Network robustness to targeted attacks. The interplay of expansibility and degree distribution

E. Estrada[a]

Complex Systems Research Group, X-Rays Unit, RIAIDT, Edifico CACTUS, University of Santiago de Compostela, Santiago de Compostela 15782, Spain

Abstract. We study the property of certain complex networks of being both sparse and highly connected, which is known as "good expansion" (GE). A network has GE properties if every subset S of nodes (up to 50% of the nodes) has a neighborhood that is larger than some "expansion factor" ϕ multiplied by the number of nodes in S. Using a graph spectral method we introduce here a new parameter measuring the good expansion character of a network. By means of this parameter we are able to classify 51 real-world complex networks — technological, biological, informational, biological and social — as GENs or non-GENs. Combining GE properties and node degree distribution (DD) we classify these complex networks in four different groups, which have different resilience to intentional attacks against their nodes. The simultaneous existence of GE properties and uniform degree distribution contribute significantly to the robustness in complex networks. These features appear solely in 14% of the 51 real-world networks studied here. At the other extreme we find that $\sim$40% of all networks are very vulnerable to targeted attacks. They lack GE properties, display skewed DD — exponential or power-law — and their topologies are changed more dramatically by targeted attacks directed at bottlenecks than by the removal of network hubs.

PACS. 89.75.Fb Structures and organization in complex systems – 89.75.Da Systems obeying scaling laws – 89.20.-a Interdisciplinary applications of physics – 89.75.Hc Networks and genealogical trees

1 Introduction

We live in a world of networked infrastructures formed by communication, power grids, transportation and fuel distribution networks whose security is of universal concern [1–4]. Such systems need to be robust and capable of surviving the loss of nodes/links that can result from random failures or intentional attacks [5–10]. Many other networks function in nature and society showing different degrees of tolerance to damage on their structures [11–13]. An understanding of the vulnerabilities of complex networks is fundamental to the design of robust systems [5,14,15] and strategies for management of their social impact [16,17]. Robust networks must be highly tolerant to errors related to random failures and as invulnerable as possible to intentional attacks against nodes/links [6]. Most investigations about the resilience of real-world networks to targeted attacks have been centered on the analysis of attacks targeted on the highest degree nodes [6–9,18–20]. It is actually known that "scale-free" networks [21], which include the World-Wide Web, the Internet, and various social and biological networks, are more robust to random failures of nodes than networks in which the node distribution is exponential. However, scale-free networks are extremely vulnerable to targeted

attacks against their hubs, the most connected nodes [6]. It is also known that the removal of certain links in a complex network due to malfunction or attack can also lead to the collapse of the system [22]. Node/link robustness in complex networks is also intimately related to a property known as "good expansion" (GE) [23,24]. Formally, a network is considered to have GE if every subset S of nodes ($S \leq 50\%$ of the nodes) has a neighborhood that is larger than some "expansion factor" ϕ multiplied by the number of nodes in S. A neighborhood of S is the set of nodes which are linked to the nodes in S. Formally, for each vertex $v \in V$ (where V is the set of nodes in the network), the neighborhood of v, denoted as $\Gamma(v)$ is defined as: $\Gamma(v) = \{u \in V \,|\, (u,v) \in E\}$ (where E is the set of links in the network). Then, the neighborhood of a subset $S \subseteq V$ is defined as the union of the neighborhoods of the nodes in S: $\Gamma(S) = \bigcup_{v \in S} \Gamma(v)$ and the network has GE if $\Gamma(v) \geq \phi|S| \; \forall S \subseteq V$. Good expansion networks (GENs) show excellent communication properties due to the absence of bottlenecks, which are small sets of nodes/links whose elimination leads to fragmentation of the network into at least two large connected components.

While the determination of network robustness with respect to hub removal can be determined by studying the degree distribution (DD) of the nodes [6–9], the exploration of GE properties is an NP-hard computational problem [25]. An indirect exploration of GE properties

[a] e-mail: estrada66@yahoo.com

can be carried out by analysis of the spectrum of the adjacency matrix of the network [26]. A necessary condition for a network to be a good expander is that the gap between the first and second eigenvalues of the adjacency matrix ($\Delta\lambda = \lambda_2 - \lambda_1$) be sufficiently large. The problem is to determine how large this *spectral gap* needs to be for a network be a GEN. We can solve this problem by assessing the degree of correlation between the largest eigenvector of the adjacency matrix of the network and the weighted sum of all odd-length walks that start and end at the same node. The theoretical justification of this method is given below.

2 2 GE networks vs. non-GE networks

The simplest way of defining informally a good expansion network (GEN) is by defining a network lacking this property. A non-GEN is a graph having at least two parts that can be isolated from one another by disconnecting a "small" number of nodes or links. These nodes or links which make the function of bridges between these parts are known as bottlenecks. Consequently, a GEN is a network without bottlenecks.

Here we designate the number of walks of length L starting at node i by $N_L(i)$ and the total number of walks of this length existing in the network by $N_L(G)$. The probability that a walk selected at random in the network has started at node i is simply:

$$P_L(i) = \frac{N_L(i)}{N_L(G)}. \tag{1}$$

We will consider walks of extremely large lengths, in such a way that we can consider $L \to \infty$. This probably increases with the density of the network. That is, in a network having a large proportion of links to nodes there is a high probability that a walk selected at random has started in an arbitrary node i. $P_L(i)$ is equal to one when all nodes are interconnected to each other, the so-called complete graph. In this case every walk of infinite length starting at any node of the network will visit necessarily at least one of the other nodes, but because all nodes are equivalent, it guarantees that the walk will necessarily visit node i.

There is a graph spectral measure that fulfil the requirements for being considered as the probability $P_L(i)$. The "spectrum" of a network is a listing of the *eigenvalues* of the adjacency matrix of such network. It is well-known that every $n \times n$ real symmetric matrix **A** has a spectrum of n orthonormal eigenvectors $\overrightarrow{\gamma}_1, \overrightarrow{\gamma}_2, \ldots, \overrightarrow{\gamma}_n$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ [26]. It is known that for non-bipartite connected network such as those studied in this work, with nodes $1, 2, \ldots, n$, for $L \to \infty$, the vector $[P_L(1) \quad P_L(2) \quad \cdots \quad P_L(n)]$ tends toward the eigenvector corresponding to the largest eigenvalue of the adjacency matrix of the network. Consequently, if λ_1 designates the largest eigenvalue of the network and γ_1 is the eigenvector corresponding to λ_1, then the elements of γ_1 represent the probabilities of selecting at random

a walk of length L starting at node i when $L \to \infty$: $\gamma_1(i) = P_L(i)$ [27].

On the other hand, for dense and homogenous networks it is natural to expect that the nodes with the largest probabilities $P_L(i)$ correspond to highly clustered nodes. This is well illustrated by the example of the complete graph, in which the nodes have the largest possible clustering. Highly clustered nodes are those implicated in a large number of clusters. However, if a network is formed by two dense parts, A and B, connected with a "small" number of links the situation is radically different. In this case, one particular node can be highly clustered but it can displays a very low probability that a walk selected at random in the network has started at that node. This is because a walk can infinitely visit the nodes of one of the two chunks forming the network, which are not necessarily where our "target" node is located. For a walk starting at a node located at A visits the part B of the network it has to cross the narrow bridge, or bottleneck, connecting both branches.

