Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Stochasticity in complex networks: a random matrix analysis

by

Jayendra Bandyopadhyay, and Sarika Jalan

Stochasticity in Complex Networks: A Random Matrix Analysis

Jayendra N. Bandyopadhyay[1] and Sarika Jalan[2]

[1] Max-Planck Institute for the Physics of Complex Systems, Nöthnitzerstr. 38, D-01187 Dresden, Germany[*]
[2] Max-Planck Institute for Mathematics in the Sciences, Inselstrasse 22, D-04103 Leipzig, Germany[†]

Following random matrix theory, we study nearest neighbor spacing distribution (NNSD) of the eigenvalues of the adjacency matrix of various model networks, namely scale-free, small-world and random networks. Our analysis shows that, though spectral densities of these model networks are different, their eigenvalue fluctuations are same and follow Gaussian orthogonal ensemble (GOE) statistics. Secondly we show the analogy between the onset of small-world behavior (quantified by small diameter and large clustering coefficients) and the transition from Poisson to GOE statistics (quantified by Brody parameter). We also present our analysis for a protein-protein interaction network in budding yeast.

Regular and random networks are the two limiting cases of network topology. For regular networks, each node is connected in a fixed pattern to the same number of neighboring nodes; on the other hand, for random networks, each node is randomly joined with any other node. Real-world networks show the properties which are intermediate of the regular and the random one [1–3]. To model randomness and regularity, Watts and Strogatz proposed an algorithm to generate popularly known as 'small-world network', which has the properties of small diameter and high clustering [1]. In addition to above mentioned properties, Barabási and Albert show that many real-world networks are scale-free, which means that the degree distribution $p(k)$, fraction of nodes that have k number of connections with other nodes, decays as $p(k) \propto k^{-\gamma}$, where γ depends on the topology of the networks. The scale-free nature of networks suggests that some nodes are much more connected than the rest [4].

The structure of networks is described by its associated adjacency matrix A. It is defined in following way: $A_{ij} = 1$ if i and j nodes are connected and zero otherwise. We consider only undirected networks. In this case, the adjacency matrix is symmetric and consequently has real eigenvalues. These eigenvalues give information about some basic topological properties of the underlying network. There exists extensive literature demonstrating that the properties of graphs (or networks) and the associated adjacency matrices are well characterized by spectral methods, that provide global measure of the network properties [5–7]. Spectral density of random matrices, whose elements are Gaussian distributed random numbers, follows Wigner's semicircular law [8]. Interestingly, the spectral density of the adjacency matrix of random graphs, whose elements are randomly 0 or 1, also follows the semicircular law [9].

With the increasing availability of large maps of real-world networks, in the past few years lot of work has been done on the spectral densities of adjacency matrix of these real-world networks and models network having real-world properties [9–11]. These analysis show that the spectral densities of real-world networks are not semi-circular, instead they have some specific features depending on the minute details of the corresponding model. For example, small-world model networks show very complex spectral density with many sharp peaks, while the spectral density of scale-free model networks exhibits triangular distribution [9, 11]. Real-world networks may have all or one of the above mentioned properties, that is small diameter, high clustering coefficients and degree distribution showing power law [2, 12]. Whatever properties real-world networks, studied in the recent literature [3], have, one thing commonly existing in all of them, that is certain amount of *randomness or disorder* among network connections.

In this paper we propose to quantify this randomness following tools of random matrix theory (RMT). So far we are aware of only one relevant paper where the authors have studied the eigenvalue fluctuations in the microarray data for discovering functional gene modules [13]. In the present paper, we show our RMT analysis for different model networks studied vastly in the recent network literature and also for a real-world network. We find that in spite of having differences (in terms of various local and global properties, which are being used to characterize networks) in all these networks, there exists a common underlying universal features shown in the fluctuation of the eigenvalues of the adjacency matrix. From now onwards, the eigenvalues of the adjacency matrix of a network will be referred to as the eigenvalues of the network.

RMT was proposed by Wigner to explain the statistical properties of nuclear spectra [8]. Later this theory was successfully applied in the study of different complex systems including disordered systems, quantum chaotic systems, spectra of large complex atoms, etc [14]. The eigenvalue fluctuation is generally obtained from the nearest neighbor spacing distribution (NNSD) of the eigenvalues. The NNSD follows two universal properties depending upon the underlying correlations among the eigenvalues. For correlated eigenvalues, the NNSD follows Wigner-

Dyson formula of Gaussian orthogonal ensemble (GOE) statistics of RMT; whereas, the NNSD follows Poisson statistics of RMT for uncorrelated eigenvalues.

