link removed in $G_{\cup SPT}$ will definitely impact at least those flows of traffic that pass over that link. Since all the traffic traverses only the overlay $G_{\cup SPT}$ and all the nodes in the substrate also appear in the overlay $G_{\cup SPT}$, the betweenness of a node in the substrate is equal to the betweenness of that node in the overlay $G_{\cup SPT}$. A link in the substrate has betweenness 0 if it does not belong to the overlay $G_{\cup SPT}$. Otherwise, its link

betweenness is the same as that in the overlay $G_{\cup SPT}$.

In this paper, we study the link betweenness of the overlay $G_{\cup SPT}$. The study of betweenness usually deals with scale-free trees [5–8] or scale-free networks [9,10] whose degree distribution follow a power law, i.e., $\Pr[D=k] \sim k^{-\gamma}$. However, the overlay $G_{\cup SPT}$ that we are going to examine possesses different degree distribution, e.g., uniform, exponential or power law distribution. The structure of the overlay network $G_{\cup SPT}$ can be controlled, e.g., by tuning the extreme value index α of the independent and identically distributed (IID) polynomial link weights [11]. In the strong disorder limit ($\alpha \to 0$), the overlay $G_{\cup SPT}(\alpha \to 0)$ becomes the minimum spanning tree (MST), a tree which has the minimum total weight of all possible spanning trees. The betweenness of the MST for various network models and real-world complex networks are surprisingly found to follow a power law. This power law betweenness distribution for MST holds more generally than in Erdös-Rényi random graph and scale-free networks as found in Ref. [12]. In addition, the relationship between the structural characteristics and its betweenness distribution is investigated. We study the correlation between the link weights and the corresponding link betweenness when the system is in weak disorder, in-
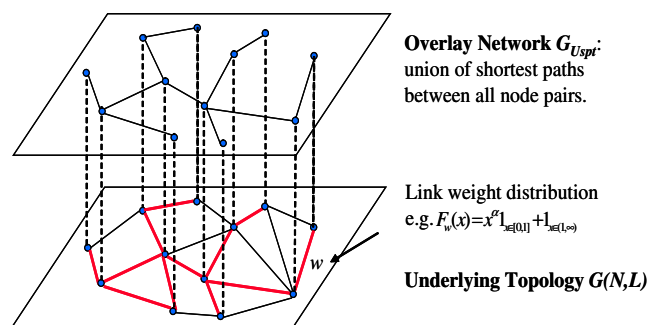

FIG. 1. (Color online) The overlay network $G_{\cup SPT}$.

stead of the correlation between the node betweenness and the node degree as in Refs. [12,13].

In Sec. II, we explain the notions of structural changes in the overlay $G_{\cup SPT}(\alpha)$ as we tune the extreme value index α. Simulation scenarios are mentioned. The correlation between link weight and its betweenness is investigated in Sec. III. Furthermore, the link betweenness distribution of the overlay $G_{\cup SPT}$ that characterizes the traffic distribution is examined in Sec. IV. If $\alpha \to 0$, the overlay $G_{\cup SPT}$ becomes the MST. The link betweenness of such overlay trees on top of network models as well as real-world networks are compared together with other classes of trees in Sec. V. Finally, our results are summarized in Sec. VI.

II. BASIC NOTIONS AND SIMULATION SCENARIOS

We restrict ourselves to additive link weights and nondirected graphs. Hence, the shortest path between two nodes is the path that minimizes the sum of the weights along the path. Since the shortest path (SP) is mainly sensitive to the smaller, non-negative link weights, the simplest distribution of the link weight w with a distinct different behavior for small values than a regular distribution (Ref. [14], Chap. 16) is the polynomial distribution

$$F_w(x) = x^\alpha 1_{x \in [0,1]} + 1_{x \in (1,\infty)}, \quad \alpha > 0, \tag{1}$$

where the indicator function 1_x is one if x is true else it is zero. The corresponding density is $f_w(x) = \alpha x^{\alpha-1}$, $0 \le x \le 1$. The exponent

$$\alpha = \lim_{x \downarrow 0} \frac{\ln F_w(x)}{\ln x}$$

is called the extreme value index of the probability distribution of w and $\alpha = 1$ for uniform distributions. The link weights in a network are IID according to Eq. (1).

In Ref. [11], a transition is observed around a critical extreme value index α_c [15], that is defined by $\Pr[G_{U \ SPT(\alpha=\alpha_c)} = MST] = \frac{1}{2}$: When $\alpha \to 0$ (or in the $\alpha < \alpha_c$ regime for large networks), all flows are transported over the MST. Hence, $G_{\cup SPT(\alpha \to 0)} = MST$ is also called an overlay tree. When $\alpha > \alpha_c$, transport in the network traverses many links. The $\alpha \to 0$ (or $\alpha < \alpha_c$ for large networks) regime corresponds to the strong disorder limit, where the total weight of a path is characterized by the maximum link weight along the path. The shortest path in this case is the path with the minimum value of the maximum link weight. When all links contributes to the total weight of the shortest path, the system is weak disordered, e.g., $\alpha > \alpha_c$. In fact, other distributions that could lead to strong disorder [16] would arrive at similar betweenness behavior, because the MST is probabilistically the same for various IID link weights distributions [17].

For the underlying topology, called the substrate, we consider the following complex network models: the Erdös-Rényi random graph $G_p(N)$, the square and the cubic lattice, and the Barabási-Albert (BA) power law model [18]. Traditionally, complex networks have been modeled as Erdös-Rényi random graphs $G_p(N)$, which can be generated from a set of N nodes by randomly assigning a link with probability
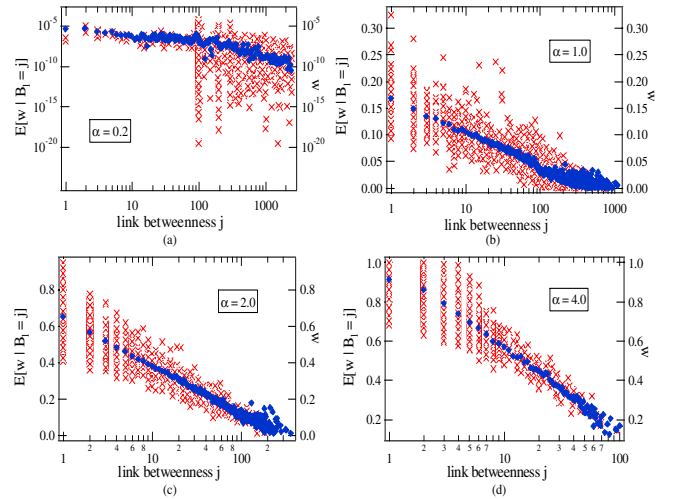

FIG. 2. (Color online) The link weight w (cross) versus its link betweenness j and $E[w|B_l=j]$ (square) the average link weight of links with given betweenness j in the overlay $G_{\cup SPT}$ on top of Erdös-Rényi random graph $G_{0.4}(100)$.

p to each pair of nodes. In addition to their analytic tractability, the Erdös-Rényi random graphs are reasonably accurate models for peer-to-peer networks and *ad hoc* networks. The square lattice, in which each node has four neighbors, is the basic model of a transport network (Manhattan grid) as well as in percolation theory [19] and is frequently used to study the network traffic [20]. The power law degree distribution is followed by many natural and artificial networks such as the scientific collaborations [21], the world-wide web, and the Internet [22].

