

Scaling theory of transport in complex biological networks

Lazaros K. Gallos*, Chaoming Song*, Shlomo Havlin†, and Hernán A. Makse**

*Levich Institute and Physics Department, City College of New York, New York, NY 10031; and †Minerva Center and Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved March 5, 2007 (received for review January 10, 2007)

Transport is an important function in many network systems and understanding its behavior on biological, social, and technological networks is crucial for a wide range of applications. However, it is a property that is not well understood in these systems, probably because of the lack of a general theoretical framework. Here, based on the finding that renormalization can be applied to bionetworks, we develop a scaling theory of transport in self-similar networks. We demonstrate the networks invariance under length scale renormalization, and we show that the problem of transport can be characterized in terms of a set of critical exponents. The scaling theory allows us to determine the influence of the modular structure on transport in metabolic and protein-interaction networks. We also generalize our theory by presenting and verifying scaling arguments for the dependence of transport on microscopic features, such as the degree of the nodes and the distance between them. Using transport concepts such as diffusion and resistance, we exploit this invariance, and we are able to explain, based on the topology of the network, recent experimental results on the broad flow distribution in metabolic networks.

metabolic networks | modularity | protein-interaction networks

Transport in complex networks is a problem of much interest in many aspects of biology, sociology, and other disciplines. For example, the study of metabolic fluxes in organisms is crucial for a deeper understanding of how the cell carries its metabolic cycle (1). The use of metabolic flux analysis can provide important cellular physiological characteristics by using the network stoichiometry and predict optimal flux distributions that satisfy a defined metabolic objective. Similarly, information flow between the molecules of a biological network provides insight for both the network structure and the functions performed by the network. Such an example is the concept of the “diffusion distance” in a protein–protein interaction network which is used to predict possible interactions between proteins, simply by studying diffusion in the existing network (2). In food webs, energy transfer between different levels of the web is crucial for the organism survival, whereas spreading of a disease between different organisms may affect the regular operation of the equilibrated system. Moreover, applications of transport in complex networks extend to a plethora of other systems, such as, for instance, transport of information in the Internet, spreading of diseases and/or rumors in social networks, etc. Despite its significance, the laws of transport in such a complex substrate are yet unclear compared with transport in random media (3). This is due to the complexity added by the heterogeneous degree distribution in such networks.

We study transport in real-world biological networks and via a model, which possess both self-similar properties and the scale-free character in their degree distribution. We explain our results with theoretical arguments and simulation analysis. We use approaches from renormalization theory in statistical physics that enable us to exploit the self-similar characteristics of the fractal networks and develop a scaling theory of transport, which we use to address the effects of the modularity and the degree inhomogeneity of the substrate.

Because of the existence of a broad degree distribution, transport on a network is different when it is between two hubs with a large number of connections k or between low-degree nodes (4). We therefore characterize the transport coefficients by their explicit dependence on k_1 , k_2 , and ℓ , where k_1 and k_2 denote the degree of two nodes ($k_1 > k_2$), separated by a distance ℓ (distance is measured by the minimum number of links, i.e., it is the chemical distance). We study the diffusion time $T(\ell; k_1, k_2)$ and the resistance $R(\ell; k_1, k_2)$ between any two nodes in the system. The dependence on k_1 and k_2 is not significant in homogeneous systems, but is important in networks where the node degree spans a wide range of values, such as in biological complex networks. In fact, this dependence is critical for many other properties as has been already shown for e.g., fractality, where traditional methods of measuring the fractal dimension may fail because they do not take into account this inhomogeneity (5).

Modularity is one of the most important aspect of these networks with direct implications to transport properties. Here, we quantify the modular character of complex networks according to our box-covering algorithm and reveal a connection between modularity and flow. Our results are consistent with recent experiments and metabolic flux studies, and provide a theoretical framework to analyze transport in a wide variety of network systems.

