

Network as (0,1)-embedded system aka unweighted adjacency matrix

Ph.D. Thesis

By

ALOK YADAV



**DISCIPLINE OF PHYSICS
INDIAN INSTITUTE OF TECHNOLOGY, INDORE, INDIA
JANUARY 2019**

Network as (0,1)-embedded system aka unweighted adjacency matrix

A THESIS

Submitted in partial fulfilment of the
requirements for the degree of
of
DOCTOR OF PHILOSOPHY

by
ALOK YADAV



DISCIPLINE OF PHYSICS
INDIAN INSTITUTE OF TECHNOLOGY INDORE, INDIA
JANUARY 2019

Dedicated

to

My Beloved Mother

Acknowledgements

I wish to place on record my deep sense of gratitude and reverence to my supervisor, **Dr. Sarika Jalan** for providing me with an opportunity to be part of the Complex Systems Lab. She has always been a steady source of inspiration to me, and her constant encouragement, support as well as reprehension has enabled me to draw out the best in me. I have thoroughly enjoyed learning, interacting and co-operating during my Ph.D. at Complex Systems Lab, which has helped me evolve not only as a better researcher but also as a better person. I have enjoyed learning physics with her and think myself very lucky for having been associated with such a good researcher.

I humbly thank the Director, IIT Indore for providing the necessary amenities required for involved research. I am grateful to my PSPC committee members, **Dr. Manavendra Mahato**, and **Dr. Santosh K Vishvakarma** for scrutinizing my work and lending insightful suggestions at different phases of my Ph.D. I am also thankful to the **DPGC convener, Head (Physics)**, all the **faculty members** and **staff** of Discipline of Physics, IIT Indore for their timely help and support throughout. I humbly acknowledge the sincere efforts of all the staff members of Academic Office, Accounts Section, IT Section, and Central Library, IIT Indore whose constant support ascertained smooth functioning of my research work. I would also take this opportunity to thank my collaborators from Complex Systems Lab (CSL) IIT Indore, **Dr. Sanjiv Kumar Dwivedi**, **Dr. Aparna Rai**, **Dr. Camellia Sarkar**, **Pramod Shinde**, **Priodyuti Pradhan** and my external collaborators, **Prof. Stefano Boccaletti** (CNR-Institute of Complex Systems, Italy), **Dr. Obul Reddy Bandapalli** (German Cancer Research Center, Germany) and **Krishna kanhaiya** (Åbo Akademi University, Finland) for their help and support during my research work.

My heartfelt gratitude to all my colleagues at Complex Systems Lab - **Dr. Aradhana**, **Saptarshi**, **Anil**, **Ankit Mishra**, **Rahul**, **Vasundhara** and **Dr. Ajay** for their love, care, support and for sharing his knowledge and expertise. I especially thank Pramod for his timely help in every place it was needed as a big brother. I would also like to sincerely thank other, past and present members of Complex Systems

Lab, namely **Vipin, Amit, Ankush, Palak, Parth, Madhoolika, Sudeep, Loïc** for providing a challenging yet conducive atmosphere for my research.

I am indebted to my friends Dr. Rohit Rai, Kuber Singh, Pramod Gavel, Dr. Prabhat, Dr. Thaksen, Rohit, Jivan, Anirudh, Anupama, Dr. T. Sonamani Singh, Pramod Yadav, Dr. Anurag Gupta, Dr. Prashant Singh for helping, motivating and enlightening me in all phases of my Ph.D.

I am indebted to my all my **teachers**, who laid the foundation of education for me, enabling me to reach this far.

Words fall short for expressing my gratitude towards my parents, my loving Late **mother**, inspiring **father**, caring **younger two sisters, Geeta and Reeta** and brave **younger brother, Yogesh**, who have showered their unconditional love, believed in me and encouraged me to chase my dreams. My special thanks to my beloved **mother**, who would say to me "*beta, padha maan laga ke. baki chejo ke lye mai hu naa.*". This statement has always been my energy buster through all ups and downs. Without them, I wouldn't have been able to embark upon this journey. Finally, I thank **Almighty God** for sending all these wonderful people to my life. Thank you!!

Alok Yadav

Synopsis

Introduction

Quest of humans to understand things around them is as old as the human itself. In doing so, the human has tried to understand the complicated phenomena and things in term of smaller and known subparts. The example, in various civilizations, it can be found that five fundamental elements make all the things around us and that we need divine force to join them in some combination to get everything. This five fundamental theory concept has gone through many scrutinies, and modifications and ultimately leading to the Mendeleev's periodic table of elements and the force needed to join them has become the electromagnetic force. Similarly, to understand heavenly bodies, various models have been given from earth-centric to helio-centric, and the present-day universe model does not provide any preference for any location or direction in it and is homogeneous and isotropic, while the Gravitational force holds them together.

Nevertheless, how gravitational, electromagnetic, strong, and weak forces together give rise to the world around us, is understood to a great length, that is, standard model. However, there are many things still around us to be explored and researched. Like, how the group of a single atom is still a matter of research, that is, properties as a function of atom number in a group of atoms and are studied separately in atomic physics, nano-particle physics, condensed matter physics. It is clear here that the number of interacting partners changes can also change the overall output behavior of systems. It can be seen that such a long search has helped us to live in the present age of science and technology.

There are many other systems, natural and human-made, which need a new set of tools and studies. One such elusive system is a group of sentinel beings. Stand for one day at any busy crossing, and you will find that some people are walking, some use personal vehicles to commute, and some are using public or

private transport for travel. In any human community, few people are known and appreciated by many (or sometimes by nearly all), while a vast majority remains to be identified and interact with a tiny group of people around them. Nowadays, we can easily find this on facebook, twitter, that few celebrities have millions of followers, and a majority have friends and followers in hundreds or thousands. Also, synchronized on-off behavior of Firefly observed at night. North India blackouts on 30-31 July, 2012 due to power-distribution failure to curb the excess rise in demand for electricity caused by extreme heat and late monsoon in summer and how few healthy people became cancer patients? All these examples have interacting units but with a range of interacting partners and hence making the output behaviors not so visible to predict.

