EUROPHYSICS LETTERS

Maximum flow and topological structure of complex networks

* * *

D.-S. Lee and H. Rieger

1986
2006
20 years of successful publishing

Europhysics Letter was launched more than fifteen years ago by the European Physical Society, the Société Française de Physique, the Società Italiana di Fisica and the Institute of Physics (UK) and owned now by 17 National Physical Societies/Institutes.

Europhysics Letters aims to publish short papers containing non-trivial new results, ideas, concepts, experimental methods, theoretical treatments, etc. which are of broad interest and importance to one or several sections of the physics community.

Europhysics letters provides a platform for scientists from all over the world to offer their results to an international readership.

Maximum flow and topological structure
of complex networks

D.-S. Lee and H. Rieger

Theoretische Physik, Universität des Saarlandes - 66041 Saarbrücken, Germany

Abstract. – The problem of sending the maximum amount of flow q between two arbitrary nodes s and t of complex networks along links with unit capacity is studied, which is equivalent to determining the number of link-disjoint paths between s and t. The average of q over all node pairs with smaller degree $k_{\min}$ is $\langle q \rangle_{k_{\min}} \simeq c\, k_{\min}$ for large $k_{\min}$ with c a constant implying that the statistics of q is related to the degree distribution of the network. The disjoint paths between hub nodes are found to be distributed among the links belonging to the same edge-biconnected component, and q can be estimated by the number of pairs of edge-biconnected links incident to the start and terminal node. The relative size of the giant edge-biconnected component of a network approximates to the coefficient c. The applicability of our results to real-world networks is tested for the Internet at the autonomous system level.

The analysis of network structures of complex systems has turned out to be extremely useful in exploring their large-scale organization and unveiling their evolutionary origin [1–4]. Anomalous features found by their graph-theoretical analyses [5–7] are the fingerprints of their hidden organization principles as well as the key to predicting their behaviors. Real-world networks usually serve a particular purpose which frequently can be expressed in terms of flow problems, in particular in the context of transport between the nodes along links (edges) with restricted capacities. For instance, if one is interested in the maximum possible flow that can be sent from one node to another, one has to solve a maximum-flow problem, for which polynomial algorithms exist. It also arises *per se* in a variety of situations such as assignment and scheduling problems [8]. In real-world networks one is often interested in the question of how many link-disjoint paths do exist between a particular node pair —for instance, to establish as many independent transportation routes as possible between two nodes in the occasion of a sudden demand (like in the event of a natural catastrophe or in a military context). This problem is again identical to the maximum-flow problem between those two nodes in the same network with unit-capacity links.

Properties of complex networks related to transport are inherently connected to their topological structure, in particular to various aspects of connectedness. The existence of a single path from one node to another is guaranteed if both nodes belong to the same connected component of the network. Moreover, nodes that belong to a single biconnected component

have two disjoint paths between them. We will show in this paper that the arrangement of the biconnected components of a given network is the essential determinant for the number of disjoint paths between two nodes of large degree, the hubs (which is non-trivial as soon as more than two disjoint paths between nodes exist). In this way we also establish a specific connection between the topological structure and the flow properties of complex networks.

To be specific, we consider undirected networks of heterogeneous connectivity pattern and calculate the number of link-disjoint paths between two nodes s and t. We assign a unit flow capacity to each link and compute numerically the maximum flow, denoted by $q_{s,t}$, between s and t [9]. Since only one unit of flow can be sent along each link, $q_{s,t}$ is a non-negative integer and is equal to the number of link-disjoint paths connecting s and t.

We will show that on average the maximum flow between two nodes is proportional to the minimum of their degree (= number of incident links), with a proportionality constant c that is asymptotically independent of this degree. As a consequence, we obtain that in scale-free (SF) networks, a class of networks where the degree k follows a power law distribution $P_d(k) \sim k^{-\gamma}$ asymptotically, the maximum flow obeys a power law distribution with the exponent $2\gamma - 1$ while in completely random networks it follows a Poisson distribution. Moreover we will demonstrate that the coefficient c reflects the effect of the global connectivity pattern on transport properties such that the constant c varies depending on the total number of links as well as the degree exponent. Finally, we will show that the edge-biconnectedness is crucial in the transport among hub nodes and that the maximum flow between hub nodes can be estimated by the number of pairs of edge-biconnected links incident to the start and terminal node. This implies also that the coefficient c is related to the relative size of the giant edge-biconnected component.

