Fractal growth of complex networks: repulsion between hubs

Chaoming Song[1], Shlomo Havlin[2], and Hernán A. Makse[1]

[1] *Levich Institute and Physics Department,*

City College of New York, New York, NY 10031, US

[2] *Minerva Center and Department of Physics,*

Bar-Ilan University, Ramat Gan 52900, Israel

(Dated: July 7, 2005)

Abstract

The emergence of universal properties such as the scale-free property self-similarity, and modularity, as key features of complex networks raises the fundamental question of the governing growth process according to which these structures evolve. The possibility of a unique growth mechanism for biological and social networks, as well as computers in the Internet, is of interest to the specialist and the laymen alike, as it promises to uncover the origins of collective behavior. Here, we bring the concept of renormalization from critical phenomena as a mechanism for the growth of fractal and non-fractal modular networks. We show that the key principle that gives rise to the fractal architecture of the networks is a strong effective "repulsion" between the most connected nodes (hubs) on all length scales, i.e. the hubs tend to be very disperse in the network (and not clump together). We show that the renormalization growth naturally explains to the emergence of modules in biological networks, which is crucial in understanding the structure of the biochemical functional classes. More importantly, we find that the self-similar property of networks significantly increases the robustness of such networks against targeted attacks on hubs, as compared to the very vulnerable non fractal scale-free networks.

I. INTRODUCTION

Recent developments in the field of complex networks have led to the development of new methods to examine the structure of a wide variety of systems ranging from sociological, technological to biological networks which identify universal properties [1, 2, 3, 4, 5] with important unifying conclusions. In particular the emergent similar properties of systems such as the network of protein-protein interactions, the map of metabolic pathways and the hyperlinked pages in the web, suggest that there is a self-organizing principle according to which the networks have evolved. They all exhibit the small-world property [6, 7] and a power-law probability distribution of the number of links per node (the degree) which earned them the name "scale-free" [8]. The fact that networks are scale-free was shown to have important implications on network robustness to random failures and vulnerability to attack [9, 10, 11]. Further investigation of the network topology then showed that there are two universality classes in complex networks: while some networks such as in biology and the WWW are length-scale invariant or self-similar fractals, other networks such as the Internet are not fractals [12, 13].

The "democratic" rule of the seminal Erdös-Rényi model [14] (where the nodes are connected at random) was first invoked to explain the small world effect [7]. It was then replaced by the "rich-get-richer" principle of preferential attachment [8] to explain the scale-free property [15]. However these rules do not capture the fractal topologies found in diverse complex networks. Here, we present a new view of network dynamics in which, rather than creating new nodes and links in an *additive* random way such as in models of preferential attachment [8], new nodes are generated *multiplicatively* in a self-similar modular fashion.

We formalize these ideas by borrowing the concept of "length scale renormalization" from critical phenomena [16] which naturally reproduces the modular structure of complex networks. We show that the emergence of self-similar fractal networks is due to the strong repulsion (anticorrelations) between the hubs at all length scales. In this paradigm, the rich get richer but at the expense of the "poor". In other words, the hubs prefer to grow by connections with less-connected nodes rather to another hubs, which can be viewed as an effective repulsion between hubs. On the other hand, weakly anticorrelated or uncorrelated structures generate non-fractal topologies. What are the practical implications of such in-depth understanding of network structures? We show that the renormalization ap-

proach explain the emergence of the functional modules in metabolic networks and that self-similarity increases the robustness of networks under intentional attacks.

II. THE RENORMALIZATION GROWTH APPROACH

The renormalization scheme tiles a network of N nodes with $N_B(\ell_B)$ boxes using the so-called box-covering algorithm [12, 17], as shown in Fig. 1a. The boxes contain nodes separated by a distance ℓ_B (distance is measured as the length of the shortest path between nodes). Each box is subsequently replaced by a node, and the process is repeated until the whole network is reduced to a single node. The way to distinguish between fractal and non fractal networks is represented in their scaling properties as seen in Fig. 2a and 2b. Fractal networks satisfy (Fig. 2a) :

$$N_B(\ell_B)/N \sim \ell_B^{-d_B}, \qquad k_B(\ell_B)/k_{hub} \sim \ell_B^{-d_k}. \tag{1}$$

Here k_{hub} is the degree of the largest hub inside the box and $k_B(\ell_B)$ is the degree of the box (Fig. 1a). The two exponents d_B and d_k are the fractal dimension and the degree exponent of the boxes, respectively. For a non-fractal network (Fig. 2a), we have $d_B \to \infty$ and $d_k \to \infty$; the scaling laws in Eq. (1) are replaced by exponential functions.

Based on this, we propose a network growth dynamics as the inverse of the renormalization procedure. Thus, the coarse-grained networks of smaller size are network structures appearing earlier in time, as exemplified in Fig. 1a. A present time network with $\tilde{N}(t)$ nodes is tiled with $N_B(\ell_B)$ boxes of size ℓ_B. Each box represents a node in a previous time step, so that $\tilde{N}(t-1) = N_B(\ell_B)$. The maximum degree of the nodes inside a box corresponds to the present time degree: $\tilde{k}(t) = k_{hub}$, which is renormalized such that $\tilde{k}(t-1) = k_B(\ell_B)$.

This dynamics consequently leads to a self-similar modular structure where modulus are represented by the boxes. In real networks, modularity is evident in the biochemical classes in cellular networks, the diverse communities within social networks, domains of nodes in the WWW, just to name a few [18, 19, 20, 21]. Modularity has been identified with the scaling of the clustering coefficient [19]. Here we propose an alternative definition of "modular network" as one whose statistical properties remain invariant under renormalization. In particular a modular network has an invariant scale-free degree probability to find a node with degree k, $P(k) \sim k^{-\gamma}$ [8], i.e. the exponent γ remains the same.

III. CORRELATION PROFILES

A critical question remains: how are the new self-similar nodes and boxes connected with the ancestral ones? We can answer this question by investigating the statistics of the links between the nodes and the boxes. Studying the correlation profile in real networks similar to that considered in [22, 23, 24] supports the relation between fractality and anticorrelations. The correlation profile compares the joint probability distribution, $P(k_1, k_2)$, of finding a node with k_1 links connected to a node with k_2 links with their random uncorrelated counterpart, $P_r(k_1, k_2)$. This is obtained by random swapping of the links, yet preserving the degree distribution [22]. A plot of the ratio $R(k_1, k_2) = P(k_1, k_2)/P_r(k_1, k_2)$ provides evidence of any correlated topological structure deviating from the random uncorrelated case.

At first glance, a qualitative classification can be obtained by comparing the correlation profiles of several networks by normalizing the ratio $R(k_1, k_2)$ to that of a given network, for instance the WWW. This serves as a signature of the strength of the hubs tendency to connect with each other or not. Figure 2a and 2b shows the correlation profiles of the fractal metabolic cellular network of E. *coli* [25] and the non-fractal Internet at the router level [26]: the fractal network poses a higher degree of anticorrelations compared to the WWW; nodes with a large degree tend to be connected with nodes of a small degree. On the other hand, the non-fractal Internet is less anticorrelated than the WWW. Thus, fractal topologies seem to display a higher degree of hubs repulsion in their structure than the non-fractals. However, for this property to be the real hallmark of fractality it is required that the anticorrelations appear not only in the original network (captured by $R(k_1, k_2)$) but also for the renormalized networks at different length scales.