Network science studies these diverse systems and to address these different systems; network science maps these into networks. These networks have been of extreme interest of study after the two landmark papers in 1998 by Watts and Strogatz and 1999, Albert and Barabási, coupled with easy access of high computational power and extensive data on complex systems, the internet, www, facebook, and soon. Networks are graphs combined with the properties of the system they represent. The use of a graph to solve such a problem can be seen as early as the 1700s when Euler used a graph to answer the famous question of the seven bridges of Königsberg; In the 1900s, Quantum Chemistry, communication engineering, mathematics, and social science used graphs. In this period, the graphs were expressed as an adjacency matrix or list, and the use of eigenvalues was more profound in Huckle's theory for the calculation of energy levels. Later, the 1970s Laplacian and other matrices were also used to express graphs. However, for networks, the eigenvalues of the adjacency matrix were more focused on the largest eigenvalues, Restrepo, et al., 2005, and spectral density distribution, Farkas, et al., 2001 [1]. The use of random matrix theory (RMT) started in 2007, with the work of Jalan and Bandyopadhyay [1].

In this thesis, we will focus on two questions on complex networks: what can individually and collectively, eigenvalues of the adjacency matrix of a network tell

about network structure? Moreover, can we use the adjacency matrix as representative of a network for the study of the network?

Objectives

The following three workable objectives made from the two questions mentioned in the introduction :

1. How is the assortativity reflected in eigenvalues of the adjacency matrix of a network?
2. What do the zero eigenvalues of the adjacency matrix of a network mean regarding the structure of a network? Why there is degeneracy observed at zero eigenvalues in spectra of real-world networks adjacency matrix?
3. Can the adjacency matrix be treated as the picture of a network on the matrix plane?

Model

This thesis is an aggregation of three network studies based on an unweighted adjacency matrix [2–4]. We have also validated our studies of model networks on biological real-world protein-protein interaction (PPI) network of six model organisms, namely *C. elegans*, *D. melanogaster*, *H. pylori*, *H. sapiens*, *S. cerevisiae*, and *E. coli*. For abstractions of a complex system in the framework of networks, we need to first precisely define the units of a network and pairwise interaction of these units, i.e., nodes and edges of a network. In the case of real-world systems, we use empirical data for network construction, with an ultimate aim towards mapping the complexity in the underlying complex systems into a network with preserving most of the information the system. However, In our case, we have restricted our model by following conditions imposed that our network is connected. That is, the only largest connected component of the finally generated network was taken. Also, our

network is unweighted, our network is undirected, and there is not self-connection in the systems, that is, a simple, undirected and unweighted graph. All the nodes can reach each other without leaving the network itself, and all pairwise interactions are of the same strength and nature. If i^{th} -node is connected to j^{th} -node, then it also means that j^{th} -node is also connected to i^{th} -node and vice-versa.

For the construction of any model network, we need necessary information, network size, N , and the average degree of a network, $\langle k \rangle$. For model networks we have used Erdős-Rényi (ER) random network and Barabási-Albert (BA) scale-free random network. In the case of the ER random network, every pair of nodes is connected with equal probability, $p = \frac{\langle k \rangle}{N}$ and $\langle k \rangle \geq \ln(N)$, for there always exists a connected network with N nodes and average degree, $\langle k \rangle$. Also, for BA scale-free random network, start with m -seed nodes (they can be connected or not according to the final network required and $m = \frac{\langle k \rangle}{2}$) and at each time step a new node is introduced with m new links to the network, and this new node connects with any other existing nodes with probability, $p_j = \frac{k_j}{\sum_i k_i}$ and this process is repeated till all the m new connections are completed. We keep introducing $N - m$ new nodes, one at a time. Note that the connection probability of the new node to the existing node is proportional to the degree of the existing node. The tendency of new nodes to have a high probability of attaching to a high degree node is termed as preferential attachment, and leads to an outcome popularly known as rich get richer.

As already mentioned, the real-world model system for the validation of our results was the PPI networks, where the nodes are the protein, and edges denote interactions between proteins. A pairwise interaction between proteins is said to be existing if there exists a direct relation (i.e., physical), indirect relation (i.e., functional), or both the relation exist. We extracted the PPI information for the six species from a highly curated publicly available data source, DIP (database of interacting proteins). DIP (database of interacting proteins) integrates experimentally determined, reviewed, and cited protein interactions information extracted from different sources to create a single set of protein-protein interactions. With the knowledge of nodes and interactions, we construct a PPI network for each species

A is the adjacency matrix with elements a_{ij} taking values 1 and 0 depending upon whether or not there is a connection between nodes i and j . We consider undirected networks which lead to the symmetric adjacency matrix, i.e., $A_{ij} = A_{ji}$. We represent all such networks as an adjacency matrix, A with elements defines by,

$$A_{ij} = \begin{cases} 1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases} .$$

Adjacency matrix such constructed is composed of only '0' and '1' entries symmetrically distributed about the diagonal, which itself is composed of only 0's (because of no self-loop). Sum of row or column (because of the undirected nature of network) of the adjacency matrix is equal to the degree of the node corresponding to the index of row or column. The degree distribution, $P(k)$ of ER random network is Poisson distribution, $P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}$ and of BA scale-free random network is power-law distribution, $P(k) \sim k^{-\gamma}$. On the basis of Pearson degree-degree correlations ($-1 \leq r \leq +1$), networks are broadly divided in three categories assortative ($0 < r \leq +1$), neutral ($r \simeq 0$), and dis-assortative ($-1 \leq r < 0$) networks, where $r = \frac{[\frac{1}{N_c} \sum_{i=1}^{N_c} j_i k_i] - [\frac{1}{N_c} \sum_{i=1}^{N_c} \{\frac{1}{2}(j_i + k_i)\}]^2}{[\frac{1}{N_c} \sum_{i=1}^{N_c} \{\frac{1}{2}(j_i^2 + k_i^2)\}] - [\frac{1}{N_c} \sum_{i=1}^{N_c} \{\frac{1}{2}(j_i + k_i)\}]^2}$. First two of our studies were based on the eigenvalues of the network adjacency matrix. This calculation of eigenvalues from the adjacency matrix was done numerically through FORTRAN code using standard LAPACK subroutines (dsyev, ssyev, etc.) for real symmetrical matrix in single and double precisions. It is known that for our case the eigenvalues will show following relations, (i) $\sum \lambda_i = 0$, (ii) $\{\lambda_1 > \lambda_2 \geq \lambda_3 \dots \geq \lambda_i \geq \dots \geq \lambda_N\}$, where N is the number of nodes, (iii) eigenvalues spectral density distribution, $\sum P(\lambda_i) = 1$.

