EUROPHYSICS
LETTERS

OFFPRINT

Vol. 73 • Number 4 • pp. 649–655

Spectral scaling and good expansion properties in complex networks

* * *

E. Estrada

1986
2006 20 years of successful publishing

Published under the scientific responsibility of the
EUROPEAN PHYSICAL SOCIETY

Incorporating
JOURNAL DE PHYSIQUE LETTRES • LETTERE AL NUOVO CIMENTO

EUROPHYSICS LETTERS 15 February 2006

Europhys. Lett., **73** (4), pp. 649–655 (2006)
DOI: 10.1209/epl/i2005-10441-3

Spectral scaling and good expansion properties in complex networks

E. Estrada

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

received 30 September 2005; accepted in final form 2 January 2006
published online 11 January 2006

PACS. 89.75.Fb – Structure and organization in complex systems.
PACS. 89.75.Da – Systems obeying scaling laws.
PACS. 89.75.Hc – Networks and genealogical trees.

Abstract. – The existence of a scaling between the principal eigenvector and the subgraph centrality of a complex network indicates that the network has "good expansion" (GE) properties. GE is the important but counterintuitive property of being both sparsely populated and highly connected. We have detected GE properties in half of the 16 real-world networks studied, which include communication, information and biological networks. Most of social networks studied do not show GE properties as a consequence of the existence of communities with low number of inter-community links. However, the majority of food webs represent ecosystems that are not composed of separate communities with low interconnections among them and possess GE properties.

The study of complex networks represents a unifying language describing systems in disparate real-world contexts ranging from biological to technological systems [1–7]. Important global characteristics of complex networks such as *small-worldness* [8,9], *scale-freeness* [10,11], the existence of network motifs [12,13] and self-similarity characteristics [14] have been derived from these studies. Most of these analyses have been limited to static structural characteristics or to statistical parameters which are only loosely connected to structural properties. It is known however, that global measures of network properties are well characterized by spectral methods [15,16]. One of these properties is known as "good expansion" (GE) [17,18]. It is the apparently contradictory property of a network of being both sparse and highly connected.

A network is considered to have GE if every (up to 1/2 the number of nodes) subset of nodes S has a neighborhood that is larger than some factor ϕ multiplied by $|S|$ [17,18]. A neighborhood of S is the set of vertices which have an endpoint in S and the other in $\bar{S}$. The factor ϕ is known as the expansion and its computation is NP-hard [19]. It is known that for a network to be a *good expander*, the second eigenvalue, λ_2, of the adjacency matrix must, compared to the index λ_1, be as small as possible [20,21]. In particular, random regular networks are expected to have big spectral gaps $(\lambda_1 - \lambda_2)$ with high probability, and thus are GE [22]. Good expansion networks (GENs) are desired as they show excellent communication properties due to the absence of bottlenecks. A bottleneck is a small set S for which $G\backslash S$ has at least two large connected components. GENs also yield good error-correcting codes and exhibit certain desired pseudo-random properties [23]. Thus it is interesting to investigate which real-world complex networks exhibit GE characteristics.

One important question in the analysis of GE properties of real-world networks is the problem of determining how large the spectral gap must be for the network to have GE properties. Here we avoid this problem by analyzing the scaling between two graph spectral properties of a network, the principal eigenvector and the subgraph centrality. The subgraph centrality, $SC(i)$, is based on the sum of all closed walks (CWs) of different lengths in the network starting (and ending) at node i. A walk of length l is any sequence of (not necessarily) different vertices $v_1, v_2, \cdots, v_l, v_{l+1}$ such that for each $i = 1, 2, \cdots, l$ there is an edge from v_i to v_{i+1}. A CW of length l is a walk in which $v_{l+1} = v_1$. Using spectral graph theory we have shown [24] that $SC(i)$ can be obtained as follows:

$$SC(i) = \sum_{j=1}^{N} [\gamma_j(i)]^2 e^{\lambda_j}, \tag{1}$$

where $\gamma_j(i)$ is the i-th component of the j-th eigenvector of the adjacency matrix $\boldsymbol{A}$ and λ_j is the corresponding j-th eigenvalue [25]. It should be noted that $SC(i)$ counts all CWs in the network, which can be of even or odd length. CWs of even length might be trivial on moving back and forth in acyclic subgraphs, *i.e.*, those that do not contain cycles, while odd CWs do not contain contributions from acyclic subgraphs. It is easy to show [26] that

$$SC(i) = \sum_{j=1}^{N} [\gamma_j(i)]^2 \cosh(\lambda_j) + \sum_{j=1}^{N} [\gamma_j(i)]^2 \sinh(\lambda_j) = SC_{even}(i) + SC_{odd}(i) \tag{2}$$

which means that the term $SC_{odd}(i)$ only accounts for subgraphs containing at least one odd cycle. In this way $SC_{odd}(i)$ can be considered as a local property of order in networks that characterise the odd-cyclic wiring of a typical neighbourhood. We can write $SC_{odd}(i)$ in the following form:

$$SC_{odd}(i) = [\gamma_1(i)]^2 \sinh(\lambda_1) + \sum_{j=2} [\gamma_j(i)]^2 \sinh(\lambda_j), \tag{3}$$

where $\gamma_1(i)$ is the i-th component of the principal eigenvector and λ_1 is the principal eigenvalue of the network. Here we will consider only non-bipartite networks as for them $SC_{odd}(i) = 0$ [26] and the use of $SC(i)$ instead of $SC_{odd}(i) = 0$ is recommended.

In cases where the network is a GEN it is known that $\lambda_1 \gg \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N$ and the first term of (3) 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

$$SC_{odd}(i) \approx [\gamma_1(i)]^2 \sinh(\lambda_1) \tag{4}$$

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

$$\gamma_1(i) \propto A \left[SC_{odd}(i)\right]^{\eta}, \tag{5}$$

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

$$\log\left[\gamma_1(i)\right] = \log A + \eta \log\left[SC_{odd}(i)\right]. \tag{6}$$

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. A lack of this scaling in a network indicates that the spectral gap $(\lambda_1 - \lambda_2)$ is not "sufficiently large" and that the corresponding network is not a GEN.

The existence of this spectral scaling in GENs is a consequence of the structural characteristics that both spectral measures are accounting for. On the one hand, $\gamma_1(i)$, represents the probability of choosing a walk of length l emanating from node i for $l \to \infty$ (see theorem 2.2.4 in [27]). On the other hand, $SC_{odd}(i)$ accounts for the local wiring around a node as a consequence of the higher weight given to short CWs. As a first example we will consider this scaling in Barabasi-Albert (BA) networks [10]. Gksanditis *et al.* have shown that networks which obey power law degree distribution have GE properties [28]. We have explored this scaling for a BA network having 3000 nodes, which was obtained from $m_0 = 3$ starting nodes by adding $m = 5$ nodes at each step. The linear correlation equation obtained for the scaling is $\gamma_1(i) = -2.8538 + 0.503\,[SC_{odd}(i)]$ with a correlation coefficient, $r = 0.999$. The values of the largest and second-largest eigenvalues for this BA network are $\lambda_1 = 13.834$ and $\lambda_2 = 9.913$, respectively. It can be seen that $\log A = \log\left\{[\sinh(\lambda_1)]^{-0.5}\right\} = -2.8535$, which coincides very well with the intercept of the regression model (-2.8538) and the slope of the regression model (0.503) is very close to the expected value of $\eta = 0.5$.