A. Assortative parameter

Properties such as the assortative or dissortative tendency of hubs to attract or repel each other have been identified as important to unravel the structural properties of complex networks. In particular Newman [24] considers the Pearson coefficient r of the degrees at the end of an edge, and finds that most networks such as the Internet, WWW, and biological networks are dissasortative with $r < 0$, while collaboration networks are assortative with $r > 0$.

These results are in general agreement with our findings. In general, most networks are dissasortative, since the hubs, having a large number of connections, will necessarily connect to less connected nodes. This is the property that is captured by the assortative parameter. However, the parameter r is not able to capture the strength of the anticorrelations needed to differentiate between fractals and non fractals networks. One of the main points of the present work is to propose new quantities to provide a measurement of the strength of the anticorrelations needed to develop a fractal topology.

For instance, in the case of the non fractal model Mode I, we find that the Pearson coefficient could be positive or negative depending on the growth rate of the network which is specified by the factor m of the multiplicative growth. For instance in the case of Mode I we find that for $m = 1.2$ the networks is assortative while for $m = 1.8$ the network became dissortative.

Thus, this parameter is not able to distinguish between fractals and non fractals since for the same model we find different values of r. Moreover, we find that the parameter r is not invariant under renormalization. Thus, it cannot identify the dissasortative property at all length scales. However, it captures the same general trend as found in the present work. For all fractal topologies found in Nature we find that $r < 0$.

IV. THEORY

To quantitatively link the anticorrelations at all length scales with the emergence of fractality we next develop a mathematical framework and propose a mechanism for fractal network growth. In the case of modular networks, owning to Eqs. (1), we require that the different quantities grow as:

$$\begin{aligned} \tilde{N}(t) &= n\tilde{N}(t-1), \\ \tilde{k}(t) &= s\tilde{k}(t-1), \\ \tilde{L}(t) + L_0 &= a(\tilde{L}(t-1) + L_0), \end{aligned} \tag{2}$$

where $n > 1$, $s > 1$ and $a > 1$ are time-independent constants. The first equation is analogous to the multiplicative process naturally found in many population growth systems [27]. The second relation is analogous to the preferential attachment rule [8], which gives rise to the scale-free topology. The third equation describes the growth of the diameter of the

TABLE I: Relation between time evolution and renormalization.

Quantity	Time evolution	Renormalization
diameter	$\tilde{L}(t_1)+L_0$	$L/(\ell_B+L_0)$
	$\tilde{L}(t_2)+L_0$	L
number of	$\tilde{N}(t_1)$	$N_B(\ell_B)$
nodes	$\tilde{N}(t_2)$	N
degree	$\tilde{k}(t_1)$	$k(\ell_B)$
	$\tilde{k}(t_2)$	k_{hub}
hub-hub	$\tilde{k}(t_1)$	$k(\ell_B)$
links	$\tilde{n}_h(t_2)$	$n_h(\ell_B)$

network ($\tilde{L}(t)$, as defined by the largest distance between nodes) and determines whether the network is small-world [6, 7] and/or fractal [12]. Here we introduce the characteristic size L_0, the importance of which lies in describing the non-fractal networks. Since every quantity increases by a factor of n, s and a, we first derive the scaling exponents in terms of the microscopic parameters: $d_B = \ln n/\ln a$, $d_k = \ln s/\ln a$. The exponent of the degree distribution satisfies $\gamma = 1 + \ln n/\ln s$.

Next we fully develop the theoretical framework of renormalization and its analogy with the time evolution of networks.

We obtain the relation between two times $t_2 > t_1$ as

$$\begin{aligned} \tilde{L}(t_2)+L_0 &= a^{t_2-t_1}(\tilde{L}(t_1)+L_0), \\ \tilde{N}(t_2) &= n^{t_2-t_1}\tilde{N}(t_1), \\ \tilde{k}(t_2) &= s^{t_2-t_1}\tilde{k}(t_1), \\ \tilde{n}_h(t_2) &= e^{t_2-t_1}\tilde{k}(t_1). \end{aligned} \tag{3}$$

The relationship between the quantities describing the time evolution and the renormal-

ization is shown in Table I. They are formalized as follows:

$$\begin{aligned}
\ell_B + L_0 &= (\tilde{L}(t_2) + L_0)/(\tilde{L}(t_1) + L_0) = a^{t_2 - t_1} \\
\mathcal{N}(\ell_B) &\equiv N_B(\ell_B)/N = \tilde{N}(t_1)/\tilde{N}(t_2) = n^{t_1 - t_2}, \\
\mathcal{S}(\ell_B) &\equiv k_B(\ell_B)/k_{hub} = \tilde{k}_B(t_1)/\tilde{k}_B(t_2) = s^{t_1 - t_2}, \\
\mathcal{E}(\ell_B) &\equiv n_h(\ell_B)/k_B(\ell_B) = \tilde{n}_h(t_2)/\tilde{k}(t_1) = e^{t_2 - t_1}.
\end{aligned} \tag{4}$$

Here we define the additional ratios, $\mathcal{N}$ and $\mathcal{S}$. Replacing the time interval $t_2 - t_1$ by $\ln(\ell_B + L_0)/\ln a$, as obtained from the first equation in (4), we obtain:

$$\begin{aligned}
\mathcal{N}(\ell_B) &= (\ell_B + L_0)^{-\ln n/\ln a}, \\
\mathcal{S}(\ell_B) &= (\ell_B + L_0)^{-\ln s/\ln a}, \\
\mathcal{E}(\ell_B) &= (\ell_B + L_0)^{-\ln(1/e)/\ln a},
\end{aligned} \tag{5}$$

or

$$\begin{aligned}
\mathcal{N}(\ell_B) &= (\ell_B + L_0)^{-d_B}, d_B \equiv \ln n/\ln a, \\
\mathcal{S}(\ell_B) &= (\ell_B + L_0)^{-d_k}, d_k \equiv \ln s/\ln a, \\
\mathcal{E}(\ell_B) &= (\ell_B + L_0)^{-d_e}, d_e \equiv \ln(1/e)/\ln a.
\end{aligned} \tag{6}$$

Notice that we have considered $L_0 = 0$ in Eqs. (1) for simplicity. Equations (6) are more general and accommodate the case of non fractal networks which are characterized by exponential functions:

$$\begin{aligned}
\mathcal{N}(\ell_B) &\sim \exp(-\ell_B/\ell_0), \\
\mathcal{S}(\ell_B) &\sim \exp(-\ell_B/\ell_0'),
\end{aligned} \tag{7}$$

These expressions arise from Eqs. (6) by taking the limit of $d_B \to \infty$, $d_k \to \infty$, and $L_0 \to \infty$ while $L_0/d_B \to \ell_0$ and $L_0/d_k \to \ell_0'$, where ℓ_0 and ℓ_0' are characteristic constants of the network.

V. GROWTH MODES

In order to incorporate the growth modes in the dynamical Eqs. (2) we consider, without loss of generality, two different modes of connectivity between boxes. *(i)* Mode I with

probability e (Fig. 1b): two boxes are connected through a direct link between their hubs leading to hub-hub attraction. *(ii)* Mode II with probability $1-e$ (Fig. 1c): two boxes are connected via non-hubs leading to hub-hub repulsion or anticorrelations. We will show that Mode I leads to non-fractal networks while Mode II leads to fractal networks.

