NETWORKS AND HETEROGENEOUS MEDIA Website: http://aimSciences.org

Volume **3**, Number **2**, June **2008** pp. **395–411**

SPECTRAL PLOT PROPERTIES: TOWARDS A QUALITATIVE CLASSIFICATION OF NETWORKS

Anirban Banerjee and Jürgen Jost

Max Planck Institute for Mathematics in the Sciences
Inselstr. 22, 04103 Leipzig, Germany

Abstract. We introduce a tentative classification scheme for empirical networks based on global qualitative properties detected through the spectrum of the Laplacian of the graph underlying the network. Our method identifies several distinct types of networks across different domains of applications, indicates hidden regularity properties and provides evidence for processes like node duplication behind the evolution or construction of a given class of networks.

1. **Introduction.** Real world networks tend to be irregular and complicated and, because of their size, also difficult to visualize. It is therefore important to develop methods to identify qualitative properties that can characterize specific classes of networks and that can be easily visualized. Recent research has identified certain universal properties shared by large classes, if not most, of empirical graphs, like randomness [10], small-world property [35], scalefreeness [31, 7]. Here, we want to introduce a possible path towards a qualitative classification of different classes of networks. This classification on one hand is robust towards fluctuations and perturbations within a given class, and on the other hand, can readily distinguish different types. It can also easily and directly be visually inspected.

Even though empirical networks typically have directed and weighted edges, we here consider only the underlying undirected and unweighted graph. The methods utilized, however, easily extend to the directed and weighted case, but it turns out that already that reduced graph carries a lot of structural information that is quite informative about the network. This, as well as space constraints, is our rationale for that simplification.

Our essential tool is the spectrum of an (undirected and unweighted) graph Γ (representing our network) with N vertices or nodes. For functions v from the vertices of Γ to $\mathbb{R}$, we define the Laplacian as

$$\Delta v(i) := \frac{1}{n_i} \sum_{j, j \sim i} v(j) - v(i) \tag{1}$$

where n_i is the degree of the vertex i, that is, the number of its neighbors (the vertices with which it is connected by an edge).

In the graph theoretical literature, one usually considers the algebraic graph Laplacian which employs a different normalization and therefore leads to a different spectrum. That Laplacian has also been applied to the investigation of complex

2000 *Mathematics Subject Classification.* Primary: 05C75, 47A75; Secondary: 90C35, 68R10.
Key words and phrases. Network classification, graph Laplacian, graph spectrum spectral plots, graph evolution.

networks, but from a different perspective. In [9, 29], for instance, asymptotic results about that operator's spectrum for the limit of infinite network size are derived. – We found the Laplacian (1) better for our purposes of classifying finite networks as it gives clearer pictures of network classes. Also, this, and not the algebraic Laplacian, is the operator generating random walks on graphs, and moreover, for diffusion processes on a graph, it satisfies a conservation law. – For mathematical background, we refer to [8, 19, 18, 4, 5].

The eigenvalue equation for Δ is

$$\Delta u + \lambda u = 0. \tag{2}$$

A nonzero solution u is called an eigenfunction for the eigenvalue λ. Δ then has N eigenvalues (possibly occurring with multiplicity). The eigenvalues of Δ are real (because Δ is symmetric for the product $(u, v) = \sum_i n_i u(i)v(i)$) and nonnegative (by our convention, because Δ is a nonpositive operator). The smallest eigenvalue is $\lambda = 0$. Its multiplicity equals the number of components of Γ. As we shall only evaluate connected graphs, this eigenvalue is simple. We then order the eigenvalues as

$$\lambda_0 = 0 \leq \lambda_2 \leq ... \leq \lambda_K \text{ with } K = N - 1.$$

For the largest eigenvalue, we have

$$\lambda_K \leq 2, \tag{3}$$

with equality iff the graph is bipartite, that is, when the graph consists of two disjoint classes with the property that there are no edges between vertices in the same class. Thus, a single eigenvalue determines the global property of bipartiteness. Also, the graph is bipartite iff the spectrum is symmetric about 1, that is λ is an eigenvalue iff $2 - \lambda$ is.

The eigenvalue $\lambda = 1$ plays a special role as it gives some indication of vertex or motif duplications underlying the evolution of the graph. That is, if we have some graph Γ_0 with a node i_0 and create a new graph Γ by adjoining a new node j_0 that is given the same neighbors as i_0, we obtain an eigenfunction u with eigenvalue 1 by putting $u(i_0) = 1, u(j_0) = -1, u(k) = 0$ for all other nodes. In particular, through repeated duplication of nodes, we can create graphs with the eigenvalue 1 of very high multiplicity. In fact, the complete bipartite graph $K_{m,n}$ (the graph consisting of one class with m and another class with n nodes such that every node in the first class is connected with every node in the second one, and with no connections inside classes) has the eigenvalue 1 with multiplicity $m + n - 2$ (and in addition the eigenvalues 0 and 2 with multiplicity 1 each) because it can be constructed by iterated node duplication from the connected graph with 2 nodes. Here, we see the phenomenon of isospectral graphs, different graphs with the same spectrum, as all $K_{m,n}$ with the same sum $m + n$ share the same spectrum. Nevertheless, the spectrum is a very powerful tool for characterizing qualitative properties of graphs; for example, each graph having this spectrum then has to be of type $K_{m,n}$. Likewise, the complete graph K_N of N vertices is characterized by having the eigenvalue $\frac{N}{N-1}$ with multiplicity $N - 1$. In another direction, the smallest eigenvalue λ_1 tells us how difficult it is to disconnect the graph (splitting the graph into two large components by cutting few edges), see [8]. In the converse direction, there also exist good algorithms to reconstruct a graph from its spectrum, see [14] (of course, up to isospectrality). In essence, the spectrum of a graph yields a set of invariants that on one hand captures what is specific about that graph and on the other hand simultaneously encodes all its important properties.