Metabolism Modeling. In metabolism modeling, there exist three main approaches (6). (i) The most detailed analysis includes dynamic mechanism-based models (7), but in general it is very difficult to incorporate experimental values for the needed kinetic parameters. (ii) In the second approach one simplifies the above models, and calculates the fluxes in a metabolic network via flux balance analysis, which includes a family of static constraint-based models (8). The limiting factor in this analysis is that the problem is underconstrained (the number of unknown parameters, i.e., the fluxes, is larger than the number of metabolite conservation equations) and cannot be solved uniquely. (iii) Finally, a third approach that is widely used in metabolism modeling is to ignore stoichiometry, and focus only on the metabolites’ interactions without any thermodynamic aspects, which leads to interaction-based models (9), i.e., undirected networks where a link connects two nodes that participate in a metabolic reaction. In this article, we follow this third approach, and we use this interaction-based network to study transport on this network by drawing an analogy between the metabolic

Author contributions: L.K.G., C.S., S.H., and H.A.M. designed research; L.K.G. and C.S. performed research; L.K.G., C.S., S.H., and H.A.M. analyzed data; and L.K.G., C.S., S.H., and H.A.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviation: PIN, protein-interaction network.

†To whom correspondence should be addressed. E-mail: hmakse@levdec.engr.cuny.cuny.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0700250104/DC1.

© 2007 by The National Academy of Sciences of the USA

network and a resistance network. If we represent the metabolites as nodes that are linked through electrical resistances and the current flow represents the flux, we can solve this problem without additional constraints, and this solution may shed additional light on the involved processes. The advantage of our approach is that it can isolate the topological effect, and we can address a broader aspect of transport in biological networks, such as whether the observed flux inhomogeneity is because of the network topology itself or because of the adopted flux constraints. Moreover, this approach enables us to carry similar studies for diffusion on such networks.

Modularity, Diffusion, and Resistance. In our work, we focus on two different examples of biological networks, namely the *Escherichia coli* metabolic network (9) and the yeast protein-interaction network (PIN) (10). We analyze the filtered yeast interactome developed by Han *et al.* (10), which removes a large number of false positives in high-throughput yeast two hybrid methods [see supporting information (SI)]. Both networks have been shown to have fractal properties and can be covered with $N_B(\ell_B)$ nonoverlapping boxes, where in each box the maximum distance between any two nodes is less than ℓ_B , the maximum distance in a box (5). For a fractal network of N nodes, N_B follows a power-law dependence on ℓ_B ,

$$N_B(\ell_B)/N \approx \ell_B^{-d_B} \quad [1]$$

and defines d_B as the fractal (or box) dimension of a network. These networks are also self-similar, i.e., their main properties, such as the degree distribution, remain invariant under a renormalization scheme where each box is replaced by a (super) node and links between boxes are transferred to the nodes of the renormalized network (see e.g., the example in Fig. 1a for a network G , tiled with $\ell_B = 3$ boxes, that yields the network G'). Many biological networks in the intermediate renormalized stages were shown to have similar properties as the original network.

This renormalization procedure also implies the presence of self-similar modularity in all length-scales, which is a central feature of these networks. The term modularity refers to the existence of sets of nodes whose links are connected preferably within this set rather than to the rest of the network. Thus, after tiling a network for a given value of ℓ_B , we introduce a measure of modularity for a network as

$$M(\ell_B) = \frac{1}{N_B} \sum_{i=1}^{N_B} \frac{L_i^{\text{in}}}{L_i^{\text{out}}}, \quad [2]$$

where L_i^{in} and L_i^{out} represent the number of links that start in a given box i and end either within or outside i , respectively. Large values of M correspond, thus, to a higher degree of modularity. Because the numerical value of $M(\ell_B)$ varies, though, with ℓ_B , a more reliable measure is the modularity fractal exponent d_M , which we define through:

$$M(\ell_B) \approx \ell_B^{d_M}. \quad [3]$$

The value of $d_M = 1$ represents the borderline case that separates modular ($d_M > 1$) from random nonmodular ($d_M < 1$) networks. For a lattice structure, the value of d_M is exactly equal to $d_M = 1$.

