Transport in Weighted Networks: Partition into Superhighways and Roads

Zhenhua Wu,[1] Lidia A. Braunstein,[2,1] Shlomo Havlin,[3] and H. Eugene Stanley[1]
[1]*Center for Polymer Studies, Boston University, Boston, Massachusetts 02215, USA*
[2]*Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, Funes 3350, 7600 Mar del Plata, Argentina*
[3]*Minerva Center of Department of Physics, Bar-Ilan University, Ramat Gan, Israel*
(Received 21 November 2005; published 13 April 2006)

Transport in weighted networks is dominated by the minimum spanning tree (MST), the tree connecting all nodes with the minimum total weight. We find that the MST can be partitioned into two distinct components, having significantly different transport properties, characterized by centrality—the number of times a node (or link) is used by transport paths. One component, *superhighways*, is the infinite incipient percolation cluster, for which we find that nodes (or links) with high centrality dominate. For the other component, *roads*, which includes the remaining nodes, low centrality nodes dominate. We find also that the distribution of the centrality for the infinite incipient percolation cluster satisfies a power law, with an exponent smaller than that for the entire MST. The significance of this finding is that one can improve significantly the global transport by improving a tiny fraction of the network, the superhighways.

DOI: 10.1103/PhysRevLett.96.148702 PACS numbers: 89.75.Hc, 05.10.−a, 89.75.Da

Recently, much attention has been focused on the topic of complex networks, which characterize many natural and manmade systems, such as the Internet, airline transport system, power grid infrastructures, and the World Wide Web [1–3]. Besides the static properties of complex networks, dynamical phenomena such as transport in networks are of vital importance from both theoretical and practical perspectives. Recently, much effort has been focused on weighted networks [4,5], where each link or node is associated with a weight. Weighted networks yield a more realistic description of real networks. For example, the cable links between computers in the Internet network have different weights, representing their capacities or bandwidths.

In weighted networks, the minimum spanning tree (MST) is a tree, including all of the nodes but only a subset of the links, which has the minimum total weight out of all possible trees that span the entire network. Also, the MST is the union of all "strong disorder" optimal paths between any two nodes [6–12]. The MST which plays a major role for transport is widely used in different fields, such as the design and operation of communication networks, the traveling salesman problem, the protein interaction problem, optimal traffic flow, and economic networks [5,13–18].

An important quantity that characterizes transport in networks is the betweenness centrality C, which is the number of times a node (or link) is used by the set of all shortest paths between all pairs of nodes [19–21]. For simplicity, we call the "betweenness centrality" here "centrality" and we use the notation "nodes" but similar results have been obtained for links. The centrality C quantifies the "importance" of a node for transport in the network. Moreover, identifying the nodes with high C enables us, as shown below, to improve their transport capacity and thus improve the global transport in the network. The probability density function (PDF) of C was studied on the MST for both scale-free (SF) [22] and Erdős-Rényi (ER) [23] networks and found to satisfy a power law

$$\mathcal{P}_{\mathrm{MST}}(C) \sim C^{-\delta_{\mathrm{MST}}}, \tag{1}$$

with δ_{MST} close to 2 [21,24].

Here we show that a subnetwork of the MST [25], the infinite incipient percolation cluster (IIC), has a significantly higher average C than the entire MST—i.e., the set of nodes inside the IIC is typically used by transport paths more often than other nodes in the MST. In this sense, the IIC can be viewed as a set of *superhighways* (SHW) in the MST. The nodes on the MST which are not in the IIC are called *roads*, due to their analogy with roads which are not superhighways (usually used by local residents). We demonstrate the impact of this finding by showing that improving the capacity of the superhighways (IIC) is surprisingly a better strategy to enhance global transport compared to improving the same number of links of the highest C in the MST, although they have higher C [26]. This counterintuitive result shows the advantage of identifying the IIC subsystem, which is very small compared to the full network [27]. Our results are based on extensive numerical studies for centrality of the IIC and comparison with the centrality of the entire MST. We study ER, SF, and square lattice networks.