In the present study, we find that the NNSD of random networks follow GOE. The spectral density of random network and of the Gaussian distributed random matrix are both semicircular, so it was expected that their spacing distribution would be identical. However, though the spectral density of the scale-free network is triangular, very interestingly its NNSD follows GOE statistics. In addition to these model networks, we also analyze a protein-protein interaction network in budding yeast. Our analysis shows that this real-world network has scale free degree distribution and its spacing distribution follows GOE.

Secondly, we study the change of NNSD with the transition from regular to small-world network. Watts-Strogatz model of small-world network is constructed by rewiring the edges of regular ring lattice with probability p. This rewiring procedure generates a network with some random connections, without altering the number of vertices or edges. For $p = 0$, structure of the regular lattice or k-nearest neighbor coupled network remains same; on the other hand, for $p = 1$, the regular lattice becomes random network. For the intermediate values of p, the graph is a small-world network. We start with a regular lattice ($p = 0$) for which NNSD follows Poisson statistics. For $p = 1$, NNSD follows GOE. We find that for the intermediate value of p the NNSD shows intermediate statistics of Poisson and GOE. Moreover we show that NNSD changes from Poisson to GOE with a very small increment in p, and most importantly, the transition to GOE takes place exactly at the onset of small-world transition. We establish the relation between the small-world transition and the GOE transition in the NNSD of the networks by comparing the diameter and the clustering coefficients of the networks with a parameter (β), coming from the semiempirical eigenvalues spacing distributions studied extensively in RMT to model Poisson→GOE transition.

Here we briefly describe some aspects of RMT which we have used in our network analysis. We denote the eigenvalues of networks by λ_i, $i = 1, \ldots, N$, where N is the size of the network. In order to get universal properties of the fluctuations of the eigenvalues, it is customary in RMT to unfold the eigenvalues by a transformation $\overline{\lambda}_i = \overline{N}(\lambda_i)$, where $\overline{N}$ is the averaged integrated eigenvalue density [8]. Since we do not have any analytical form for $\overline{N}$, we have numerically unfolded the spectrum by polynomial curve fitting. After the unfolding, the average spacings will be *unity* which is independent of the system. Using the unfolded spectra, we calculate the spacings as $s_i = \overline{\lambda}_{i+1} - \overline{\lambda}_i$. The NNSD $P(s)$ is defined as the probability distribution of these s_i's. In case of Poisson statistics, $P(s) = \exp(-s)$; whereas for GOE
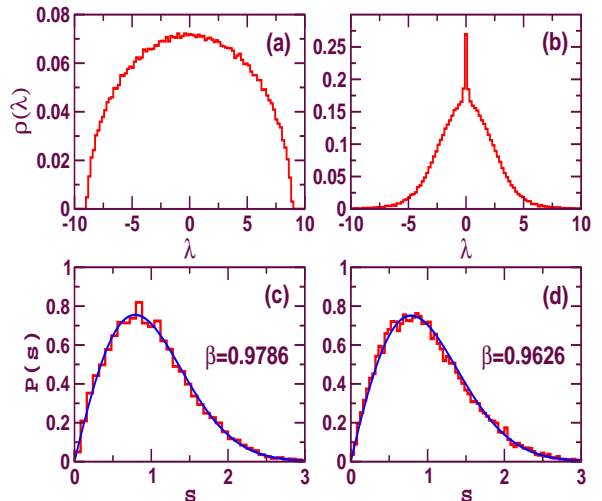

FIG. 1: (Color online) (a)-(b) Spectral density ($\rho(\lambda)$) of random and scale-free network, respectively. (c)-(d) Corresponding spacing distribution ($P(s)$). Both follow GOE statistics. All networks have $N = 2000$ nodes and an average degree $k = 20$ per node. Figures are plotted for average over 10 random realizations of the networks.

