Inhomogeneous substructures hidden in random networks

Dong-Hee Kim and Hawoong Jeong
Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea
(Received 15 November 2005; published 10 March 2006)

We study the structure of the load-based spanning tree (LST) that carries the maximum weight of the Erdös-Rényi (ER) random network. The weight of an edge is given by the edge-betweenness centrality, the effective number of shortest paths through the edge. We find that the LSTs present very inhomogeneous structures in contrast to the homogeneous structures of the original networks. Moreover, it turns out that the structure of the LST changes dramatically as the edge density of an ER network increases, from scale free with a cutoff, scale free, to a starlike topology. These would not be possible if the weights are randomly distributed, which implies that topology of the shortest path is correlated in spite of the homogeneous topology of the random network.

DOI: 10.1103/PhysRevE.73.037102 PACS number(s): 89.75.Hc, 89.75.Da, 89.75.Fb, 64.60.−i

Complex network theories have attracted much attention because of their usefulness to analyze diverse complex systems in the real world [1,2]. The most representative measure characterizing a network is the *degree distribution*, $P(k)$, which indicates probability for a vertex to be directly connected to k neighboring vertices with edges. Specifically, while the power-law distribution, $P(k) \sim k^{-\gamma}$, is widely found in the most of real-world networks including technological, biological, and social networks, such as the Internet [3], the World Wide Web [4], the metabolic networks [5], the protein interaction networks [6], and the coauthorship networks [7], there also exist homogeneous networks, such as the U.S. highway network and the U.S. power-grid network, which are explained by the bell-shaped and exponential degree distributions. Recently it has been claimed that an apparent scale-free network can be originated from a sampling result of a underlying homogeneous network [8].

While the degree distribution gives valuable knowledge of local structures of networks around us, it is also necessary to know global structures of networks to understand dynamics on the networks properly. The information transport between two vertices occurs along an optimal path connecting them, defined as the path minimizing the total cost [9–11], which is usually determined by using the global knowledge of the network. For instance, full information of connections in the network is required to determine a shortest path defined as a path consisting of a minimum number of edges, which would be an optimal path if the costs of all edges are identical. The scale-free network has been revealed to have very inhomogeneous shortest path topology so that one can find extremely important vertices or edges that a huge number of shortest paths are passing through, which has been supported by the power-law distribution of the betweenness centrality (BC) [12] and the existence of the transport skeleton structure [13] that makes it possible to understand the origin of the difference between the BC exponent classes of real-world networks [12,13] and the universal properties of the fractal scaling [14,15]. However, in the non-scale-free networks, even though it has been known that there are no such vertices or edges used heavily in the shortest paths, the topology of the shortest paths has not been intensively studied so far.

Our main interest is to find out how the topology of shortest paths is correlated with the network topology in the Erdös-Rényi (ER) random-network model [16], where two arbitrary vertices in the network are randomly connected to each other by an edge with a given probability p, which gives the Poisson degree distribution. In order to systematically study the shortest path topology, it is necessary to treat the network as a weighted network in which the contribution of the shortest paths on each edge is assigned as the weight of the edge. We use the *edge-betweenness centrality* (EBC) [17–19] to represent the contribution of the shortest paths on each edge of the network, which is a convenient quantity counting the effective number of shortest paths through the edge and thereby gives the average traffic though the edges. The EBC of an edge e_{ij} between vertices i and j is the total contribution of the edge on the shortest paths between all possible pairs of vertices, which is defined as follows:

$$b(e_{ij}) = \sum_{m \neq n} b(m,n;i,j) = \sum_{m \neq n} \frac{c(m,n;i,j)}{c(m,n)}, \qquad (1)$$

where $c(m,n;i,j)$ denotes the number of shortest paths from a vertex m to n through the edge e_{ij}, and $c(m,n)$ is the total number of shortest paths from m to n.