In ref. 5, we had introduced a fractal network model where a network grows by adding m new offspring nodes to each existing network node, resulting in well defined modules. In that version of the model, modules are connected to each other through $x = 1$ links, which leads to a tree structure. A generalization of this model (presented in detail in SI) allows us to tune the degree of modularity in the network by assigning a larger number of links $x > 1$ between modules. Although for $x = 1$ all of the modules

are well defined, increasing x leads to the presence of loops and to a progressive merging of modules, so that for large x values a node cannot be assigned unambiguously in a given module and modularity is destroyed. A straightforward analytical calculation in this case leads to (see SI)

$$d_M = \frac{\ln(2\frac{m}{x} + 1)}{\ln 3}. \quad [4]$$

In this article, we use this value of d_M for the model and calculate d_M for real networks to study the influence of modularity on network transport.

In general, the problem of transport is expressed in terms of $T(\ell; k_1, k_2)$ and $R(\ell; k_1, k_2)$, where T is the average first passage time needed by a random walker to cover the distance ℓ between two nodes with degrees k_1 and k_2 , respectively, and R represents the resistance between these two nodes. For homogeneous systems [with very narrow degree distribution $P(k)$] such as lattices and regular fractals, there is no dependence on k_1 and k_2 , and the average is only over the distance ℓ . One of the goals of this work is to find the scaling of T and R in heterogeneous networks with a broad degree distribution and self-similar properties.

In the general case of a renormalizable network, T and R scale in the renormalized network G' (the primes always denote a quantity for the renormalized network) as

$$T'/T = \ell_B^{-d_w}, \quad R'/R = \ell_B^{-\zeta}. \quad [5]$$

The exponents d_w and ζ are the random walk exponent and the resistance exponent, respectively. This equation is valid as an average of T and R over the entire system, applying for example in different generations when growing or renormalizing a fractal object. Thus, this relation holds true for both homogeneous and inhomogeneous systems.

For homogeneous systems, the above exponents d_w and ζ are related through the Einstein relation (3)

$$d_w = \zeta + d_B, \quad [6]$$

where d_B is the fractal dimension of the substrate on which diffusion takes place. This relation is a result of the fluctuation-dissipation theorem relating spontaneous fluctuations (diffusion) with transport (resistivity) and the underlying structure (dimensionality) (3). Although the validity of this relation for scale-free networks is not yet clear, our following analysis shows that it also applies for these systems as well.

Renormalization and Scaling Theory. The renormalization procedure on self-similar biological networks provides an as yet unexplored method for estimation of the dynamical exponents in these systems. Because such a network is left invariant after substituting all nodes in a box with a single node, we can calculate the transport properties on the networks during successive renormalization stages. With this method, we can also study transport in biological networks before (R and T) and after renormalization (R' and T'). The results are presented in Fig. 1 for the yeast PIN, the *E. coli* metabolic network, and the fractal network model with $d_M = 1.46$ (tree, highly modular) and $d_M = 1.26$ (network with loops and lower modularity). Fig. 1 *b* and *c* suggest a linear relation between R' and R (T' and T) for a given value of ℓ_B , so that the ratio R'/R (and T'/T) is almost constant for all boxes in the system for this ℓ_B value. For a different value of the box diameter ℓ_B , this ratio is again constant for all boxes in the network but assumes a different value. We can plot the values of this ratio as a function of the network size ratio N'/N for different values of ℓ_B (Fig. 1 *d* and *e*). The data indicate the