We then plot the spectrum not directly, that is, as the collection of the eigenvalues λ_j, but we convolve them with a smoothing kernel, here a Gaussian, that is, we plot

$$f(x) = \sum_{\lambda_j} \frac{1}{0.025\sqrt{2\pi}} \exp(-\frac{|x-\lambda_j|^2}{0.00125}) \tag{4}$$

The particular value .025 for the width of the Gaussian has been chosen as a compromise between high resolution with too many small scale fluctuations and low resolution with a blurring of qualitative patterns.

2. **Network classes.** In order to get some orientation, we start with spectral plots of artificial, that is, simulated networks. Our first examples come from two classes of regular networks. The first one consists of regular 2d square grids, with 10,000 nodes. As we see in Fig.1, when we make the grid narrower and longer, the spectrum shows characteristic side peaks. The spectral plot is symmetric about 1 as all these graphs are bipartite. When we add one of the two possible diagonals (always the same) in each square, we destroy bipartiteness and get a systematic shift in the spectral plot (Fig.2), again with the side peaks when the grid gets narrower. The other regular graphs originate from a circular arrangement of 1,000 nodes where we connect each node with the the 2, 4, 6, 10, 20, or 50 closest nodes on the circle (Fig.3). When thus progressing to higher degrees, we see the eventual merging at 1 of the two peaks that start out at 0 and 2 for small degrees.

We next turn to stochastically constructed graphs (Fig.4), all with 1000 nodes. We first have an Erdös-Renyi random graph;[1] here, a single realization and the average of 100 such graphs will not exhibit substantially different spectral plots, that is, each realization already shows the typical spectral properties. This is an indication of the robustness of our scheme against random fluctuations – which, of course, are at the heart of the idea of a random graph. Next, we have scale free graph constructed by the algorithm of Barabási-Albert; here, averaging over 100 realizations smoothes the spetral plot out a bit. This is even more evident for a small-world graph a la Strogatz-Watts. We construct them by rewiring a regular graph, either of the square grid or of the circle type, both with rewiring probability 0.3 (Fig.5). The spectral plot becomes characteristically different from the regular one. (A systematic investigation of the spectrum of a small-word graph as the superposition of a regular ring and a random graph has been carried by Monasson [25].)

We now turn to empirical networks and compare their spectral plots both with each other and with the model types presented above. We shall have to keep in mind below, however, that some of the empirical networks are quite small, on the order of 100 nodes only, and so, obviously random fluctuations may have stronger effects that suggest some caution concerning the robustness of our classification. – The first type comprises several classes of biological networks at the molecular level, including metabolic, transcription, signal transduction, and protein-protein interaction networks, as well as word adjacency and internet topology graphs (Fig.[6 - 10]). The characteristic features are the very high peak at or near 1, the shallow rest

[1]Because we have normalized our Laplacian, we do not get Wigner's semicircle law for the spectrum of a random graph here. See also [2] for some investigations of the spectrum of the unnormalized Laplacian.

with two secondary peaks, and the high degree of symmetry about 1. As we recall from the mathematical discussion above, these graphs then come close in spectral terms to a complete bipartite graph which, as we discussed, arises through repeated node duplication. Simulations that we present elsewhere [3] indicate that the secondary peaks arise from random deletion of edges after the node duplications. Node duplication with subsequent random edge deletion has been proposed in different application fields as a mechanism for network growth that can reproduce qualitative properties of empirical networks, e.g. for the internet [21], for protein-interaction networks [32, 33, 34] or citation networks [22], although the precise rules can differ between those investigations, for instance, whether the duplicated node and its copy are connected or not.

For each of these empirical classes, one can then try to find an explanation of their evolution or construction through such processes, like gene duplication in the biological ones. Our second class contains weblog hyperlink graphs (in US politics), conformation spaces of polypeptides, food webs, and, with less confidence, email interchanges (Fig.[11-13]). Neuronal connectivity graphs of C.elegans constitute a borderline case (Fig.14). This class is characterized by a concentration near 1, but not as sharply peaked as in the first case, and, except for the neuronal network, again symmetry about 1. This class is different from all the model types, but shows a little similarity with the scalefree type. The third class contains power grids, coauthorships between scientists, copurchasing of books, and US football games (Fig.[15-17]). They all resemble the square grid with diagonal class, moving from the less narrow to the very narrow ones. Finally, the electronic circuit graph spectra (Fig.18) resemble those of a narrow square grid without diagonals.

3. **Conclusion.** We have presented a scheme for the rough classification of empirical networks in terms of their qualitative spectral properties. Since we can also understand from mathematical theory how some of those characteristic spectral properties are caused by topological properties of the underlying graph or can emerge from processes like node duplication, random rewiring, random edge deletion etc., this scheme also offers the potential for systematic insights into the evolution or the emergence of global properties of specific classes of empirical networks. As usual with mathematical structures, structural similarities can be shared across empirical domains.

Of course, this represents at best the first step towards a systematic theory of complex networks. Perhaps the state at this moment is a little similar to the one of cellular automata about 25 years ago when also classifications were proposed in terms of visually representable global features. Not all of what was proposed then could be consolidated by subsequent research, but it nevertheless opened up a fruitful perspective.

REFERENCES

[1] L. A. Adamic and N. Glance, *The political blogosphere and the 2004 US election: divided they blog*, LinkKDD '05: Proceedings of the 3rd international workshop on Link discovery, (2005), 36–43.

[2] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Rev. Mod. Phys., **74** (2002), 47–97.

[3] A. Banerjee and J. Jost, *Laplacian spectrum and protein-protein interaction networks*, preprint, arXiv:0705.3373v1.