To generate a ER network of size N with average degree $\langle k\rangle$, we pick at random a pair of nodes from all possible $N(N-1)/2$ pairs, link this pair, and continue this process until we have exactly $\langle k\rangle N/2$ edges. We disallow multiple connections between two nodes and self-loops in a single node. To construct SF networks with a prescribed power law distribution $\mathcal{P}(k) \sim k^{-\lambda}$, with $k \geq k_{\min}$ [22], we use the Molloy-Reed algorithm [30]. We assign to each node i a random number k_i of links drawn from this power law distribution. Then we choose a node i and connect each of its k_i links with randomly selected k_i different nodes.

To construct a *weighted* network, we next assign a weight w_i to each link from a uniform distribution between 0 and 1. The MST is obtained from the weighted network using Prim's algorithm [31]. We start from any node in the largest connected component of the network and grow a treelike cluster to the nearest neighbor with the minimum weight until the MST includes all the nodes of the largest connected component. Once the MST is built, we compute the value of C of each node by counting the number of paths between all possible pairs passing through that node. We normalize C by the total number of pairs in the MST $N(N-1)/2$, which ensures that C is between 0 and 1 [32].

To find the IIC of ER and SF networks, we start with the fully connected network and remove links in descending order of their weights. After each removal of a link, we calculate $\kappa \equiv \langle k^2\rangle/\langle k\rangle$, which decreases with link removals. When $\kappa < 2$, we stop the process because, at this point, the largest remaining component is the IIC [33]. For the two-dimensional (2D) square lattice, we cut the links in descending order of their weights until we reach the percolation threshold p_c ($=0.5$). At that point, the largest remaining component is the IIC [29].

To quantitatively study the centrality of the nodes in the IIC, we calculate the PDF $\mathcal{P}_{\rm IIC}(C)$ of C. In Fig. 1, we show for nodes that for all three cases studied, ER, SF, and square lattice networks, $\mathcal{P}_{\rm IIC}(C)$ satisfies a power law

$$\mathcal{P}_{\rm IIC}(C) \sim C^{-\delta_{\rm IIC}}, \tag{2}$$

where

$$\delta_{\rm IIC} \approx \begin{cases} 1.2 & [\text{ER, SF}] \\ 1.25 & [\text{square lattice}]. \end{cases} \tag{3}$$

Moreover, from Fig. 1, we find that $\delta_{\rm IIC} < \delta_{\rm MST}$, implying a larger probability to find a larger value of C in the IIC compared to the entire MST. Our values for $\delta_{\rm MST}$ are consistent with those found in Ref. [24]. We obtain similar results for the centrality of the links. Our results thus show that the IIC is like a network of *superhighways* inside the MST. When we analyze centrality for the entire MST, the effect of the high C of the IIC is not seen, since the IIC is only a small fraction of the MST. Our results are summarized in Table I.

To further demonstrate the significance of the IIC, we compute for each realization of the network the average C over all nodes $\langle C\rangle$. In Fig. 2, we show the histograms of $\langle C\rangle$ for both the IIC and for the other nodes on the MST. We see that the nodes on the IIC have a much larger $\langle C\rangle$ than the other nodes of the MST.

Figure 3 shows a schematic plot of the SHW inside the MST and demonstrates its use by the path between pairs of nodes. The MST is the "skeleton" inside the network, which plays a key role in transport between the nodes. However, the IIC in the MST is like the "spine in the skeleton," which plays the role of the superhighways inside a road transportation system. A car can drive from the entry node A on roads until it reaches a superhighway and finds the exit which is closest to the exit node B. Thus, those nodes which are far from each other in the MST should use the IIC superhighways more than those nodes which are close to each other. In order to demonstrate this, we compute f, the average fraction of pairs of nodes using the IIC, as a function of $\ell_{\rm MST}$, the distance between a pair of nodes on the MST (Fig. 4). We see that f increases and approaches 1 as $\ell_{\rm MST}$ grows. We also show that f scales as $\ell_{\rm MST}/N^{\nu_{\rm opt}}$ for different system sizes, where $\nu_{\rm opt}$ is the percolation connectedness exponent [9,10].