We carried out 10^4 iterations for each simulation. Within each iteration, we randomly generate an underlying topology. Polynomial link weights with parameter α are assigned independently to each link. The overlay $G_{\cup SPT}$ as well as its betweenness is found by calculating the shortest paths between all node pairs with Dijkstra's algorithm [23] for weak disorder regime. For the strong disorder limit $\alpha \to 0$, $G_{\cup SPT} = MST$ is found by Kruskal's algorithm [24] on the corresponding network with uniform link weights, because with IID link weights, the structure of the MST is probabilistically the same for various link weights distributions [17].

III. LINK WEIGHT VERSUS LINK BETWEENNESS

Does a lower link weight implies a high link betweenness B_l? When polynomial link weights are independently assigned to links in the substrate, we randomly choose a link in each overlay network $G_{\cup SPT}$. The betweenness of this link and the corresponding link weight are plotted in Fig. 2. According to Ref. [11], $\alpha_c = 0.2$ [15] for Erdös-Rényi random graph with $N=100$ nodes. When the system is weakly disordered, i.e., $\alpha > \alpha_c$ [Figs. 2(b)–2(d)], a link with lower link weight is more likely to have higher betweenness. However, when $\alpha = 0.2$ [Fig. 2(a)], where link weights possess relatively strong fluctuations, the correlation between link weight and betweenness disappears. Hence, a negative correlation exists between the link weight and its betweenness

TABLE I. The correlation coefficient between weight and betweenness of a link.

α	0.2	1.0	2.0	4.0	8.0	16.0
$G_{\cup SPT}$ on $G_{0.4}(100)$	−0.06	−0.61	−0.70	−0.78	−0.84	−0.84
$G_{\cup SPT}$ on square lattice $N=100$	−0.22	−0.53	−0.54	−0.53	−0.53	−0.53
$G_{\cup SPT}$ on cubic lattice $N=125$	−0.18	−0.60	−0.66	−0.67	−0.68	−0.68
$G_{\cup SPT}$ on BA $N=100, m=3$	−0.12	−0.53	−0.66	−0.60	−0.50	−0.49

for the weak disorder regime. The correlation becomes stronger as α increases, as illustrated in Table I where the linear correlation coefficient is equal to the covariance between the two random variables divided by the product of their standard deviations. The increasing strength of the correlation for larger α is also reflected by Fig. 2, where as α increases, the plot of link weights become narrower.

The correlation between the weight and the betweenness of the link is shown to be dependent on the underlying graph as well as on the extreme value index α of the polynomial link weight distribution. For homogeneous network such as the Erdös-Rényi random graph and lattice, the correlation coefficient increases monotonically as α increases. However, in a the nonhomogeneous topology such as the BA power law substrate, the correlation coefficient decreases after a maximum has been reached. In a homogeneous network, when α is large, a link with lower link weight tends to attract more traffic. While in a nonhomogeneous topology, the relative importance of a link or its connectivity in substrate is also an determinant factor for its betweenness. In short, both the nonhomogeneity of the underlying topology and the link weight disorder (e.g., a smaller α) contribute to the nonhomogeneity of the overlay $G_{\cup SPT}$, which reduces the correlation between link weight and betweenness.

IV. LINK BETWEENNESS DISTRIBUTION OF OVERLAY $G_{\cup SPT}$

The link betweenness represents the total traffic passing through a link if a unit packet is transmitted between each node pair. Hence, the link betweenness distribution reflects how the traffic is distributed over the network.

A. Overlay $G_{\cup SPT}$ on top of complex network models

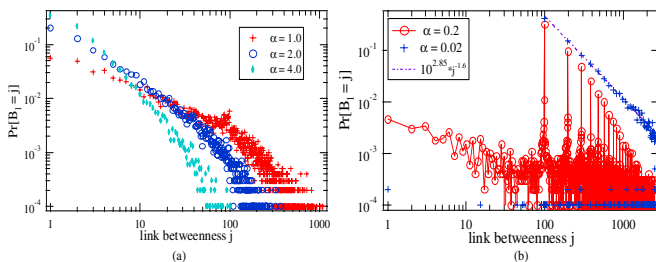As shown in Fig. 3(a), the traffic on the overlay $G_{\cup SPT}$ on

FIG. 3. (Color online) The probability density function (PDF) of link betweenness B_l in the overlay $G_{\cup SPT}$ on top of $G_{0.4}(100)$. The PDF for $\alpha=0.02$ is linear fitted by the dashed line.

top of $G_{0.4}(100)$ varies less for large α, because the betweenness is distributed within a small range. When α is small, as shown in Fig. 3(b), the betweenness is ranging between approximately $1-2500$ for $N=100$ and peaks appear on the betweenness at $n(N-n)$, where $1 \leq n \leq N-1$.

A link is called critical if its removal will disconnect the overlay $G_{\cup SPT}$ into two clusters with n and $N-n$ nodes. The betweenness of such critical link is $n(N-n)$, because all the traffic with source and destination separated in these two clusters will traverse this link. However, if a link has betweenness $n(N-n)$, the removal of this link does not necessarily disconnect the overlay graph.