With the objective of studying the existence of GE properties in real-world complex networks we study 16 complex networks of different types and sizes. These datasets include two semantic networks based on Roget's Thesaurus of English (Roget) ($N = 994$, $E = 3640$) and the Online Dictionary of Library and Information Science (ODLIS) ($N = 2898$, $E = 16376$); two social networks that include inmates in prison ($N = 67$, $E = 142$) and injecting drug users (IDUs) ($N = 616$, $E = 2012$); two bibliographic citation (information) networks, one consisting of papers published in the *Proceedings of Graph Drawing* in the period 1994–2000 ($N = 249$, $E = 635$), papers published in the field of "Network Centrality" ($N = 118$, $E = 613$); the airport transportation network in the US in 1997 ($N = 332$, $E = 2126$); the Internet at the autonomous systems (AS) level from September 1997 ($N = 3015$, $E = 5156$); one network of secondary-structure elements adjacency for a large protein ($N = 95$, $E = 213$); the protein-protein interaction network (PINs) for *Saccharomyces cerevisiae* (yeast) ($N = 2224$, $E = 6608$); the transcription interaction network of yeast ($N = 662$, $E = 1062$); a neural network in *C. elegans* ($N = 280$, $E = 1973$); 4 food webs: Grassland ($N = 75$, $E = 113$), Ythan Estuary with parasites ($N = 134$, $E = 593$), El Verde Rainforest ($N = 156$, $E = 1439$), and St Marks Seagrass ($N = 48$, $E = 218$).

Figure 1 shows plots of $\gamma_1(i)$ *vs.* $SC_{odd}(i)$ for each of the 16 networks studied. The first 8 plots illustrate networks showing GE properties. In general, we have explored a total of 40 real-world networks and found that about half of them are GENs (results not shown). Among these GENs there are networks with known scale-free characteristics, such as ODLIS, Internet and USAir97 and others which have been shown to have exponential decay like Ythan1, El Verde or uniform distribution of node degrees like St Marks [29]. The last 8 plots show complex networks which are not GENs, and which include almost all biological and social networks studied, with the exception of food webs. Curiously, the only one food web that is not GEN is Grassland, which has a power law distribution of node degrees but the lowest average degree $\langle k \rangle$ among all food webs. However, high $\langle k \rangle$ is not a necessary requisite for complex networks to show GE properties. For example, the Internet is GEN despite it exhibits one of the lowest $\langle k \rangle$ in all the networks studied. In general, complex networks can naturally grow to form GENs as demonstrated by the fact that the BA model of preferential attachment [10] generates networks with GE properties [28]. Furthermore, a complex network that originally grew without GE properties can evolve into a GEN by two possible evolving