Formally, for a node with $\tilde{k}(t)$ links at time t, we define $\tilde{n}_h(t+1)$ as the number of links which are connected to hubs in the next time step (see Fig. 5a). Then the probability e representing the hub-hub attraction satisfies:

$$\tilde{n}_h(t) = e\ \tilde{k}(t-1). \tag{8}$$

Using the analogy between time evolution and renormalization we introduce the analogous quantity, $n_h(\ell_B)$, as a function of the length of boxes that defines the ratio $\mathcal{E}(\ell_B) \equiv n_h(\ell_B)/k_B(\ell_B)$ (for the nonlinear relation between t and ℓ_B, leading to the ℓ_B dependence on $\mathcal{E}$. In the extreme case of strong hub attraction with the hubs of the boxes connected at all length scales we have $\mathcal{E}(l_B) \sim$ constant. On the other hand, hubs repulsion leads to decreasing $\mathcal{E}(\ell_B)$ with ℓ_B. From scaling we obtain a new exponent $d_e = -\ln e/\ln a$ characterizing the strength of the anticorrelations in a scale-invariant way:

$$\mathcal{E}(\ell_B) \sim l_B^{-d_e}. \tag{9}$$

Figure 2e shows $\mathcal{E}(\ell_B)$ for two real fractal and non fractal networks: a map of the WWW domain nd.edu consisting of 352,728 web-sites [28] and a map of the Internet at the router level consisting of 284,771 nodes [26]. We find that for the fractal WWW, $d_e = 1.5$, indicating that it poses strong anticorrelations. On the other hand, the non-fractal Internet shows $\mathcal{E}(\ell_B) \sim$ constant. These results confirm that indeed fractal networks have strong hub repulsion at all length scales and non fractal networks have weak hub repulsion or none.

A. Definition of $h(bk)$. Calculation of d_e for biological networks

The exponent d_e is the key to understanding the scale-invariant hub repulsion in networks. However, since it is based on the scaling with the size of the boxes, it requires a large network to be precisely calculated. The biological networks contain a only a relative small number of nodes in comparison with the WWW and the Internet and therefore require a different method to evaluate the exponent d_e. We introduce a new scale-invariant quantity, related

to Eq. (9), that allows the measurement of d_e for small networks since it is based on the scaling of the degree rather than on the distance ℓ_B which is very small in these biological networks.

For the biological networks we obtain $d_e = 1.1$ for the protein-protein interaction network of H. *sapiens* [32] and $d_e = 4.5$ for the metabolic network of E. *coli* [19, 29]. This implies that the biological networks are even more anticorrelated than the WWW. Since the biological networks have smaller d_B as well, this result further suggests that repulsion between hubs are at the origin of the fractal structure of networks.

We consider the ratio $h(bk)/k$, where $h(bk)$ is the number of neighbors with degree larger than bk of a node with degree k (b is an arbitrary positive number, large b corresponds to the identification of the hubs). Thus, for a given b, a fast decay with k of $h(bk)$ corresponds to stronger anticorrelations. This quantity is capable of determining the degree of hub repulsion needed to produce fractal topologies. From scaling we obtain

$$h(bk)/k \sim k^{-\epsilon} \tag{10}$$

with

$$\epsilon = 1 + d_e/d_k. \tag{11}$$

To prove Eq. (11) we assume t^* as the current time in the evolution of the network, we set $b = s^{-\Delta}$, and the new generated degree at a time t satisfies: $k \sim s^{t^*-t}$, so that $bk \sim s^{t^*-(\Delta+t)}$. At time $\Delta + t$, we have $s^{(\Delta+t)-t} = b^{-1}$ neighbors whose degrees will be larger than k. Moreover, only a fraction $e^{t^*-(t+\Delta)}$ of the links s^{Δ} are left in the network. These are the links connecting boxes via Mode I. Then we obtain $h(bk) \sim s^{\frac{\ln e}{\ln s}(t^*-t)} = k^{-d_e/d_k}$, or $h(bk)/k \sim k^{-\epsilon}$, with $\epsilon = 1 + d_e/d_k$. We note that Eq. (11) is analogous to the relation $\gamma = 1 + d_B/d_k$ found in [12].

Therefore, the exponent ϵ can be used to obtain an estimate of d_e if we know the exponent d_k. The advantage of this method to obtain d_e over the direct calculation using Eq. (9) is that the latter requires a large network since it is based on the scaling with the distance ℓ_B, and the distances are very small unless the networks is very large. On the other hand, Eq. (10) is based on the scaling with the degree k which in general has a larger range, as measured in decades of available data, than the distance.

We first test the prediction Eq. (11) by calculating $h(bk)$ for the WWW and the Internet for which we know d_e independently. Figure 3a shows the results for the WWW and the

Internet confirming our previous calculations. We find $\epsilon = 1.5$ which together with $d_k = 2.5$ [12] gives $d_e = (\epsilon - 1)d_k = 1.5$ consistent with the previous estimation using Eq. (9) in the main text. The calculations for the Internet also confirm the values of the exponents. We find $\epsilon = 1$ in agreement with the value $d_e = 0$ found in the main text.

We also test that $h(bk)/k$ is independent of the value of b as shown in Fig. 3b. Moreover, $h(bk)/k$ is invariant under renormalization since we find the same scaling exponent ϵ for different renormalized networks at several scales ℓ_B as shown in Fig. 3c for the WWW.

Next, we apply this procedure to calculate d_e for the smaller biological networks (lees than 1000 nodes) of the protein-protein interactions of H. *sapiens* [32] and the metabolic network of E. *coli* [19, 29]. Fig. 3d shows the results. We obtain $\epsilon = 1.5$ ($d_k = 2.2$) and $\epsilon = 2.8$ ($d_k = 2.5$) for the protein interaction network and the metabolic network indicating that $d_e = 1.1$ and $d_e = 4.5$, respectively. In general, we obtain larger values of d_e than in the WWW, which implies that the biological networks are even more anticorrelated. Since the biological networks have smaller d_B as well, this result further shows that anticorrelations are at the origin of the fractal structure of networks.

In Fig. 4 we classify many real networks and all the available models in terms of their fractality and correlated properties in the plane (d_B, ϵ). The general trend follows our prediction that stronger hub repulsion (larger d_e) induces more pronounced fractal properties (smaller d_B), as exemplified by the case of biological networks. Conversely, the loss of fractality ($d_B \to \infty$) occurs as the anticorrelations disappear, e.g. in the Internet.

VI. PREDICTIONS OF THE MINIMAL MODEL

In Fig. 5 we investigate the predictions of the different modes separately. While each mode leads to scale-free renormalized structures, they differ in their fractal and small-world properties. Mode I exhibits the small-world effect, but is not fractal due to its strong hub-hub attraction. On the other hand, Mode II alone gives rise to a fractal topology. However, the anticorrelations in Mode II are strong enough to push the hubs far apart, leading to the breakdown of the small-world property.