$P(s) = \frac{\pi}{2} s \exp\left(-\frac{\pi s^2}{4}\right)$. For the intermediate cases, the spacing distribution is described by Brody distribution [15]:

$$P_\beta(s) = As^\beta \exp\left(-\alpha s^{\beta+1}\right),$$

where

$$A = (1 + \beta)\alpha \text{ and } \alpha = \left[\Gamma\left(\frac{\beta+2}{\beta+1}\right)\right]^{\beta+1}.$$

This is a semiempirical formula characterized by the parameter β. As β goes from 0 to 1, the Brody distribution smoothly changes from Poisson to GOE. We fit the spacing distributions of different networks by the Brody distribution $P_\beta(s)$. This fitting gives an estimation of β, and consequently identifies whether the spacing distribution of a given network is Poisson or GOE or intermediate of these *two*.

In Fig. 1, we present the ensemble averaged spectral density ($\rho(\lambda)$) and spacing distribution ($P(s)$) of random and scale-free networks. Figs. 1(a) and 1(b) respectively show the well known semicircular and triangular distribution of the spectral density of random and scale-free networks. Though the spectral densities of these two networks are different, Figs. 1(c) and 1(d) show that the spacing distribution of both the networks follow GOE very closely ($\beta \simeq 1$). Following RMT, these results imply that even though the spectral density of the scale-free network is different from the random network, but the correlations among the eigenvalues of the scale-free network is as strong as that of the correlations among the eigenvalues of the random network.
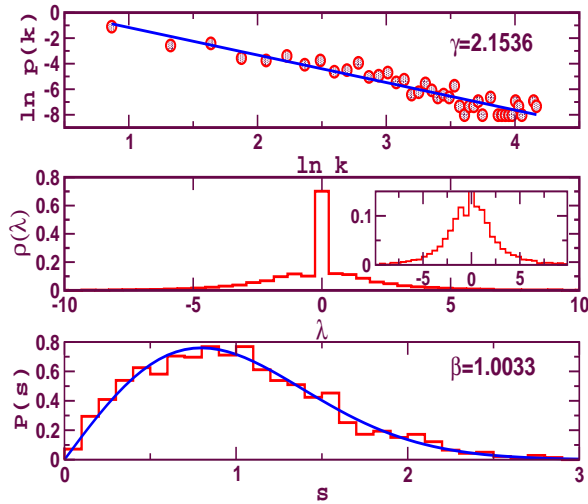
FIG. 2: (Color online) Figure shows different properties of a protein-protein interaction network in budding yeast. (a) Degree distribution : the scale-free nature of the network is clearly observed. (b) Spectral density : large value of $\rho(0)$ (Inset : besides large $\rho(0)$, overall spectral density follows well-known triangular distribution. (c) Spacing distribution : it follows GOE, estimated value of β is ~ 1.

To show that our analysis exhibiting universality of GOE statistics for the model random networks are generic, we studied some real-world networks also and here we present our results for a protein-protein interaction network in budding yeast [16]. Results are presented in Fig. 2, top panel showing that the degree distribution $p(k)$ of the network follows power-law, i.e., $p(k) \propto k^{-\gamma}$, with $\gamma \simeq 2.1536$, which satisfies the scale-free condition $2 \leq \gamma \leq 3$. The middle panel of this figure shows that the spectral density of this network is overall triangular (see also the inset of this panel for magnified figure) but with very large $\rho(0)$. Large value of $\rho(0)$ is one of the characteristics of many real-world networks [11]. Due to the large value of $\rho(0)$, it is very difficult to numerically unfold the spectra. Therefore, in this case, we divide the spectra into two parts : one part contains only negative eigenvalues with values less than -0.1 and the other part contains positive eigenvalues with values greater than 0.1. We assume these two sets of eigenvalues as the ensemble of two realizations, and calculate ensemble averaged spacing distribution. The bottom panel of Fig. 2 is showing that the spacing distribution of the protein-protein interaction network of budding yeast also follows GOE.

Now we discuss our results for Watts-Strogatz model of small-world network. In Figs. 3(a) and 3(e), we present respectively the spectral density and spacing distribution of the regular ring lattice with each node having 20 edges. These subfigures show that the spectral density of the lattice is complicated without having any known analytical form; but its spacing distribution clearly follows Poisson statistics ($\beta \sim 0$). Then we randomize a