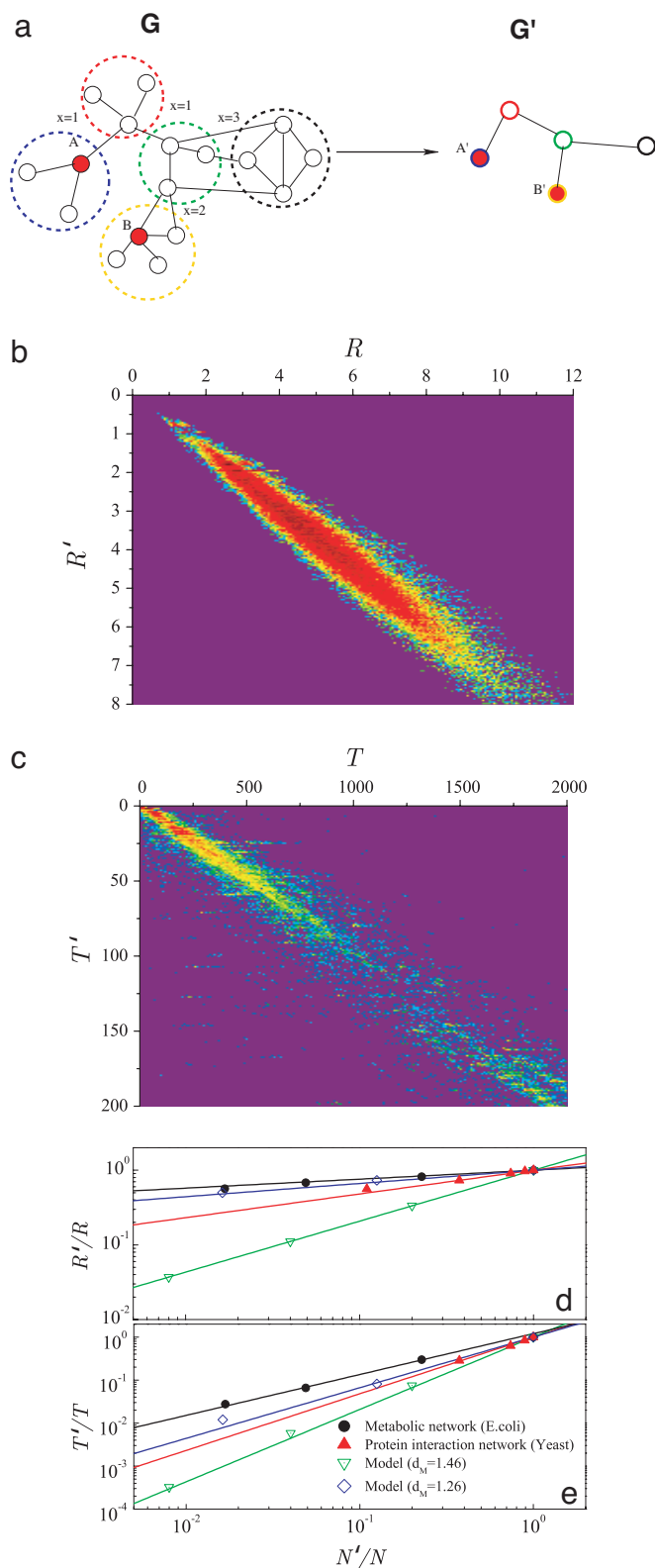


Fig. 1. Scaling of resistance and transport time in complex networks. (a) Example of a network G that undergoes renormalization to a network G' . For this example, the maximum possible distance between any two nodes in the same box has to be less than $\ell_B = 3$. We compare the resistance $R(A,B)$ and diffusion time $T(A,B)$ between the two local-hub nodes A and B in G to the resistance $R'(A',B')$ and $T'(A',B')$ of the renormalized network G' . The values of R' and T' are found to always be proportional to R and T , respectively, for all pairs of nodes. (b and c) Typical behavior of the probability distributions for the resistance R vs. R' and the

existence of a power-law dependence, and a comparison with the model networks shows that the results are consistent with PIN exhibiting a more modular structure compared with the metabolic network.

Although, in principle, we can use our numerical results to directly calculate the exponents d_w and ζ through Eq. 5, this method is not practical because the variation of ℓ_B is very small. We can overcome this difficulty by using the system size N instead, where we combine Eqs. 1 and 5 to get

$$\frac{T'}{T} = \left(\frac{N'}{N}\right)^{d_w/d_B}, \quad \frac{R'}{R} = \left(\frac{N'}{N}\right)^{\zeta/d_B} \quad [7]$$

Thus, the slopes in Fig. 1 *d* and *e* correspond to the exponent ratios d_w/d_B and ζ/d_B , respectively.

Notice also that the verification of the above equation through Fig. 1 validates the relation in Eq. 5 for inhomogeneous systems. The numerical values for the calculated exponents are shown in Table 1. These ratios are consistent in all cases, within statistical error, with the Einstein relation, Eq. 6.