These results suggest that the simultaneous appearance of both small-world and fractal properties in scale-free networks is due to a combination of both growth modes. Such a structure is visualized in Fig. 5a for $e = 0.8$, $n = 5$, $a = 1.4$, $s = 3$. Supporting evidence is

given by *(i)* Fig. 5b showing how the model with $e = 0.8$ is more anticorrelated than the $e = 1$ model (Mode I) by plotting the correlation profile of the ratio $R_{e=0.8}(k_1, k_2)/R_{e=1}(k_1, k_2)$, *(ii)* Fig. 5c displaying the power law dependence of N_B versus ℓ_B for the fractal structure of $e = 0.8$ and the exponential dependence of the non fractal structure of $e = 1$, and *(iii)* Fig. 5d showing how Mode I reproduces $\mathcal{E}(\ell_B) \sim$ constant while the $e = 0.8$ model gives rise to $\mathcal{E}(\ell_B) \sim \ell_B^{-d_e}$ with $d_e = -\ln 0.8/\ln 1.4 = 0.66$, as predicted by the analytical formula and in agreement with the empirical findings of Fig. 2e on real networks. Although simplistic (for instance, we did not consider the introduction of loops, although this can be easily done in a self-similar way), this minimal mechanism clearly captures the essential properties of networks, one of which is the relationship between anticorrelations and fractality.

Additional supporting evidence is given in Fig. 6 by (a) calculating the mean number of nodes (mass) of the boxes tiling the network using the box covering methods, $\langle M_B(\ell_B)\rangle \equiv N/N_B(\ell_B)$, and (b) by averaging over boxes of size ℓ_c around a randomly chosen node to obtain the local mass $\langle M_c(\ell_c)\rangle$ (the so-called cluster growing method, see [12, 17]). The results show how the minimal model reproduces one of the main properties of fractal networks [12]: the power-law relation for the global average mass $\langle M_B(\ell_B)\rangle \sim \ell_B^{d_B}$, with $d_B = \ln 5/\ln 1.4 = 4.8$, as a signature of fractality consistent with Eq. (1), and the exponential dependence of the local mass $\langle M_c(\ell_c)\rangle \sim e^{\ell_c/\ell_0}$ as a signature of the local small-world effect: $\ell_c \sim \ln\langle M_c\rangle$. Note that the cluster growing method is actually a way to measure the distance while the box covering method measures the fractality [12] (the global small-world properties are in the next section).

A. Global small world: short cuts in the network.

An important factor in the dynamics of real networks is the existence of randomness or noise in the growth process. The simplest type of noise is the appearance of random connections between nodes as exemplified in the Watts-Strogatz model of small world networks [6]. To investigate how this noise affects the fractality of networks, we use the model and modify the dynamical law as following: at each time step, $pK(t)$ number of links will be added in at random, here $K(t)$ is the total number of links at time t, and p is a constant to control the fraction of the noise. We build a model which is fractal, small-world and scale-free with parameter $m = 1.5$, $e = 0.5$, and add $p = 1\%$ random connections at each time

step. Our analytical considerations predict a box dimension $d_B = 2$ in the case of absence of noise. The numerical simulation (see Fig. 7a) shows that this theoretical prediction of d_B still fits very well the simulated data except for a small deviation at large box sizes, i.e. the added noise appears as an approximately exponential tail for large distances. Thus the fractal properties would be still valid in presence of this kind of noise as long as the noise is not very large. Interestingly, this method could be used to test the appearance of noise in real complex networks, or to asses the quality of the data in for instance protein interaction networks obtained by yeast two hybrid methods which are known to give rise to many false positives; a study that we will perform in the future.

The most interesting results is that the addition of noise leads to the small world effect at the global level. In principle the existence of fractality seems to be at odds with the small-world effect. Fractality implies a power-law dependence on the distance, while the small-world implies a exponential dependence [12]. Above, we show how the combination of Mode I and Mode II of growth leads to a fractal global property and a local small-world effect. Here we show that by adding a fraction of short-cuts in a fractal network, we reproduce also the small-world effect at the global level. Using the algorithm explained above to add noise, we find that the average distance $< d >$ over all pairs of vertices is

$$< d > \sim 2.61 \ln N, \tag{12}$$

(Fig. 7b), indicating that the fractal model predicts also the small-world effect. This is akin to the ideas presented in Wattz-Strogatz, where starting with a regular network, the small-world is obtained by adding a fraction of short cuts in the system. We notice that the fraction of short cuts is very small.

VII. MODULARITY

The scale-invariant properties naturally lead to the appearance of a hierarchy of self-similar nested communities or modules. In this new view, boxes represent nested modules of different length scales. The importance of modular structures is stressed in biological networks, where questions of function and evolutionary importance are put to the test [18, 19, 20, 21]. The relevant question is whether the self-similar hierarchy of boxes encodes the information about the functional modules in biological networks. To answer this

question we analyze the fractal metabolic network of E. *coli* [25] which has been previously studied using standard clustering algorithms [19]. Here we show that by repeatedly applying the renormalization we produce a renormalized tree whose branches are closely related to the annotated biochemical classes of the products of the metabolic reactions such as carbohydrates, lipids, amino acid biosynthesis, etc. We renormalize the network at a given box size and cluster the substrates which belong to the same box and repeat the procedure to generate the hierarchical tree shown in Fig. 8a and explained in detail in Fig. 8b. The main known biochemical classes of the substrates emerge naturally from the renormalization tree indicating that the boxes capture the modular structure of the metabolic network of E. *coli*.

VIII. VULNERABILITY TO INTENTIONAL ATTACKS

Finally our results suggest the importance of self-similarity in the evolution of the topology of networks. Understanding the growth mechanism is of fundamental importance as it raises the question of its motivation in Nature. For instance, given that systems in biology are fractals, could there be an evolutionary drive for the creation of such networks? A parameter relevant to evolution is the robustness of the network, which can be compared between fractal and non fractal networks.

Non-fractal scale-free networks, such as the Internet, are extremely vulnerable to targeted attacks on the hubs [9, 10]. When a small fraction of nodes (those with highest degree) are removed the network collapses. In such non-fractal topologies, the hubs are connected and form a central compact core such that the removal of few largest hubs has catastrophic consequences for the network. Here we show that the fractal property of networks significantly increases the robustness against targeted attacks since the hubs are more dispersed in the network (see Fig. 8c). This could explain why evolutionary constrained biological networks have evolved into a fractal architecture.

IX. CONCLUSIONS

Our results paint a picture of nodes organized around disperse hubs in self-similar nested modules [21] characterized by different functionalities that compartmentalize the hubs, isolate them [22], and protect them from a failure at the system level [33]. These modules

function relatively autonomously so that a failure in one module may not propagate so easily to the next. This may provide significant higher protection against intentional attacks reducing the high vulnerability— the Achilles' heel— of non fractal scale-free networks.