We can also account for the probability that a "short-range" walk selected at random in the network has started and ended at node i. This probability can be written as follow:

$$P_S(i) = \frac{SC_{odd}(i)}{SC_{odd}(G)} \tag{2}$$

where the term SC_{odd} refers to the *odd subgraph centrality* [28,29], which accounts for the participation of a node in all subgraphs containing at least one odd cycle but giving more weight to closest neighbours. This weighting scheme guarantees the "short-range" nature of this parameter, because walks visiting very distant nodes receive a very small weight in the sum [28,29]:

$$SC_{odd}(i) = \sum_{l=0}^{\infty} \frac{\mu_{2l+1}(i)}{(2l+1)!}. \tag{3}$$

We have previously shown that this expression can be calculated by using the graph spectrum in the following form [29]:

$$SC_{odd}(i) = \sum_{i=1}^{N} [v_j(i)]^2 \sinh(\lambda_j). \tag{4}$$

Because $SC_{odd}(G)$ is simply the sum of $SC_{odd}(i)$ for all nodes in the network we can use the values of $SC_{odd}(i)$ directly instead of using $P_S(i)$. The approach that study the existence of a correlation between $P_L(i)$ and $SC_{odd}(i)$ is designated as *spectral scaling method*.

3 Spectral scaling method

If we consider the extreme case of a network consisting of two separated (identical) components such as the one illustrated in Figure 1a we can see that the difference between the largest and second largest eigenvalues (spectral gap) is zero: $\Delta\lambda = 0$. The inclusion of a "small" number
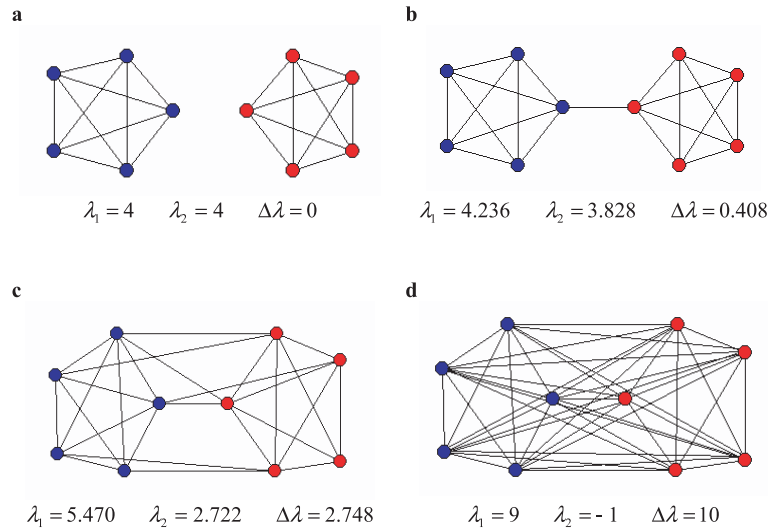
Fig. 1. Illustration of an artificial network formed by two isolated components (a), which shows a spectral gap $\Delta\lambda = 0$. The connection of these two components by a single link (b) represents an example of a network lacking good expansion properties due to the presence of a bottleneck. The spectral gap in this network is close to zero. The inclusion of more links between the two components of the network increases the spectral gap for network in (c), which reach the maximum for the complete graph (d), where every pair of nodes is connected to each other.

of links joining together these two components will change the values of the first two eigenvalues but we can see that the spectral gap remains close to zero (see Fig. 1b). However, when the number of links connecting both clusters increases, the spectral gap also increases, reaching its maximum when every pair of nodes is connected to each other (see Figs. 1c and 1d). Consequently, it is known that when a network has bottlenecks like those in Figure 1b, $\lambda_1 \approx \lambda_2$, but on the contrary, in cases where the network is a GEN $\lambda_1 \gg \lambda_2$.

Then, if we write the expression for $SC_{odd}(i)$ in the following form:

$$SC_{odd}(i) = [\gamma_1(i)]^2 \sinh(\lambda_1) + \sum_{j=2} [\gamma_j(i)]^2 \sinh(\lambda_j) \quad (5)$$

where $\gamma_1(i)$ is the ith component of the principal eigenvector and λ_1 is the principal eigenvalue of the network. We will have that for networks having good expansion properties the first term of (5) will be much larger than the second one, $[\gamma_1(i)]^2 \sinh(\lambda_1) \gg \sum_{j=2} [\gamma_j(i)]^2 \sinh(\lambda_j)$. Thus:

$$SC_{odd}(i) \approx [\gamma_1(i)]^2 \sinh(\lambda_1) \quad (6)$$

and the principal eigenvector of the network will be directly related to the subgraph centrality in GENs according to the following expression:

$$\gamma_1(i) \propto A[SC_{odd}(i)]^\eta \quad (7)$$

where $A \approx [\sinh(\lambda_1)]^{-0.5}$ and $\eta \approx 0.5$. This means that a linear correlation exists between the $\gamma_1(i)$ and $SC_{odd}(i)$ for GENs, which in a log-log scale can be written as [30]:

$$\log[\gamma_1(i)] = \log A + \eta \log[SC_{odd}(i)]. \quad (8)$$

Consequently, a log-log plot of $\gamma_1(i)$ vs. $SC_{odd}(i)$ has to show a linear fit with slope $\eta \approx 0.5$ and intercept $\log A$ for GENs. This "ideal" straight line is illustrated in all the Figs. showing spectral scaling along this paper. A case where this ideal case coincides with the plot of data points for a real network is illustrated in Figure 2a. However, the attachment of a small cluster to this network with a few links connecting the two major chunks of the network makes the GE properties disappear. This case is illustrated in Figure 2b. This lack of GE properties is immediately obvious from the lack of scaling observed because the spectral gap $(\lambda_1 - \lambda_2)$ is not "sufficiently large". In this case it is observed that despite node A showing a great deal of cliquishness, and consequently a large value of $SC_{odd}(i)$, it displays a low value of the largest eigenvector, $\gamma_1(i)$. This means that the probability that a walk selected at random in the network has started at this node is very low. The explanation is, of course, due to the fact that most of the walks can visit the nodes at the right part of the network (blue nodes) an infinite number of times. This node A is the most distant from the largest cluster of the network, i.e., it is separated by three steps from the blue nodes. The situation is different for node B, which is directly connected to the nodes in the larger cluster and the probability $\gamma_1(i)$ is proportional to its cliquishness. The rest of nodes, which are at distance two are in an intermediate situation compared to nodes A and B.