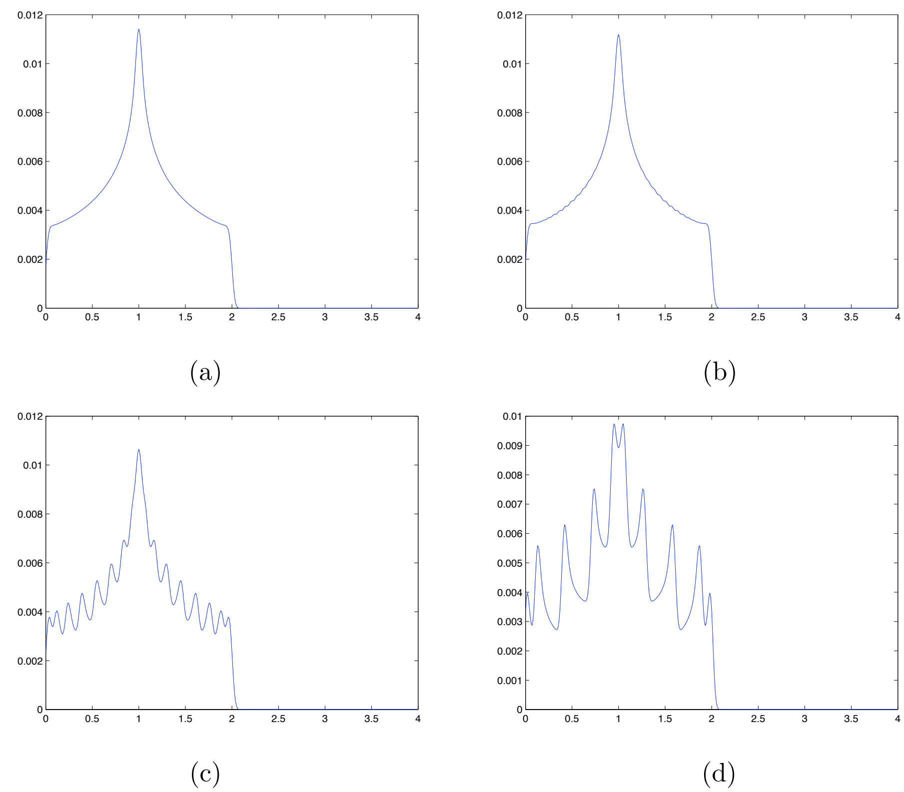


FIGURE 1. 2-dimensional grid with dimension m by n. (a) m= 100, n= 100. (b) m= 25, n= 400. (c) m= 10, n= 1000. (d) m= 5, n= 2000.

[4] A. Banerjee and J. Jost, *On the spectrum of the normalized graph Laplacian*, Lin. Alg. Appl., in press.

[5] A. Banerjee and J. Jost, *Graph spectra as a systematic tool in computational biology*, Discr. Appl. Math, to appear.

[6] A. Banerjee and J. Jost, *Spectral plots and the representation and interpretation of biological data*, Theory Biosc., **126** (2007), 15–21.

[7] A.-L. Barabási and R. A. Albert, *Emergence of scaling in random networks*, Science, **286** (1999), 509–512.

[8] F. Chung, "Spectral Graph Theory," AMS, 1997.

[9] S. N. Dorogovtsev et al., *Spectra of complex networks* Phys. Rev. E, **68** (2003), 046109.

[10] P. Erdös and A. Rényi, *On random graphs*, Publ. Math. Debrecen, **6** (1959), 290–297.

[11] M. Faloutsos et al., *On power-law relationships of the Internet topology*, SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication, (1999), 251–262.

[12] M. Girvan and M. E. J. Newman, *Community structure in social and biological networks* Proc. Natl. Acad. Sci. USA, **99** (2002), 7821–7826.

[13] R. Guimerá et al., *Self-similar community structure in a network of human interactions*, Physical Review E, **68** (2003), 065103(R).

[14] M. Ipsen and A. S. Mikhailov, *Evolutionary reconstruction of networks*, Phys. Rev. E, **66** (2002).

[15] H. Jeong et al., *The large-scale organization of metabolic networks*, Nature, **407** (2000), 651–654.