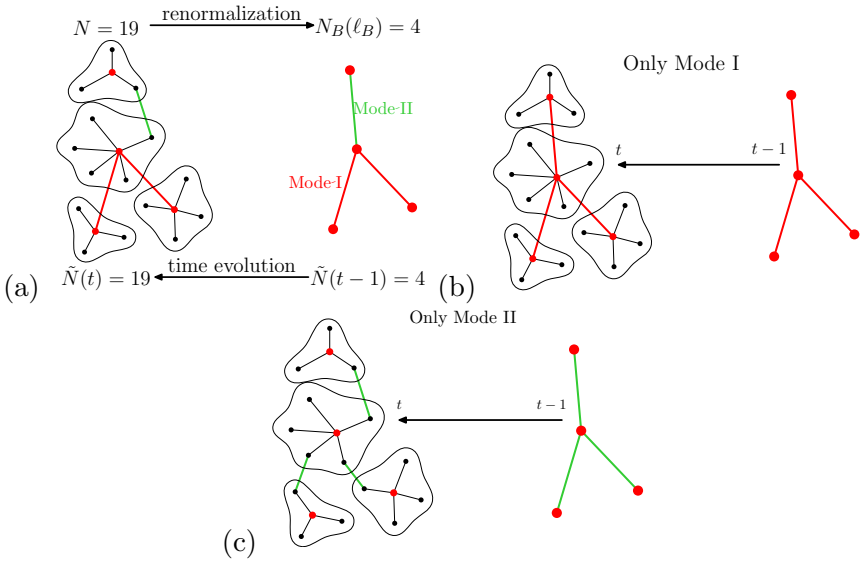


FIG. 1: Self-similar dynamical evolution of networks. (a) The dynamical growth process can be seen as the inverse renormalization procedure with all the properties of the network being invariant under time evolution. In this example $\tilde{N}(t) = 19$ nodes are renormalized with $N_B(\ell_B) = 4$ boxes of size $\ell_B = 3$ so that $\tilde{N}(t-1) = 4$ nodes in the previous time step. The central box has a hub with $k_{hub} = 7$ links, then $\tilde{k}(t) = 7$. After renormalization, $k_B(\ell_B = 3) = 3$ so that $\tilde{k}(t-1) = 3$ for the central box. Out these three links, two are via a hub-hub connection (Mode I), thus $n_h(\ell_B) = 2$ and $\mathcal{E}(\ell_B) = 2/3$, for this case. (b) Analysis of Mode I, only. Starting with the four nodes at $t-1$ in the figure, at the next time step $m\tilde{k}(t-1)$ ($m = 2$ in this example) new nodes are generated for each node with degree $\tilde{k}(t-1)$. In Mode I this implies $n = 2m+1$ (since $\tilde{N}(t+1) = \tilde{N}(t)(1 + m < k >)$ and $s = m+1$. The diameter of the network grows additively as: $\tilde{L}(t+1) = \tilde{L}(t) + 2$ ($a = 1$ and $L_0 \to \infty$) and therefore $\tilde{L}(t) \sim 2t$. For this mode we obtain a non-fractal topology: $N_B(l_B)/N \sim \exp(-\frac{\ln n}{2} l_B)$ and $k_B(l_B)/k_{hub} \sim \exp(-\frac{\ln s}{2} l_B)$. This is a direct consequence of the linear growth of the diameter $\tilde{L}(t)$. In fact since $a = 1$ for this mode, then $d_B = \ln n / \ln a \to \infty$ and $d_k = \ln s / \ln a \to \infty$. Moreover, the additive growth in the diameter with time assures that the network is small-world. This mode is similar to a class of models called pseudo-fractals [30, 31]. (c) Mode II alone produces a scale-free with a fractal topology but not the small-world effect. Here the boxes are connected via non-hubs. In this case the diameter increases multiplicatively $\tilde{L}(t+1) = 3\tilde{L}(t)$, with $n = 2m+1$, $s = m$, $a = 3$ leading to an exponential growth in the diameter and therefore to a fractal topology with finite d_B and d_k since $a \neq 1$. However, the multiplicative growth of $\tilde{L}(t)$ leads to the disappearance of the small-world effect, which is replaced by a power-law dependence.

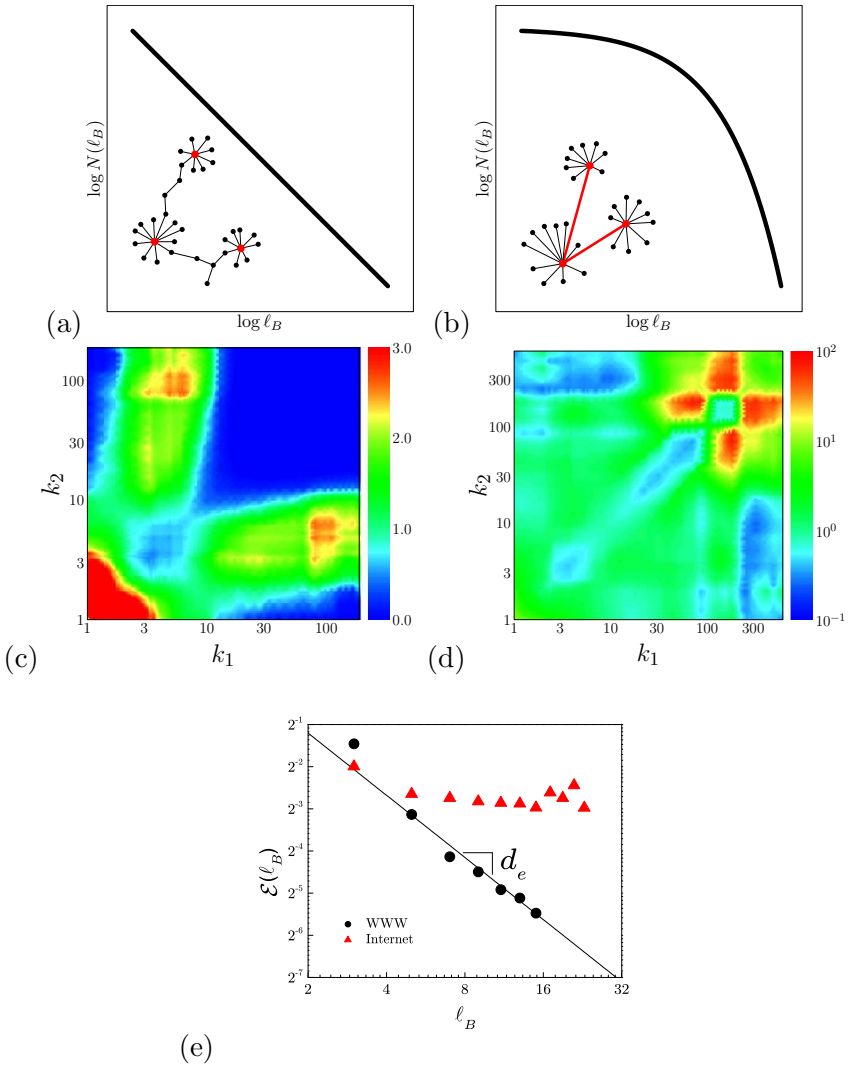


FIG. 2: Empirical results on real complex networks. (a) Fractal networks are characterized by a power law dependence between N_B and ℓ_B while (b) non fractal networks are characterized by a exponential dependence. The main contention of this paper is that fractality arises due to the repulsion between hubs while non fractal networks the hubs are not disperse. (c) We compare the correlation profiles of the fractal topology of the metabolic network of E. *coli* $R_{\mathrm{E.}coli}(k_1,k_2)/R_{\mathrm{WWW}}(k_1,k_2)$ and (d) the non-fractal topology of the Internet $R_{\mathrm{Int}}(k_1,k_2)/R_{\mathrm{WWW}}(k_1,k_2)$, to the profile of the WWW in search of a signature of the origin of fractality. The patterns reveal that the fractal cellular networks are strongly anticorrelated. (e) Scale-invariant correlations in complex networks. Scaling of $\mathcal{E}(\ell_B)$ as defined in Eq. (9) for the fractal topology of the WWW with $d_e = 1.5$, and the non-fractal topology of the Internet showing that fractal topologies are strongly anticorrelated at all length scales. In order to calculate $\mathcal{E}$ (and in all the calculations in this study) we tile the network by first identifying the nodes which are center of the boxes with the largest mass and sequentially centering the boxes around these nodes.