The situation is identical if we consider one network as illustrated in Figure 3a and then rearrange it into two clusters with very few connections between them (Fig. 3b). In this case, despite both networks display the same degree distribution, the first shows GE properties and the second not as illustrated by their spectral scaling. Note in Figures 2 and 3 that the spectral scaling is always able to discern between GENs and non-GENs despite the spectral
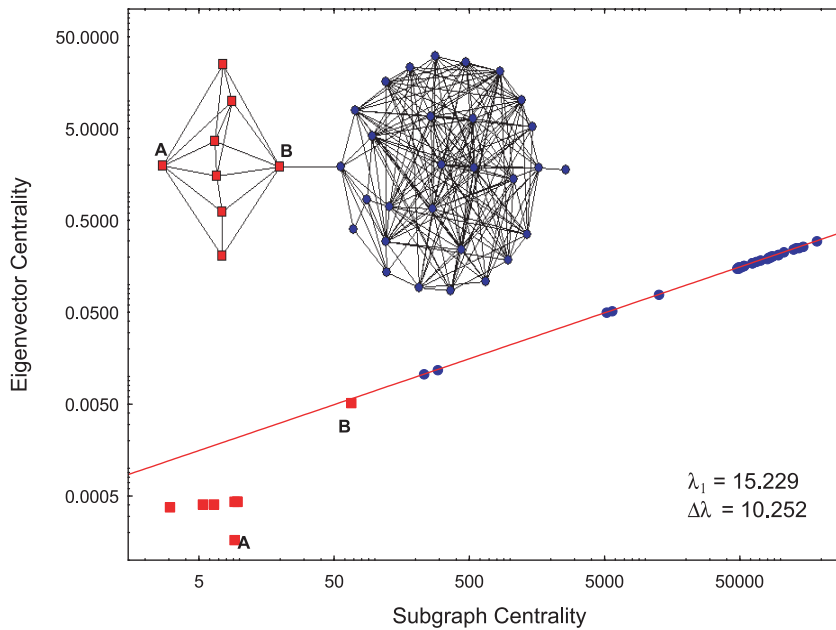
a)
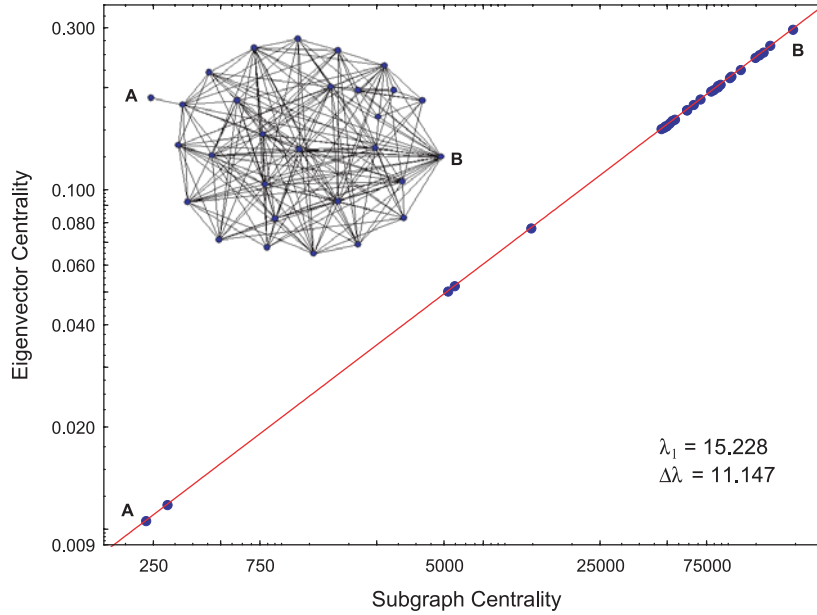
b)

Fig. 2. Scaling plot of a network with good expansion properties (a) corresponding to the food web of Benguela. Nodes A and B correspond to the those with lowest and largest cliquishness in the network, respectively, which also show the lowest and largest probabilities that a walk selected at random in the network have started at one of them. The network illustrated in (b) is formed by attaching a small cluster to the network given in (a). This network lacks good expansion properties due to the presence of a bottleneck connecting the nodes in blue to those in red.

gap not always distinguish between these two cases. For instance, in Figure 2 the difference between $\Delta\lambda$ for a GEN and a non-GEN is not significant.

We will introduce here a measure of the GE character of a network. The "*good expansion character*" of a complex network is given by the deviation from this scaling plot, in such a way that perfect GENs will have $\xi(G) = 0$; the higher the value of $\xi(G)$ the larger the departure of the network from GE properties:

$$\xi(G) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\{\log\left[\gamma_1(i)\right] - \left[\log A + \eta\log\left[SC_{odd}(i)\right]\right]\right\}^2}.$$

(9)

4 Network expansibility

Our investigation of the expansibility of complex networks starts with the analysis of 51 real-world networks. They include the following naturally-evolving networks: two protein-protein interaction networks (PINs), one for *Saccharomyces cerevisiae* (yeast) and other for the bacterium *Helicobacter pylori*; three transcription interaction networks concerning *E. coli*, yeast and sea urchins; a neural network in *C. elegans*; three networks of adjacency of secondary-structure elements for large proteins; and 17 food webs representing a wide range of species numbers, linkage densities, taxa, and habitat types. These webs are Benguela, Bridge Brook, Canton Creek, Chesapeake Bay,
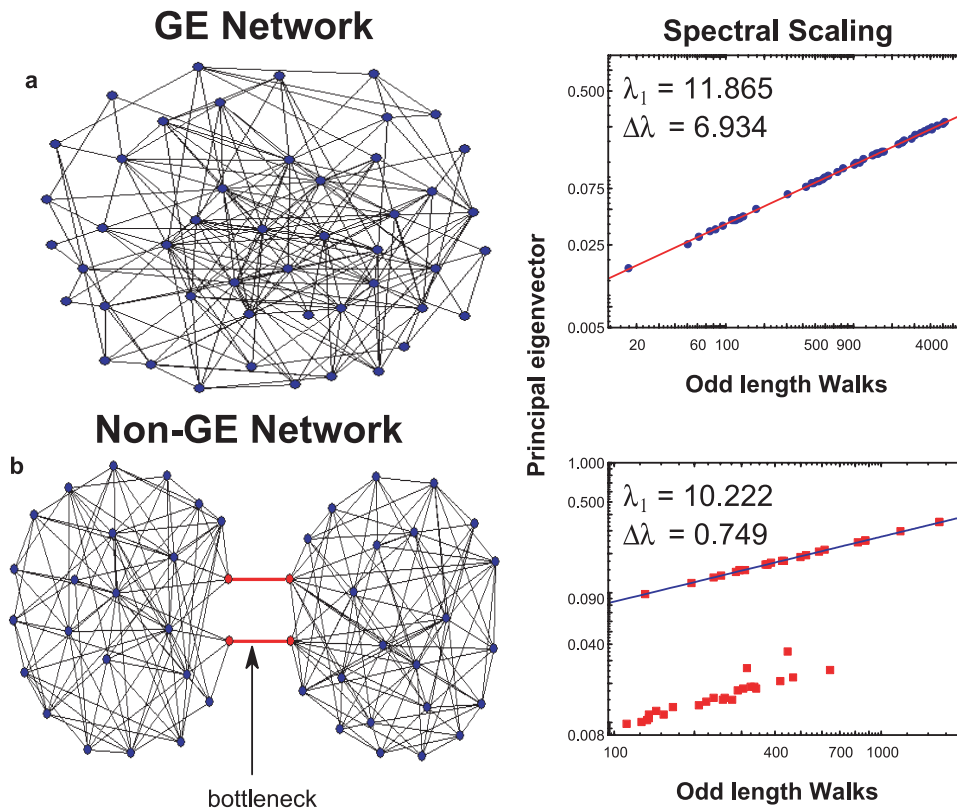
Fig. 3. Illustration of the differences between networks with and without good expansion (GE) properties and having the same degree distribution. (a) GEN; (b) non-GEN.