Using these scaling arguments and the renormalization property of these networks we next predict the dependence of both R and T on the distance ℓ between two nodes and their corresponding degrees k_1 and k_2 . After renormalization, the network becomes smaller, so that both the degrees and the distances in the network decrease. A distance ℓ in G is scaled by a factor ℓ_B in G' , so that $\ell' = \ell/\ell_B$, whereas in earlier work (5), it had been shown that the degree k of the largest hub in a box transforms to a degree k' for the renormalized box, where $k' = \ell_B^{-d_k} k$, and d_k is an exponent describing the scaling of the degree. According to the result of Fig. 1 and Eq. 5, it follows,

$$R'(\ell'; k'_1, k'_2) = \ell_B^{-\zeta} R(\ell; k_1, k_2) \quad [8]$$

$$T'(\ell'; k'_1, k'_2) = \ell_B^{-d_w} T(\ell; k_1, k_2). \quad [9]$$

Using dimensional analysis (see SI), we can show that

$$R(\ell; k_1, k_2) = k_2^{\zeta/d_k} f_R\left(\frac{\ell}{k_2^{1/d_k}}, \frac{k_1}{k_2}\right) \quad [10]$$

$$T(\ell; k_1, k_2) = k_2^{d_w/d_k} f_T\left(\frac{\ell}{k_2^{1/d_k}}, \frac{k_1}{k_2}\right), \quad [11]$$

where $f_R()$ and $f_T()$ are undetermined functions. In the case of homogeneous networks where there is almost no k -dependence, these functions reduce to the forms $f_R(x,1) = x^\zeta$, $f_T(x,1) = x^{d_w}$, leading to the classical relations $R \approx \ell^\zeta$ and $T \approx \ell^{d_w}$.

The scaling in Eqs. 10 and 11 is supported by the numerical data collapse shown in Fig. 2. For the data collapse, we used the values of the exponents ζ and d_w as obtained from the renormalization method above (Table 1) confirming the scaling in Eqs. 10 and 11.

The functions f_R and f_T introduced in Eqs. 10 and 11 have two arguments, so we first need to fix the ratio k_1/k_2 and in the plot (Fig. 2), we present different ratios using different symbols. We

diffusion time T vs T' , respectively, for a given ℓ_B value. Similar plots for other ℓ_B values verify that the ratios of these quantities during a renormalization stage are roughly constant for all pairs of nodes in a given biological network. (d and e) We present the average value of this ratio for the resistance R/R' and the diffusion time T/T' , respectively, as measured for different ℓ_B values (each point corresponds to a different value of ℓ_B). Results are presented for both biological networks, and two fractal network models with different d_M values. The slopes of the curves correspond to the exponents ζ/d_B in *d* and d_w/d_B in *e* (see Table 1).

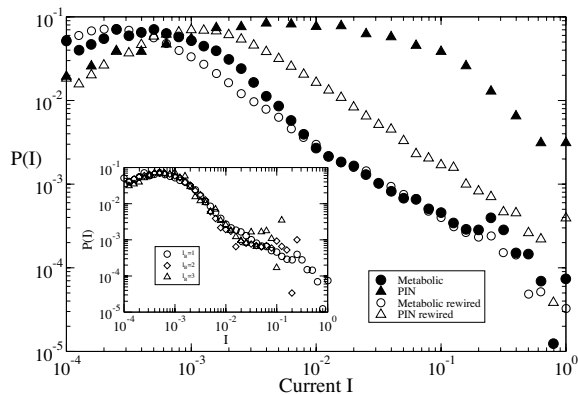


Fig. 4. Probability distribution $P(I)$ of current magnitudes I flowing through the links in PIN (filled triangles) and metabolic networks (filled circles). Open symbols are the corresponding results for the randomly rewired networks. (Inset) Invariance of $P(I)$ for the metabolic network under renormalization with different ℓ_B values.

of offspring m . For each pair of m and x , we calculate numerically the exponent d_w from the slope of figures similar to Fig. 1 *d* and *e* and use Eq. 4 for the value of d_M . The results are fully consistent with Eq. 16, and all of the points lie on the predicted line. Subdiffusion ($d_w > 2$) is observed for $d_M > 1$, in accordance with our observation that modularity slows down diffusion. On the contrary, for nonmodular networks diffusion is accelerated remarkably ($d_w < 2$), which is also in agreement with previous work on random networks. When $d_M = 1$, we recover classical diffusion ($d_w = 2$), even though the structure is still that of a scale-free network. For the biological networks, we find that the PIN network follows very closely the proposed scaling relation, whereas the metabolic does not. This indicates that the model captures very well the modular structure of the PIN, whereas more structure is found in the metabolic compared with the above model.