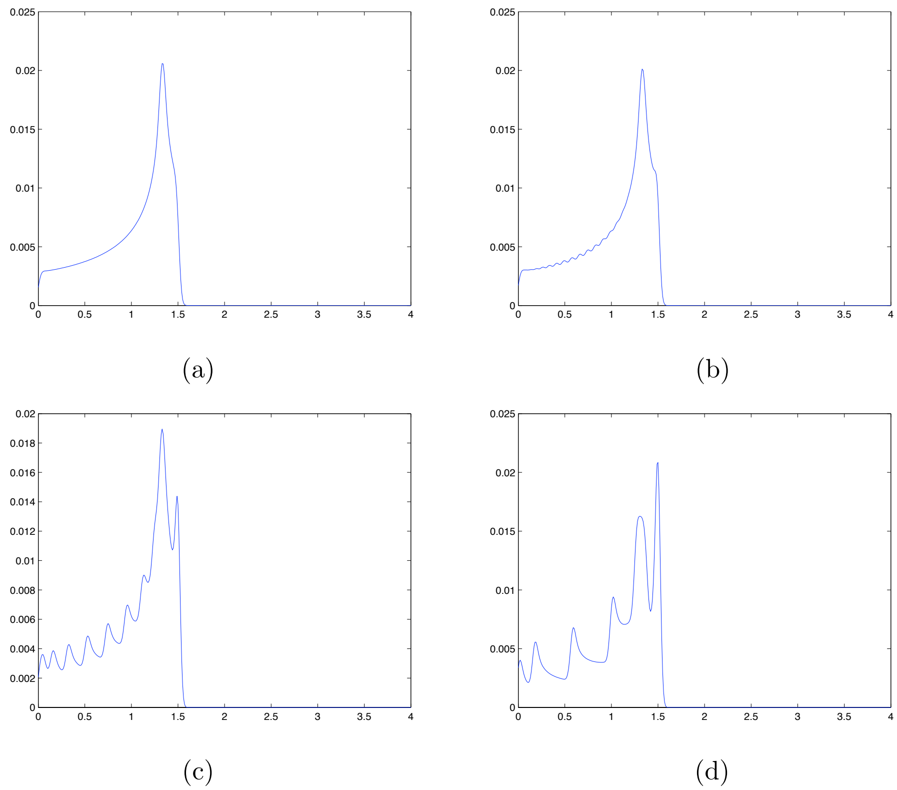


FIGURE 2. 2-dimensional grid with one diagonal in each square with dimension m by n. (a) m= 100, n= 100. (b) m= 25, n= 400. (c) m= 10, n= 1000. (d) m= 5, n= 2000.

[16] H. Jeong et al., *Lethality and centrality in protein networks*, Nature, **411** (2001), 41–42.

[17] J. Jost, "Mathematical Methods in Biology and Neurobiology," monograph, to appear.

[18] J. Jost, *Dynamical networks*, in: "J. F. Feng, J. Jost, M. P.Qian (eds.), Networks: from Biology to Theory," Springer, (2007), 35–62.

[19] J. Jost and M. P. Joy, *Spectral properties and synchronization in coupled map lattices*, Phys. Rev. E, **65** (2001), 16201–16209.

[20] J. Jost and M. P. Joy, *Evolving networks with distance preferences*, Phys. Rev. E, **66** (2002), 36126–36132.

[21] J. Kleinberg et al., *The web as a graph: measurements, models, and methods*, LNCS, **1627** (1999), 1–17.

[22] P. Krapivsky and S. Redner, *Network growth by copying*, Phys. Rev. E, **71** (2005), 036118.

[23] R. Milo et al., *Network motifs: Simple building blocks of complex networks*, Science, **298** (2002), 824–827.

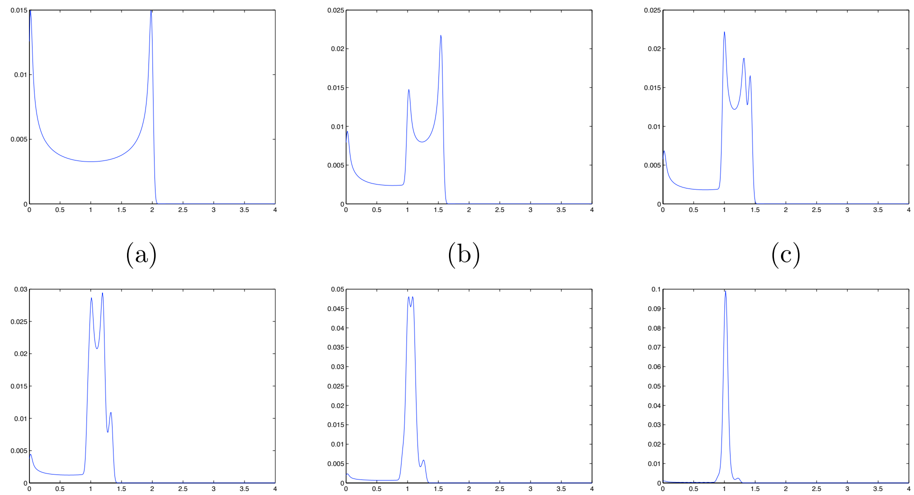
[24] R. Milo et al., *Superfamilies of evolved and designed networks*, Science, **303** (2004), 1538–1542.

[25] R. Monasson, *Diffusion, localization and dispersion relations on "small-world" lattices*, Europ. Phys. J. B, **12** (1999), 555–567.

[26] M. E. J. Newman, *The structure of scientific collaboration networks*, Proc. Natl. Acad. Sci. USA, **98** (2001), 404–409.

[27] M. E. J. Newman, *Finding community structure in networks using the eigenvectors of matrices*, Phys. Rev. E, **74** (2006), 036104.

[28] F. Rao and A. Caflisch, *The protein folding network*, J. Mol. Biol., **342** (2004) 299–306.



(a) (b) (c)

(d) (e) (f)

Figure 3. 1-dimensional regular ring lattice of size 1000 with degree of each vertex (a) 2 (b) 4 (c) 6 (d) 10 (e) 20 (f) 50.

[29] A. N. Samukhin, S. N. Dorogovtsev and J. Mendes, *Laplacian spectra of complex networks and random walks on them: are scale-free architectures really important? and random walks on them: are scale-free architectures really important?*, preprint, arXiv:0706.1176.

[30] S. S. Shen-Orr et al., *Network motifs in the transcriptional regulation network of Escherichia Coli*, Nature Genetics, **31** (2002), 64–68.

[31] H. Simon, *On a class of skew distribution functions*, Biometrika, **42** (1955), 425–440.

[32] R. Solé et al., *A model of large scale proteome evolution*, Adv.Compl.Syst., **5** (2002), 43–54.

[33] A. Vazquez et al., *Modelling of protein interaction networks*, ComPlexUs, **1** (2003), 38–44.

[34] A. Wagner, *How the global structure of protein interaction networks evolves*, Proc. Roy. Soc. B, **270** (2003), 457–466.

[35] D. J. Watts and S. H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature, **393** (1998), 440–442.

[36] J. G. White et al., *The structure of the nervous-system of the nematode Caenorhabditis-Elegans*, Phil. Trans. Royal Soc. of London Series B-Bio. Sc., **314** (1986), 1–340.