We investigate various properties of the adjacency matrix and their eigenvalues spectra under the network theory, spectral graph theory, and random matrix theory (RMT), to identify and study the relation between network structure and network's adjacency matrix eigenvalues spectra. We discuss the results under the following titles, 1) (Dis)assortative mixing investigated using the spectra of graphs, and 2) Origin and implications of zero degeneracy in a network spectra. We also validated these results on the model biological network of six species PPI networks.

In the third work, the adjacency matrix itself was treated as the picture of the

network (figure-3). As the indexing of nodes in a given network is random, i.e., any node can freely have any index. Therefore, for a network of N nodes, there are a large number ($o(N^2)$), of the maximum possible number of indexing schemes, i.e., the same network can have many adjacency matrix appearances. For our study of the multi-fractal nature of the model and real-world networks, we found that degree-based reshuffling keeps the multi-fractal nature of network intact. We applied the box-counting method on a large number of such degree-based reshuffling realizations. The results of all three studies have been discussed below.

Summery of the work

(Dis)assortative mixing investigated using the spectra of graphs

We investigate the impact of degree-degree correlations on the spectra of networks [2]. Even though density distributions exhibit drastic changes depending on the (dis)assortative mixing and the network architecture, the short-range correlations in eigenvalues exhibit universal RMT predictions. The long-range correlations turn out to be a measure of randomness in (dis)assortative networks. The analysis further provides insight into the origin of high degeneracy at the zero eigenvalues displayed by a majority of the biological networks. In figure 1, it is shown how the eigenvalues give information about assortativity in the network over the full range of possible changes.

Origin and Implications of Zero Degeneracy in Networks Spectra

Spectra of real-world networks exhibit properties that are different from those of the random networks. One such property is the existence of a very high degeneracy at zero eigenvalues. In this work [3], we provide all the possible reasons behind the occurrence of the zero degeneracy in the network spectra, namely the complete and partial duplications, as well as their implications. The power-law degree sequence and the preferential attachment are the properties which enhance the occurrence of

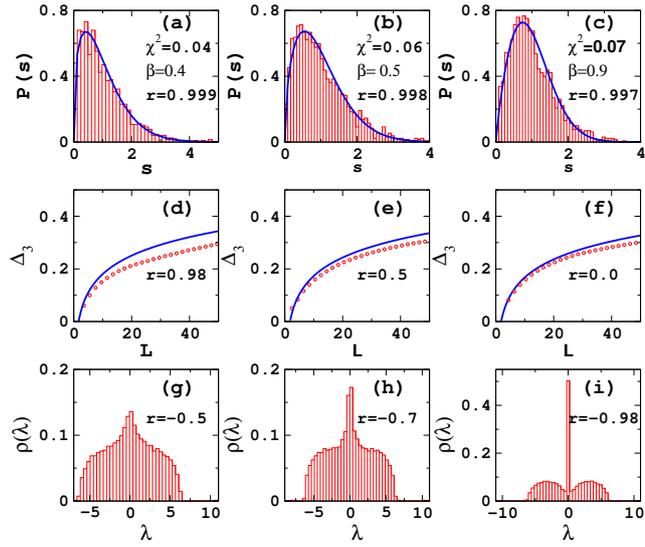


Figure 1: (a)-(c) represent the NNSD, (d)-(f) present the statistic, and (g)-(i) depict the spectral density distribution of the networks ER random networks. All graphs are for the networks with size $N = 2000$, $k = 10$, and for average, over twenty different realizations of the network.

such duplications and hence leading to the zero degeneracy. The comparison of zero degeneracy in protein-protein interaction networks of six different species and their corresponding model networks indicates the importance of degree sequences and the power-law exponent on the occurrence of zero degeneracy. Figure 2 is a schematic diagram of how the two rank reduction reasons of the matrix are equivalent to complete and partial duplications of nodes.

Unveiling the Multi-fractal Structure of Complex Networks

Traditional investigation of the fractal nature of graphs uses the network's nodes as the basic units. In this work [4], instead, we propose to concentrate on the graph's edges and introduce a practical and computationally not demanding method for revealing changes in the fractal behavior of networks, and particularly for allowing distinction between mono-fractal, quasi mono-fractal, and multi-fractal structures. We show that degree homogeneity plays a crucial role in determining the fractal nature of the underlying network, and report on six different protein-protein interaction networks along with their corresponding random networks. Our analysis

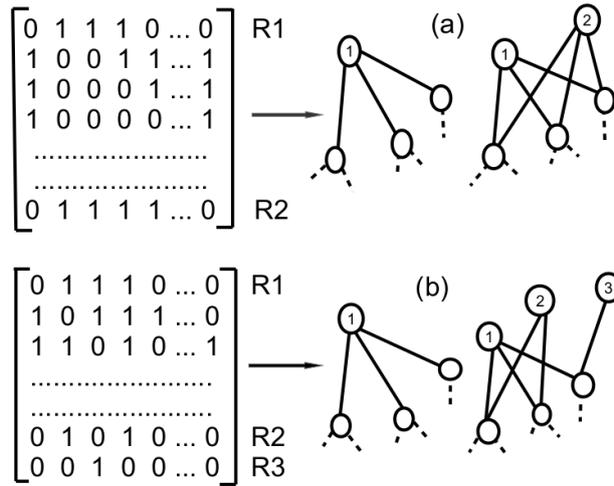


Figure 2: Schematic diagram representing (a) complete node duplication and (b) partial node duplication in networks.

allows for identifying varying levels of complexity in the species. Figure 3 shows the comparative adjacency matrix appearances for three model networks, namely ER random network, BA scale-free random network, and 1d- regular network.