Flow Distribution Across the Network. In our scaling theory above, we derived results for the average values of the current flowing in a complex network. The inherent inhomogeneity and modularity of biological networks is expected, however, to strongly influence the distribution of flow throughout the network. Using flux balance analysis, it was recently shown that the distribution of fluxes in the metabolic network is highly uneven, and a small number of reactions have the largest contribution to the overall metabolic flux

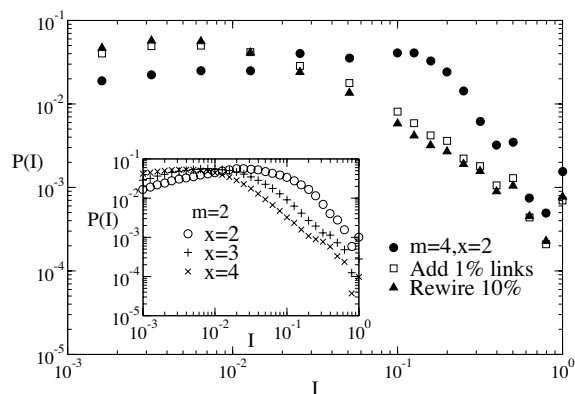


Fig. 5. Probability distribution $P(I)$ for the fractal model before and after randomly adding 1% of links or rewiring 10% of the network. (Inset) $P(I)$ for the fractal model with varying d_M values, where $m = 2$, and x varies from 2 to 4.