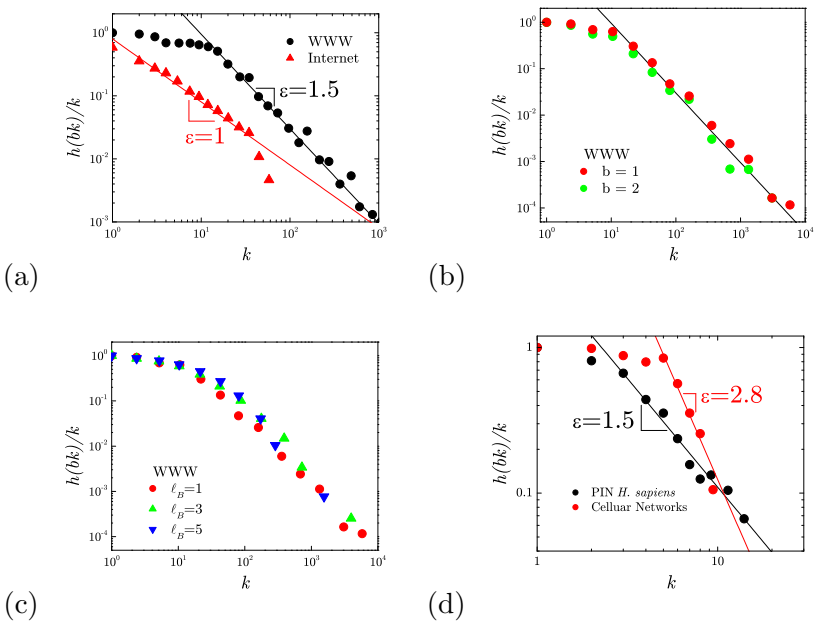


FIG. 3: Quantifying the scale-invariant correlated topology of complex networks. (a) Factor $h(bk)/k$ as defined in the text for the fractal topology of the fractal WWW and the non-fractal topology of the Internet, showing the two different scaling, confirming the values of the exponent d_e. (b) Factor $h(bk)/k$ for the WWW for different values of b showing that its definition is independent of b. (c) Factor $h(bk)/k$ for the WWW at different scales showing that it is scale-invariant. (d) Same as (a) but for the biological networks of protein interactions and metabolic cellular. These networks show stronger anticorrelations than the WWW.

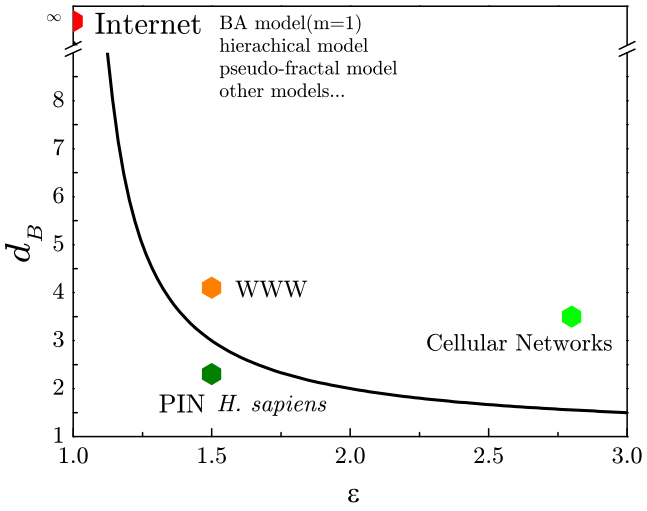


FIG. 4: Classification of complex networks. We quantify the fractality with the exponent d_B and the strength of anticorrelations with the exponent $\epsilon = 1 + d_e/d_k$ (which is analogous to $\gamma = 1 + d_B/d_k$, and can be calculated more accurately in particular for small networks. In general when $\epsilon = 1$ we have non fractal networks while for $\epsilon > 1$ we have fractal networks. When the hubs are dispersed in the network a fractal topology results. The available scale-free models are non fractals with $\epsilon = 1$ while the available regular fractals are not scale-free [17]. We also plot the prediction of the minimal model for $e = 1/2$. We are motivated to do this, since the anticorrelated properties of this model are intermediate between Mode I ($e = 1$) and Mode II ($e = 0$). For this model we find analytically that $\gamma = 2 + d_e/d_k$ and $d_B = \epsilon/(\epsilon - 1)$. Thus $d_B \to \infty$ when $\epsilon \to 1$, signaling the lost of fractality as the anticorrelations disappear in the system.