Conclusions and Discussions

In our first work for the thesis, we have found numerically using spectral graph theory and random matrix theory (RMT), that eigenvalues as a whole also show changes captured in three different segments of the assortativity coefficient, r . First by network's nearest neighbor spacing distribution (NNSD)(in terms of Broody parameter, β), then by Δ_3 -statistics(in terms of long-range correlations) and then finally by spectral density distribution, $P(\lambda_i)$ (in terms of single bulk distribution to double humped distribution of eigenvalues with a peak at zero.) This numerical study has provided with numerical observations in support of the spectral based tool development for network studies for first hand and cross-verifications of results from structural properties measure tools. Our second work for the thesis has used the matrix-rank relation to eigenvalues for understanding the structural root of zero degeneracy in the adjacency matrix eigenvalues. We find that the complete

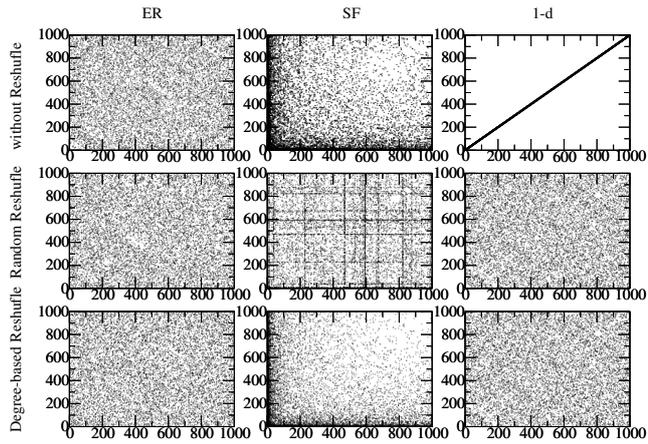


Figure 3: The effect of random reshuffling and degree-based reshuffling on the adjacency matrix. (Left) ER random networks, (middle) SF networks, and (right) 1-d lattices. In each panel, each dot is the position of an entry in the adjacency matrix plane.

duplication and collective or partial duplication of the node are the reasons for this occurrence. Also, we did numerical studies about how the two types of duplications contribution to the zero eigenvalues changes with network size, N , and average degree of network, $\langle k \rangle$. We also studied six models of biological species network and their corresponding ER random, BA scale-free, and configuration networks eigenvalues spectra. We numerically found that biological networks have more contribution dominantly from complete duplication of nodes than from partial duplication. This result for the model species network is in agreement with the evolutionary gene duplication growth mechanism already established in the networks. This result also allows us to pick into the growth processes of the network. As node duplications-adaptation and preferential-attachment growth mechanism, As both give heterogeneous degree distribution with power-law, but show different dominant reasons for zero degeneracy. For node duplications-adaptation, the dominant contribution comes from complete node duplication, but for preferential-attachment, the dominant contribution is from the partial node duplications. Our third work for the thesis probes the possibility of treating the adjacency matrix itself as a system for study rather than a representative of systems. We found numerically through multi-fractal analysis that the adjacency matrix retains the multi-fractal properties information

of the system it represents and has provided the same nature as already reposted through traditional studies of the model networks.

This thesis shows that eigenvalues could be used for the study of the networks and also provide the alternative for cross-checking the results found from non-eigenvalue study tools. The adjacency matrix plane preserves the many properties and patterns of the networks and can be a substitute for the qualitative study of the networks.

Keywords : Complex Networks, Spectra, Random Matrix Theory (RMT), Multi-fractal, Zero Degeneracy

Bibliography

- [1] Sarkar, C., Jalan, S., Spectral properties of complex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28, 10:102101(2018).
- [2] Jalan,S. , Yadav, A., Assortative and disassortative mixing investigated using the spectra of graphs, *Physical Review E* 91, 1:012813 (2015).
- [3] Yadav, A., Jalan, S., Origin and implications of zero degeneracy in networks spectra, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25, 4:043110 (2015).
- [4] Jalan, S., Yadav, A., Sarkar, C., Boccaletti, S., Unveiling the multi-fractal structure of complex networks, *Chaos, Solitons and Fractals* 97:11 (2017).

List of Publications

(A) Publications from PhD thesis work:

- [1] Sarika Jalan, and **Alok Yadav**. *Assortative and disassortative mixing investigated using the spectra of graphs*. Physical Review E 91.1 : 012813 (2015).(Impact factor: 2.284 (2017)) (citation : 09)
- [2] **Alok Yadav**, and Sarika Jalan. *Origin and implications of zero degeneracy in networks spectra*. Chaos: An Interdisciplinary Journal of Nonlinear Science 25.4: 043110 (2015). (Impact factor: 2.415 (2017)) (citation : 15)
- [3] Sarika Jalan, **Alok Yadav** , Camellia Sarkar, and Stefano Boccaletti. *Unveiling the multi-fractal structure of complex networks*. Chaos, Solitons & Fractals, 97, 11-14 (2017).(Impact Factor: 2.213 (2017))(citation : 06)

(B) Other publications during PhD :

- [1] Pramod Shinde, **Alok Yadav**, Aparna Rai, and Sarika Jalan. *Dissortativity and duplications in oral cancer*. The European Physical Journal B 88, no. 8 : 197 (2015).(Impact Factor : 1.536 (2017))(citation : 09)
- [2] Sarika Jalan, Krishna Kanhaiya, Aparna Rai, Obul R. Bandapalli, and **Alok Yadav**. *Network topologies decoding cervical cancer*. PloS one 10, no. 8 : e0135183 (2015).(Impact factor: 2.766 (2017))(citation : 06)
- [3] Camellia Sarkar, **Alok Yadav**, and Sarika Jalan. *Multilayer network decoding versatility and trust*. EPL (Europhysics Letters) 113, no. 1 : 18007 (2016).(Impact Factor : 1.834 (2017))(citation : 05)
- [4] Priodyuti Pradhan, **Alok Yadav**, Sanjiv K. Dwivedi, and Sarika Jalan. *Optimized evolution of networks for principal eigenvector localization*. Physical Review E, 96(2), 022312 (2017).(Impact factor: 2.284 (2017))(citation : 11)

Reference :

[https://scholar.google.co.in/citations?user=sVwG4UcAAAAJ &hl=en&authuser=1](https://scholar.google.co.in/citations?user=sVwG4UcAAAAJ&hl=en&authuser=1)
on 24-01-2019