fraction $p = 5 \times 10^{-5}$ of the edges of the regular lattice. For this value of p, the spectral density and the spacing distribution are plotted respectively in Fig. 3(b) and Fig. 3(f). These figures reveal that, for this very small value of p, the spectral density does not show any noticeable change as compared to the spectral density of the regular lattice, whereas the spacing distribution shows different property ($\beta \sim 0.0838$). As we further increase the parameter p from 5×10^{-5} to $p = 2 \times 10^{-4}$ and thereafter to $p = 5 \times 10^{-4}$, the spectral density shows hardly any changes in its features (Figs. 3(c)-3(d)), but very interestingly, according to Figs. 3(g) and 3(h), for these two values of p the spacing distributions show significantly different properties as compared to the Poissonian spacing distribution of the regular lattice. Now the spacing distributions are looking like intermediate of Poisson and GOE. By fitting the spacing distribution corresponding to these two p values with the Brody formula, we estimate the Brody parameter β respectively as 0.628 and 0.785. These values indicate that we are already at the onset of Poisson→GOE transition. From this trend, we expect that the Brody parameter β will increase with the increment of p and will reach asymptotically to unity with $p \to 1$. Note that we take regular lattice with average degree $k \simeq 20$, for this value of k we have NNSD showing Poisson statistics, for other values of k where we do not have Poisson statistics there also we get transition to GOE statistics. We choose this value of k just to make RMT analogy clear.
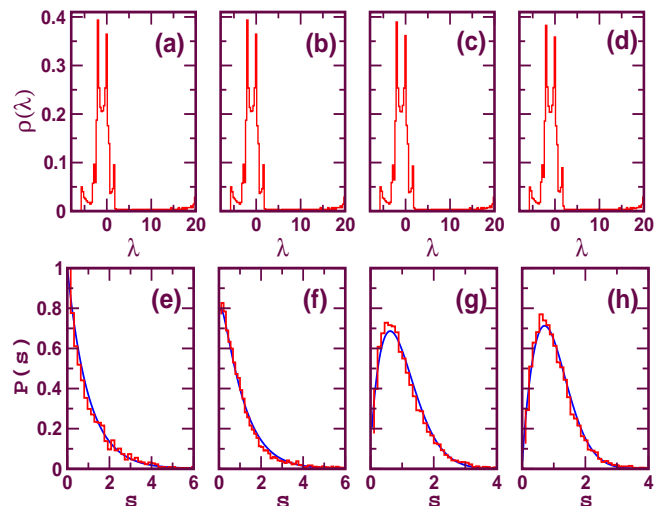

FIG. 3: (Color online) Figure shows the transition from ring regular lattice to the small-world network. (a)-(d) show the spectral densities and (e)-(h) show the corresponding spacing distributions for $p = 0, 5 \times 10^{-5}, 2 \times 10^{-4}, 5 \times 10^{-4}$, respectively. See text for the corresponding values of Brody parameters. All the networks have $N = 2000$ nodes and $k = 40$ average degree per node, and data are average over 10 random realization of the rewiring process
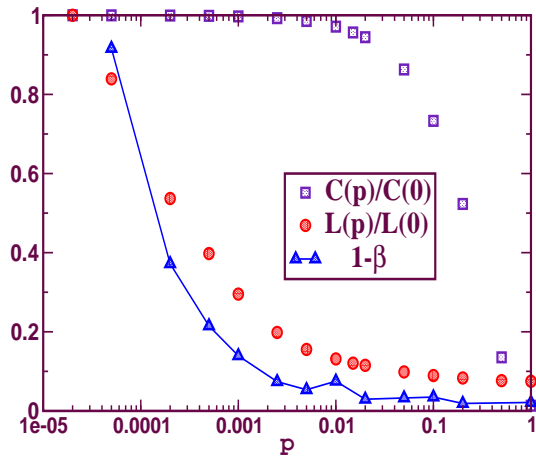
FIG. 4: (Color online) The shifted Brody parameter $1-\beta$ ($\triangle$) is compared with the two well-known network parameters, normalized characteristic length $L(p)/L(0)$ ($\bigcirc$) and normalized clustering coefficients $C(p)/C(0)$ ($\square$). The data points corresponding to the curve for β is joined by a solid line for better visibility. Network parameters are same as for Fig. 3. The data are average over 10 random realization of rewiring process for each value of p.