In this weighted network, one useful way to study the spatial correlation of the weight is to investigate the *load-based spanning tree* (LST) [13], which consists of a set of selected edges to maximize its total weight, which corresponds to the skeleton of the network [13]. From the degree distribution of the LST, we can check whether the spatial distribution of the weights is correlated with the original network topology. If the weights are randomly distributed on the edges of the network, the degree distribution of the original network would be preserved in its LST [20]. On the other hand, if there exists significant topological correlation in the distribution of the weights on the edges, the degree distribution is expected to show a large deviation from the Poisson distribution, the degree distribution of the ER network. In this paper, we investigate the structure properties of the LST of the ER model in this respect. We find that the LSTs show very inhomogeneous structures in contrast to the homoge-

 ©2006 The American Physical Society

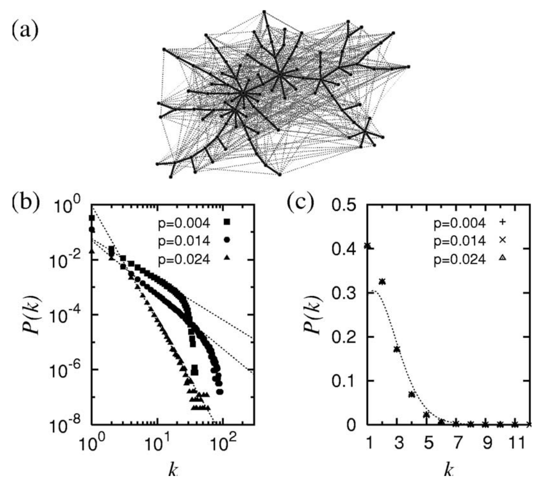

FIG. 1. (a) An illustration of topologies of the LST (solid line) and the original ER network (dotted line) with 100 vertices and $p=0.1$. (b) The degree distributions of the LSTs for various constructions of the ER networks with $N=4000$ vertices and connection probability of edges $p=0.004$, 0.014, 0.024. The degree distributions are guided by the curve-fits (dotted lines) to $P(k)–k^{-\gamma}$ with $\gamma=1.47$, 1.97, 4.20 for $p=0.004$, 0.014, 0.024, respectively. (c) The degree distribution of the LST when the weights are randomly redistributed on the edges. The Poisson degree distribution with average degree two of trees (dotted line) is given for comparison. All data points are obtained from the averages over 256 network ensembles.

neous structures of the original networks. It is found that the degree distribution of the LST shows rich characteristics depending on the connection probability and the size of the ER network, which turns out to be very different from the Poisson distribution.

Figure 1(a) shows an illustration of the inhomogeneous LST structure obtained from the ER network with $N=100$ vertices and $p=0.1$, in which the hub vertices, having a significant number of degree, are found. The detailed characteristics of degrees can be found in Fig. 1(b) which displays the degree distributions of the LSTs obtained from the ER networks generated with various connection probabilities. It is the most interesting feature that the right-skewed degree distributions are observed in the LSTs for the wide range of connection probability p of the ER network whose degree distributions follow the narrow Poisson distribution. For the examined LSTs, it is found that a power law with a cutoff fits well to the degree distribution of the LST, where the exponent and the cutoff degree depend on p. The emergence of these inhomogeneous degree distributions in the LSTs indicates that there exist non-negligible correlations between the weights of neighboring edges sharing a common vertex at their ends, since a vertex shared by the edges having higher weights becomes a hub in the LST. If we remove the correlation by the random redistribution [21] of the weights on the edges, the degree distributions of the LSTs become the Poisson distribution as expected [see Fig. 1(c)]. These imply that the shortest paths, which give the weights of its constituting edges, are not randomly distributed on the ER network but strongly correlated enough to generate an inhomogeneity in spite of the homogeneous topology of the ER network.