Table of Contents

1	Introduction to Networks	1
1.1	Introduction with historical background	1
1.2	Network Representation	5
1.2.1	Pictorial Representation of Network	6
1.2.2	List Representation of Network	6
1.2.3	Matrix Representation of Network	6
1.3	Basic definitions and tools of network theory	7
1.4	Local network properties	8
1.5	Global Network Property	9
1.6	Network models	11
1.6.1	Erdős-Rényi (ER) random networks	11
1.6.2	Barabási-Albert (BA) scale-free networks	12
1.7	Watts-Strogatz (WS) small-world networks	13
1.7.1	Configuration model	14
1.8	Network spectra	14
1.8.1	Spectral density	15
1.8.2	Degeneracies in a network spectra	15
1.9	Random matrix theory and networks - What is the connection? . . .	16
1.9.1	Tools of RMT used in networks' investigations	16
1.9.2	Randomness: structural and spectral perspective	19
1.10	motivation behind this thesis	19
1.11	Readers Guide to the thesis	20
2	Network Structure and Eigenvalues Spectra of Networks	23
2.1	Introduction	23
2.2	Methods and Techniques	24
2.3	Results	27

2.4	Discussion and Conclusions	35
2.5	Appendix	38
3	Network Structure and Zero-Eigenvalue of Networks	41
3.1	Introduction	41
3.2	Methods and Techniques	42
3.3	Results	45
3.4	Discussion and Conclusions	50
4	Network Structure and (0,1)-Embedded Matrix Image of Networks	53
4.1	Introduction	53
4.2	Methods and Techniques	54
4.3	Results	57
4.4	Discussion and Conclusions	61
5	Conclusions and Scope for Future Work	63
5.1	Conclusions and Discussions	63
A	Appendix	67
A.1	Random reshuffling and degree-based reshuffling of the nodes	67
A.1.1	Random reshuffling	67
A.1.2	Degree-based reshuffling	70
A.2	Edge distribution of different networks on adjacency matrix plane	74
A.2.1	Case A: Strictly Homogeneous distribution	74
A.2.2	Case B : Approximately Homogeneous distribution	76
A.3	τ_q versus q for PPI Networks	77
	Bibliography	79

List of Figures

1	(a)-(c) represent the NNSD, (d)-(f) present the statistic, and (g)-(i) depict the spectral density distribution of the networks ER random networks. All graphs are for the networks with size $N = 2000$, $k = 10$, and for average, over twenty different realizations of the network.	xv
2	Schematic diagram representing (a) complete node duplication and (b) partial node duplication in networks.	xvi
3	The effect of random reshuffling and degree-based reshuffling on the adjacency matrix. (Left) ER random networks, (middle) SF networks, and (right) 1-d lattices. In each panel, each dot is the position of an entry in the adjacency matrix plane.	xvii
1.1	Symetic diagram to depict how much information about the network was used to calculate local properties.	5
2.1	Spectral density for Erdős-Rényi random networks with different values of assortativity coefficient r . All graphs are plotted for the networks with size $N = 1000$ and connection probability $p = 0.01$, averaged over twenty different realizations of the networks.	28
2.2	The NNSD for Erdős-Rényi random networks with different values of assortativity coefficient r . All graphs are plotted for the networks with size $N = 1000$ and connection probability $p = 0.01$. Histograms are from the data points, and the solid line is for fitting with Brody distribution (Eq. 2.3).	29
2.3	The $\Delta_3(L)$ statistic for Erdős-Rényi random networks with different values of assortativity coefficient r . All graphs are plotted for the networks with size $N = 1000$ and connection probability $p = 0.01$. The solid line is the prediction from GOE statistics (Eqs. 2.4 and 2.5) and open circle are calculated from the network.	30

- 2.4 (a)-(c) represent the NNSD, (d)-(f) present the $\Delta_3(L)$ statistic and (g)-(i) depict the spectral density distribution of the networks ER random networks. All graphs are plotted for the networks with size $N = 2000$, $\langle k \rangle = 10$ and for average over twenty different realizations of the network. 31
- 2.5 Spectral density for scale-free networks with different values of assortativity coefficient r . All graphs are plotted for the networks with size $N = 1000$, $\langle k \rangle = 10$, for twenty different realizations. 33
- 2.6 (a), (b) and (c) plot average NNSD for an ensemble of twenty realizations for different values of r , whereas (a'), (b') and (c') plot the ensemble having different number of the network realizations. The histogram is drawn using the data from the networks, and the solid line is the fitted Brody distribution. For all the graphs $N = 1000$ and $\langle k \rangle = 10$ 39
- 3.1 Schematic diagram representing (a) complete node duplication (Eq. 3.1) and (b) partial node duplication (Eq. 3.2) in networks. 45
- 3.2 The density distribution of Erdős-Rényi (ER) random networks and scale-free (SF) networks for different average degrees and $N = 1000$. Δ , \square , \circ and $*$ represent the data points of density distribution for $\langle k \rangle = 2, 4, 6$ and 8 , respectively. All values are averaged over ten realizations of the networks. 47
- 3.3 (Effect of the change in the network parameters, namely, size (N) and the average degree ($\langle k \rangle$) on the number of the duplicates and zero eigenvalues in different model networks. All values are averaged over 10 random realizations of the networks. Note that for ER random networks, connected component could not be obtained at $\langle k \rangle = 4$ for N above 1000. Here the Δ , \circ and \square represent the D_c of the ER, SF and configuration model networks, respectively. The solid Δ , \circ and \square represent the λ_0 of the ER, SF and configuration model networks, respectively. 49
- 4.1 (Color Online). Schematic representation of the used box counting method. Left: original network. Right: partition of the adjacency matrix \mathcal{A} 54
- 4.2 (Color Online). P_ϵ^q vs. ϵ in a double logarithmic scale with $q \in (-10, 10)$ for (a) ER and (b) SF networks. $N = 1,000$ and $\langle k \rangle = 20$. 55