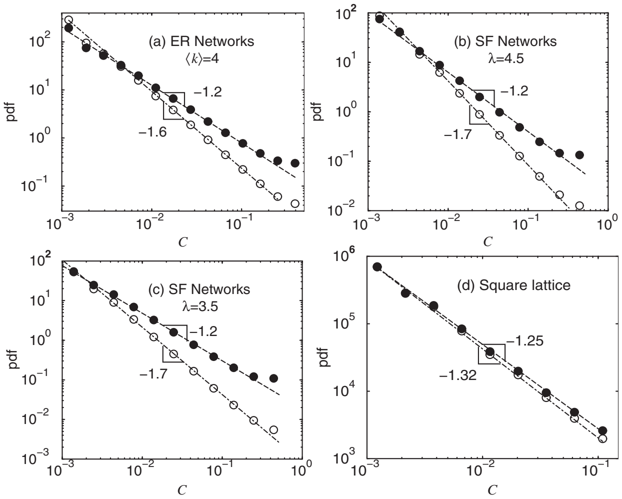

FIG. 1. The PDF of the centrality of nodes for (a) an ER graph with $\langle k\rangle = 4$, (b) a SF with $\lambda = 4.5$, (c) a SF with $\lambda = 3.5$, and (d) a 90×90 square lattice. For ER and SF $N = 8192$, and for the square lattice $N = 8100$. We analyze 10^4 realizations. For each graph, the solid circles show $\mathcal{P}_{\rm IIC}(C)$; the unfilled circles show $\mathcal{P}_{\rm MST}(C)$.

The next question is how much the IIC is used in transport on the MST. We define the IIC *superhighway usage*

$$u \equiv \frac{\ell_{\rm IIC}}{\ell_{\rm MST}}, \tag{4}$$

where $\ell_{\rm IIC}$ is the number of the links in a given path of length $\ell_{\rm MST}$ belonging to the IIC superhighways. The average usage $\langle u\rangle$ quantifies how much the IIC is used by the transport between all pairs of nodes. In Fig. 5(a), we show $\langle u\rangle$ as a function of the system size N. Our results suggest that $\langle u\rangle$ approaches a constant value and becomes independent of N for large N. This is surprising, since the average value of the ratio between the number of nodes on

TABLE I. Results for the IIC and the MST.

	ER	SF ($\lambda = 4.5$)	SF ($\lambda = 3.5$)	Square lattice
$\delta_{\rm IIC}$	1.2	1.2	1.2	1.25
$\delta_{\rm MST}$	1.6	1.7	1.7	1.32
$\nu_{\rm opt}$	1/3	1/3	0.2	0.61
$\langle u\rangle$	0.29	0.20	0.13	0.64

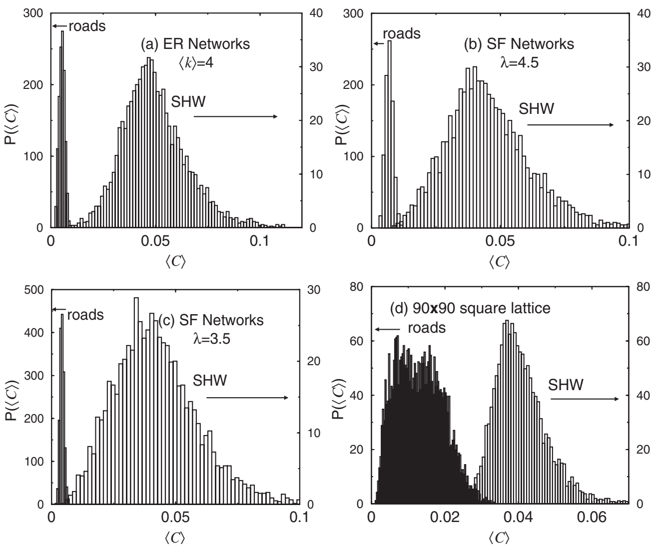

FIG. 2. The normalized PDF for superhighway and roads of $\langle C \rangle$, the C averaged over all nodes in one realization. (a) An ER network, (b) a SF network with $\lambda = 4.5$, (c) a SF network with $\lambda = 3.5$, and (d) a square lattice network. To make each histogram, we analyze 1000 network configurations.