Received August 2007; revised February 2008.

E-mail address: banerjee@mis.mpg.de
E-mail address: jost@mis.mpg.de

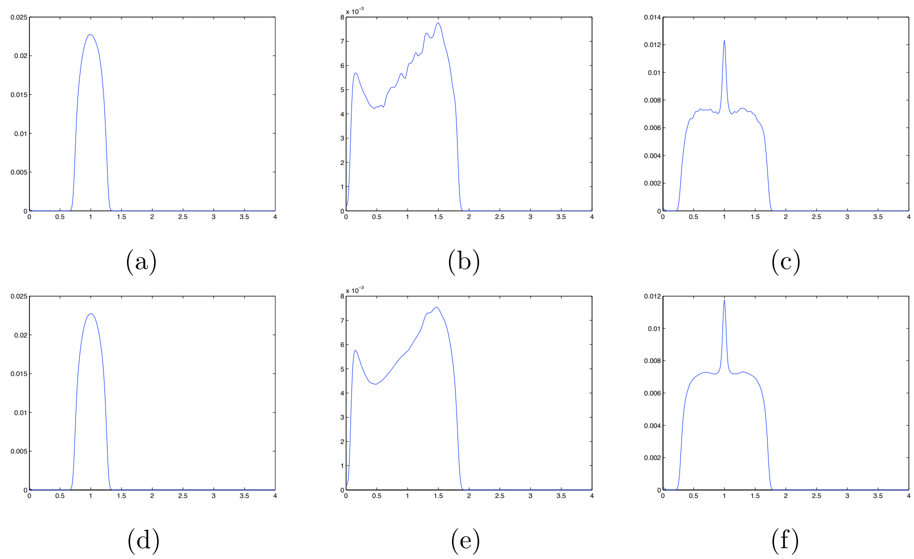


Figure 4. Specral plots of generic networks. (a) Random network by the Erdös-Rényi model [10] with $p = 0.05$. (b) Small-world network by the Watts-Strogatz model [35] (rewiring a regular ring lattice of average degree 4 with rewiring probability 0.3). (d) Scale-free network by the Albert-Barabási model [7] ($m_0 = 5$ and $m = 3$). *Figures (a-c) obtained from a single realization, (d-f) represent the averages of* 100 *realizations. Size of all networks is* 1000*.*

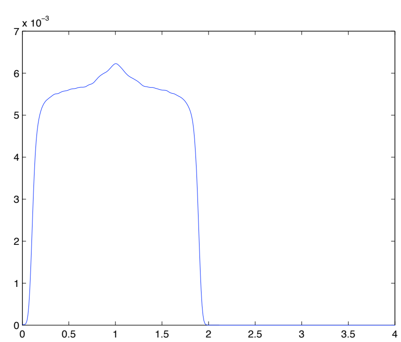
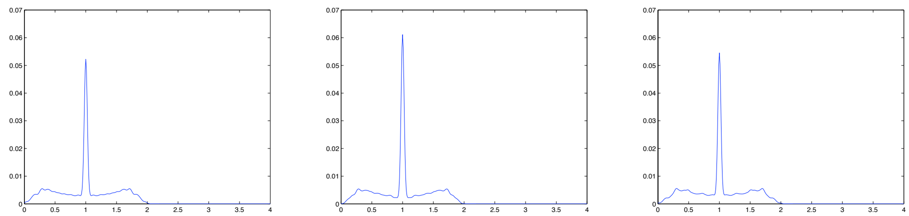
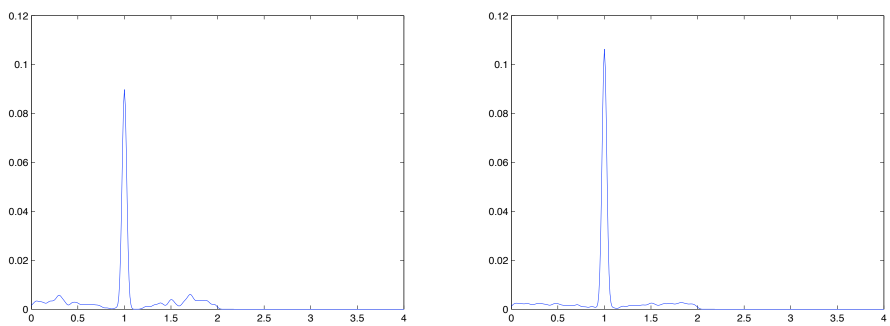


Figure 5. Specral plot of a small-world network created by rewiring a 2-dimensional grid of dimension 100 by 100 with rewiring probability 0.3. *Single realization.*



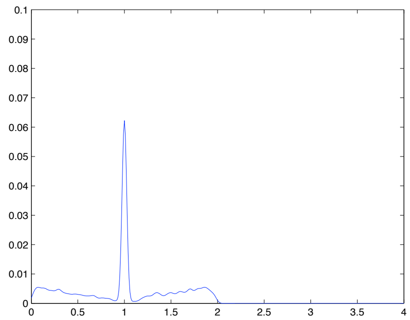
(a) (b) (c)

FIGURE 6. Metabolic networks. Nodes represent substrates, enzymes and intermediate complexes. Data used in [15]. Data Source: http://www.nd.edu/~networks/resources.htm. [Download date: 22nd Nov. 2004] (a) *Archaeoglobus fulgidus*. Network size 1268. (b) *Escherichia coli*. Network size 2268. (c) *Saccharomyces cerevisiae*. Network size 1511.