4.3	$(\tau(q)$ (a) and D_q (b) vs. q . (c) $f(\alpha)$ vs. α . In all plots results are reported for 1-d lattices, ER and SF networks with $N = 1,000$ and $\langle k \rangle = 10$. All quantities are calculated for 50 random realizations of the networks generated by reshuffling indices 100,000 times for each realization. The left and right region slopes of $\tau(q)$ for the 1-d lattices, ER and SF networks are, respectively, 2.1 and 1.9; 2.4 and 1.9; 2.3 and 1.2.	56
4.4	D_q vs. q for PPI networks (open circles) of six species, along with their corresponding ER random networks (closed circles) having the same size and average degree. The results for each of the corresponding ER random networks are presented for 100,000 node reshufflings, and for 50 realizations.	59
A.1	The effect of random reshuffling and degree-based reshuffling on the adjacency matrix. (Left) ER random networks, (middle) SF networks and (right) 1-d lattices. In each panel, each dot is the position of an entry in the adjacency matrix plane.	68
A.2	Various multi-fractal measures plotted for the case of random reshuffling.	69
A.3	Graph representation of Heaviside step function used in Eq. A.7. . .	71
A.4	(Color Online) Comparison between, respectively, $\ln(P_i^q(\epsilon))$ vs $\ln(\epsilon)$ plot, $\tau(q)$ vs q plot, $D(q)$ vs q plot and $f(\alpha)$ vs α plots for strictly (top panel) and approximately (bottom panel) homogeneous distribution of edges.	75
A.5	(Color Online) Plots of τ_q as a function of q for PPI networks (circles) of six species along with their corresponding ER random networks (triangles) having the same size and average degree. The results for each of the corresponding ER random networks are presented for 100,000 node reshuffling done for 50 realizations.	78

Introduction to Networks

Chapter 1

Introduction

1.1 Introduction with historical background

From the scientific point of view, the study of complex natural systems is done by abandoning some of the details and focus on the significant features. When such abstractions are reasonable enough, then the loss of the knowledge of the details is abundantly compensated by the gain of a better insight into the systems. One of such abstract procedure is graph or network. In a graph of a system, nodes or vertices represent interacting units, and edges or links are pairwise relation or interaction between units. Few such complex systems are friendships in a group of people, neural networks, and the internet. The use of graphs to solve the real-world problem can be traced back to the 18th century when Leonhard Euler proposed his brilliant solution to the Seven Bridges of Königsberg problem [5]. Also, Graph theory gained popularity in the 19th and 20th centuries because of its power to describe complex systems in many different fields such as communication infrastructure, path planning, coloring maps, and social structure, in terms of much smaller interacting unit phenomena [6].

The advent of the information age and increase in the computational power of computers since the early 1990s has led to an increasing interest in the fundamental properties of real networks. A striking observation of these real networks is that complex topologies and very different structures characterize them. Many of them are scale-free, meaning that there are a much small number of nodes, usually called hubs, that are much more connected than the average. Typical examples are the internet, facebook, and protein-protein interaction network. Some of them are small-world: a short chain of links separates most of the pairs nodes. This feature has been well known in social science since the 1960s as six degrees of separation [7, 8]. Another highlight for many real networks is that the entities tend to form groups with dense internal connections, i.e., such a system shows a high clustering coefficient.

Many models were proposed to capture these features in the model networks or graphs. The simplest model was introduced by Pál Erdős and Alóert Rényi, [9, 10] and with the assumption that the connection probability of each pair of nodes in a graph is equal and fixed. The resulting network shows small-world effects but with a small clustering coefficient. In 1998, a new proposed model by Watts and Strogatz was able to capture both the small-world effect and the high clustering coefficient [11]. This small-world network is observed at the small-world transition region, which is achieved through a small number of random reconnections in a regular network. In 1999, the famous preferential attachment model was proposed by Barabasi and Albert [12], which is a growth model with features of a scale-free connectivity distribution, high clustering, and small-world effects. Since these two most influential papers on small-world networks and scale-free networks in the last two decades, the network theory has been intensively used to investigate various real-world and model systems. The network theory framework applies to any complex system as long as it can be mapped into a suitable complex network structure composed of nodes and connections. Therefore, any system which is composed of interacting units can be studied by network theory framework.

While recently, the networks are the driving force behind the investigation of

the graphs and their spectra. Mathematically, a network is a graph $G(N, N_c)$, where N is a set of nodes, and N_c is a set of pairs of nodes called edges. So, what's the difference between graphs and networks? A generally accepted, all-encompassing definition of a network does not seem to be available. Instead, networks are understood by instantiation: The World Wide Web, The Internet, Citation networks., Language networks, Food webs., Economic networks., Metabolic and protein networks, Social networks, transportation (car, train, airplane) and infrastructural (electricity, gas, water, sewer) networks, the human brain network, software dependency networks are examples of networks. Some of these networks are made by nature; others are built by humans. So, while a graph is an abstract mathematical object, a network is a real-world graph with specific structural properties based on a real-world system. These properties have been exploited to investigate the origin and evolution of networks and to study the processes taking place on them. The majority of these networks exist for many years, some of them (biological networks) are here for millions of years.

Also, the use of eigenvalues or spectral graph theory for the study of physical systems can be traced back in the 1700s, and were for the study of the rigid bodies, Leonhard Euler showed the importance of principal axes, and Joseph-Louis Lagrange later found that angular momentum $\mathbf{L} = \mathbf{I}\bar{\omega} = \lambda\bar{\omega}$,

Where the (scalar) eigenvalue λ is called a principal moment of inertia corresponding to these principal axes are the eigenvectors, ω of the inertia matrix, I . In the 1800s, Augustin-Louis Cauchy extended the work of Euler and Lagrange and further proves that real symmetric matrices have real eigenvalues. Cauchy also introduced the term "characteristic roots" for eigenvalues. In 1904, David Hilbert used term eigenvalues driven from German word Eigen meaning characteristic. In 1929, Richard von Mises developed First numerical algorithms, i.e., a powerful method for finding eigenvalues. In the 1950s and 1960s, spectral graph theory was treated as a branch of algebraic graph theory, with mainly focused on the adjacency matrix of regular graphs. Also, about the same period, the research was independently begun in quantum chemistry [13–17], as eigenvalues of a graphical

representation of atoms correspond to energy levels of electrons.

While the eigenvalues of matrices with entries following some pattern and or based on some real system were a very successful tool of study, the collective and statistical survey of eigenvalues is done in random matrix theory (RMT). The original motivation for RMT research was to understand the statistics of the distribution of spacings of energy levels of heavy nuclei, measured in nuclear reactions (Wigner, 1957), who was probably inspired by the compound nuclear model, proposed by N. Bohr, in 1937. RMT was developed in the 1960s, notably by Wigner, Dyson, Mehta, and Gaudin. RMT deals with the statistical properties of large matrices with randomly distributed elements. In the case of RMT, The probability distribution of the matrices are inputs, from which the correlation functions of eigenvalues and eigenvectors are outputs. From the correlation functions, one then computes the physical properties of the system.