As we decrease the extreme value index α, the overlay $G_{\cup SPT}$ contains less links and it becomes tree-like or even an exact tree. Any link in a tree is critical. We consider, for example, the Erdös-Rényi random graph $G_{0.4}(100)$. When $\alpha=0.2$, the average number of links in the overlay is 107.2. Within such a sparse overlay topology, a link is very likely to be critical, which contributes to the peaks in Fig. 3(b). A sparse overlay $G_{\cup SPT}$ is composed of the minimum spanning tree and few shortcuts, that direct a small part of the traffic. The largest link betweenness 2500 comes from the critical link which could separate the overlay network into two clusters each with $\frac{N}{2}=50$ nodes. A link has higher betweenness if it is critical and the maximal link betweenness is achieved when $n=\left[\frac{N}{2}\right]$. Hence, the betweenness of any link in a graph with N nodes obeys

$$B_l \leq \left\lfloor \frac{N}{2} \right\rfloor \left(N - \left\lfloor \frac{N}{2} \right\rfloor \right). \tag{2}$$

When the overlay becomes a tree, the magnitude of peaks at $n(N-n)$ also depends on the structure of the tree. For example, if the overlay network is a star with N nodes, the link betweenness is always $N-1$. And if $G_{\cup SPT}$ is a line graph, the betweenness of a link is $n(N-n)$ with n uniformly distributed over $[1, N-1]$. We find that the betweenness distribution of the overlay tree $G_{\cup SPT(\alpha \to 0)}$ on top of the Erdös-Rényi random graph $G_p(N)$ follows a power law

$$\Pr[B_l=j] = c_0 j^{-c}, \quad N-1 \leq j \leq \left\lfloor \frac{N}{2} \right\rfloor \left(N - \left\lfloor \frac{N}{2} \right\rfloor \right) \tag{3}$$

with exponent $c=1.6$. Further, we observe that the overlay tree $G_{\cup SPT(\alpha \to 0)}$ on top of other complex network models such as the lattice, cubic lattice or a BA model also seems to possess a power law betweenness distribution as illustrated in Fig. 4. The lower bound $N-1$ of the betweenness in a tree is attained at a link connected to a degree 1 node while the upper bound obeys Eq. (2). The exponent c we found for
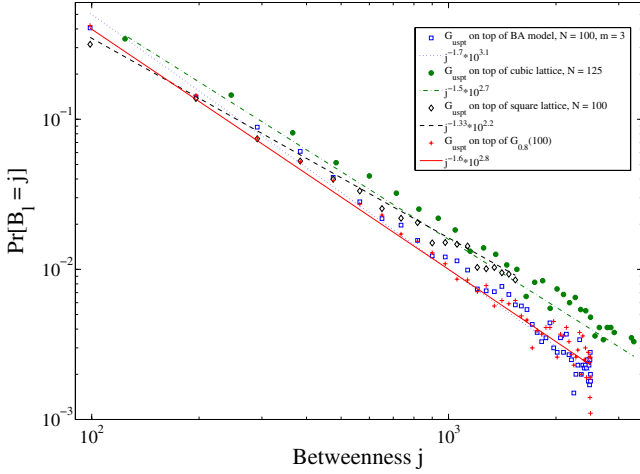
FIG. 4. (Color online) Link betweenness distribution (markers) of overlay tree $G_{\cup SPT(\alpha \to 0)}$ on top of complex network models and the corresponding linear fitting (dashed lines).

FIG. 5. (Color online) Betweenness distribution (+) of $G_{\cup SPT(\alpha \to 0)}$ on top of real network topologies. The line is the linear curve fitting.

Erdös-Rényi ($c=1.6$) lattice ($c=1.33$) and BA model ($c=1.7$) with $N \sim 100$ are the same as observed in Ref. [25] with $N \sim 8100$. The scaling exponent c seems insensitive to the size N of the network. Additional simulations for Erdös-Rényi random graph suggest that the exponent c is independent of the size N of the underlying graph as well as the link density p, if p is larger than the disconnectivity threshold $p_c \sim \ln N/N$. For example, the power exponent $c=1.6$ remains the same for the substrate $G_{0.4}(100)$, $G_{0.4}(50)$, $G_{0.8}(100)$ and the Erdös-Rényi random graph in [12] with $N=10^4$ nodes and $L=2N$ links.

B. Overlay tree $G_{\cup SPT(\alpha \to 0)}$ on top of real networks

As found in Sec. IV A and Fig. 4, an overlay tree $G_{\cup SPT(\alpha \to 0)}$ follow a power law betweenness distribution when the substrate is an Erdös-Rényi random graph, a square or cubic lattice or a BA power law graph. It would be especially interesting to examine whether the power law link betweenness distribution still holds for overlay trees $G_{\cup SPT(\alpha \to 0)}$ on top of real-world networks. Hence, we perform a statistical analysis of real data sets, representing the topology of different real-world networks. On top of each, usually large network, 100 realizations of IID uniform link weights assignments are carried out. Within each realization, the MST, equivalent to the overlay tree $G_{\cup SPT(\alpha \to 0)}$, is found with the Kruskal algorithm [24]. The complex networks come from a wide range of systems in nature and society: The Internet network at the level of autonomous systems [26]; the Gnutella [27] snapshots (Crawl2) retrieved from firewire.com; the air transportation network representing the world wide airport connections, documented at the Bureau of Transportation Statistics (http://www.bts.gov) database; the Western States Power Grid of the United States [28]; the coauthorship network [29] between scientists posting preprints on the High-Energy Theory E-Print Archive between Jan 1, 1995 and December 31, 1999; two citation networks [30] created using the Web of Science database (Kohonen and SciMet); the coauthorship network [31] of scientists
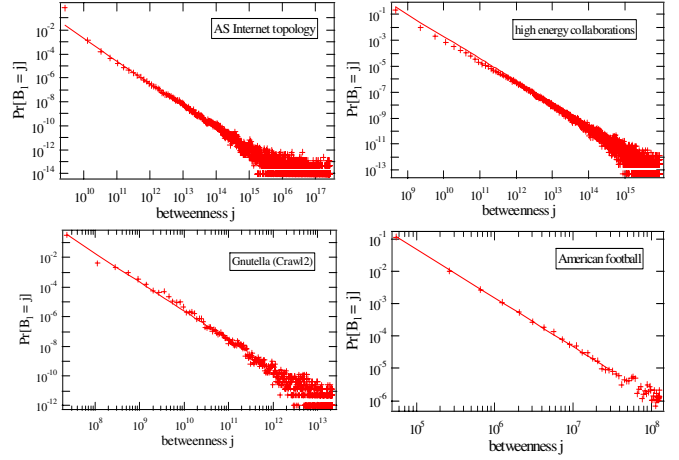
working on network theory and experiment; the network representing soccer players association to Dutch soccer teams [32]; the network of American football games between division IA colleges during regular season Fall 2000 [33]; and the adjacency network [34] of common adjectives and nouns in the novel *David Copperfield* by Charles Dickens. As shown in Fig. 5 as well as Figs. 6 and 7, the betweenness distribution of these overlay trees on top of real networks follows, surprisingly, for almost all a power law, while their corresponding degree distribution of the tree (see Figs. 8–10) may differ significantly.