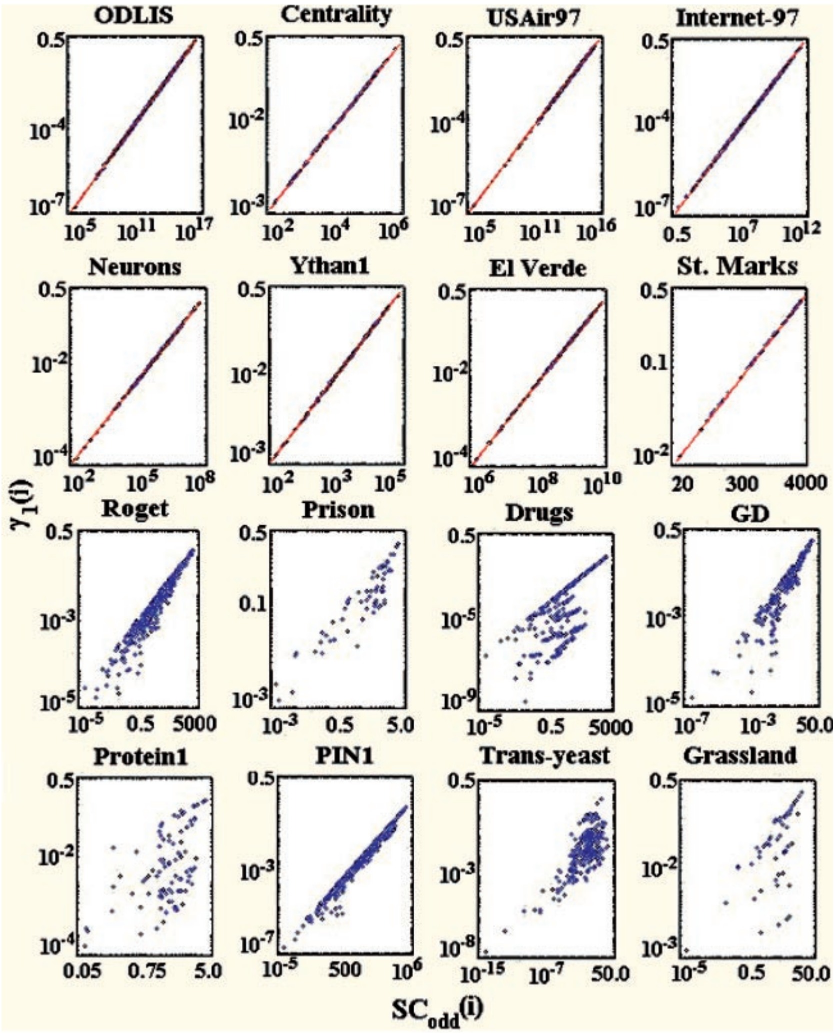

Fig. 1 – Plot of the eigenvector centrality of node i, $\gamma_1(i)$, *vs.* the odd-cyclic subgraph centrality, $SC_{odd}(i)$, in a log-log plot for some networks with GE properties (first eight) and networks that do not exhibit GE properties (last eight).

mechanisms. The first consists of the addition of new links between the nodes which are in separated clusters, a process which leads to the evolved network showing a higher $\langle k \rangle$ than the original one. The second process consists in rewiring the links which are connecting nodes inside clusters to joint nodes in different clusters. This rewiring process leaves $\langle k \rangle$ of the network unaltered. For instance, the network of injecting drug users (IDU) shows several clusters with many internal connections and few inter-cluster links (see fig. 2 top). A random rewiring process which keeps the same degree distribution of the original network produces a GEN with the same number of nodes and $\langle k \rangle$ as the original network (see fig. 2 top). This is a clear indication that when random networks are used as models of complex networks, even if we maintain the same degree distribution, there are important topological properties of the real-world systems, such as GE, which are not retained.

Real-world networks that expand well permit the selection of arbitrary sets of at most $N/2$