The topological correlation of the shortest path that gives rise to the inhomogeneous LST structure can be specified by the correlation between the degree of the vertex and the weights of the edges connected to the vertex. In order to find out more about the spatial correlation in the distribution of the weights, we measure the average weight rank R_k, an averaged value of weight-rank r over the edges attached to a vertex having degree k. The rank r_{ij} of the edge between i and j is graded for its weight w_{ij}, i.e., the largest weight gives $r=1$, the second largest weight gives $r=2$, and so on. Mathematically R_k is defined as follows:

$$R_k = \left\langle \frac{1}{|\mathbf{V}_k|} \sum_{i \in \mathbf{V}_k} \frac{1}{k} \sum_j r_{ij} a_{ij} \right\rangle, \tag{2}$$

where $\mathbf{V}_k$ and $|\mathbf{V}_k|$ are the set of vertices having degree k and the number of those vertices, respectively, a_{ij} is the adjacency matrix element; $a_{ij}=1$ if i and j is connected and $a_{ij}=0$ otherwise, and $\langle \cdots \rangle$ denotes an average over network ensembles. Consequently, the small (large) value of R indicates that a vertex has the edge with high (low) ranks. The reason why we attach great importance to R_k is that it gives an insight into how the structure of the network changes in the LST because the edges of the network are picked in the rank order for the LST. While the average value of the weights can give similar information, since the ER network has a rather homogeneous distribution of weight values, the average weight rank shows more clear changes depending on the spatial distribution of the weights as compared with the average value of the weights. In counting ranks, we do not admit a tie in ranks for the edges with the same weights but assign ranks according to random priority if the values of weights are the same. However, we note that averaging over the network ensembles removes the dependence on the counting method of the tie in ranks.

In order to figure out how the weights are distributed on the network, we compare the average weight rank R_k with that for the uncorrelated one in which the weights are randomly redistributed on the edges to remove the correlation between the edge and the weight on it. For instance, if R_k is smaller than the value for uncorrelated weights, the vertex with degree k can be regarded to have an edge with a higher weight as compared with the uncorrelated situation. This comparison gives an insight into the topological correlation of the weights on the network. From Fig. 2(a), we find that the extreme vertices, which belong to the tail of either the smaller or larger degree part in the Poisson degree distribution, have significantly higher average weight ranks (small value of R_k) than the average rank for the uncorrelated spatial distribution of the weights, given by $\sim L/2$, where L is the total number of edges of the original network. This indicates that these extreme vertices are more likely to preserve their degree in the LSTs. More directly, we also measure the degree correlation between the LST and its original network. In Fig. 2(b), for the extreme vertices, it is confirmed that the degrees of the LST show strong correlations with the degrees of the original network. Consequently, the extreme vertices become hubs that constitute the heavy tail of the inhomogeneous degree distribution of the LST.

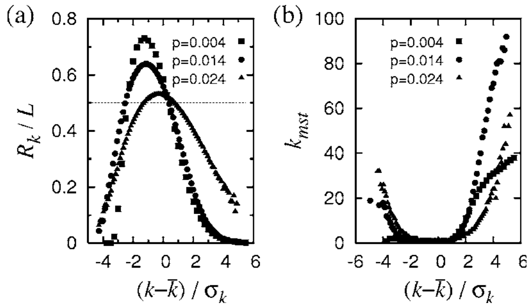

FIG. 2. (a) The comparison of the average weight rank R_k of the weighted ER network with that of the uncorrelated network (dotted line) generated by the random redistribution of weights on the edges. (b) The degree correlation between the LST and the original ER network. ER networks with $N=4000$ vertices and connection probability of edges $p=0.004$, 0.014, 0.024 are examined. $\bar{k}$ and σ_k^2 denote the average and the variance of degrees in the original ER network, respectively. The data points are obtained from the averages over 256 network ensembles.

The structural homogeneity of the ER network varies with the edge density represented by using the connection probability p. As well as the structural properties, the characteristics of the shortest path topology also depend on the edge density of the ER network. Specifically, the change of the connection probability p in the ER network gives rise to the change of the degree distribution of its LST. At small values of p's, as seen in Fig. 1(b), the degree distribution of the LST can be characterized by the power-law exponent γ and the average maximum degree k_{max}. Thus, we can monitor the structural change of the LST by looking at the dependence of γ and k_{max} on the connection probability p and the network size N.