The power law betweenness distribution of the overlay tree $G_{\cup SPT(\alpha \to 0)}$ or MST implies that a set of links in the MST possess a much higher betweenness. In Ref. [25], it is found that the infinite incipient percolation cluster (IIC), a subgraph of the MST has a significantly higher average betweenness than the entire MST, and the betweenness distribution of the IIC also satisfies a power law. But why does the betweenness distribution of a MST follow a power law? Is that due to the network topology, a particular link weight
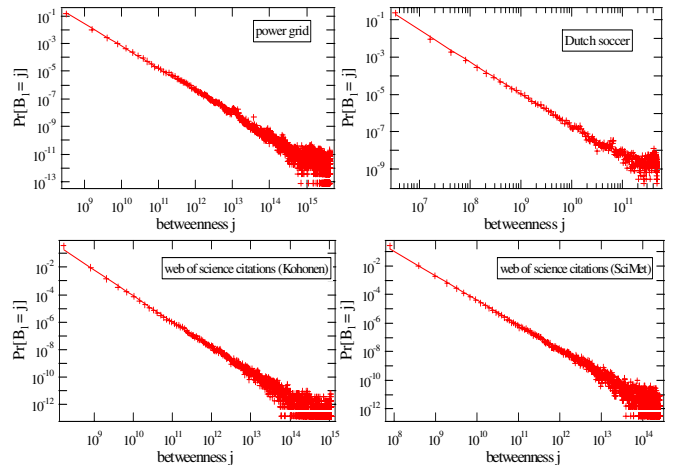

FIG. 6. (Color online) Betweenness distribution (+) of $G_{\cup SPT(\alpha \to 0)}$ on top of real network topologies. The line is the linear curve fitting.
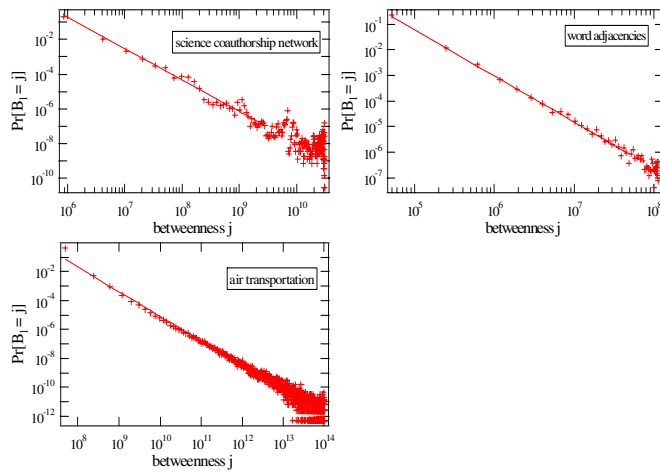
FIG. 7. (Color online) Betweenness distribution (+) of $G_{\cup SPT(\alpha \to 0)}$ on top of real network topologies. The line is the linear curve fitting.

distribution function or the fact that link weights are independently and identically distributed? The betweenness of the overlay tree follows a power law distribution no matter the substrate is a traditional complex network model or a real network, provided the substrate is denser than a tree. When the substrate is close to a tree, the overlay tree is almost the same as the substrate and the corresponding betweenness distribution does not necessarily follow a power law. Hence, the power law betweenness distribution does not hold for any tree structure but seems to hold for the overlay tree $G_{\cup SPT(\alpha \to 0)}$ on top of a substrate which is not too sparse. With IID link weights, the structure of the overlay tree or MST is probabilistically the same for various link weight distributions, because the ranking of the link weights suffices to construct the MST. Therefore, the IID link weights compared to the network topology and link weight distribution, contribute more to the power law betweenness distribution of the MST for various networks. In fact, with IID link weights, the equivalent Kruskal growth process of the MST starts from N individual nodes and in each step an arbitrary link in the substrate is added while links generating loops are forbidden. However, the power exponent c of the betweenness
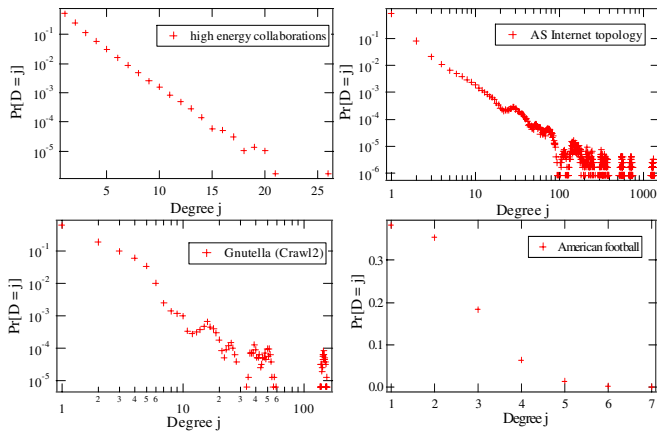
distribution of a MST is determined by the network topology, due to the exclusion of links that generating loops in the growth process of the MST. The relationship between the topological characteristics of a network and the exponent c of the betweenness distribution of the corresponding MST is studied in Sec. V B.

V. BETWEENNESS DISTRIBUTION OF TREES

Since the path between each node pair is unique in a tree and is independent of link weights, the betweenness of a tree depends purely on its tree structure. In the strong disorder limit ($\alpha \to 0$), the betweenness distribution depends on the structure of $G_{\cup SPT(\alpha \to 0)}$ or MST. In this way, we are able to compare the tree structure of overlay $G_{\cup SPT(\alpha \to 0)}$ to other classes of trees via the link betweenness distribution. Although trees are special graphs, real-world networks such as the autonomous systems in the Internet [35] can be modeled by trees or treelike graphs with a negligible number of shortcuts.
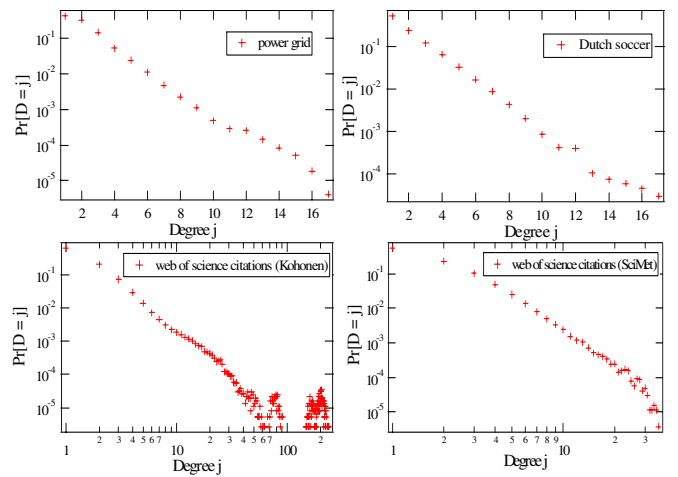

FIG. 9. (Color online) Degree distribution of $G_{\cup SPT(\alpha < \alpha_c)}$ on top of real network topologies.