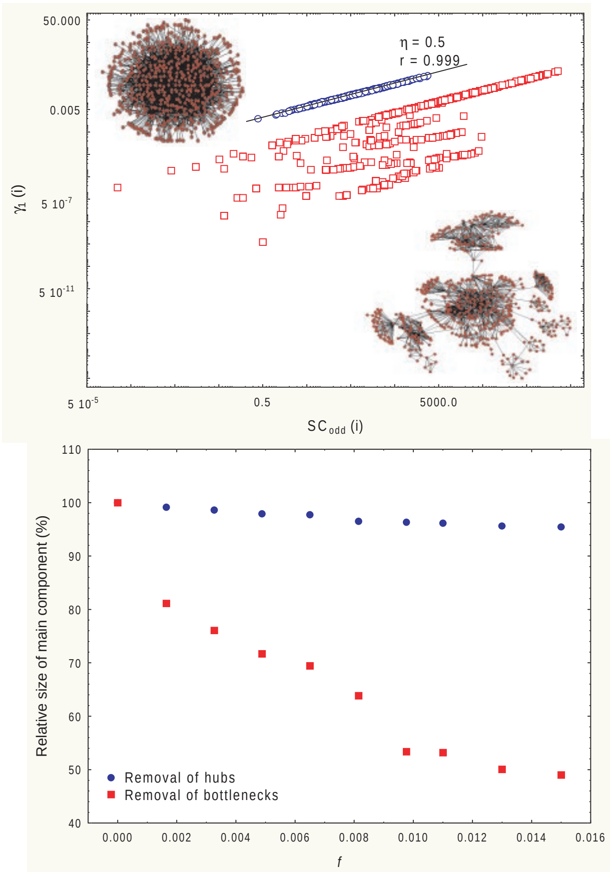

Fig. 2 – (Top) Transformation of a network that does not show GE properties into another which is GEN by a rewiring process. The social network of IDUs has several clusters and does not show scaling between $\gamma_1(i)$ and $SC_{odd}(i)$ as a consequence of the lack of GE properties (bottom correlation and network). The links of this network are rewired at random keeping the same degree distribution but obtaining a network with good scaling between $\gamma_1(i)$ and $SC_{odd}(i)$ and GE properties (top correlation and network). (Bottom) Vulnerability of GENs and not GENs under intentional attack, measured by the size of the main connected component *vs.* the fraction of removed nodes (f). For the random network with the same degree distribution than the IDUs network, which is GEN, we have removed the most connected nodes (hubs) while for the real-world network, which is not GEN, we removed the nodes connecting different clusters in the network (bottleneck nodes).

nodes in such a way that for every set there are relatively large numbers of edges with precisely one endpoint in this set. If a network contains two or more clusters with low number of inter-cluster links then there will be at least one set S with very few links that have one endpoint in S. In other words, the critical topological characteristic which is necessary for the existence of GENs is the absence of separate clusters in the network in which the number of intra-clusters links is significantly higher than the number of inter-clusters links. These elements are important in analyzing the results observed here for real-world networks. For instance, the lack of GE properties observed for proteins is due to the well-known fact that proteins are built up from domains in a modular architecture [30]. Thus, the secondary structure elements manifest a higher number of within-domain than between-domain interactions, a situation that is translated into a highly clustered structure for the network and the existence of bottlenecks in these networks. It has also been shown that most metabolic networks are organised into many small clusters of highly interconnected nodes [31]. In general, this kind of modularity can be present [4] or not in biochemical networks depending on whether or not the degree distribution is a trivial structure of such networks [32]. On the other hand, all social networks that have been analyzed have been shown to contain social communities, which is probably a natural characteristic of the process of social organization, where several highly cohesive groups are distinguished as clusters with small interactions between each other [33]. The presence of this type of topological structure in these networks is revealed in the lack of GE properties. In contrast, most food webs, with the exception of Grassland, are GENs and represent ecosystems that are not composed of separate communities with low interconnections among them.

An immediate consequence of the presence of GE properties in complex networks is related to the robustness of the network against node/link failures. It is possible that certain connections/nodes in a network are malfunctioning for whatever reason, resulting in random failures. Also the intentional attack that removes certain nodes/links in the network can produce significant damages on the network structure and functioning. For instance, those networks without GE properties are very vulnerable to the removal of certain nodes/links, *e.g.*, those connecting the separate clusters, the so-called bottleneck nodes/links. The removal of bottleneck nodes/links will separate non-GENs into different isolated components. A good example is provided by the IDUs network shown in fig. 2 (top) and its random analogue with the same degree distribution. As can be seen in fig. 2 (bottom) the random network, which shows GE properties, is very robust against the removal of the most connected nodes under an intentional attack. However, the real-world network, which is not GEN, is very vulnerable to the removal of bottleneck nodes.

To conclude, in this work we have introduced a theoretical methodology for exploring the global topological structure of complex networks using spectral graph theory. The existence of a scaling between principal eigenvector and subgraph centrality is an unequivocal evidence of the existence of GE properties in the network. The importance of GE properties for complex networks is evident as GENs are, as Sarnak has observed [18] "explicit superefficient communication networks" and permit the "construction of error-correcting codes with very efficient encoding and decoding algorithms". In contrast, the absence of GE in other complex networks, such as social networks, is an important characteristic related to the presence of different communities with low extra-community connections, which are essential for understanding the evolution of such systems as well as the propagation of "something", *i.e.*, information, infection, goods, etc., across such systems.