The degree of a node i is the number of incident links and is denoted as k_i. For illustrational purposes we first consider $D > 1$-dimensional regular lattices where all nodes have the same degree $2D$ except for those at the boundary. Due to the homogeneous connectivity pattern, the maximum flow or the number of link-disjoint paths between two distinct nodes s and t is $2D$ unless either node is at the boundary. If s or t is at the boundary, the maximum flow is given by the smaller value of k_s and k_t. Therefore $q_{s,t}$ in regular lattices can be expressed as

$$q_{s,t} = k_{\min}, \tag{1}$$

where $k_{\min} = \min\{k_s, k_t\}$.

In contrast, real-world networks typically have a heterogeneous connectivity pattern: Hub nodes with a very large degree as well as isolated nodes are simultaneously present. The following question then naturally arises: Does eq. (1) hold for such heterogeneous networks? To answer this question, we computed the maximum flow for each pair of nodes in random SF networks [10] using the MAX_FLOW algorithm [11]. Hundreds of network realizations were generated for given numbers of nodes N and links L, and the degree exponent γ, for which we were able to identify that the average of q over the node pairs that have $k_{\min}$ as smaller degree, $\langle q \rangle_{k_{\min}}$, satisfies a relation that is analogous to eq. (1):

$$\langle q \rangle_{k_{\min}} \simeq c\, k_{\min}, \tag{2}$$

for large $k_{\min}$, with a coefficient c less than 1 (see fig. 1(a)).

If we define the distribution of the maximum flow as $P(q) \equiv \langle \sum_{s \neq t} \delta_{q_{s,t},q} / (\sum_{s \neq t} 1) \rangle$ with $\langle \cdots \rangle$ denoting the ensemble average, it can be decomposed in terms of the smaller degree as $P(q) = \sum_k P_{\min}(k) P(q|k)$, where $P_{\min}(k) \equiv \langle \sum_{s \neq t} \delta_{\min\{k_s,k_t\},k} / (\sum_{s \neq t} 1) \rangle$ and $P(q|k) \equiv \langle \sum_{s \neq t} \delta_{q_{s,t},q}\, \delta_{\min\{k_s,k_t\},k} \rangle / \langle \sum_{s \neq t} \delta_{\min\{k_s,k_t\},k} \rangle$. The conditional distribution $P(q|k)$ is sharply peaked around the average $q = \langle q \rangle_k$ as shown in fig. 1(b), which implies $P(q) \approx P_{\min}(k = q/c)$
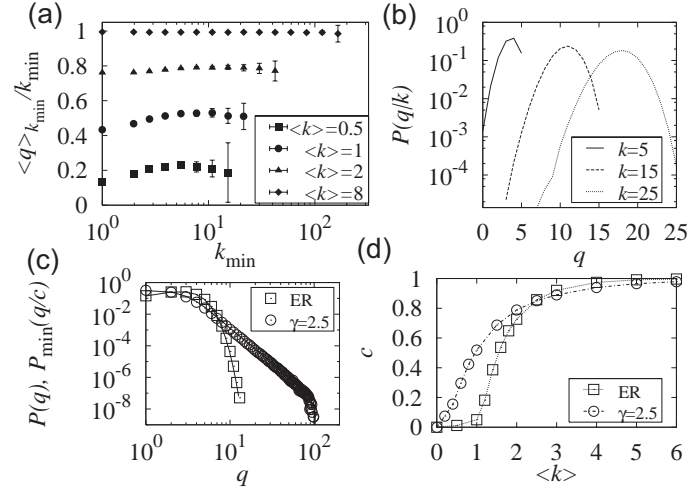
Fig. 1 – Maximum flow in model networks. (a) $\langle q \rangle_{k_{\min}}/k_{\min}$ *vs.* $k_{\min}$ in SF networks with $\gamma = 2.5$ and different values of $\langle k \rangle = 2L/N$. (b) Conditional distributions $P(q|k)$ in SF networks with $N = 1000$, $\langle k \rangle = 1.6$, and $\gamma = 2.5$. (c) Distribution of the maximum flow $P(q)$ for the ER graphs and SF networks with $\gamma = 2.5$, both consisting of 1000 nodes and 2000 links ($\langle k \rangle = 4$). Each line represents $P_{\min}(q/c)$ with $c \simeq 0.97$ used for the ER graphs and $c \simeq 0.94$ for the SF networks, which is in good agreement with $P(q)$. (d) Flow efficiency c *vs.* $\langle k \rangle$ for the ER graphs and SF networks with $\gamma = 2.5$. The statistical error is smaller than the data point size and $N = 1000$ in both networks.