Later the same techniques were applied to describe the microwave absorption by granular metals using the level statistics of small metal particles (Gor'kov and Eliashberg, 1965). In recent years there has been a revival of interest in RMT, mainly because of two developments. The first was the discovery that the Wigner-Dyson ensemble applies generically to chaotic systems (Bohigas, Giannoni, and Schmit, 1984; Berry, 1985). The second was the discovery of a relation between the universal properties of large random matrices and global conductance fluctuations in disordered conductors (Altshuler and Shklovskié, 1986; Imry, 1986a).

In network theory research the use of network eigenvalues or spectrum sparked after the work of Farkas et al., in 2001, showed that how spectra of scale-free network differ from random network's semicircular law predictions and follow the triangular shape. Restrepo et al., in 2005, where they found that synchronization dynamics transition from incoherence to coherence in a vast network of coupled phase oscillators is governed by the largest eigenvalue of the adjacency matrix of the network. While in 2007, Sarika et al., showed that nearest-neighbor spacing distribution of the eigenvalues of the adjacency matrices of various model networks follow universal Gaussian orthogonal ensemble statistics of RMT, and an analogy

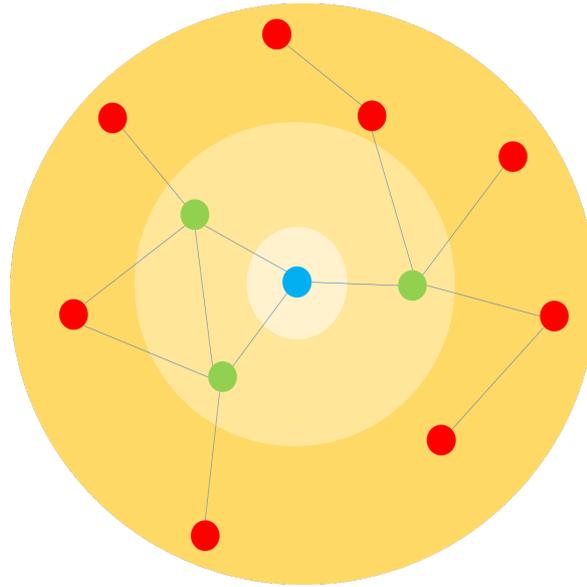


Figure 1.1: Symetic diagram to depict how much information about the network was used to calculate local properties.

between the onset of small-world behavior, quantified by the structural properties of networks, and the transition from Poisson to Gaussian orthogonal ensemble statistics, quantified by Brody parameter characterizing a spectral property.

In this thesis, we will focus on three questions on complex networks: 1. Do eigenvalues collectively store information about network degree-degree correlations? 2. What can degenerate eigenvalues at $\lambda = 0$ of the adjacency matrix of a network tell about network structure? And finally, 3. Can the adjacency matrix of a network be treated as 0,1-embedded images of the network in a matrix plane?

1.2 Network Representation

The network is made of N nodes, and N_c connections. Therefore, (N, N_c) pair is the most basic information needed for any network. Networks thus formed can be directed or undirected, also weighted and unweighted. In work presented here, we will only talk about undirected and unweighted networks, until unless stated explicitly.

However, even with a given pair of (N, N_c) , a large set of networks can be formed, and this pair becomes a very loose characterization property for any sys-

tems or model network. Therefore, we use the other one or more properties to characterize the network further. This narrow down the previous (N, N_c) set of networks to $(N, N_c, \alpha, \beta, \dots)$, where α, β, \dots are characterization properties for networks. Therefore, one of the best ways is to provide a connection list of a pair of nodes in the network. From this list, we can reproduce the network structure.

There are three main ways to represent a network, namely (i) a pictorial representation of a network, (ii) list representation of a network, and (iii) matrix representation of a network.

1.2.1 Pictorial Representation of Network

Pictorial representation of a network offers an exact and visual way to study networks. Therefore, it has been a preferred and frequently used method to analyze and represent networks by social scientists, management, and media people. However, as the size, N , of the networks grows, it becomes highly impossible to study networks through pictorial representation, and hence, people switch to other methods for network study.

1.2.2 List Representation of Network

List representation of a network used to map network into a relation list. The most frequently used relation is the connection between a pair of nodes, and this gives us a connections list, also known as an adjacency list. In a network, such links can be studied through many structural properties like degree, clustering coefficient, betweenness centrality, duplication, and so forth.

1.2.3 Matrix Representation of Network

Networks can also be represented as a matrix, example adjacency matrix, Laplacian matrix, etc. The adjacency matrix of a network is defined by Eq. 1.1

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \text{ is connected to } v_j \\ 0 & \text{otherwise} \end{cases} \quad (1.1)$$

In undirected graphs, the relation $a_{ij} = a_{ji}$ holds good, whereas same is not the case for directed graphs. With this representation spectral study of network is possible.

In case of undirected adjacency matrix all eigenvalues are real. Several statistical measures are proposed to understand specific features of the network [18, 19].

1.3 Basic definitions and tools of network theory

If we are interested in knowing the behavior of any node in a network, then local properties tell us more about it than global properties. For example, knowing the degree of the node is more useful or informative than the average degree of the network. However, an average degree does tell us about the overall behavior of that network, like is network sparsely connected or dense. Like if we randomly select a node in a network than what could be the degree of this node. Example, in the given network (shown in figure 1.1) average degree, $\langle k \rangle$ is approximately 2.4 (number of connection, $N_e = 13$ and number of nodes in the network, $N = 11$) and 3 is the degree of the blue node in the smallest white circle. As the fractional degree is not possible for an unweighted network, the only way to say anything about the network by $\langle k \rangle$ is that the majority of nodes in the network are connected with a connection 2 or 1 with few nodes with a degree higher than 2. Out of 13 nodes in the network, eight nodes have only a degree of 1 or 2 with five nodes with a degree of 3 or 4. However, the degree of the blue node is not predictable by $\langle k \rangle$. So, global properties tell us about the overall behavior of the network, but local features depict the local response. To calculate local features, it can be further divided on the bases of the amount of information needed.