Coachella Valley, El Verde rainforest, Grassland, Little Rock Lake, Reef Small, Scotch Broom, Shelf, Skipwith Pond, St. Marks Seagrass, St. Martin Island, Stony, Ythan Estuary (1) with and without parasites (2). The other systems are formed by man-made complex networks including two semantic networks, one based on Roget's Thesaurus of English (Roget) and the other on the Online Dictionary of Library and Information Science (ODLIS) and three citation networks: one consisting of papers published in the *Proceedings of Graph Drawing* in the period 1994–2000 (GD), papers published in the field of "Network Centrality" (Centrality) and papers published or citing articles from *Scientometrics* for the period 1978–2000 (SciMet); three electronic sequential logic circuits parsed from the ISCAS89 benchmark set, where nodes represent logic gates and flip-flops, the airport transportation network in the US in 1997, the Internet at the autonomous systems (AS) level as from April 1998 and five software networks: Abi, Digital, MySQL, VTK and XMMS; 10 social networks that include a network of the corporate elite in US, a scientific collaboration network in the field of computational geometry (Geom), inmates in prison, injectable drug users (IDUs), Zachary karate club, college students on a course about leadership, a sexual network in Colorado Springs and collaboration between Jazz musicians. The results of the spectral scaling method are given in Table 1 together with the values of the spectral gap and principal eigenvalue for the 51 networks studied as well as their degree distribution. In Figure 4 we illustrate the re-

sults of both the spectral scaling and degree distribution for some of 51 networks studied.

These networks are classified as GENs or non-GENs according to their values of the spectral scaling: correlation coefficient, slope and expansion character. A network is considered here to be GEN if $\xi(G) < 10^{-2}$, $r > 0.999$ and Slope = 0.5, simultaneously. We find that only 39% of the complex networks studied here show GE properties; of these, 65% are naturally-evolving networks and only 35% correspond to man-made systems. Natural systems tend indeed to be GENs: only 42% of the complex networks which are not GENs represent natural systems. We have to remark that we have considered here nonweighted networks for the analysis of their GE properties. However, it is known that some of the networks analyzed here have different weights for the links representing information valuable for the context in which such systems exist in the real-world. Consequently, we need to distinguish between two different kinds of GE properties of networks. The first, which we have considered here, refers to the "purely" topological expansibility character of a network. These GE properties are concerned with the lack of topological bottlenecks, whose elimination will separate the network into large separated clusters. However, if we consider weighted complex network then the analysis of GE properties take a different dimension. In this case the existence of GE represents a sort of "functional" expansibility of the network, which is mainly concerned with the lack of functional bottlenecks in such system. For instance,

Table 1. Results obtained for the 51 complex networks studied in this work. The number of nodes (N) and links (E) as well as the node degree distribution (DD) displayed by each network are given. The spectral parameters, λ_1 (largest eigenvalue) and $\Delta\lambda$ (spectral gap) are also given together with the parameters of the spectral scaling: correlation coefficient, r, slope of the regression plot in log-log scale and the expansion character, $\xi(G)$. The networks are classified according to their good expansion properties as determined by the classification method.

No.	Network	N	E	DD*	λ_1	$\Delta\lambda$	r	Slope	$\xi(G)$
				Good expansion networks					
1	Benguela	29	191	U	15.228	11.147	0.9998	0.50	6.30×10^{-3}
2	Coachella	30	241	U	18.147	13.058	1.0000	0.50	7.18×10^{-5}
3	Skipwith	35	353	U	22.076	18.648	1.0000	0.50	6.16×10^{-5}
4	St. Martin	44	218	E	12.531	5.559	0.9999	0.50	1.50×10^{-3}
5	St. Marks	48	218	U	11.865	6.934	0.9999	0.50	2.90×10^{-3}
6	Reef Small	50	503	U	23.756	15.174	1.0000	0.50	4.31×10^{-5}
7	Bridge Brook	75	542	U	20.639	7.943	1.0000	0.50	9.00×10^{-3}
8	Shelf	81	1451	U	41.916	30.193	1.0000	0.50	6.75×10^{-5}
9	Ythan2	92	416	E	15.771	9.635	0.9999	0.50	2.92×10^{-3}
10	Centrality	118	613	E	19.331	10.896	1.0000	0.50	2.26×10^{-5}
11	Ythan1	134	593	E	16.737	9.272	0.9999	0.50	1.50×10^{-3}
12	El Verde	156	1439	E	31.494	22.082	1.0000	0.50	4.50×10^{-5}
13	Little Rock	181	2318	E	40.816	14.643	1.0000	0.50	3.72×10^{-5}
14	Small World	233	994	P	20.962	6.239	0.9999	0.50	3.26×10^{-3}
15	Neurons	280	1973	E	23.293	9.227	1.0000	0.50	2.70×10^{-4}
16	USAir97	332	2126	E	41.233	23.924	1.0000	0.50	9.04×10^{-5}
17	Jazz	1265	38356	P-E	171.490	58.900	1.0000	0.50	1.03×10^{-3}
18	ODLIS	2898	16376	P	44.184	15.677	0.9999	0.50	1.55×10^{-5}
29	Internet97	3015	5156	P	31.754	11.673	0.9999	0.50	1.18×10^{-3}
20	Internet98	3522	6324	P	34.892	13.195	0.9999	0.50	9.90×10^{-4}
				Not good expansion networks					
21	College	32	96	E	5.971	2.161	0.9210	0.59	0.073
22	Chesapeake	33	71	EU	5.745	1.215	0.9824	0.48	0.094
23	Zachary	34	78	E	6.726	1.749	0.9490	0.55	0.066
24	Trans Urchins	45	80	E	6.684	3.735	0.8850	0.63	0.430
25	GD	49	635	E	9.210	1.480	0.9070	0.57	0.465
26	Protein2	53	123	U	5.809	0.723	0.8060	0.76	0.504
27	Prison	67	142	U	5.591	0.975	0.8290	0.40	0.228
28	Grassland	75	113	P	5.524	1.274	0.6150	0.31	0.518
29	Protein1	95	213	U	5.361	0.314	0.5220	0.74	1.100
30	Protein3	97	212	U	6.531	0.327	0.6940	0.64	1.200
31	Canton	108	707	E	19.559	12.524	0.9930	0.49	0.183
32	Stony	112	830	U	22.702	16.423	0.9905	0.50	0.219
33	Electronic1	122	189	U	4.106	0.476	0.6587	0.22	0.743
34	Digital	150	198	P	6.702	1.860	0.9852	0.44	0.323
35	Scotch Broom	154	366	P-E	14.714	8.714	0.9997	0.50	0.020
36	Electronic2	252	399	U	4.360	0.398	0.7321	0.23	0.746
37	ColoSpring	324	347	P	4.876	0.031	0.9147	0.38	0.802
38	Trans Ecoli	328	456	P	9.064	2.766	0.6970	0.52	0.764
39	Electronic3	512	819	U	5.010	0.889	0.6359	0.25	1.030
40	Drugs	616	2012	E	18.010	3.776	0.7410	0.71	1.390
41	Trans Yeast	662	1062	P	9.976	1.524	0.7240	0.40	0.796
42	PIN-2	710	1396	P	10.466	2.216	0.9890	0.50	0.104
43	VTK	771	1357	P	11.458	2.751	0.9868	0.50	0.140
44	XMMS	971	1802	E	10.28	1.352	0.7466	0.57	1.218
45	Roget	994	3640	PU	12.027	2.218	0.9770	0.58	0.230
46	Abi	1035	1719	P	11.945	4.269	0.9569	0.49	0.281
47	MySQL	1480	4140	P	21.740	7.300	0.9163	0.90	1.666
48	Corporate elite	1586	11540	UE	23.229	3.999	0.9937	0.51	0.053
49	PIN-1	2224	6608	P	19.039	3.355	0.9920	0.52	0.141
50	SciMet	2678	10368	P	20.429	3.525	0.9930	0.52	0.102
51	Geom	3621	9461	P	29.026	10.004	0.9320	0.56	0.452