with a good accuracy, where c is the coefficient appearing in eq. (2). $P_{\min}(k)$ is related to the degree distribution $P_d(k)$ by $P_{\min}(k) = 2P_d(k)\sum_{k' \geq k} P_d(k')$. Thus the asymptotic behavior of the degree distribution determines the large-q behavior of $P(q)$ (fig. 1(c)):

$$P(q) \sim e^{-(\text{const.})\, q \ln(q/\langle k \rangle)}, \tag{3}$$

for the Erdős-Rényi (ER) graph that have $P_d(k) \sim e^{-k \ln(k/\langle k \rangle)}$ [12], and

$$P(q) \sim q^{-(2\gamma-1)}, \tag{4}$$

for SF networks with the degree exponent γ. Recently, López *et al.* reported the same asymptotic behavior for the conductance distribution in complex networks [13].

Contrary to $D > 1$-dimensional regular lattices, even when $k_{\min}$ is large, the maximum flow in heterogeneous networks may be very small due to the collision of connecting paths at some critical links. For example, in fig. 2(a), the maximum flow between the node 2 and 14 is only 1 because every connecting path has to pass the link $(3, 13)$. The value of c in eq. (2) is thus a measure for how efficiently such collisions are avoided in the underlying path structure. From now on, we call the coefficient c *flow efficiency*, which depends on the network topology as shown in fig. 1(d). While c is higher with more links, its dependence on the degree exponent γ is non-trivial. The example in fig. 1(d) shows that c is higher in SF networks for $\langle k \rangle \lesssim 2.5$ while it is higher in ER graphs otherwise. This illuminates ambivalent effects of heterogeneity on transport. When links are abundant, most node pairs can be connected and have as many connecting paths as $k_{\min}$. What matters is then to avoid collisions of those connecting paths, which can be achieved more efficiently in networks closer to a regular one. On the contrary, when links are deficient but as many pathways as possible are required between each pair of connected nodes, concentration of links within hub nodes is preferable to forming a chain of nodes all of which have degree 2.
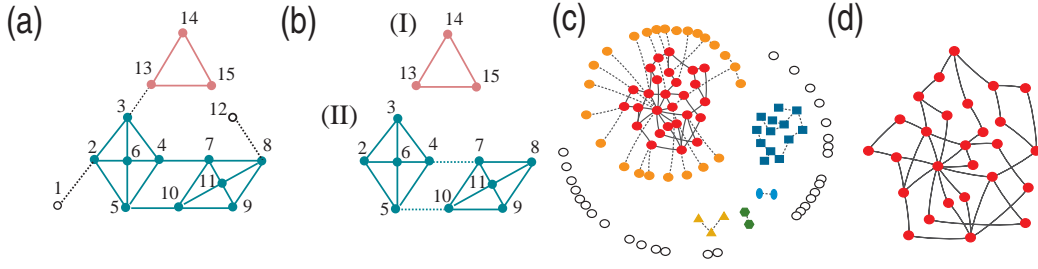
Fig. 2 – (Color online) *Bridge*, *separation pair*, and EBCC. (a) The links $(1, 2)$, $(8, 12)$ and $(3, 13)$ are *bridges* that would increase the number of CCs if removed [15]. The network in (a) has two EBCCs (I) and (II) shown in (b). The pair of links $\{(4, 7), (5, 10)\}$ is a *separation pair* that disconnects the EBCC (II). An example of the EBCC in a SF network is shown in (c), which has 100 nodes, 80 links, and $\gamma = 2.5$. It has 5 CCs that have more than one node and the largest CC has an EBCC presented in (d).