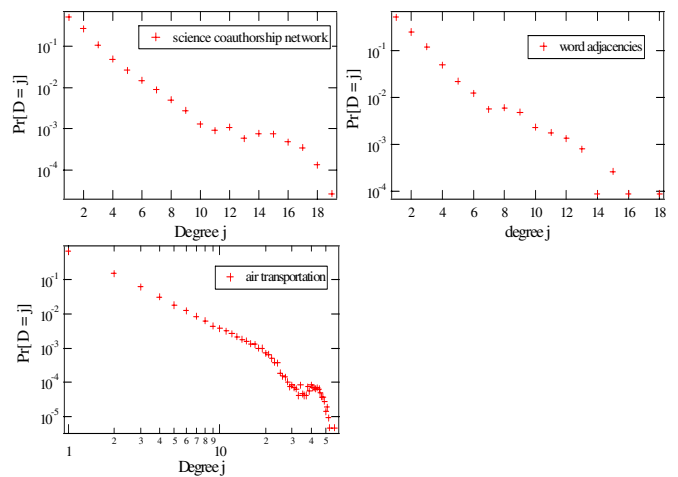FIG. 8. (Color online) Degree distribution of $G_{\cup SPT(\alpha \to 0)}$ on top of real network topologies.

FIG. 10. (Color online) Degree distribution of $G_{\cup SPT(\alpha \to 0)}$ on top of real network topologies.

In this section we compare the following trees. (a) Three tree models: the k-ary tree, the scale-free trees, and the uniform recursive tree URT. (b) The overlay tree $G_{\cup SPT(\alpha \to 0)}$ on top of complex network models: the Erdös-Rényi random graph, the square or cubic lattice, and the BA power law model. (c) The overlay tree $G_{\cup SPT(\alpha \to 0)}$ on top of real complex networks. The class (b) and (c) have been shown to possess power law betweenness distribution. Hence, it is interesting to first examine whether the class (a) has such power law betweenness distribution.

A link l in any tree connects two clusters with size $1 \le |C_l| \le \frac{N}{2}$ and $N - |C_l|$. The betweenness of a link l is $B_l = |C_l|(N - |C_l|)$, because traffic traverses the link l if and only if the source and destination lie in the two clusters separated by l. If $|C_l| = o(N)$, which holds for all but a few large clusters, then we have $B_l \sim |C_l| \cdot N$ for large N.

A. Betweenness distribution of tree models

1. k-ary tree

We investigate the k-ary tree [14] of depth [36] D, where each node has exactly k children. In a k-ary tree the total number of nodes is

$$N(D) = 1 + k + k^2 + \cdots + k^D = \begin{cases} \dfrac{k^{D+1} - 1}{k - 1}, & k \ne 1, \\ 1 + D, & k = 1. \end{cases}$$

A link is called the jth level link if it connects two nodes which is j and $j-1$ hops away from the root. The removal of a jth level link disconnect the graph into two clusters: one is a k-ary tree of depth $D-j$ with $N(D-j)$ nodes and the other cluster has $N(D) - N(D-j)$ nodes. Since there are k^j jth ($1 \le j \le D$) level links

$$\Pr[|C_l| = N(D - j)] = \frac{k^j}{N(D) - 1}.$$

Hence,

$$\Pr[|C_l| = n] = \frac{k^{D+1}}{(N(D) - 1)(kn - n + 1)},$$

$$n = N(D - j) \text{ and } 1 \le j \le D.$$

The approximate betweenness distribution

$$\Pr[B_l \sim n \cdot N] = \frac{k^{D+1}}{[N(D) - 1](kn - n + 1)},$$

$$n = N(D - j) \text{ and } 1 \le j \le D$$

follows an inverse power law with exponent $c = 1$. Two exceptions are the line graph, where $k = 1, \Pr[B_l = n(N-n)] = \frac{1}{N-1}, 1 \le n \le N - 1$ and the star where $k = N - 1, \Pr[B_l = N - 1] = 1$. A rigorous analysis based on

$$\Pr\{B_l = N(D - j)[N(D) - N(D - j)]\} = \frac{k^j}{N(D) - 1}, \quad 1 \le j \le D \tag{4}$$

is given in Appendix B.

2. Scale-free trees

A scale-free tree contains initially only one node, the root. Then, at each step a new node is attached to one of the existing node. The probability that a new node connects to a certain existing node is proportional to the attractiveness of the old node, defined as

$$A(v) = a + q,$$

where $a > 0$ denotes the initial attractiveness and q is the in-degree of node v, the number of links connected to the node. The corresponding in-degree distribution [37] is

$$\Pr[D_{in} = q] = (q + a)^{-(2+a)}.$$

Early in 2002, the power law betweenness distribution with $c = 2$ for the scale-free trees is solved analytically by Goh *et al.* [38]. Here, we relate the betweenness distribution to the subtree size distribution, which is derived by Fekete and Vattay [5]. In our notation, the probability distribution of the size of a subtree rooted at a random node in a scale-free tree with N nodes is

$$\Pr[|\mathcal{T}^{(N)}| = k] = \frac{N - \beta}{N - 1} \frac{1 - \beta}{(k - \beta)(k + 1 - \beta)}, \tag{5}$$

$$\approx (1 - \beta)\frac{1}{k^2}, \tag{6}$$

where $\beta = \frac{1}{1+a} \in [0, 1]$. When $\beta = \frac{1}{2}$, the scale-free tree is exactly the BA tree, with $m = 1$ in the BA model. When $\beta = 0$, the tree becomes a uniform recursive tree URT. Hence, the probability that a link has load approximately kN will be

$$\Pr[B_l = kN] \approx (1 - \beta)\frac{1}{k^2}.$$

The inverse square power law betweenness distribution with $c = 2$ holds for the class of scale-free trees where the scaling property of the degree can be finely tuned by the initial attractiveness a. Further as shown in Ref. [6], if $N \ll B_l \ll (\frac{N}{2})^2$, its complementary distribution can be approximated by the power law $\Pr[B_l \ge x] = (1 - \alpha)N\frac{1}{B_l}$ which leads to our $c = 2$ scaling for the probability distribution of B_l. The link and node betweenness distribution is considered to be same in a tree [39]. Szabó *et al.* [7] found the scaling exponent $c = 2$ for node betweenness in a BA tree with a "mean-field" approximation. The rigorous proof of the heuristic result of [7] has been provided by Bollobás and Ridordan in Ref. [8].

TABLE II. Topological characteristics of tree models and overlay tree on top of network models.