*U = Uniform; E = Exponential; P = Power-law. Networks denoted by P-E display power-law degree distributions with an exponential drop-off of the tail.
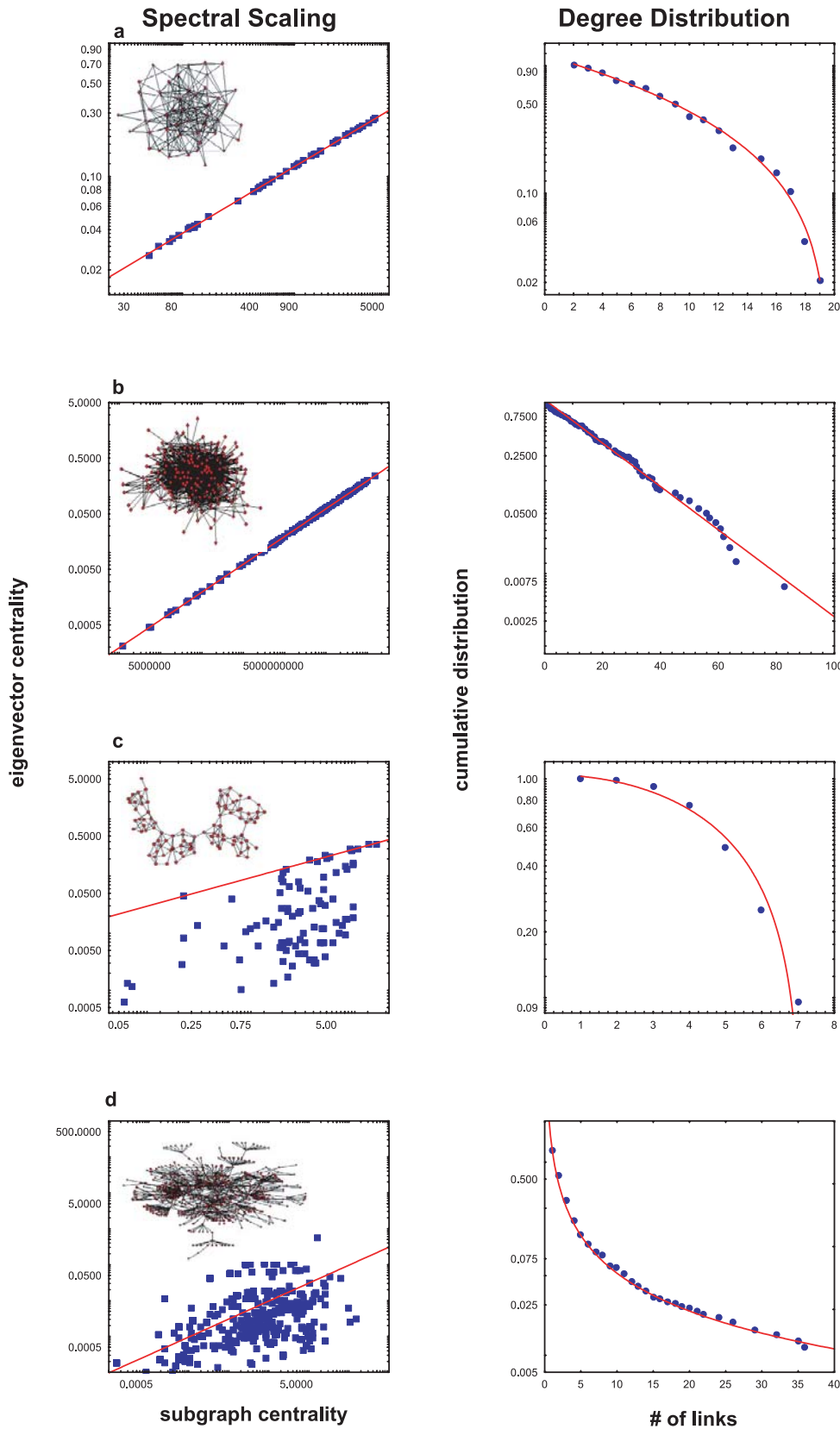
Fig. 4. Classes of networks. (a) GE network with uniform DD (St. Marks seagrass food web); (b) GE network with exponential/power-law DD (El Verde rainforest food web); (c) non-GE network with uniform DD (Secondary structure elements in the immunoglobulin with PDB 1A4J); (d) non-GE network with exponential/power-law DD (Direct transcriptional regulation between genes in *Saccaromyces cerevisae*). Graphs on the left of this figure show that a network is GE if there is scaling between the principal eigenvector and the sum of odd-closed walks starting and ending at a given node in a log-log scale. The graphs on the right side are linear-log plots of cumulative distribution vs. the number of links in the network.

consider a network formed by two large chunks connected only by a link. For the sake of simplicity let consider that this network represent two parts of a city connected by a bridge. If we consider the non-weighted network it is evident that this bridge constitute a bottleneck of the network. However, let consider the same system as a weighted network in which every link is weighted, for instance by the number of lanes that the corresponding streets have. In the case that the bridge has a large number of lanes it is not necessarily a functional bottleneck. Thus, the GE properties of this weighted network will depend on the weights the links have. This lack of correspondence between topological and functional expansibility can be observed in Figure 5 for the collaboration network studied here for authors in Computational Geometry. When the network is considered as a non-weighted one it lacks GE properties as can be seen in the plot showed at the top. However, if we consider a weighted network in which every link represents the number of papers that the corresponding two author have published together the network is a GEN (see bottom plot). This means that in this particular case the topological and functional expansibility are different. However, for the objectives of this work which is mainly concerned with network robustness we think that the most important characteristic is the topological expansibility. Think for instance in the consequences of closing the bridge in the example of the city given below. In this case, no mater how many lanes this bridge has the two major parts of the city will be incommunicated by roads.