* * *

The author thanks J. A. Dunne, R. Milo, U. Alon, J. Moody, V. Batagelj and D. J. Watts for providing datasets. Useful comments made by A. Vázquez, S. Iztkovitz and J. A. Rodríguez-Velázquez are also acknowledged, as are clarifying discussions with S. N. Dorogovtsev and S. Valverde. This work was partially supported by the "Ramón y Cajal" program, Spain.

REFERENCES

[1] Strogatz S. H., *Nature*, **410** (2001) 268.
[2] Albert R. and Barabási A.-L., *Rev. Mod. Phys.*, **74** (2002) 47.
[3] Newman M. E. J., *SIAM Rev.*, **45** (2003) 167.

[4] Barabási A.-L. and Oltvai Z. N., *Nature Rev. Genet.*, **5** (2004) 101.
[5] Ferrer i Cancho R. and Solé R. V., *Proc. R. Soc. London, Ser. B*, **268** (2001) 2261.
[6] Liljeros F. *et al.*, *Nature*, **411** (2001) 907.
[7] Dodds P. S., Muhamad R. and Watts D. J., *Science*, **301** (2003) 827.
[8] Watts D. J. and Strogatz S. H., *Nature*, **393** (1998) 440.
[9] Watts D. J., *Small Worlds* (Princeton University Press, Princeton) 1999.
[10] Barabási A.-L. and Albert R., *Science*, **286** (1999) 509.
[11] Barabási A.-L., Albert R. and Jeong H., *Physica A*, **272** (1999) 173.
[12] Milo R. *et al.*, *Science*, **298** (2002) 824.
[13] Milo R. *et al.*, *Science*, **303** (2004) 1538.
[14] Song C., Havlin S. and Makse H. A., *Nature*, **433** (2005) 392.
[15] Farkas I. J., Derényi I., Barabási A.-L. and Viscek T., *Phys. Rev. E*, **64** (2001) 026704.
[16] Farkas I. J. *et al.*, *Physica A*, **314** (2002) 25.
[17] Bollobás B., *Extremal Graph Theory* (Academic Press, New York) 1978.
[18] Sarnak P., *Not. Am. Math. Soc.*, **5** (2004) 762.
[19] Mohar B., *J. Comb. Theor. B*, **47** (1989) 274.
[20] Tanner R. M., *SIAM J. Alg. Disc. Math.*, **5** (1984) 287.
[21] Alon N., *Combinatorica*, **6** (1986) 83.
[22] Friedman J., *Combinatorica*, **11** (1989) 331.
[23] Reingold O., Vadhan S. and Wigderson A., *Ann. Math.*, **155** (2002) 157.
[24] Estrada E. and Rodríguez-Velázquez J. A., *Phys. Rev. E*, **71** (2005) 056103.
[25] Biggs N., *Algebraic Graph Theory* (Cambridge University Press, Cambridge) 1993.
[26] Estrada E. and Rodríguez-Velázquez J. A., *Phys. Rev. E*, **72** (2005) 055510.
[27] Cvetković D., Rowlinson P. and Simić S., *Eigenspaces of Graphs* (Cambridge University Press, Cambridge) 1997.
[28] Gkantsidis C., Mihail M. and Saberi A., *Proceedings of the 2003 ACM SIGMETRICS International Conference on Measurement Modeling of Computer Systems, San Diego, June 11-14 2003* (ACM Press, New York) 2003, p. 148.
[29] Dunne J. A., Williams R. J. and Martinez N. D., *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002) 12917.
[30] Ponting C. and Russell R., *Annu. Rev. Biophys. Biomed. Struct.*, **31** (2002) 45.
[31] Ravasz E. *et al.*, *Science*, **297** (2002) 1551.
[32] Holme P., Huss M. and Jeong H., *Bioinformatics*, **19** (2003) 532.
[33] Girvan M. and Newman M. E. J., *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002) 82711.