	N	c	$E[H]$	sdev$[D]$
BA tree	100	2.3	4.7	2.38
URT	100	2.1	6.6	1.36
$G_{\cup \mathrm{SPT}(\alpha<\alpha_c)}$ on BA model $(m=3)$	100	1.7	9.6	1.04
$G_{\cup \mathrm{SPT}(\alpha<\alpha_c)}$ on $G_{0.8}(100)$	100	1.6	9.8	1.04
$G_{\cup \mathrm{SPT}(\alpha<\alpha_c)}$ on cubic lattice	125	1.5	12.8	0.92
$G_{\cup \mathrm{SPT}(\alpha<\alpha_c)}$ on square lattice	100	1.3	13.4	0.81
$G_{\cup \mathrm{SPT}(\alpha<\alpha_c)}$ on square lattice	144	1.3	16.8	0.82
k-ary tree [40]	100	1	$E[H(k)]$	$\sqrt{k-1}$

An URT ($\beta=0$) possesses in fact exponential degree distribution. A rigorous derivation of link betweenness distribution for URT is given in Appendix A.

B. Comparison of betweenness distribution of overlay trees and tree models

All the three classes of trees have been shown to follow approximately a power law betweenness distributions. The power law betweenness distribution has been proved for class (a) tree models in Sec. V A, while for class (b) overlay tree on top of network models and (c) overlay tree on top of real networks it seems to arise from the random sampling of the overlay tree (caused by the IID link weights) as explained in Sec. IV B.

The slope of the betweenness distribution in log-log scale or, equivalently, the power exponent c of the corresponding power law distribution (3), characterizes the variance of the traffic carried along links in the network. High values of c can be interpreted as a high concentration of traffic on the most important links. The betweenness distribution of a tree depends purely on the structure of the tree. Hence, we further examine the relationship between the scaling exponent c and

the corresponding tree structure which can be partially characterized by the average hop count $E[H]$ (or the average number of links) of the shortest path and the standard deviation sdev$[D]$ of the degree, because the average degree in any tree is $E[D]=2(N-1)/N=2-\frac{2}{N}$.

We compare class (a) and (b) in Table II and class (c) in Table III. With a similar number of nodes in Table II we find the following. The scaling exponent c seems to be negatively correlated with the $E[H]$ except for the k-ary tree. The scaling exponent c seems to be positively correlated with the sdev$[D]$ standard deviation of the degree except for the k-ary tree. The higher the variance of the degree is, the more traffic among links varies. The scaling exponent c seems to be insensitive to the size N of the tree. A same slope c is obtained for different substrate size, e.g., the k-ary tree, and the $G_{\cup \mathrm{SPT}(\alpha \to 0)}$ on top of network models as mentioned in Sec. IV A. However, the $E[H]$ behaves as a function of N and the sdev$[D]$ can slightly depend on N with the fixed average $E[D] \approx 2$. Hence, the correlation between c and $E[H]$ as well as sdev$[D]$ may become weaker or even disappear when networks with different sizes are considered, which will be further examined for real-world networks in Table III. The URT and the class of scale-free trees (e.g., the BA tree) discussed in Sec. V A 2, have $c \to 2.0$ for large N and $N \ll B_l \ll (\frac{N}{2})^2$. Compared to URT, the degree of the BA tree varies more and has higher scaling exponent c (see Table II), when the complete range $B_l \in (N-1, \lfloor \frac{N}{2} \rfloor(N-\lfloor \frac{N}{2} \rfloor))$ is taken into account.

The scaling exponent c of betweenness distribution varies from $c=1$ for the k-ary tree to $c=2$ for scale-free trees. For overlay trees on top of real networks $G(N,L)$ with N nodes and L links in Table III. The exponent c ranges from 1.5 to 1.9, while the network size varies from $N=112$ to $N=12254$. The scaling exponent c does not seem to be dependent on the size N of the topology. The negative correlation between hop count $E[H]$ and c disappear because $E[H]$ is positively correlated with N. The positive correlation between sdev$[D]$ and c still holds for most of the considered networks. The overlay trees possess different degree distributions as plotted in Figs. 6–10. The overlay tree of networks that are marked with a star in Table III possesses a

TABLE III. Topological characteristics of overlay tree on top of real-world networks. The overlay tree of networks that are marked with a star possesses a power law degree distribution.

|	N	L	c	$E[H]$	sdev$[D]$
Internet As topology*	12254	25319	1.9	12.2	16
Web of Science citations (koh)*	3704	12673	1.9	14.6	6.0
Gnutella Crawl2*	1568	1906	1.9	11.6	4.3
Science coauthorship network	379	914	1.8	14.1	1.6
Word adjacencies	112	850	1.8	7.7	1.6
Air Transportation*	2179	31326	1.7	17.9	2.8
Web of Science citations(scimet)*	2678	10385	1.7	22.18	1.9
High Energy collaborations	5835	13815	1.7	31.1	1.5
Dutch soccer	685	10310	1.7	22.7	1.4
Power grid	4941	6594	1.6	58.9	1.2
American football	115	613	1.5	11.8	1.0
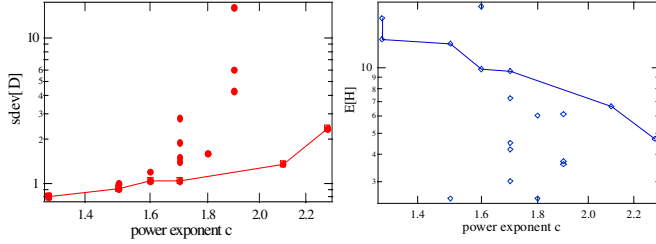
FIG. 11. (Color online) Relationship between the power exponent c of the betweenness and the standard deviation sdev[D] and the average hop count $E[H]$ of the tree.

power law degree distribution. Hence, the betweenness distribution of scale-free networks does not necessarily follow the same power law exponent c, while a similar exponent c can be obtained in networks with different degree distributions.

The relationship between the sdev[D] as well as the $E[H]$ and the scaling exponent c of betweenness distribution is given in Fig. 11. Points lying on the line are for networks listed in Table II with similar topology size N. The approximately positive correlation between sdev[D] and c can be observed for all the three classes. Since the average degree $E[D]=2(N-1)/N \approx 2$ in a tree is almost constant, a higher degree variance implies more nodes with higher degree or/and more nodes with degree 1. The betweenness of links connected to a degree 1 node is always the minimum $N-1$ while the traffic passing through a high degree node is split by links connected to this node. Both contribute to a higher scaling exponent c.