We present in Fig. 4 the variation of Brody parameter β as a function of p over the whole range $0 \leq p \leq 1$. Moreover, in this figure, we show the correspondence between β and *two* important network parameters - the characteristic path length $L(p)$ and the clustering coefficient $C(p)$ - as a function of p. L measures the number of connections in the shortest path between two nodes, averaged over all pairs of nodes. Clustering coefficient C measures the cliquishness of a typical neighborhood, averaged over all the nodes. In this figure we have normalized L and C by the values $L(0)$ and $C(0)$ for a regular lattice. Due to this normalization, at $p = 0$, the normalized L and C both are *one*; whereas when $p \to 1$, both network parameters will be closer to *zero*. However, β behaves completely opposite way at the two extreme values of p. Therefore, in Fig. 4, we have compared $1 - \beta$ with normalized L and C. This figure shows that the Brody parameter β and the normalized characteristic length $L(p)/L(0)$ display similar trends and strong correspondence. The most important result of this study is that the Poisson→GOE transition and the small-world transition take place at the same rewiring probability p. Note that here all the results are presented for the adjacency matrix, however we have done the similar analysis for Laplacian matrix also and for Fig. 3 and Fig. 4 qualitatively same results are obtained.

In summary, we study eigenvalues spacing distribution of various model networks and a real-world network. We study the effect of the randomness in the network architecture on the eigenvalue fluctuation of the network, and use Brody parameter to quantify this randomness. We show that though the spectral densities of the random,

the scale-free and the small-world networks are different, their eigenvalues spacing distributions are same and follow GOE statistics. The interesting finding is that *GOE transition* is observed at the onset of the small-world transition. We find that after this transition all networks behave similar as of the completely random graph. We also study a protein-protein interaction network in budding yeast and find that the spacing distribution for this network also follows GOE statistics. Following the interpretation of RMT, these results imply that at the onset of small-world transition, eigenvalues of the network show strong correlations among themselves which could be understood as some kind of correlation among the nodes of the network. Or we feel that it is some kind of spreading over the randomness in the whole network.

According to many recent studies, randomness in the network connections is one of the most important and desirable ingredients for the proper functionality or the efficient performance of the system. For instance, information processing in the brain [17], is considered to be because of many random long range connections among different modular structures. Based on our study we feel that we are at the proper ground to quantify the role of randomness, and we could also pose the questions like, can we compare the randomness in the different networks, or how much *minimal* randomness in connections is required for the information sharing and spreading in the whole system. According to our results, even very small randomness (corresponding to $\beta \sim 1$) is enough to introduce correlations among the nodes, after this network behaves as the completely random graph. Our results are based on some model networks, more general remarks and complete understanding of randomness in the real-world networks would require analysis of the networks having specific features [12, 18].

So far many local and global measures are available in literature to characterize complex networks. Following RMT we introduce a new measure, and we feel that in this framework real-world networks can be characterized [19] and compared by the amount of randomness they have.

* Electronic address: jayendra@mpipks-dresden.mpg.de
† Electronic address: jalan@mis.mpg.de
[1] D. J. Watts and S. H. Strogatz, Nature **440**, 393 (1998).
[2] S. H. Strogatz, *Nature* **410**, 268 (2001).
[3] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002) ; S. Boccalettia *et al.*, Phys. Rep. **424**, 175 (2006).
[4] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[5] F. R. K. Chung, *Spectral Graph Theory*, Number 92, AMS (1997).
[6] H. Minc and M. Marcus, *A Survey of Matrix Theory and Matrix Inequalities* (Prindle, Weber & Schmidt, 1964).
[7] M. Doob in *Handbook of Graph Theory*, edited by J. L. Gross and J. Yellen (Chapman & Hall/CRC, 2004).

[8] M. L. Mehta, *Random Matrices*, 2nd ed. (Academic Press, New York, 1991).

[9] I. J. Farkas *et al.*, Phys. Rev. E **64**, 026704, (2001).

[10] S. N. Dorogovtsev *et al.*, Phys. Rev. E **68**, 046109 (2003).

[11] M. A. M. de Aguiar and Y. Bar-Yam, Phys. Rev. E **71**, 016106 (2005).

[12] E. Ravsaz *et al.*, Science **297**, 1551 (2002).

[13] F. Luo *et al.*, Phys. Rev. E **73**, 031924 (2006).

[14] T. Guhr *et al.*, Phys. Rep. **299**, 189 (1998).

[15] T. A. Brody, Lett. Nuovo Cimento **7**, 482 (1973).

[16] http://vlado.fmf.uni-lj.si/pub/networks/data/bio /Yeast/Yeast.htm.

[17] J. D. Cohen and F. Tong, *Science*, **293**, 2405 (2001).

[18] R. Guimerá and L. A. N. Amaral, *Nature* **433**, 895 (2005).

[19] R. Milo *et al.*, Science **303**, 1538 (2004).