Another important network attribute which is relevant to the analysis of robustness is the node DD. It is known that the response of scale-free networks to attacks is similar to, but more pronounced than the response of exponential networks to attack and random failure [6]. However, because uniform DD is less skewed that exponential and power-law we expect that these networks display lower vulnerability to intentional attacks, which will be explored further on in this work. By considering GE properties and node DD together we can classify complex networks into four different groups: (i) GENs with uniform DD; (ii) GENs with power-law or exponential DD; (iii) non-GENs with uniform DD; and (iv) non-GENs with power-law or exponential DD. Most of the GENs (65%) show exponential or power-law DD. This group of networks included several of the "classical" complex networks such as the Internet, the neural networks of *C. elegans*, and the US airport transportation network, all of which are known to be vulnerable to attacks at their hubs. The remaining 35% of the GENs are complex networks with uniform DDs. These networks do not contain bottlenecks or show a skewed DD of nodes, in which the DD obeys a power law or exponential law. On the contrary, in a network with uniform DD there is the same probability of selecting a node with degree $a \leq k \leq b$ on the interval $[a, b]$ and this probability is zero outside this interval. As a consequence networks with uniform DD can be considered as power-law networks, $P(k) \propto k^{-\gamma}$, with exponent $\gamma = 0$. Due to these particular characteristics of GE networks with uniform DD distribution we expect them to
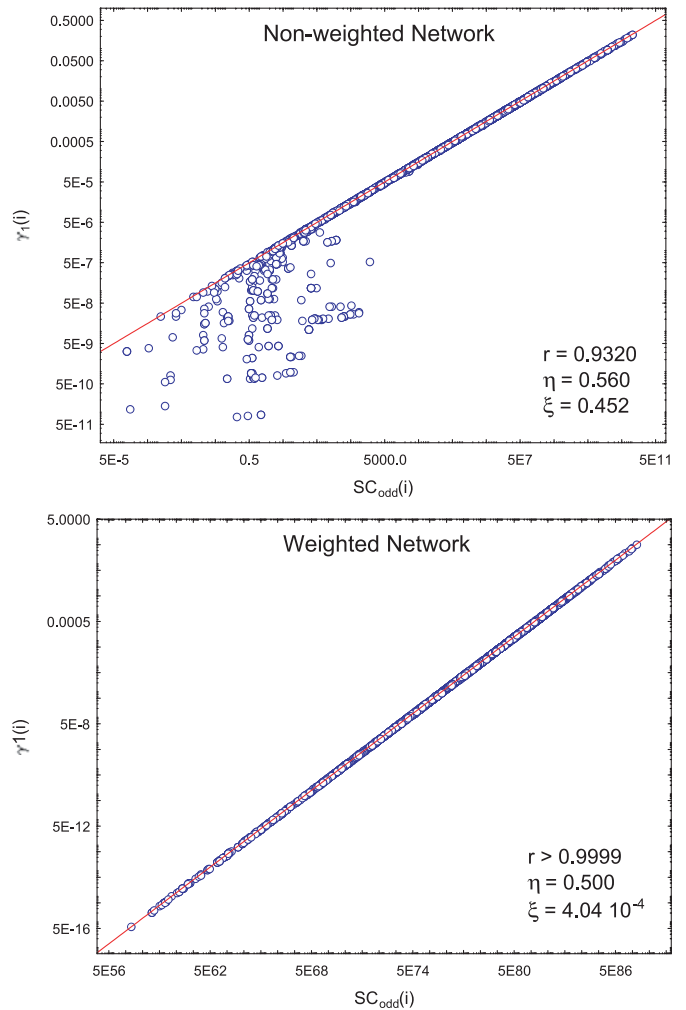

Fig. 5. Comparison of the expansibility characteristics of the collaboration network in Computational Geometry when links are considered as non-weighted (top) and weighted (bottom). The study of the non-weighted network considers the purely topological expansibility properties of this network. However, the analysis of the weighted network is concerned with the existence of functional expansibility in this complex system.

be more robust to intentional attacks against the most connected nodes than the more skewed networks in the other groups. In fact, the most robust food webs identified by Dunne et al. [31], which coincide with the ones having the highest average degree — Bridge Brook, Coachella and Skipwith — display uniform degree distribution and have been identified here as having GE properties.

The totality of uniform GE networks are exemplified by food webs in different ecological systems [32]. At the other extreme, 21 networks ($\sim$41% of the total networks studied) are not GENs and show power-law or exponential node DD. Consequently, these networks should be vulnerable to node/link attacks which operate by removing either the hubs or the vertices/links which form bottlenecks. In this most vulnerable group we find 6 social and 5 technological networks. The absence of GE properties in social networks reveals one of their most prominent topological
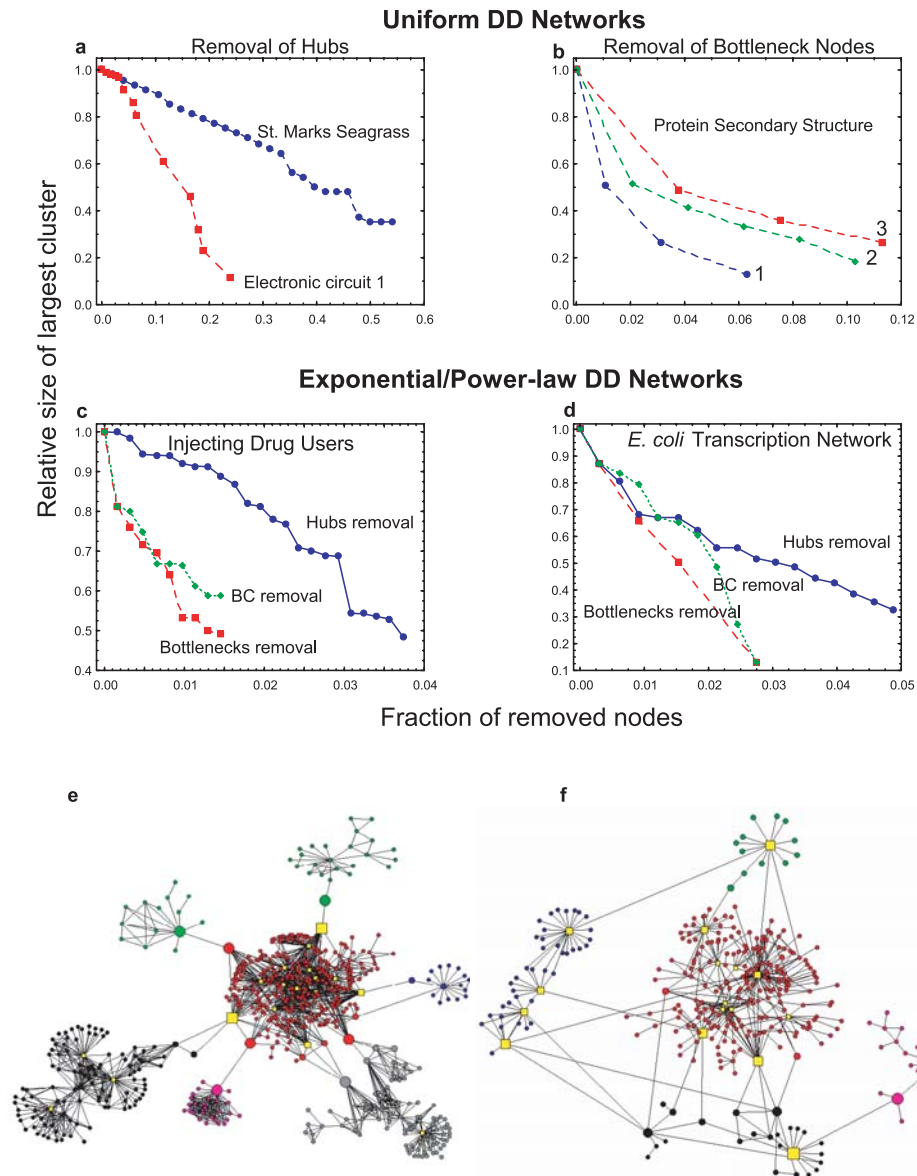
Fig. 6. Network fragmentation resulting from removal of nodes. (a) Removal of the most connected nodes from networks with uniform DD. GE networks (St. Mark's Seagrass) tolerate this well but non-GE networks (Electronic Circuit 1) are fragmented catastrophically by removal of only 20% of their nodes; (b) removal of bottleneck nodes from non-GE networks with uniform DD results in decomposition of the network after removal of only 10% of the nodes; (c, d) removal of hubs or bottleneck nodes from non-GE networks with exponential/power law DD. The networks are more vulnerable to attack at the bottleneck nodes, because many of the hubs in these networks, the yellow squares in e (injecting drug users) and f (*E. coli* transcription networks), are not bottleneck nodes. A node is more important (larger circles/squares in e and f) when its removal isolates a larger part of the network.