In the following we will show that the flow efficiency is related to particular topological properties of the network that can be described by various aspects of connectedness. One unit of flow can be sent from a node s to another node t if and only if a link e_{sv} incident to s and a link e_{tw} incident to t belong to the same connected component (CC) or, equivalently, both links are connected. A CC of a graph is a maximal subgraph in which at least one connecting path exists between each pair of nodes. Next, the second unit of flow requires another path that does not share any link with the first path. This condition is fulfilled if and only if at least two links e_{sv} and $e_{sv'}$ that are incident to s and at least two links e_{tw} and $e_{tw'}$ incident to t belong to the same edge-biconnected component (EBCC) or, equivalently, there exist two pairs of edge-biconnected links, (e_{sv}, e_{tw}) and $(e_{sv'}, e_{tw'})$. An EBCC is a maximal subgraph which cannot be disconnected by removing a single link. Each node of a graph either belongs to a unique EBCC or does not belong to any EBCC.

One could now naively expect that $k > 2$ units of flow between s and t require that s and t belong to the same k-edge-connected component (a k-edge-connected graph is a maximal subgraph which cannot be disconnected by removing any $(k-1)$ links [14] and thus, a graph is $(k-1)$-edge-connected if it is k-edge-connected). This is a sufficient condition, but not a necessary one for $k > 2$ units of flow: The nodes on k disjoint paths connecting s and t may have degree 2 and thus the nodes s and t may belong to different k-edge-connected components even though they have $k > 2$ link-disjoint paths.

It turns out that already the structure of the connected and the biconnected components determines the existence of flow values $q_{s,t} > 1$ between two nodes s and t. Obviously for $q_{s,t} \geq 1$, s and t must belong to the same CC. For $q_{s,t} > 1$, s and t should have two pairs of edge-biconnected links. When s and t belong to the same CC but to different EBCCs, the maximum flow will only be one, $q_{s,t} = 1$. There exist one or more *bridges* [15] between different EBCCs belonging to the same CC, which would disconnect the EBCCs if removed (fig. 2). Analogously if there exists a *separation pair* of links that disconnect s and t if removed (fig. 2), $q_{s,t}$ will not be larger than 2 even when s and t belong to the same EBCC and $k_{\min} > 2$.

To study this relation between flow and topology statistically, we determined all CCs and EBCCs in a given network and defined for all node pairs (s, t)

$$\Theta_{s,t} \equiv \sum_{n=1}^{N_c} \min\{f_s^{(n)}, f_t^{(n)}\}, \qquad \Sigma_{s,t} \equiv \sum_{m=1}^{N_{ebc}} \min\{g_s^{(m)}, g_t^{(m)}\}. \tag{5}$$

Here N_c (N_{ebc}) is the total number of CCs (EBCCs) in a network, $f_i^{(n)}$ ($g_i^{(m)}$) is the number
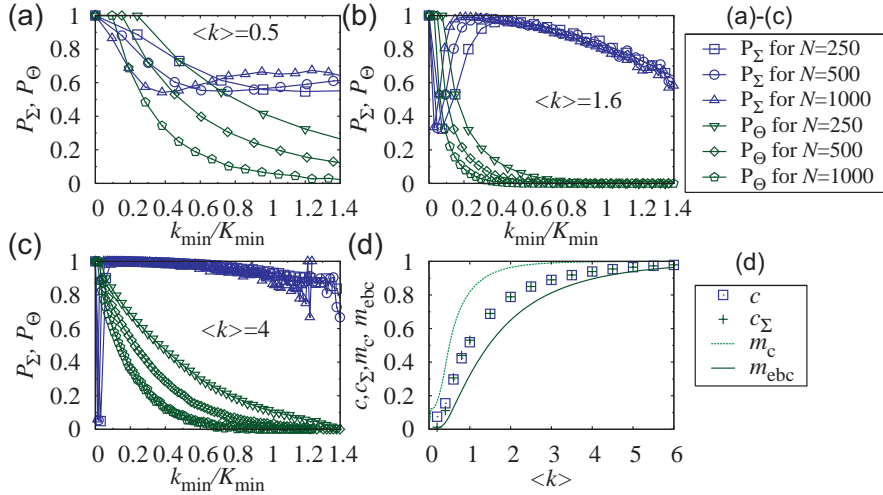
Fig. 3 – (Color online) Relation between the maximum flow and edge-biconnectedness. (a)-(c) Conditional probabilities P_Σ and P_Θ *vs.* $k_{\min}/K_{\min}$ in SF networks with $\gamma = 2.5$ and (a) $\langle k \rangle = 0.5$, (b) $\langle k \rangle = 1.6$, and (c) $\langle k \rangle = 4$. Here $K_{\min}$ is the ensemble-averaged largest value of $k_{\min}$. Note that P_Σ approaches one for increasing average connectivity $\langle k \rangle$. (d) Plots of c, c_Σ, m_c, and m_{ebc} *vs.* $\langle k \rangle$ for SF networks with $\gamma = 2.5$ and $N = 1000$.