VI. CONCLUSION

In this paper, we examine the traffic in a weighted network via the link betweenness distribution of the corresponding transport overlay network $G_{\cup SPT}(\alpha)$, the union of all shortest paths. In the strong disorder regime, all transport flows over the overlay tree $G_{\cup SPT(\alpha \to 0)}=\text{MST}$. Important new findings are the power law betweenness distribution specified in Eq. (3) of trees: tree models such as scale-free trees and k-ary trees; overlay trees on top of traditional network models; overlay trees on top of real-world complex networks. The scaling exponent $1 < c \leq 2$ for large networks is shown to be positively correlated with the sdev[D] of the corresponding tree and is insensitive to the network size N. Equipped with IID link weights, the overlay tree is, in fact, a random minimum spanning tree (RMST). We conjecture that the scaling exponent c may be used to characterize these tree structures and probably the underlying topology. First, recall that any link in a tree connects two clusters with size $1 \leq |C_l| \leq \frac{N}{2}$ and $N-|C_l|$ and $B_l \sim |C_l| \cdot N$ in Sec. V. The power law betweenness distribution implies approximately a same power law scaling for $\Pr[|C_l|=n] \sim n^{-c}$ the probability distribution of cluster size. Second, for the Internet As topology, our power law scaling of betweenness with $c=1.9$ is the same as $\Pr[S=n] \sim n^{-1.9 \pm 0.1}$ the probability of finding n

points downhill [41], a signature of the intrinsic fractal properties of webs. And recently, Kitsak *et al.* [9] have brought fractal properties of networks into the betweenness analysis.

In the weak disorder regime, traffic flows over more links than that of the MST. The negative correlation between link weight and betweenness also depends on α, the strength of link weight disorder and the structure of the underlying topology. Both a stronger disorder in link weights and the non-homogeneity of the substrate reduce the correlation.

ACKNOWLEDGMENTS

APPENDIX A: LINK BETWEENNESS DISTRIBUTION OF URT

A URT [14] of size N is a random tree rooted at some node A. At each stage a node is attached uniformly to one of the existing nodes until the total number of nodes is equal to N. When the jth node is attached, the corresponding jth attached link is also added except that no link is added when we start from the root or the first node. In a tree, the traffic traverses the link if and only if the source and destination lies in different clusters separated by this link. In a URT, we define $|\mathcal{T}_j^{(N)}|$ as the size of the subtree rooted at the jth attached node. The removal of the jth ($2 \leq j \leq N$) attached link will separate the graph into two clusters with size $|\mathcal{T}_j^{(N)}|$ and $N-|\mathcal{T}_j^{(N)}|$. Correspondingly, the betweenness of the jth ($2 \leq j \leq N$) attached link is $|\mathcal{T}_j^{(N)}| \cdot (N-|\mathcal{T}_j^{(N)}|)$. The probability distribution of the size of the subtree [42] equals

$$\Pr[|\mathcal{T}_j^{(N)}|=k]=\frac{(j-1)(N-j)!(N-k-1)!}{(N-1)!(N-j-k+1)!}=\frac{\binom{N-k-1}{j-2}}{\binom{N-1}{j-1}}$$

(A1)

Using the law of total probability [14], we have for the URT that

$$\Pr[B_l=k(N-k)]=\sum_{j=2}^{N}\Pr[B_l=k(N-k)|l=j]\Pr[l=j],$$

$$1 \leq k \leq \left\lfloor \frac{N}{2} \right\rfloor.$$

A random link l is the jth attached link or attaches the jth node to the URT with probability $\Pr[l=j]=\frac{1}{N-1}$.

For $k \in [1, \lfloor \frac{N}{2} \rfloor]$ and $k \neq \frac{N}{2}$, the conditional probability is

$$\Pr[B_l = k(N-k)|l=j] = \Pr[|\mathcal{T}_j^{(N)}| = k] + \Pr[|\mathcal{T}_j^{(N)}| = N-k]$$

because only if the size of the subtree rooted at node j is of size $|\mathcal{T}_j^{(N)}|=k$ or of size $|\mathcal{T}_j^{(N)}|=N-k$, the betweenness of the link $l=j$ equals $k(N-k)$. Combining both yields

$$\Pr[B_l = k(N-k)] = \frac{1}{N-1}\sum_{j=2}^{N} \Pr[|\mathcal{T}_j^{(N)}| = k] + \Pr[|\mathcal{T}_j^{(N)}| = N-k].$$

Substituting Eq. (A1) gives

$$\Pr[B_l = k(N-k)] = \frac{(N-k-1)!}{(N-1)(N-1)!}\sum_{j=2}^{N} \frac{(j-1)(N-j)!}{(N-j-k+1)!}$$
$$+ \frac{(k-1)!}{(N-1)(N-1)!}\sum_{j=2}^{N} \frac{(j-1)(N-j)!}{(k+1-j)!}.$$

We use the identity

$$\sum_{j=n}^{m} j \binom{a-j}{b-j} = n\binom{a+1-n}{b-n} + \binom{a+1-n}{b-1-n} - m\binom{a-m}{b-1-m}$$
$$- \binom{a+1-m}{b-1-m} \tag{A2}$$

and obtain

$$\sum_{j=2}^{N} \frac{(j-1)(N-j)!}{(N-j-k+1)!} = (k-1)! \sum_{j=1}^{N-1} j\binom{N-1-j}{N-k-j} = (k$$
$$-1)! \left[\binom{N-1}{N-k-1} + \binom{N-1}{N-k-2} \right]$$
$$= (k-1)! \binom{N}{N-k-1}$$

$$= \frac{(k-1)!\,N!}{(k+1)!\,(N-k-1)!}.$$

Similarly,

$$\sum_{j=2}^{N} \frac{(j-1)(N-j)!}{(k+1-j)!} = (N-1-k)! \sum_{j=1}^{N-1} j\binom{N-1-j}{k-j}$$
$$= (N-1-k)! \left[\binom{N-1}{k-1} + \binom{N-1}{k-2} \right]$$
$$= (N-1-k)! \binom{N}{k-1}$$
$$= \frac{(N-1-k)!\,N!}{(k-1)!\,(N-k+1)!}.$$

Hence,

$$\Pr[B_l = k(N-k)] = \frac{(N-k-1)!}{(N-1)(N-1)!}\frac{(k-1)!\,N!}{(k+1)!\,(N-k-1)!}$$
$$+ \frac{(k-1)!}{(N-1)(N-1)!}\frac{(N-1-k)!\,N!}{(k-1)!\,(N-k+1)!}$$
$$= \frac{N}{(N-1)}\left(\frac{1}{(k+1)k} + \frac{1}{(N-k+1)(N-k)} \right)$$
$$= \frac{N}{(N-1)k(N-k)}\left(\frac{N-k}{k+1} + \frac{k}{N-k+1} \right).$$

While for $k = \lfloor \frac{N}{2} \rfloor = \frac{N}{2}$, the probability has to be halved,

$$\Pr[B_l = k(N-k)] = \frac{N}{(N-1)k(k+1)}.$$

Hence,

$$\Pr[B_l = k(N-k)] = \begin{cases} \dfrac{N}{(N-1)k(N-k)}\left(\dfrac{N-k}{k+1} + \dfrac{k}{N-k+1} \right), & k \in \left[1, \left\lfloor \dfrac{N}{2} \right\rfloor \right] \text{ and } k \neq \dfrac{N}{2}, \\[4mm] \dfrac{N}{(N-1)k(k+1)} & \text{if } k = \left\lfloor \dfrac{N}{2} \right\rfloor = \dfrac{N}{2}. \end{cases} \tag{A3}$$