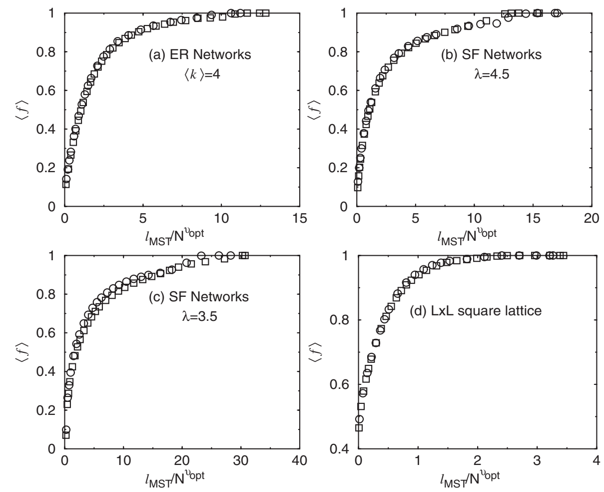

FIG. 4. The average fraction $\langle f \rangle$ of pairs using the SHW, as a function of ℓ_{MST}, the distance on the MST. (a) An ER graph with $\langle k \rangle = 4$, (b) a SF with $\lambda = 4.5$, (c) a SF with $\lambda = 3.5$, and (d) a square lattice. For ER and SF: (○) $N = 1024$ and (□) $N = 2048$ with 10^4 realizations. For a square lattice: (○) $N = 1024$ and (□) $N = 2500$ with 10^3 realizations. The x axis is rescaled by $N^{\nu_{\mathrm{opt}}}$, where $\nu_{\mathrm{opt}} = 1/3$ for ER and for SF with $\lambda > 4$, and $\nu_{\mathrm{opt}} = (\lambda - 3)/(\lambda - 1)$ for SF networks with $3 < \lambda < 4$ [9]. For the $L \times L$ square lattice, $\ell_{\mathrm{MST}} \sim L^{d_{\mathrm{opt}}}$, and since $L^2 = N$, $\nu_{\mathrm{opt}} = d_{\mathrm{opt}}/2 \approx 0.61$ [7,8].

the IIC and on the MST $\langle N_{\mathrm{IIC}}/N_{\mathrm{MST}} \rangle$ approaches zero as $N \rightarrow \infty$ [27], showing that, although the IIC contains only a tiny fraction of the nodes in the entire network, its usage for the transport in the entire network is constant. We find that $\langle u \rangle \approx 0.3$ for ER networks, $\langle u \rangle \approx 0.2$ for SF networks with $\lambda = 4.5$, and $\langle u \rangle \approx 0.64$ for the square lattice. The reason why $\langle u \rangle$ is not close to 1.0 is that, in addition to the IIC, the optimal path passes through other percolation clusters, such as the second largest and the third largest percolation clusters. In Fig. 5, we also show for ER networks the average usage of the two largest and the three largest percolation clusters for a path on the MST, and we see that the average usage increases significantly and is also independent of N. However, the number of clusters used by a path on MST is relatively small and proportional to $\ln N$ [34], suggesting that the path on the MST uses only a few percolation clusters and a few jumps between them ($\sim \ln N$) in order to get from an entry node to an exit node on the network. When $N \rightarrow \infty$, the average usage of all percolation clusters should approach 1.