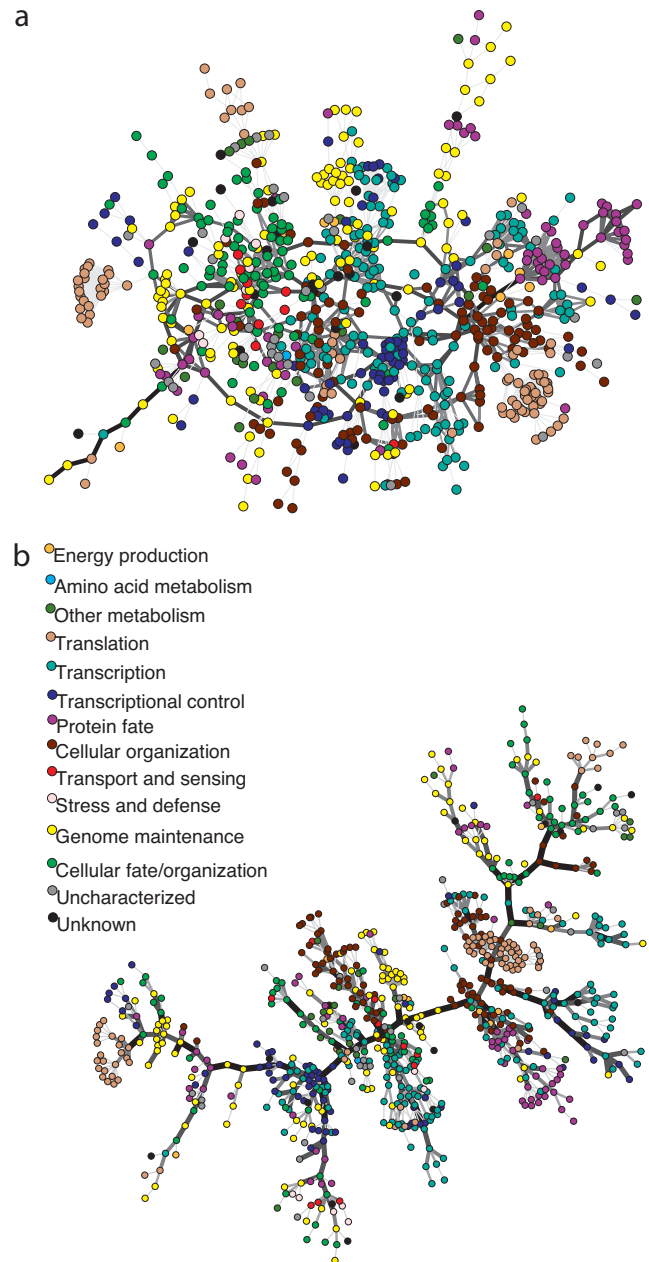


Fig. 6. Schematic illustration of flow in the PIN. (a) Current flow through the links of the yeast PIN network for one random selection of the two nodes acting as current input/output. (b) Minimum spanning tree for the PIN. The thickness of a link corresponds to the current flowing through this link. Different node colors correspond to different protein functions.

activity (1). To study the influence of the complex substrate on the flow distribution we calculate the probability $P(I)$ of current I_{ij} flowing between the two neighboring nodes i and j .

The probability distribution $P(I)$ for the magnitude of the current I across a link of the metabolic network decays according to a power law (see Fig. 4). The form of the curve and the exponent in the range 1.0–1.5 of the decay is very similar to those found in previous studies of the metabolic flux, both experimental (12) and theoretical (1). The decay suggests that only a small fraction of link carries high current. For the yeast PIN, the distribution is even broader, and its form is different from the metabolic network. The variation of $P(I)$ is smaller in PIN compared with the metabolic network, meaning that in PIN

there is a larger number of important links that carry large currents. The self-similar character of the biological networks is also verified for the distribution $P(I)$ as well, which remains invariant after renormalizing the network, as seen in Fig. 4 *Inset*.

Next we explore the connection between $P(I)$ and topology. The information contained in $P(I)$ can be better understood if we compare these results with a surrogate random case. The random case is obtained by rewiring the original networks, preserving the degree distribution $P(k)$, but destroying any correlations between neighboring nodes. Thus, we remove all traces of the initial network organization. The distribution $P(I)$ for the metabolic network remains almost the same under this rewiring, indicating that, despite the modular character of the network, the original structure behaves similar to uncorrelated networks, because the degree correlations do not affect $P(I)$ in the metabolic network. In contrast, the distribution $P(I)$ for the rewired PIN is very different from the original distribution and is similar to that of the metabolic network. We can, thus, conclude that the original PIN has a much richer structure that deviates from the random case, corroborating thus the results on modularity from the previous section on diffusion.

The above results for the biological networks can be understood in terms of the fractal network model, where we can control the number of links that form loop structures in a network. In Fig. 5, we calculate $P(I)$ for the model network with $m = 4$ and $x = 2$, which is a highly modular structure ($d_M = 1.46$). The form of the $P(I)$ distribution is similar to that of the PIN, but if we add a small number of random links (or equivalently rewire a small part of the network), this distribution is significantly influenced in a similar way as observed in random PIN rewiring. This suggests again that in the case of PIN, modularity is high. In Fig. 5 *Inset* we can also see that as x increases, i.e., more loops appear in the structure, the distribution has a longer tail, which shows that there is a smaller number of high-current links.

Because the number of added links in Fig. 5 is small, the modularity is preserved. We verified that the main factor that influences $P(I)$ is the number of loops in the network, rather than modularity itself, by fixing d_M and only varying x . In this case (described in SI), the $P(I)$ distribution is different for networks with the same d_M exponent. This can also be seen through Eq. 13, where the resistance exponent depends only on x .

Using the information of I_{ij} , we can also construct the “backbone” of the network in the form of the minimum spanning tree (MST) (13). The importance of such a tree is that it identifies the substructure of the network that is dominant for transport. Starting from a completely empty network, we insert links in decreasing order of current magnitude, provided that they do not form a loop. The resulting MST tree for the PIN is presented in Fig. 6, where the thickness of the links in the drawing increases logarithmically with increasing current through a link. The color of a node corresponds to the function performed by a protein in the network. This tree (created solely on the base of current flowing through a network) exhibits a large degree of modularity where nodes that perform similar functions are close to each other. It is also possible through this construction to identify the most critical links in the network in terms of the largest current flowing through them.