APPENDIX B: LINK BETWEENNESS DISTRIBUTION OF A k-ARY TREE

If the link betweenness distribution (4) of a k-ary tree follows a power law of the form $y = c_0 x^c$, then for any two points (x_1, y_1) and (x_2, y_2) on this curve, we have $\frac{y_1}{y_2} = (\frac{x_1}{x_2})^c$. Two nodes are selected: $[N(D)-1, \frac{k^D}{N(D)-1}]$ corresponding to $j=D$ in Eq. (4) and a random node $\{N(D-j)[N(D)-N(D-j)], \frac{k^j}{N(D)-1}\}$.

$$\frac{N(D-j)[N(D)-N(D-j)]}{N(D)-1} = \frac{(k^{D-j+1}-1)(k^{D+1}-k^{D-j+1})}{(k^{D+1}-1-k+1)(k-1)}$$

$$= \frac{(k^{D-j+1}-1)(k^{D+1}-k^{D-j+1})}{k(k^D-1)(k-1)}$$

$$= \frac{k^D(k^{D-j+1}-1)(1-k^{-j})}{(k^D-1)(k-1)}.$$

For large networks with large k and D,

$$\frac{N(D-j)[N(D)-N(D-j)]}{N(D)-1} \simeq \frac{(k^{D-j+1}-1)(1-k^{-j})}{(k-1)}$$

$$= \frac{k^{D-j}(k-k^{-(j-1)}-k^{-(j-D)}+k^{-D})}{(k-1)}$$

$$\simeq k^{D-j} = \left(\frac{\frac{k^j}{N(D)-1}}{\frac{k^D}{N(D)-1}} \right)^{-1}.$$

Hence, the link betweenness distribution of a k-ary tree is not a precise power distribution, but it is close to a power law with exponent $c=-1$, especially for larger k and D. The first and last point of link betweenness corresponds to $j=1$ and $j=N$. Since

$$\frac{N(D-1)[N(D)-N(D-1)]}{N(D)-1}$$

$$= \left. \frac{k^D(k^{D-j+1}-1)(1-k^{-j})}{(k^D-1)(k-1)} \right|_{j=1}$$

$$= k^{D-1} = \left(\frac{\left. \frac{k^j}{N(D)-1} \right|_{j=1}}{\frac{k^D}{N(D)-1}} \right)^{-1}.$$

The first and the last points always lie on a power law curve with exponent $c=-1$. Hence, an exceptional case is for $D=2$, which is an exact power law although D is small.

[1] P. Van Mieghem, *Data Communications Networking* (Techne Press, Amsterdam, 2006).

[2] R. Albert, H. Jeong, and A.-L. Barabàsi, Nature (London) **406**, 378 (2000).

[3] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).

[4] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).

[5] A. Fekete, G. Vattay, and L. Kocarev, Phys. Rev. E **73**, 046102 (2006).

[6] A. Fekete, G. Vattay, and L. Kocarev, Complexus **3**, 97 (2006).

[7] G. Szabó, M. Alava, and J. Kertész, Phys. Rev. E **66**, 026101 (2002).

[8] B. Bollobás and O. Riordan, Phys. Rev. E **69**, 036114 (2004).

[9] M. Kitsak, S. Havlin, G. Paul, M. Riccaboni, F. Pammolli, and H. E. Stanley, Phys. Rev. E **75**, 056115 (2007).

[10] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).

[11] P. Van Mieghem and S. M. Magdalena, Phys. Rev. E **72**, 056138 (2005).

[12] K.-I. Goh, J. D. Noh, B. Kahng, and D. Kim, Phys. Rev. E **72**, 017102 (2005).

[13] M. Barthélemy, Eur. Phys. J. B **38**, 163 (2004).

[14] P. Van Mieghem, *Performance Analysis of Communications Systems and Networks* (Cambridge University Press, Cambridge, 2006).

[15] α_c depends on the network topology as well as the size N of the network.

[16] Y. Chen, E. Lopez, S. Havlin, and H. E. Stanley, Phys. Rev. Lett. **96**, 068702 (2006).

[17] P. Van Mieghem and S. van Langen, Phys. Rev. E **71**, 056113 (2005).

[18] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[19] R. T. Smythe and C. John, *Wierman. First-Passage Percolation on the Square Lattice* (Springer, Berlin, 1978).

[20] T. Ohira and R. Sawatari, Phys. Rev. E **58**, 193 (1998).

[21] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001).

[22] M. Faloutsos, P. Faloutsos, and C. Faloutsos, ACM Comput. Commun. Rev. **29**, 251 (1999).

[23] E. W. Dijkstra, Numer. Math. **1**, 269 (1959).

[24] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *An Introduction to Algorithms* (MIT Press, Boston, 1991).

[25] Z. Wu, L. A. Braunstein, S. Havlin, and H. E. Stanley, Phys. Rev. Lett. **96**, 148702 (2006).

[26] R. Oliveira, B. Zhang, L. Zhang, in *Proceedings of ACM SIG-COMM*, Kyoto, Japan (ACM, New York, 2007).

[27] A. Beygelzimer, G. Grinstein, R. Linsker, and I. Rish, Physica A **357**, 593 (2005).

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

[29] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001).

[30] http://vlado.fmf.uni-lj.si/pub/networks/data/.

[31] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).

[32] A. Jamakovic, R. E. Kooij, P. Van Mieghem, and E. R. van Dam, in *Proceedings of the 13th Annual Symposium of the IEEE/CVT Benelux, Liége, Belgium*, (IEEE, New York, 2006), p. 35.

[33] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).

[34] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).

[35] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52**, 386 (2000).

[36] The depth D is the number of hops (or links) from the root to a node at the leaves.

[37] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).

[38] K.-I. Goh, E. Oh, H. Jeong, B. Kahng, and D. Kim, Proc. Natl. Acad. Sci. U.S.A. **99**, 12583 (2002).

[39] D.-H. Kim, J. D. Noh, and H. Jeong, Phys. Rev. E **70**, 046126 (2004).

[40] Two exceptions (see Sec. V A 1): $c=0$ for the line graph and $\Pr[B_l=N-1]=1$ for the star.

[41] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52**, 386 (2000).

[42] R. Van der Hofstad, G. Hooghiemstra, and P. Van Mieghem, Combinatorics, Probab. Comput. **15**, 903 (2006).