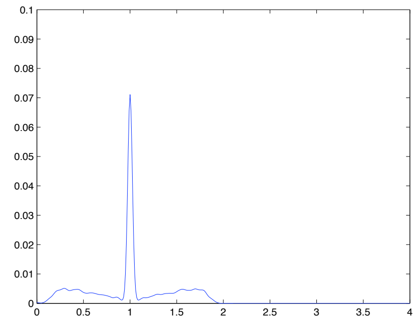


(a) (b)

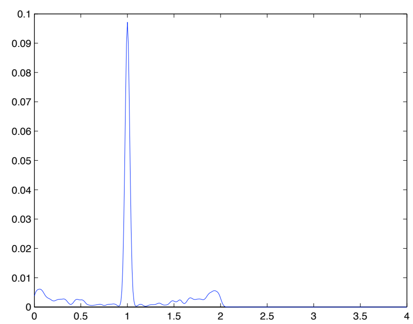
FIGURE 7. Transcription networks. Data source: Data published by Uri Alon (http://www.weizmann.ac.il/mcb/UriAlon/). [Download date: 13th Oct. 2004]. Data used in [23, 30]. (a) *Escherichia coli*. Network size 328. (b) *Saccharomyces cerevisiae*. Network size 662.



(a)

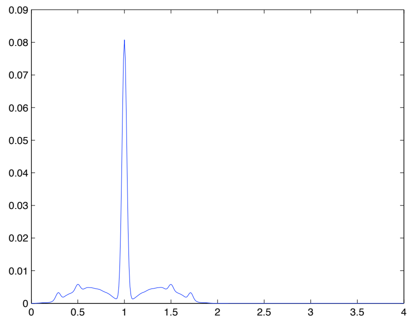


(b)

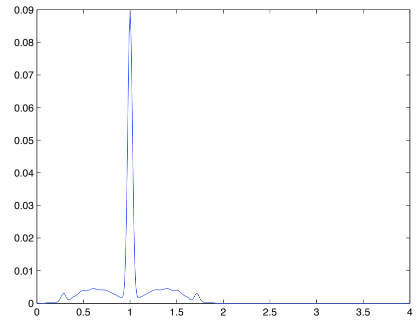


(c)

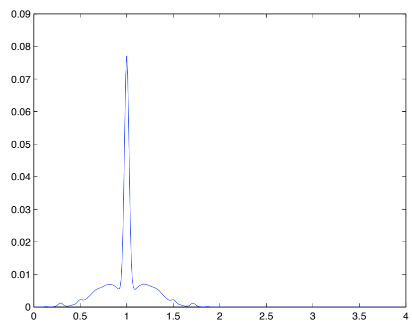
Figure 8. Protein-protein interaction networks. (a) *Saccharomyces cerevisiae*. Network size 1458. Data downloaded from `http://www.nd.edu/~networks/` and data used in [16] [download date: 17th September, 2004]. (b) *Helicobacter pylori.* Network size 710. (c) *Caenorhabditis elegans.*Network size 314. (b,c) Data collected from `http://www.cosinproject.org` [download date: 25th September, 2005].



(a)

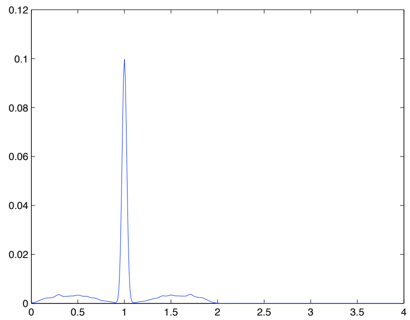


(b)

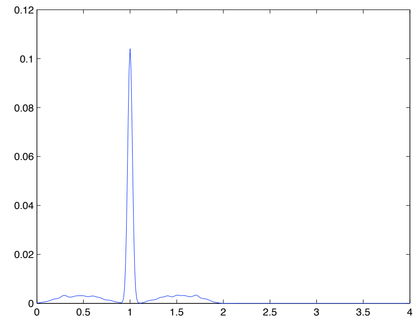


(c)

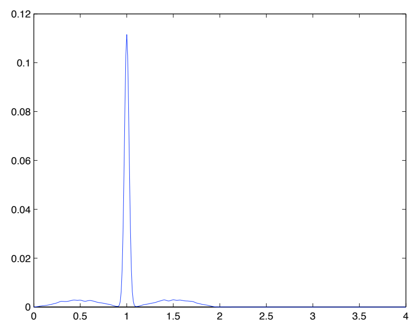
Figure 9. Word-adjacency networks of a text in (a) French. Network size 8308. (b) Japanese. Network size 2698. (c) English. SNetwork size 7377. Data downloaded from http://www.weizmann.ac.il/mcb/UriAlon/ [Download date 3rd Feb. 2005]. Data used in [24].



(a)

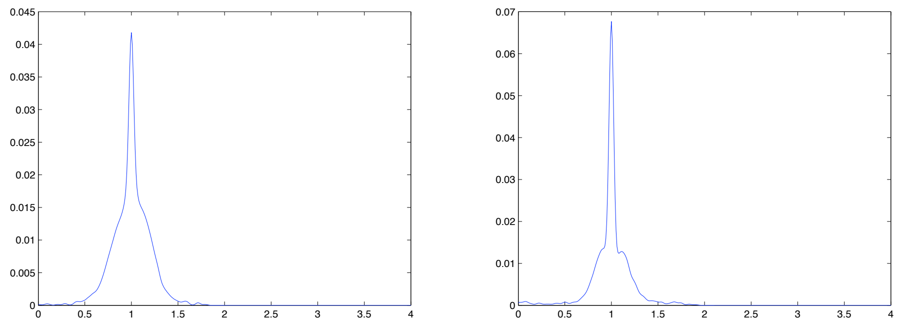


(b)