Thus, the emerging picture from the above analysis for the PIN is one of a network with a strong backbone that carries most part of the flow combined with loops organized mainly within

modules, so that flow through this backbone is not really influenced. This result highlights the strong modularity in the PIN structure. If a structure has a smaller degree of internal organization, as is the case in the metabolic network, then flow is more uniformly distributed.

Summary. Summarizing, we have presented scaling arguments and simulations on a class of self-similar complex networks, concerning transport on these networks. Diffusion and resistance in these biological networks is important, because they are both estimates of how many paths connect two nodes and how long these paths are. By using the self-similarity property and a network renormalization scheme, we have developed a scaling theory of the resistance and diffusion dependence on both the distance between two nodes and their corresponding degree. We were able to recover a relation between network modularity and transport, whereas the flow distribution in these networks was found to be consistent with earlier studies using different approaches.

Methods

Resistance Measurements. To measure the conductivity between two nodes A and B, we consider that all links in the underlying network between any two neighbor nodes i and j have unit resistances $R_{ij} = 1$. By fixing the input current to $I_A = -1$ and the output to $I_B = 1$, we can solve the Kirchhoff equations and compute the voltages in the system. The measured resistance is then $R_{A \rightarrow B} = V_A - V_B$. However, because of the required inversion of the relevant matrices, we are limited by the computer resources to networks of relatively small size, i.e., $N < 10^4$ nodes.

In principle, the magnitude of I_{ij} depends on the selection of the current input/output nodes. Upon closer inspection, however, we found that the distribution of the current magnitudes in the network links is not very sensitive to the selection of the current source and the current sink. Comparison of the result of averaging over one input and many outputs and >20 different input and output pairs of nodes for the metabolic network showed that, within statistical error, these two results are almost identical.

Diffusion Measurements. In many cases (and especially those including real-life networks), direct measurements of diffusion on complex networks exhibiting the small-world property may present significant difficulties, because of the limited time-range where diffusion takes place before settling quickly to a distance equal to the typical (very short) network diameter. The rising part of the mean-squared displacement as a function of the time is very small, and reliable measurement of the diffusion exponent is very hard to do. Moreover, we need additional information to quantify the k and ℓ dependence of the diffusion time between two nodes. For this purpose, we used the peak of the first passage time distribution as a typical diffusion time $T(A, B)$ between two points A and B in the network. Because this quantity may be asymmetric depending on which node we consider as origin, our diffusion time T represents the average of $(T_{A \rightarrow B} + T_{B \rightarrow A})/2$.

This work was supported by grants from the National Science Foundation (NSF-HSD and NSF-MCB). S.H. thanks the Israel Science Foundation, Office of Naval Research, and project DYSONET/012911 of the European Community.

- Almaas E, Kovacs B, Vicsek T, Oltvai ZN, Barabasi AL (2004) *Nature* 427:839–843.
- Paccanaro A, Trifonov V, Yu H, Gerstein M (2005) in *International Joint Conference on Neural Networks IJCNN (Jul 31–Aug 4, Montreal, Canada)*, pp 161–166.
- Havlin S, ben-Avraham D (1987) *Adv Phys* 36:695–798.
- Rozenfeld HD, Havlin S, ben-Avraham D (2006) *cond-mat/0612330*.
- Song C, Havlin S, Makse HA (2006) *Nat Phys* 2:275–281.
- Stelling J (2004) *Curr Opin Microbiol* 7:513–518.
- Kitano H (2002) *Nature* 420:206–210.

- Reed JL, Palsson BO (2003) *J Bacteriol* 185:2692–2699.
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi A-L (2000) *Nature* 407:651–654.
- Han J-DJ, Bertin N, Hao T, Goldgerg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick M, Roth FP, Vidal M (2004) *Nature* 430:88–93.
- Lopez E, Buldyrev S, Havlin S, Stanley HE (2005) *Phys Rev Lett* 94:248701.
- Emmerling M, Dauner M, Ponti A, Fiaux J, Hochuli M, Szyperski T, Wuthrich K, Bailey JE, Sauer U (2002) *J Bacteriol* 184:152–164.
- Kruskal JB (1956) *Proc Am Math Soc* 7:48–50.