As p increases, γ monotonically increases, but interestingly, k_{max} shows nontrivial behavior of decreasing even though k_{max} has a chance to increase further as p increases because a higher p generates a larger average degree in the original ER network. To find a universal description of this nontrivial behavior of the degree distributions of the LSTs, we perform a size-scaling on data obtained from the ER networks of various sizes and connection probabilities and find out a common scaling feature; k_{max} monotonically increases to the maximum value of $\sim N^{1/2}$ until p reaches at $\sim N^{-1/2}$, then it decreases as p increases further while γ monotonically increases [see Fig. 3].

This p dependence of γ and k_{max} strongly implies that there exists a critical p value at which a clear scale-free degree distribution emerges by satisfying the balance between γ and k_{max}. More precisely, because the natural cutoff k_{cutoff} scales with $\sim N^{1/(\gamma-1)}$ in the finite-sized network of N vertices with power-law degree distribution $P(k) \sim k^{-\gamma}$ [22,23], we can test the *scale-freeness* of the degree distribution of the LST depending on p by comparing k_{max} with the natural cutoff. From the plot of $k_{max}/N^{1/(\gamma-1)}$ as a function of $pN^{1/2}$ shown in Fig. 3(c), we find that k_{max} of the LST becomes comparable with the natural cutoff $N^{1/(\gamma-1)}$ when p reaches $N^{-1/2}$ so that the cutoff diminishes in the degree distribution, which leads to the emergence of the scale-free spanning tree.

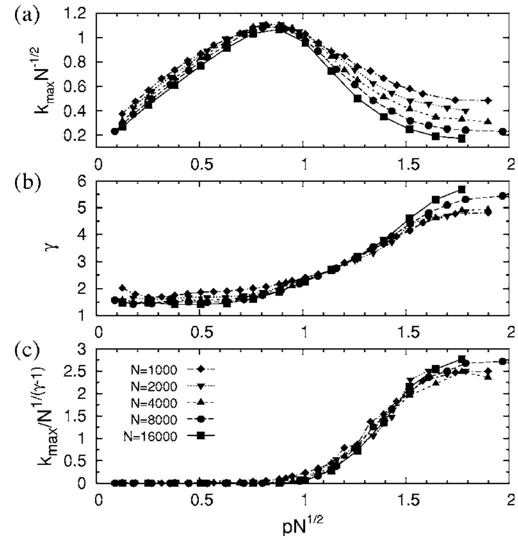

FIG. 3. The p-dependence of (a) the average maximum degree k_{max} and (b) the power-law exponent γ in the degree distribution of the LST. In (c), the comparison of k_{max} with the natural cutoff $N^{1/(\gamma-1)}$ is given. The data points are obtained from the averages over 256 ensembles.

The exponent γ of the power-law degree distribution increases as p increases in the small p region as shown in Fig. 3, but the Poisson degree distribution of the original ER network is not recovered even if p gets much larger. Interestingly, the average maximum degree k_{max} increases when p becomes far larger than $\sim N^{-1/2}$ [24]. For intermediate values of p's, the structures of LSTs show very diverse degree distribution depending on the ensemble of the ER network generation since there can exist lots of edges that have the same value of the EBC (weights), which gives large degeneracy in constructing LSTs. However, we find that a *starlike* structure finally emerges in the LST [see Fig. 4] as the ER network becomes denser to be a nearly fully connected network.

In the limiting cases, this can be understood easily. Let us consider a nearly fully connected network in which only a single pair of vertices are not connected while all the other vertex pairs are directly connected by the edges [25]. In this case, the shortest path connecting these two vertices must pass through one of their common neighbor vertices, and

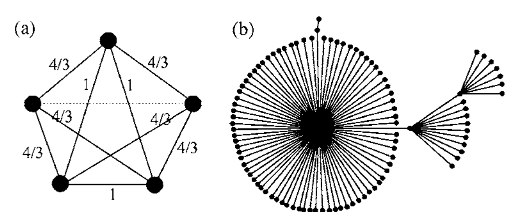

FIG. 4. (a) An example of the very dense network in which only one edge (dotted line) is eliminated from the fully connected network. The EBCs are given on the edges, which indicates that the edges of smaller degree vertices have more EBCs than the other edges. (b) Typical topology of the LST for very dense ER network. The structure is obtained from the ER network with $N=100$ vertices and connection probability $p=0.95$.

consequently this shortest path gives additional contributions to the EBCs of the edges connected to these two vertices having smaller degrees [see Fig. 4(a)]. Therefore, in the edge selection for constructing LSTs, the edges connected to the smaller degree vertices are chosen with higher priorities, which leads the emergence of the starlike LSTs.