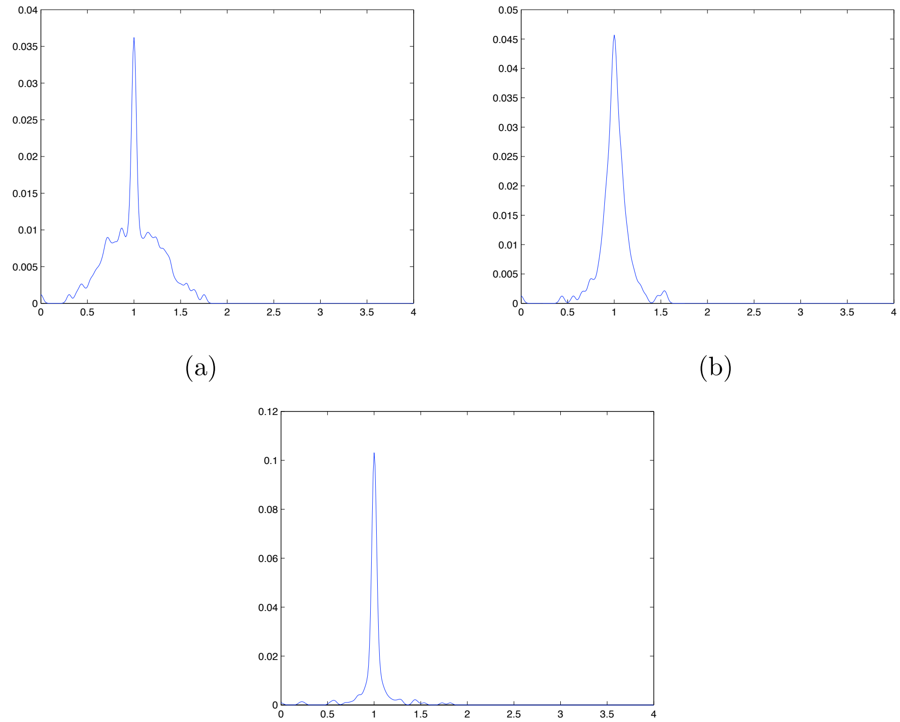
(c)

FIGURE 10. Autonomous Systems topology of the Internet. Every vertex represents an autonomous system, and two vertices are connected if there is at least one physical link between the two corresponding Autonomous Systems. (a) AS graph of 1997/11/08. Network size 3015. (b) AS graph of 1999/07/02. Network size 5357. (c) AS graph of 2001/03/16. Network size 10515. Data collected from http://www.cosinproject.org and data used in [11] [download date: 23rd September, 2005]. Main source: BGP routing data collected by University of Oregon Route Views Project, then processed and made available in various formats at the Global ISP interconnectivity by AS number page of NLANR (National Laboratory of Applied Network Research).



(a) (b)

FIGURE 11. (a) The network of hyperlinks between weblogs on US politics, recorded in 2005 by Adamic and Glance [1]. Network size 1222. Data downloaded from http://www-personal.umich.edu/~mejn/netdata/ [Download date: 23rd April 2007]. (b) Network of conformation space (Only conformations that are visited at least 20 times during the simulation are considered in the building of the network.) of a 20 residue antiparallel beta-sheet peptide sampled by molecular dynamics simulations [28]. Snapshots saved along the trajectory are grouped according to secondary structure into nodes of the network and the transitions between them are links. Network size 1199. Downloaded from Caflisch group, University of Zurich, http://www.biochem-caflisch.unizh.ch/ [Download date: 18th Dec. 2006].



(a) (b)

(c)

Figure 12. Food-web. (a) From “Ythan estuary”. Data downloaded from http://www.cosinproject.org. [Download Date 21st December, 2006]. Network size 135. (b) From “Florida bay in wet season”. Data downloaded from http://vlado.fmf.uni-lj.si/pub/networks/data/ (main data resource: Chesapeake Biological Laboratory. Web link: http://www.cbl.umces.edu/). [Download Date 21st December, 2006]. Network size 128. (c) From “Little rock lake”. Data downloaded from http://www.cosinproject.org. [Download Date 21st December, 2006]. Size of the network is 183.

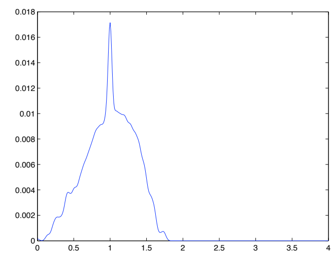
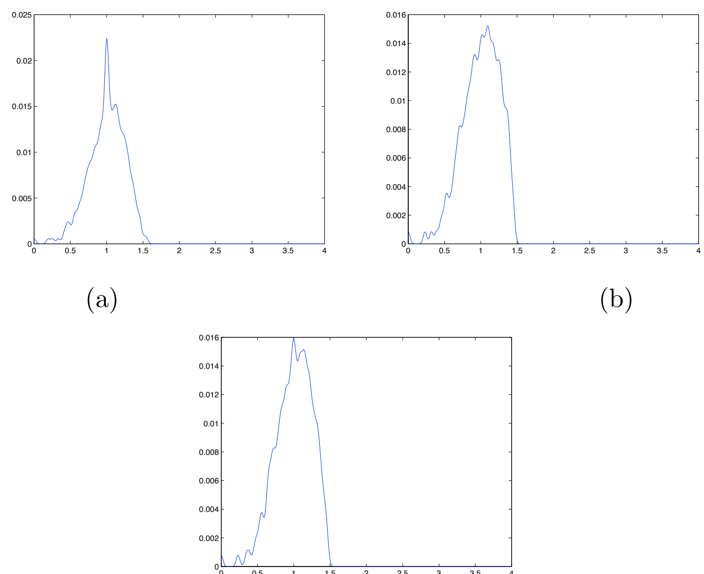


FIGURE 13. E-mail interchanges between members of the Univeristy Rovira i Virgili (Tarragona) [13]. Network size 1133. Data downloaded from http://www.etse.urv.es/aarenas/data/welcome.htm [Download date: 21st March, 2007].