of links incident to node i and belonging to the n-th CC (m-th EBCC). $\Theta_{s,t}$ ($\Sigma_{s,t}$) counts the number of disjoint pairs of connected (edge-biconnected) links incident to s and t. While $f_s^{(n)}$ ($f_t^{(n)}$) is either 0 or k_s (k_t), $g_s^{(m)}$ ($g_t^{(m)}$) can take any integer between 0 and k_s (k_t) since each link incident to a node may belong to an EBCC or not. The deviation of $q_{s,t}$ from $\Theta_{s,t}$ or $\Sigma_{s,t}$ is therefore a measure of the number of critical links, *bridges* or *separation pairs*.

We computed the conditional probability P_Θ that $q_{s,t}$ is equal to $\Theta_{s,t}$ and P_Σ that $q_{s,t}$ is equal to $\Sigma_{s,t}$ for given values of $k_{\min}$ in various model networks. It turns out that $\Sigma_{s,t}$ is in good agreement with $q_{s,t}$ for large values of $k_{\min}$, manifested via values for P_Σ close to one. In SF networks with $\gamma = 2.5$, as shown in figs. 3(a)-(c), P_Σ is larger than 0.5 ($\langle k \rangle = 0.5$), 0.8 ($\langle k \rangle = 1.6$), and 0.9 ($\langle k \rangle = 4$) for a wide range of $k_{\min}/K_{\min} \lesssim 1$ independent of the value of N. $K_{\min}$ denotes the ensemble-averaged largest value of $k_{\min}$ that can be observed in networks of N nodes, and is given by [10] $K_{\min} = \langle k \rangle (\gamma - 2)(N/2)^{1/(\gamma-1)}/(\gamma - 1)$. Whereas P_Σ approaches one for a range of $k_{\min}$ that broadens with increasing average connectivity $\langle k \rangle$, the probability P_Θ is very small except for the regime where $k_{\min}$ is small. The agreement of $q_{s,t}$ and $\Sigma_{s,t}$ for large $k_{\min}$ is observed also for other values of $\gamma > 2$ and ER graphs ($\gamma \to \infty$).

These results elucidate non-trivial features of the paths connecting hub nodes. Most dangerous links that may prevent large flow between hub nodes are *bridges*. If two hub nodes belong to the same EBCC, they can send and receive a flow nearly as large as the number of links belonging to that EBCC. In other words, *separation pairs* are very rare in the EBCC to which hub nodes belong. The agreement between the maximum flow and the number of pairs of edge-biconnected links can be of importance in practical aspects as well: The algorithm to compute EBCCs has running time $\mathcal{O}(N + L)$ [15] while the maximum flow takes $\mathcal{O}(N^3)$ time in sequential machines [16].