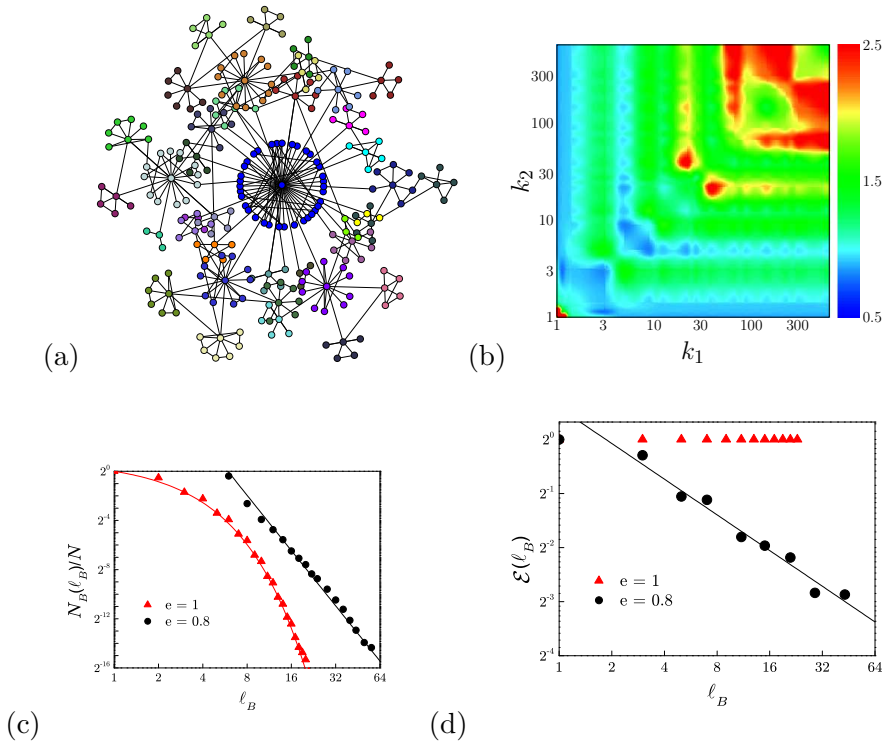


FIG. 5: Predictions of the renormalization growth mechanism of complex networks. (a) Resulting topology predicted by the minimal structure for $e = 0.8$. The colors of the nodes show the modular structure with each color representing a different box. The growth process is in general a stochastic combination of Mode I (with probability e) and Mode II (with probability $1 - e$, see Fig. 1). We obtain $\tilde{L}(t + 1) = (3 - 2e)\tilde{L}(t) + 2e$, and $a = 3 - 2e$ and $L_0 = e/(1 - e)$. This model predicts finite fractal exponents d_B and d_k for any value of e except for the singular value at $e = 1$ (Mode I) where $L_0 \to \infty$, losing the fractal properties with $d_B \to \infty$ and $d_k \to \infty$. We note that L_0 is a lower cut-off for the power-law dependence which defines the fractal exponent d_B. (b) Ratio of the correlation profiles $R_{e=0.8}(k_1, k_2)/R_{e=1}(k_1, k_2)$ between the model with $e = 0.8$ and Mode I. The profile shows how $e = 0.8$ is more anticorrelated than Mode I of growth. (c) Plot of N_B versus ℓ_B showing that Mode I is non fractal (exponential decay) and $e = 0.8$ is fractal (power-law decay) according to (b) and in agreement with the empirical results of Fig. 2. (d) Scaling of $\mathcal{E}(\ell_B)$ versus ℓ_B. The minimal model reproduces the behavior of fractal networks for $e = 0.8$ with $\mathcal{E} \sim \ell_B^{-0.66}$) and non-fractal networks Mode I, $e = 1$, $\mathcal{E} \sim$ constant) as found empirically in Fig. 2e.

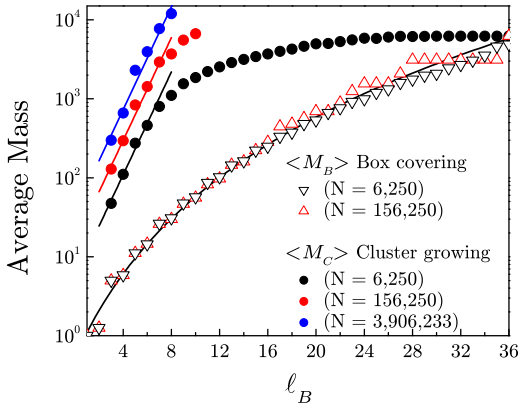
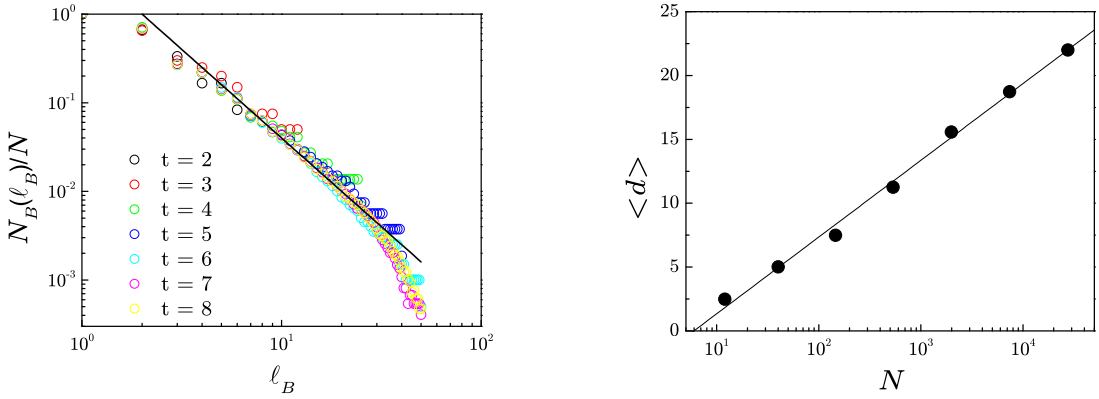


FIG. 6: Predictions of the renormalization growth mechanism of complex networks. The hallmark of fractality [12] is predicted by the minimal mechanism: the global tiling of the network evidences the fractality in the power-law dependence of the mass of the boxes, $\langle M_B \rangle$, while the local average of the cluster growing method $\langle M_c \rangle$, evidences the small-world effect. The exponential growth of $\langle M_c \rangle$ implies the small world effect as $(\ell_c \sim \ln \langle M_c \rangle)$. We show simulations for different network size. The scaling of $\langle M_B \rangle$ does not show finite size effects. The initial exponential dependence range of $\langle M_c \rangle$ increases as the size of the network increases.



(a) (b)

FIG. 7: (a) N_B vs ℓ_B shows that the fractality still holds in the presence of random noise. The straight line gives the theoretical prediction of the model $d_B = 2$. (b) Average of the shortest path between two nodes as a function of the system size. The results verify that the model predicts the small world properties: $< d > \sim 2.61 \ln N$.