Finally, we note that the emergence of the scale-free spanning tree in the ER network also has been reported in two recent works. Clauset *et al.* [8] reported that the spanning tree constructed by using tracer outes from a single source has a power-law degree distribution with a cutoff. Kalisky *et al.* [26] presented work showing that the minimum spanning tree of the percolation cluster in the ER network shows the power-law degree distribution. Together with these works, our findings indicate that the consideration of the transport pathways or the weights of the edges can drastically change the original homogeneous topology of the ER network.

In conclusion, we have investigated the structural properties of the LST of the weighted ER network in which the weight of an edge is given by the EBC. We have found that the degree distribution of the LST shows very inhomogeneous distribution, which can be described as power-law with a cutoff, scale-free, or starlike depending on the edge density of the ER network. The emergence of these inhomogeneous degree distributions in the LST implies that the shortest paths are not homogeneously distributed on the ER network but spatially correlated in spite of the homogeneous topology of the ER network.

This work was supported by Grant No. R14-2002-059-01002-0 from the KOSEF-ABRL Program (D.H.K.) and by Grant No. R01-2005-000-1112-0 from the Basic Research Program of the Korea Science and Engineering Foundation (H.J.).

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[3] M. Faloutsos *et al.*, Comput. Commun. Rev. **29**, 251 (1999).
[4] R. Albert *et al.*, Nature (London) **401**, 130 (1999).
[5] H. Jeong *et al.*, Nature (London) **411**, 41 (2000).
[6] H. Jeong *et al.*, Nature (London) **407**, 651 (2000).
[7] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[8] A. Clauset and C. Moore, Phys. Rev. Lett. **94**, 018701 (2005).
[9] L. A. Braunstein, S. V. Buldyrev, R. Cohen, S. Havlin, and H. E. Stanley, Phys. Rev. Lett. **91**, 168701 (2003).
[10] S. Sreenivasan *et al.*, Phys. Rev. E **70**, 046133 (2004).
[11] S. V. Buldyrev, S. Havlin, E. Lopez, and H. E. Stanley, Phys. Rev. E **70**, 035102(R) (2004).
[12] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001); K.-I. Goh *et al.*, Proc. Natl. Acad. Sci. U.S.A. **99**, 12583 (2002).
[13] D.-H. Kim, J. D. Noh, and H. Jeong, Phys. Rev. E **70**, 046126 (2004).
[14] K.-I. Goh, G. Salvi, B. Kahng, and D. Kim, Phys. Rev. Lett. **96**, 018701 (2006).
[15] C. Song *et al.*, Nature (London) **433**, 392 (2005).
[16] P. Erdös and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).
[17] M. L. Freeman, Sociometry **40**, 35 (1977).
[18] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).
[19] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001); **64**, 016132 (2001).
[20] G. Szabó *et al.*, Physica A **300**, 31 (2003).
[21] We break up the pairs of the edges and their weights, and randomly reassign the weights to the edges.
[22] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[23] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **63**, 062101 (2001).
[24] The value of p at which k_{max} starts to increase varies with the network size N, where we cannot find any size scaling relation or universal feature.
[25] The extreme case of $p=1.0$ is out of our interest because every edge has exactly same EBC and there is no inhomogeneity in the case of the perfectly fully connected network, in which any structure of an LST is possible, from a starlike one to a homogeneous one.
[26] T. Kalisky *et al.*, Phys. Rev. E **73**, 025103(R) (2006).