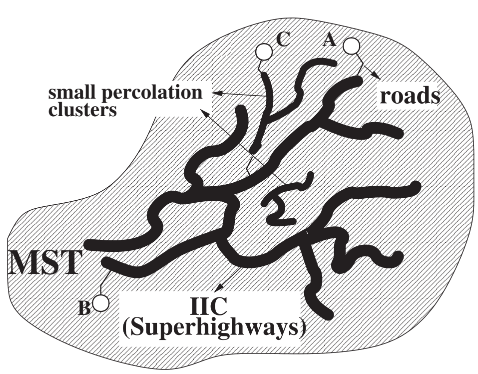

FIG. 3. Schematic graph of the network of connected superhighways (heavy lines) inside the MST (shaded area). A, B, and C are examples of possible entry and exit nodes, which connect to the network of superhighways by "roads" (thin lines). The middle size lines indicate other percolation clusters with much a smaller size compared to the IIC.

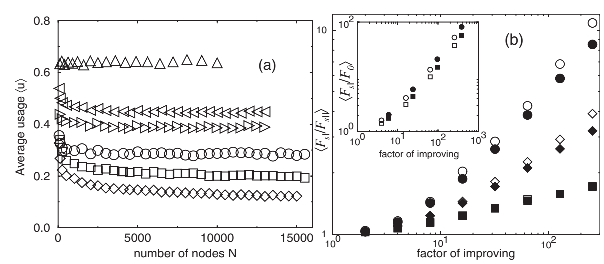

FIG. 5. (a) The average usage $\langle u \rangle \equiv \langle \ell_{\mathrm{IIC}}/\ell_{\mathrm{MST}} \rangle$ for different networks, as a function of the number of nodes N ○ (ER with $\langle k \rangle = 4$), □ (SF with $\lambda = 4.5$), ◇ (SF with $\lambda = 3.5$), △ ($L \times L$ square lattice). The symbols (▷) and (◁) represent the average usage for ER with $\langle k \rangle = 4$ when the two largest percolation clusters and the three largest percolation clusters are taken into account, respectively. (b) The ratio between the flow using strategy I, F_{sI}, and that using strategy II, F_{sII}, as a function of the factor of improving conductivity or capacity. The inset is the ratio between the flow using strategy I and the flow in the original network F_0. The data are all for ER networks with $N = 2048$, $\langle k \rangle = 4$, and $n = 50$ (○), $n = 250$ (◇), and $n = 500$ (□). The unfilled symbols are for current flow and the solid symbols are for maximum flow.

Can we use the above results to improve the transport in networks? It is clear that, by improving the capacity or conductivity of the highest C links, one can improve the transport [see Fig. 5(b) inset]. We hypothesize that improving the IIC links (strategy I), which represent the superhighways, is more effective than improving the same number of links with the highest C in the MST (strategy II), although they have higher centrality [26]. To test the hypothesis, we study two transport problems: (i) current flow in random resistor networks, where each link of the network represents a resistor, and (ii) the maximum flow problem well known in computer science [35]. We assign to each link of the network a resistance or capacity e^{ax}, where x is an uniform random number between 0 and 1, with $a = 40$. The value of a is chosen so as to have a broad distribution of disorder so that the MST carries most of the flow [10,34]. We randomly choose n pairs of nodes as sources and other n nodes as sinks and compute flow between them. We compare the transport by improving the conductance or capacity of the links on the IIC (strategy I) with that by improving the same number of links but those with the highest C in the MST (strategy II). Since the two sets are not the same and, therefore, higher centrality links will be improved in II [26], it is tempting to suggest that the better strategy to improve global flow would be strategy II. However, here we demonstrate using ER networks as an example that counterintuitively strategy I is better. We also find similar improvements of strategy I compared to strategy II for SF networks with $\lambda = 3.5$. In Fig. 5(b), we compute the ratio between the flow using strategy I ($F_{\rm sI}$) and the flow using strategy II ($F_{\rm sII}$) as a function of the factor of improving conductivity or capacity of the links. The figure clearly shows that strategy I is better than strategy II. Since the number of links in the IIC is relatively very small compared to the number of links in the whole network [27], it could be a very efficient strategy.