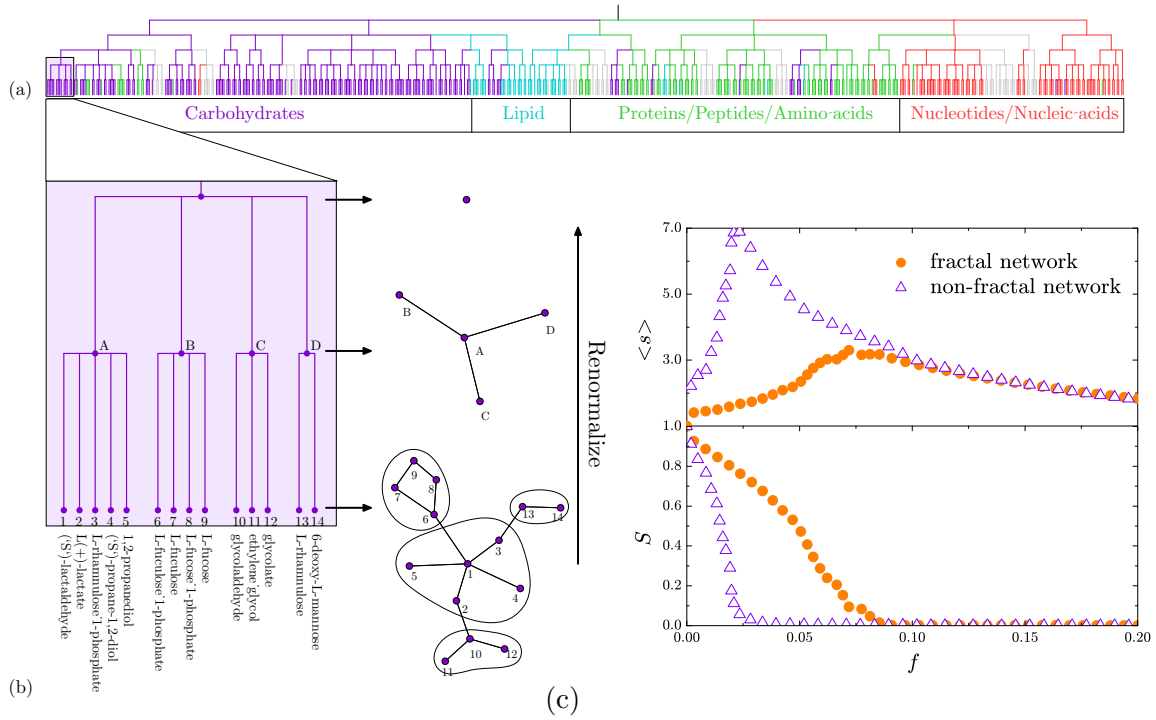


FIG. 8: Practical implications of the renormalization growth approach and self-similarity. (a) Renormalization tree of the metabolic network of E. *coli* leading to the appearance of the functional modulus. Each level of the tree represents a step in the renormalization procedure. This scheme identifies the hierarchy of self-similar modules which are in turn closely related to the known biochemical classes of the substrates in the metabolic reactions. The colors of the nodes and branches in the tree denote the main biochemical classes as: carbohydrates (violet), lipids (light blue), proteins, peptides and aminoacids (green), nucleotides and nucleic acids (red), aromatic, monocarbon compounds and essential elements (green), and coenzymes and prosthetic groups (grey). (b) Details of the construction of three levels of the renormalization tree for $\ell_B = 3$ for 14 metabolites in the carbohydrate biosynthesis class as shown in the shaded area in (a). (b) Vulnerability under intentional attack of a non fractal network generated by Mode I ($e = 1$) and a fractal network generated by Mode II ($e = 0$). The comparison is done between model networks of the same $\gamma = 2.8$, the same number of nodes (74,000), the same number of links, the same amount of loops and the same clustering coefficient. Thus, the difference in the resilience seen in this figure is attributed solely to the different degree of anticorrelations. We plot the relative size of the largest cluster, S, and the average size of the remaining isolated clusters, $\langle s \rangle$, after removing a fraction f of the largest hubs for both networks. While both networks collapse at a finite fraction f_c, evidenced by the decrease of S toward zero and the peak in $\langle s \rangle$, the fractal network has a significantly larger threshold ($f_c \approx 0.09$) compared to the non-fractal threshold ($f_c \approx 0.02$) suggesting a significantly higher robustness of the fractal modular networks to intentional attacks.

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

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

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

[4] M. E. J. Newman, *SIAM Review* **45**, 167 (2003).

[5] S. Bornholdt and H. G. Schuster, eds., *Handbook of graphs and networks* (Wiley-VCH, 2002).

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

[7] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).

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

[9] R. Albert, H. Jeong, A.-L. Barabási, *Nature* **406**, 378 (2000).

[10] R. Cohen, K. Erez, D. ben-Avraham, S. Havlin *Phys. Rev. Lett.* **85**, (2000).

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

[12] C. Song, S. Havlin, H. A. Makse, Nature **433**, 392 (2005).

[13] S. H. Strogatz, *Nature* **433**, 365 (2005).

[14] P. Erdös, A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960).

[15] Notice that the term "scale-free" coined by Barabási-Albert [8] refers to the absence of a typical number of links, as exemplified by a power-law distribution of degree connectivity, but it does not refer to the length scale invariance found in [12].

[16] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, Oxford, 1971).

[17] T. Vicsek, *Fractal Growth Phenomena*, 2nd ed., Part IV (World Scientific, Singapore, 1992).

[18] L. H. Hartwell, L. H. Hopfield, S. Leibler, A. W. Murray, *Nature* **402**, C47 (1999).

[19] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Barabasi, *Science* **297**, 1551 (2002).

[20] M. Girvan, M. E. J. Newman, *Proc. Natl. Acad. Sci.* **99**, 7821 (2002).

[21] G. Palla, I. Derenyi, I. Farkas, T. Vicsek, *Nature* **435**, 814 (2005).

[22] S. Maslov, K. Sneppen, *Science* **296**, 910 (2002).

[23] R. Pastor-Satorras, A. Vázquez, A. Vespegnani, *Phys. Rev. Lett.* **87**, 258701 (2001).

[24] M. E. J. Newman, *Phys. Rev. Lett.* **89** 208701 (2002).

[25] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, Nature **407**, 651 (2000).

[26] H. Burch, W. Cheswick, *IEEE Computer* **32**, 4 (1999). http://research.lumeta.com/ches/map

[27] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam, 1981).

[28] R. Albert, H. Jeong, A.-L. Barabási, *Nature* **401**, 130 (1999). http://www.nd.edu/∼networks

[29] R. Overbeek, *et al. Nucleic Acid Res.* **28**, 123 (2000).

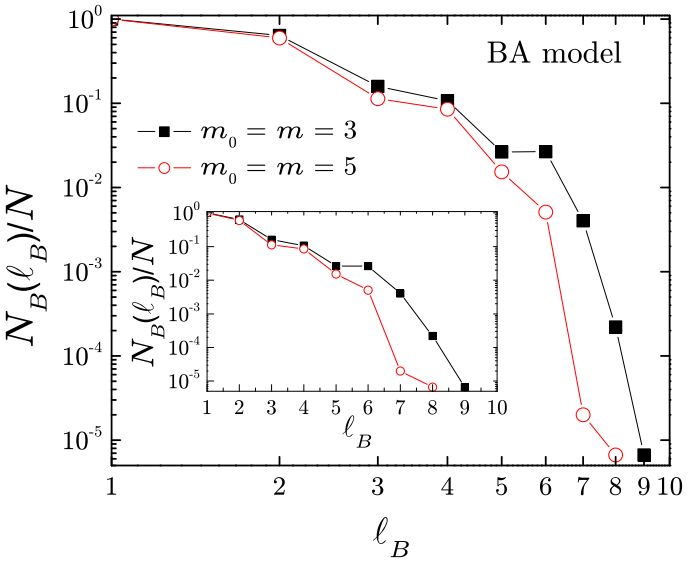
[30] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, *Phys. Rev. E* **65**, 066122 (2002).

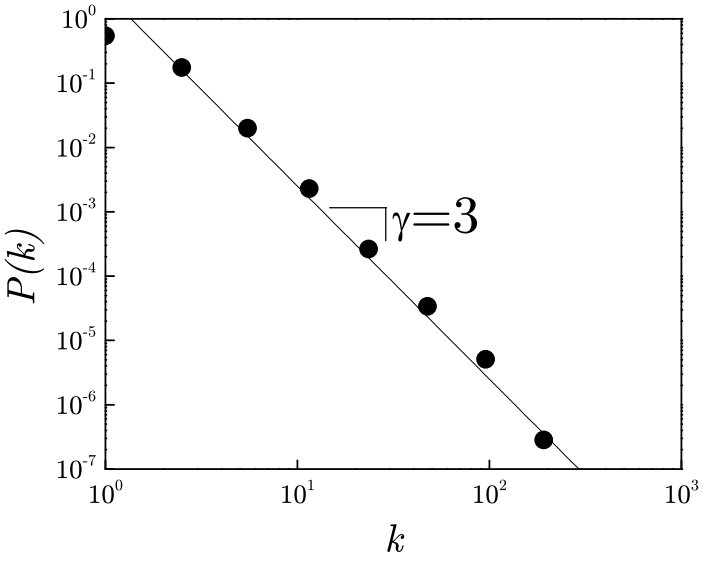
[31] S. Jung, S. Kim, B. Kahng, *Phys Rev E* **65**, 056101 (2002).

[32] *Database of Interacting Proteins* (DIP). http://dip.doe-mbi.ucla.edu

[33] H. Kitano, Science **295**, 1662 (2002).

Acknowledgements. We would like to thank J. Brujić for illuminating discussions and E. Ravasz for providing the data on the metabolic network. SH wishes to thank the Israel Science Foundation, ONR and Dysonet for support. This work is supported by the National Science Foundation, DMR-0239504 to HAM.


BA model
$m_0 = m = 3$
$m_0 = m = 5$
$N_B(\ell_B)/N$
ℓ_B


10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
$P(k)$
$\gamma=3$
10^0
10^1
10^2
10^3
k