(a) (b)

(c)

FIGURE 14. Neuronal connectivity. (a) *Caenorhabditis elegans*. Network size 297. Data used in [35, 36]. Data Source: http://cdg.columbia.edu/cdg/datasets [Download date: 18th Dec. 2006]. (b) *Caenorhabditis elegans* (animal JSH, L4 male) in the nerve ring and RVG regions. Network size 190. Data source: Data assembled by J. G. White, E. Southgate, J. N. Thomson, S. Brenner [36] and revisited by R. M. Durbin (Ref. http://elegans.swmed.edu/parts/). [Download date: 27th Sep. 2005]. (c) *Caenorhabditis elegans* (animal N2U, adult hermaphrodite) in the nerve ring and RVG regions. Network size 199. Data source: Data assembled by J. G. White, E. Southgate, J. N. Thomson, S. Brenner [36] and revisited by R. M. Durbin (Ref. http://elegans.swmed.edu/parts/). [Download date: 27th Sep. 2005].

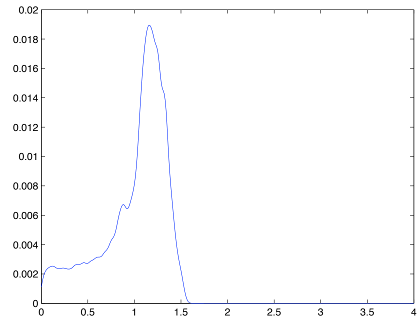
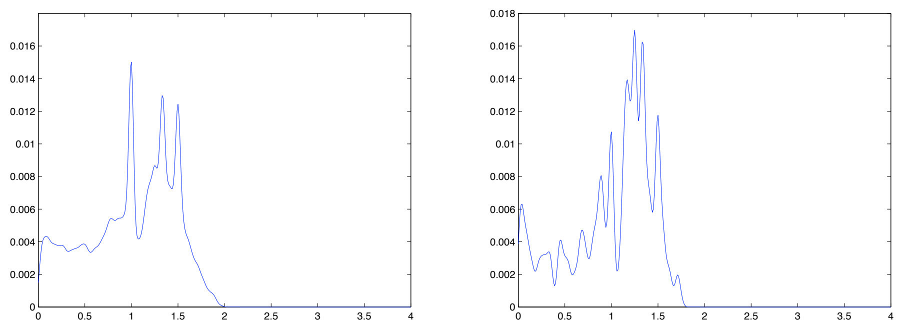
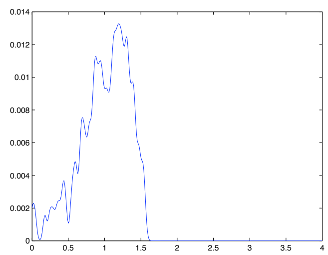


FIGURE 15. Topology of the Western States Power Grid of the United States [35]. Network size 4941. Data downloaded from http://cdg.columbia.edu/cdg/datasets [Download date: 1st March, 2007.].

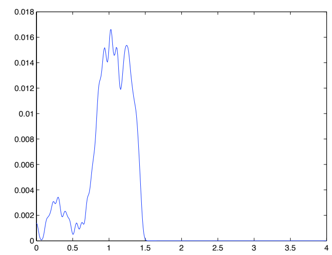


(a) (b)

FIGURE 16. (a) Coauthorships between scientists posting preprints on the High-Energy Theory E-Print Archive, http://arxiv.org/archive/hep-th between 1st Jan, 1995 and 31st December 1999 [26]. Network size 5835. (b) Coauthorships of scientists working on network theory and experiment [27]. Network size 379. Data downloaded from http://www-personal.umich.edu/~mejn/netdata/ [Download date: 23rd April, 2007].

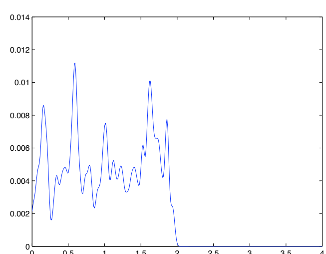


(a)

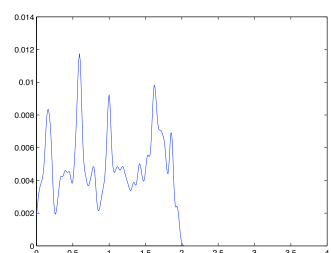


(b)

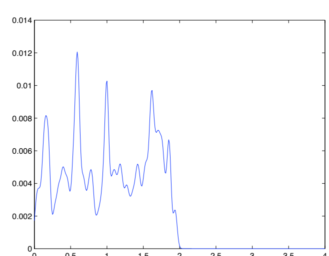
FIGURE 17. (a) Books about recent US politics sold by the online bookseller Amazon.com. Edges between books represent frequent copurchasing of books by the same buyers. Network compiled by V. Krebs (unpublished). Network size 105. Data downloaded from http://www-personal.umich.edu/~mejn/netdata/[original source http://www.orgnet.com/. Download date: 23rd April, 2007]. (b) American football games between Division IA colleges during regular season Fall 2000, as compiled by M. Girvan and M. Newman [12]. Network size 115. Data downloaded from http://www-personal.umich.edu/~mejn/netdata/ [Download date: 23rd April, 2007].



(a)



(b)



(c)

FIGURE 18. Electronic circuits. (a) With size = 122. (b) With size = 252. (c) With size = 512. Data downloaded from http://www.weizmann.ac.il/mcb/UriAlon/ [Download date: 15th March, 2005]. Data used in [23].