In summary, we find that the centrality of the IIC for transport in networks is significantly larger than the centrality of the other nodes in the MST. Thus, the IIC is a key component for transport in the MST. We demonstrate that improving the capacity or conductance of the links in the IIC is a useful strategy to improve transport.

We thank ONR, Israel Science Foundation, European NEST project DYSONET, FONCyT (PICT-O2004/370), and Israeli Center for Complexity Science for financial support.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[2] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, England, 2004).

[3] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).

[4] A. Barrat *et al.*, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).

[5] P. J. Macdonald *et al.*, Europhys. Lett. **72**, 308 (2005).

[6] By strong disorder it is meant that the weight of the path is determined by the largest weight along the path.

[7] M. Cieplak *et al.*, Phys. Rev. Lett. **72**, 2320 (1994).

[8] M. Porto *et al.*, Phys. Rev. E **60**, R2448 (1999).

[9] L. A. Braunstein *et al.*, Phys. Rev. Lett. **91**, 168701 (2003).

[10] Z. Wu *et al.*, Phys. Rev. E **71**, 045101(R) (2005).

[11] A.-L. Barabási, Phys. Rev. Lett. **76**, 3750 (1996).

[12] R. Dobrin *et al.*, Phys. Rev. Lett. **86**, 5076 (2001).

[13] M. Khan, G. Pandurangan, and B. Bhargava, Department of Computer Science, Purdue University Technical Report No. CSD TR 03-013, 2003.

[14] S. Skiena, *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica* (Addison-Wesley, New York, 1990).

[15] M. L. Fredman and R. E. Tarjan, J. ACM **34**, 596 (1987).

[16] J. B. Kruskal, Proc. Am. Math. Soc. **7**, 48 (1956).

[17] G. Bonanno *et al.*, Phys. Rev. E **68**, 046130 (2003).

[18] J.-P. Onnela *et al.*, Phys. Rev. E **68**, 056110 (2003).

[19] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001); **64**, 016132 (2001).

[20] K.-I. Goh *et al.*, Phys. Rev. Lett. **87**, 278701 (2001).

[21] D.-H. Kim *et al.*, Phys. Rev. E **70**, 046126 (2004).

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

[23] P. Erdős and A. Rényi, Publ. Math. **6**, 290 (1959).

[24] K.-I. Goh *et al.*, Phys. Rev. E **72**, 017102 (2005).

[25] For networks ($d = \infty$), the IIC has no loops and the IIC must be a subset of the MST. For lattices, we chose only the part of the IIC that belongs to the MST.

[26] The overlap between the two groups is about 30% for ER networks.

[27] The ratio between the mass of the IIC $N_{\rm IIC}$ and the system size $N \equiv N_{\rm MST}$ approaches zero for large N due to the fractal nature of the IIC. Indeed, $N_{\rm IIC} \sim N^{2/3}$ both for ER [23] and for SF with $\lambda > 4$ [28]. For SF with $\lambda = 3.5$, $N_{\rm IIC} \sim N^{0.6}$ [28], and, for the $L \times L$ lattice, $N_{\rm IIC} \sim L^{91/48} \sim N^{91/96}$ [29].

[28] R. Cohen *et al.*, in *Handbook of Graphs and Networks*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, New York, 2002), Chap. 4.

[29] *Fractals and Disordered Systems*, edited by A. Bunde and S. Havlin (Springer, New York, 1996).

[30] M. Molloy and B. A. Reed, Comb. Probab. Comput. **7**, 295 (1998); Random Struct. Algorithms **6**, 161 (1995).

[31] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms and Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1993).

[32] This C measurement is equivalent to counting the number of times a node (link) is used by the set of optimal paths linking all pairs of nodes, in the limit of strong disorder.

[33] R. Cohen *et al.*, Phys. Rev. Lett. **85**, 4626 (2000).

[34] S. Sreenivasan *et al.*, Phys. Rev. E **70**, 046133 (2004).

[35] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms* (MIT, Cambridge, MA, 1990).