The agreement of $q_{s,t}$ and $\Sigma_{s,t}$ allows a deeper understanding of the relation (2). Considering that $\Sigma_{s,t}$ is dominated by the giant EBCC, defined here as the EBCC that has a $\mathcal{O}(L)$ links, one can evaluate $\langle \Sigma \rangle_{k_{\min}}$, the average of Σ over the node pairs having $k_{\min}$ as
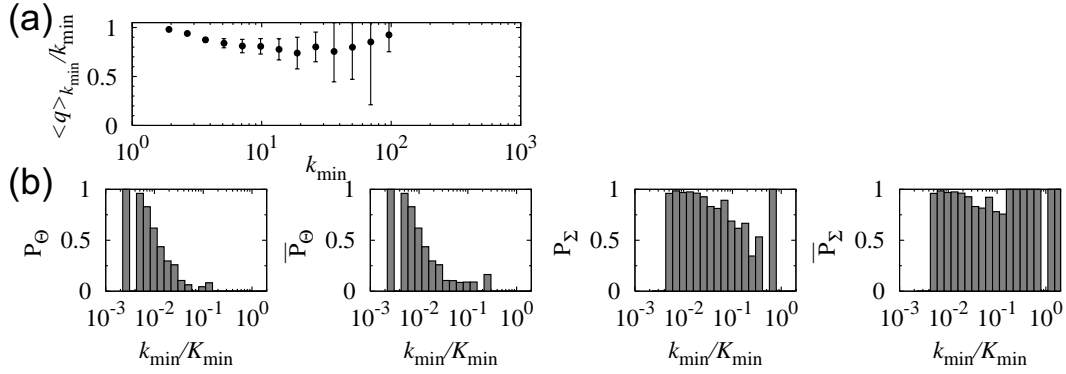
Fig. 4 – Maximum flow for the autonomous system network of the Internet with $N = 5238$, $L = 9993$, and $\gamma = 2.2(1)$. (a) $\langle q \rangle_{k_{\min}}/k_{\min}$ as a function of $k_{\min}$. It fluctuates around $c = 0.79(6)$, which is consistent with the relative size of the giant EBCC $m_{\mathrm{ebc}} \simeq 0.80$. (b) Plots of conditional probabilities P_Θ, $\overline{P}_\Theta$, P_Σ, and $\overline{P}_\Sigma$ vs. $k_{\min}/K_{\min}$ with $K_{\min} \simeq 449$.

their smaller degree. Consider the relative size of the giant EBCC m_{ebc} defined as the ratio (Number of links in the giant EBCC)/(Total number of links). Assuming that the links of a node participate in the giant EBCC statistically independently of one another, the number of links that belong to the giant EBCC, of s and t, denoted by $g_s^{(1)}$ and $g_t^{(1)}$, respectively, follow binomial distributions $B(k_s, m_{\mathrm{ebc}})$ and $B(k_t, m_{\mathrm{ebc}})$ respectively, giving

$$
\langle \Sigma \rangle_{k_{\min}, k_{\max}} \simeq \sum_{g_1=0}^{k_{\min}} \sum_{g_2=0}^{k_{\max}} \binom{k_{\min}}{g_1} \binom{k_{\max}}{g_2} m_{\mathrm{ebc}}^{g_1+g_2} \times
$$
$$
\times \ (1 - m_{\mathrm{ebc}})^{k_{\min}+k_{\max}-g_1-g_2} \min\{g_1, g_2\}, \tag{6}
$$

where $k_{\max} = \max\{k_s, k_t\}$ and $\langle \cdots \rangle_{x,y}$ means the restricted average over the pairs of nodes that have $\min\{k_s, k_t\} = x$ and $\max\{k_s, k_t\} = y$. Dominant contributions to the summations in eq. (6) come from the regime $(g_1, g_2) \simeq (m_{\mathrm{ebc}} k_{\min}, m_{\mathrm{ebc}} k_{\max})$, where $\min\{g_1, g_2\} = m_{\mathrm{ebc}} k_{\min}$. Thus we have

$$
\langle \Sigma \rangle_{k_{\min}} \simeq m_{\mathrm{ebc}} \, k_{\min}. \tag{7}
$$

Equation (7) and the agreement of $q_{s,t}$ and $\Sigma_{s,t}$ provide the basis from which eq. (2) follows. In accordance with eq. (7), we indeed observed that $\langle \Sigma \rangle_{k_{\min}} \simeq c_\Sigma \, k_{\min}$ for large $k_{\min}$, and the values of c_Σ and c are compared with m_{ebc} in fig. 3(d). Although m_{ebc} is not exactly equal to c or c_Σ, both of which are in excellent agreement, because of the correlations among different links participating in the giant EBCC, the deviation is so small that m_{ebc} can be a guide to the value of c. We can also consider m_c defined as the ratio (Number of links in the giant CC)/(Total number of links), but it turns out to deviate considerably from the value of c or c_Σ as shown in fig. 3(d).