characteristics: the appearance of highly clustered communities [33]. Only 27% of the technological, infrastructure and communication networks studied show GE properties, thus almost 73% of the man-made networks studied here, for example, the software networks and electronic circuits can be regarded as very vulnerable. The global consequences of topological structures on the robustness, functioning and strategic management of complex networks will be evident from the following analysis.

5 Network robustness

We simulate a first strategy of attack on complex networks by successive removal of the nodes in decreasing order of connectivity. We start by studying networks with uniform DDs. In Figure 6a we illustrate the effect of removal of the most connected nodes in a food web having GE properties and showing uniform DD. It can be observed that after removing almost 60% of nodes the largest cluster

retains about 40% of the total number of nodes originally in the network. However, when a network with uniform DD but lacking GE properties is attacked by this strategy the picture is quite different. As illustrated in Figure 6a for an electronic circuit the removal of only 25% of the hubs reduces the largest cluster to about 10% of the components in the original circuit. This situation occurs because the uniform distribution of node degrees ensures that many of the most connected nodes coincide with bottleneck nodes in the electronic circuit. Of course, the removal of bottleneck nodes produces catastrophic fragmentation in a network lacking GE properties. This constitutes our second strategy of attack on complex networks. By removing bottleneck nodes we also simulate the effect of removing bottleneck links. In Figure 6b we illustrate the effect of removing bottleneck nodes of the three networks formed by secondary structure elements of proteins, which represent networks in our third group. It can be seen in this case that the elimination of only 6–12% of nodes reduces the largest cluster to less than 20% of its former size.

In those networks which are not GENs and which display exponential or scale-free DDs an informed agent can use either or both of these two strategies to inflict deliberate damage on the network structure. A third strategy was considered by Holme et al. [34] by using the betweenness centrality (BC), which is defined as follows [35]. If $\rho(i,j)$ is the number of shortest paths from node i to node j, and $\rho(i,k,j)$ is the number of these shortest paths that pass through node k in the network, then the betweenness centrality of node k is given by:

$$BC(k) = \sum_i \sum_j \frac{\rho(i,k,j)}{\rho(i,j)}, \quad i \neq j \neq k. \qquad (10)$$

In Figures 6c and 6d we illustrate the consequences of using these three alternatives for a social network of injecting drug users (IDUs) with exponential DD and for the biological network of direct transcriptional regulation between operons in *E. coli*, which has scale-free characteristics. As can be seen in both cases the attack over the bottleneck nodes causes the networks to disintegrate faster than does hub removal. In the social network, removal of only 1.5% of bottleneck nodes reduces the main cluster to a size representing less than 50% of the original network. The strategy based on BC removals produces a faster disintegration of the network as compared to hubs removals but in neither case overtakes the effects produced by bottleneck removals. In fact, it could happen that the bottlenecks are those nodes having the largest betweenness, which is the case for instance of the *E. coli* network. In Figure 6e we illustrate some of this network's bottlenecks, which are formed by a small number of links connecting clusters with a large number of internal interconnections. Only 25% of these bottleneck nodes coincide with some of the hubs in the network (marked as yellow squares). In some cases, the removal of only one bottleneck node isolates a large cluster of individuals from the main drug user community. The situation is quite similar for the scale-free network for which the removal of about 3% of bottleneck

nodes produces a catastrophic fragmentation of the network (see Fig. 6f).

An interesting question that arise from the previous analysis is whether the vulnerability of non-GENs arises as a consequence of this topological characteristic or it is produced by other organizational principles. We have previously shown in this work that two networks with the same degree distribution can show different expansibility properties due to different organization of the links in the network (see Fig. 3). In order to know how these differences in expansibility affect the resilience of a network to intentional attacks we have studied the Grassland food web, which is a non-GEN with power-law DD. We have improved the good expansion character of this network by means of a random rewiring in such a way that we have created a random network with the same DD (and average degree) than the real-world web. In Figure 7 we can see that the random network with identical DD that Grassland display more robustness to targeted attack against the most connected nodes only because the number of bottlenecks have been reduced as a consequence of the rewiring. Consequently, this experiment evidences that the lack of GE properties is an important factor for understanding network resilience to targeted attacks.