Finally, we have tested our findings in a real network, the Internet at the autonomous system (AS) level. An AS includes a set of routers following a common routing strategy and the Internet may be viewed as a network of the ASs [17]. Here we used the data recorded on a particular date (June 15, 1999) and present the results in fig. 4. The network consists of 5238 nodes and 9993 links ($\langle k \rangle \simeq 3.82$) and the degree exponent is $\gamma = 2.2(1)$. All edge capacities are set to one. While all the nodes are connected with one another ($m_c = 1$), only 3251 nodes and 8006 links belong to the giant EBCC, giving $m_{\mathrm{ebc}} \simeq 0.80$. We found that the

relation in eq. (2) is observed and furthermore, the flow efficiency $c = 0.79(6)$ is very close to the value of m_{ebc}. The conditional probability P_Σ is shown to be very high in contrast to P_Θ. Additionally, we computed the relaxed probabilities, $\overline{P}_\Sigma$ and $\overline{P}_\Theta$, defined as the probability that $|1 - q_{s,t}/\Sigma_{s,t}| < 0.05$ and the probability that $|1 - q_{s,t}/\Theta_{s,t}| < 0.05$, respectively. $\overline{P}_\Sigma$ is near one for most values of $k_{\min}$, but $\overline{P}_\Theta$ differs only slightly from P_Θ.

In conclusion, we studied the relation between transport, specifically the maximum flow, and the structural organization in complex networks. The path structure of complex networks with heterogeneous connectivity is efficiently organized such that nodes with more links can send and receive larger amount of flow as expressed in eq. (2). This may explain the abundance of heterogeneous connectivity pattern in complex systems that might have evolved towards better performance of transportation and communication among their constituents. The structure of edge-biconnected components determines essentially the maximum flow between hub nodes, which therefore can be estimated in a time that is linear in the total number of nodes.

$$* \, * \, *$$

REFERENCES

[1] Watts D. J. and Strogatz S. H., *Nature*, **393** (1998) 440.
[2] Albert R., Jeong H. and Barabási A.-L., *Nature*, **401** (1999) 130.
[3] Huberman B. A. and Adamic L. A., *Nature*, **401** (1999) 131.
[4] Jeong H., Tombor B., Albert R., Oltvai Z. N. and Barabási A.-L., *Nature*, **407** (2000) 651.
[5] Albert R. and Barabási A.-L., *Rev. Mod. Phys.*, **74** (2002) 47.
[6] Dorogovtsev S. N. and Mendes J. F. F., *Adv. Phys.*, **51** (2002) 1079.
[7] Newman M. E. J., *SIAM Rev.*, **45** (2003) 167.
[8] Ahuja R. K., Magnanti T. L. and Orlin J. B., *Network Flows: Theory, Algorithms, and Applications* (Prentice-Hall, Englewood Cliffs, NJ) 1993.
[9] Hartmann A. K. and Rieger H., *Optimization Algorithms in Physics* (Wiley-VCH, Berlin) 2002.
[10] Lee D.-S., Goh K.-I., Kahng B. and Kim D., *Nucl. Phys. B*, **696** (2004) 351.
[11] LEDA is available in `http://www.algorithmic-solutions.com`.
[12] Erdős P. and Rényi A., *Publ. Math. Inst. Hung. Acad. Sci.*, **5** (1960) 17; *Bull. Inst. Int. Stat.*, **38** (1961) 343.
[13] López E., Buldyrev S. V., Havlin S. and Stanley H. E., *Phys. Rev. Lett.*, **94** (2005) 248701.
[14] Gross J. L. and Yellen J., *Graph Theory and Its Applications* (CRC Press, Boca Raton) 1999, p. 176.
[15] Gabow H. N., Technical Report CU-CS-890-99 (Department of Computer Science, University of Colorado at Boulder) (2000).
[16] Goldberg A. V. and Tarjan R. E., *J. ACM*, **35** (1988) 921.
[17] Yook S.-H., Jeong H. and Barabási A.-L., *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002) 13382.