We have shown that most technological networks analysed here are not robust to node/link removal due to their lack of GE properties or to the scale-free/exponential degree distributions. In fact, all networks which we have identified as robust against random failure and node/links attack represent naturally-evolving systems. Network designers can consider Nature and design new infrastructure and communication networks by imitating naturally-evolving robust systems. On the other hand, it is possible to achieve strategic control of certain networks by taking advantage of their lack of GE properties. The first example can be provided by drug design strategies. These can include the targeted attack on networks of protein interaction in pathogenic organisms by designing drugs that specifically prevent the interaction between bottleneck proteins. Last, but not least, are social networks lacking GE properties. Here the strategy consists in avoiding the interactions between individuals that represent bottleneck nodes in certain social networks, such as drug users or sexual partners, by appropriate educational programs or vaccination campaigns. Our results then suggest a strategy to design new robust networks, to exploit the weakness of certain complex systems for human benefit, and to protect existing vulnerable networked infrastructures. The results obtained here are based on an empirical analysis of a large pool of complex networks. However, we are optimistic that these findings can help the understanding of network robustness and can inspire more fundamental analytical results supporting these conclusions.
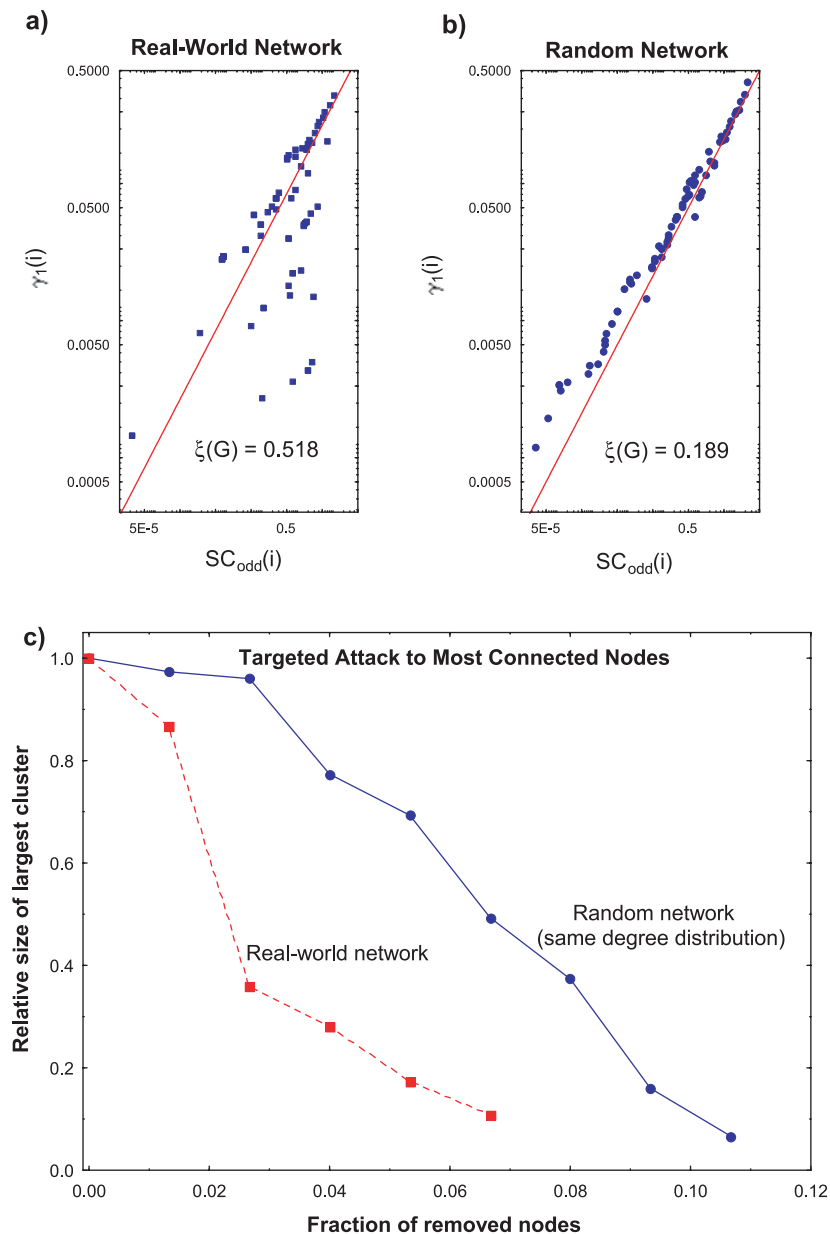
Fig. 7. Comparison of network resilience to targeted attack against the most connected nodes for networks with the same degree distribution and average degree and different expansion character. (a) Spectral scaling of the real-world network (Grassland food web) which is a non-GEN with power-law DD; (b) spectral scaling of a random network with the same DD but having an improved expansion character compared to Grassland food web; (c) resilience of both networks to targeted attack against the most connected nodes.

References

1. S.H. Strogatz, Nature **410**, 268 (2001)
2. R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
3. M.E.J. Newman, SIAM Rev. **45**, 167 (2003)
4. L.A.N. Amaral, J.M. Ottino, Eur. Phys. J. B **38**, 147 (2004)
5. *Network Science* (National Research Council, National Academy Press, Washington DC, 2005)
6. R. Albert, H. Jeong, A.-L. Barabási, Nature **406**, 378 (2000)
7. D. Callaway, M.E.J. Newman, S.H. Strogatz, D.J. Watts, Phys. Rev. Lett. **85**, 5468 (2000)
8. G. Paul, T. Tanizawa, S. Havlin, H.E. Stanley, Euro. Phys. J. B **38**, 187 (2004)
9. T. Tanizawa, G. Paul, R. Cohen et al., Phys. Rev. E **71**, 047101 (2005)
10. J. Balthrop, S. Forrest, M.J.E. Newman, M.M. Williamson, Science **304**, 527 (2004)
11. A.-L. Barabási, Z.N. Oltvai, Nature Rev. Genet. **5**, 101 (2004)
12. F. Liljeros, C.R. Edling, L.A.N. Amaral et al., Nature **411**, 907 (2001)
13. J.A. Dunne, R.J. Williams, N.D. Martinez, Ecology Lett. **5**, 558 (2002)

14. H. Jeong, S.P. Mason, A.-L. Barabási, Z.N. Oltvai, Nature **411**, 41 (2001)
15. E. Estrada, Proteomics **6**, 31 (2006)
16. R. Pastor-Satorrás, A. Vespignani, Phys. Rev. E. **65**, 036104 (2002)
17. P. Holme, Europhys. Lett. **68**, 908 (2004)
18. R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000)
19. R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001)
20. A. Valente, A. Srakar, H.A. Stone, Phys. Rev. Lett. **92**, 11872 (2004)
21. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)
22. S.N. Dorogovtsev, J.F.F. Mendes, Europhys. Lett. **52**, 33 (2000)
23. P. Sarnak, Notices of the AMS **51**, 762 (2004)
24. C. Gkantsidis, M. Mihail, A. Saberi, Perform. Eval. **63**, 241 (2006)
25. B. Mohar, J. Comb. Theor. B **47**, 274 (1989)
26. F.R. Chung, *Spectral Graph Theory* (American Mathematical Society Book Series, 1997)
27. D. Cvetkoviæ, P. Rowlinson, S. Simiæ, *Eigenspaces of Graphs* (Cambridge University Press, Cambridge, 1997)
28. E. Estrada, J.A. Rodríguez-Velázquez, Phys. Rev. E. **71**, 056103 (2005)
29. E. Estrada, J.A. Rodríguez-Velázquez, Phys. Rev. E. **72**, 055510, (2005)
30. E. Estrada, Europhys. Lett. **73**, 649 (2006)
31. J.A. Dunne, R.J. Williams, N.D. Martinez, Ecology Lett. **5**, 558 (2002)
32. J.A. Dunne, R.J. Williams, N.D. Martinez, Proc. Natl. Acad. Sci. USA **99**, 12917 (2002)
33. M. Girvan, M.E.J. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002)
34. P. Holme, B.J. Kim, C.N. Yoon, S.K. Han, Phys. Rev. E **65**, 056109 (2002)
35. L.C. Freeman, Sociometry **40**, 35 (1977)