The first type, which needs the least information, explicitly how much a node is connected, a blue node in the white circle in figure 1.1. This category has only intrinsic properties, like nodes degree, frequency, etc.. The degree of a node does not tell us to whom it is connected or how the other nodes degree are distributed on the networks. Therefore, the degree of any node is the property that requires only information about how many other node nodes it interact and nothing about the other nodes. Like, on a social network, the unit is the people who are related by the already sat conditions, will know with how much other people they are interacting or connected.

The second type is the local properties, which require the information not only about node but also about nodes with whom the node is interacting. So, the information about the node which is required is more than enclosed in the white circle. The information about the node and its neighbors is needed, i.e., in case of a blue node, we n and is enclosed in the light yellow circle in figure 1.1. These properties are like clustering coefficients, overlap, and so on. These properties usually manifest how much a node knows about with whom it is connected. As the clustering coefficient tells us about the average connectivity of the neighbors of a node, for example, $CC_i = 0$ means that none of the nodes connected to it are connected or see each other. Similarly, $CC_i = 1$ implies that all the nodes connected to it are also connected or see everyone.

Finally, the Third type of local properties. These type of local properties are the properties which require all the information about other nodes in the networks. So, they need all the information in the 'C' circle in figure 1.1. These properties are scores that either how a node is participating in the network or how it is affected by the network.

1.4 Local network properties

The properties of any node in the network are called local properties of the network. For example, the number of connections a node has in a network is called the degree of the node.

- **Degree:** In the undirected graphs, nodes have one type of degree. The degree of any node is the number of connections node have in the network. It is represented by d_i for i^{th} node in the network. The sum of the degrees of all nodes in the network gives us a number twice the number of connections in the networks. $\sum_{i=1}^N d_i = 2N_c$.

The spread in the degrees is characterized by a distribution function $P(k)$, which gives the probability that a randomly selected node has exactly k edges. The average degree of a network is a measure of its sparseness (or denseness); moreover, it is expected degrees of all the nodes in the graph.

- **Clustering coefficient:** Clustering coefficient of a node i denoted as C_i , is defined as the ratio of the number of links existing between the neighbors of the node to the possible number of links that could exist between the neighbors of that node [19] and is given by (in terms of Adjacency matrix)

$$C_i = \frac{\sum_{j_2=1}^{k_i} \sum_{j_1=1}^{k_i} (A_{ij_1} A_{j_1 j_2} A_{j_2 i})}{\frac{1}{2} k_i (k_i - 1)} \quad (1.2)$$

where i is the node of interest and j_1 and j_2 are any two neighbors of the node i and k_i is the degree of the node i .

(in terms of connections in neighbor nodes)

$$C_i = \frac{\Delta_i}{\frac{1}{2} d_i (d_i - 1)} \quad (1.3)$$

where Δ_i is the number of connections between the nodes connected with i^{th} node and $\frac{1}{2} d_i (d_i - 1)$ is the maximum possible connection between the neighbour of i^{th} node. So, the clustering coefficient is the relative connectedness of the node's neighbors.

1.5 Global Network Property

Average or global property of network is calculated or evaluated, taking the average over one kind of information/ property on nodes in a network. This property tells us by just one number what should we expect about a randomly selected node in the network without actually studying it.

- **Average Degree :** In a network with N nodes and N_c number of connections, on average any node is exact to have $\langle k \rangle$ connection. Therefore, average degree of network is given as $\langle k \rangle = \frac{2N_c}{N} = \frac{\sum_{i=1}^N d_i}{N}$.

Note: for the Erdős-Rényi random networks with connection probability, say p_c , the expected average degree of network is ($= p_c N$). This statistically will be the same as the average degree of generated random networks.

- **Degree Sequence :** Degree sequence of a network is a set of all degrees of all the nodes in the networks. $[d_1, d_2, d_3, \dots, d_i, \dots, d_N]$ This tells us exactly how many nodes have a given degree.

- **Degree distribution :** Degree distribution, $P(d_i)$ of degrees in networks are used to classify the networks in general categories like Erdos-renyi random network, scale-free network, etc. Two more commonly occurring class of degree distributions are Poisson degree distribution and Power-law degree distribution. These distributions tell us what is the probability that a given degree node will be found in a given network. Real-world networks usually have scale-free Power-law distribution of degree with power-law exponent, γ value range between (2,3).

- **Degree-degree correlation coefficient:** We quantify the degree-degree correlation of a network by considering the Pearson (degree-degree) correlation coefficient, given as [64]

$$r = \frac{[N_c^{-1} \sum_{i=1}^{N_c} j_i k_i] - [N_c^{-1} \sum_{i=1}^{N_c} \frac{1}{2}(j_i + k_i)]^2}{[N_c^{-1} \sum_{i=1}^{N_c} \frac{1}{2}(j_i^2 + k_i^2)] - [N_c^{-1} \sum_{i=1}^{N_c} \frac{1}{2}(j_i + k_i)]^2} \quad (1.4)$$

where N_c is number of connections in network, and j_i, k_i are degrees of nodes on both ends of i^{th} connection respectively. The degree-degree correlation coefficient, r , tells us that whether nodes at both ends of a connection in a network have the same degree or different, statistically. Value of r range from $(-1, 1)$.

When the value of r is more than zero, then these types of networks are called assortative or homophily networks, example k -regular network has r value 1. When r is negative for a network, then the network is called a dis-assortative network, for example, a star network has r equal to -1 . While $r = 0$ for the Erdős-Rényi random networks.

- **Average Clustering Coefficient :** The average clustering coefficient is defined as the average clustering coefficients of all nodes in the networks.

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i \quad (1.5)$$

where, C_i is the clustering coefficient of i^{th} -node.

1.6 Network models

The study of networks started off by assuming that complex networks can be faithfully represented using random graph models, where every edge has an equal likelihood of existence. Paul Erdős and Alfred Rényi pioneered the study of random graph models [9, 10]. After decades, during which the Erdős-Rényi model remained the preferred method for studying networks, Barabási and Albert observed that many of the complex real-world networks follow a power-law degree distribution and consequently proposed the scale-free network model [12]. This came as a revolutionizing change in the perspective of network analysis leading to extensive investigations of various real-world systems. Some of the networks that were found to follow the power-law degree distribution include the internet, the world-wide-web, cellular networks, mobile communication networks, scientific collaboration networks, to mention a few [18].

Meanwhile, another category of networks that gained prominence was small-world networks, proposed by Watts and Strogatz [11], motivated from the analogy of *six degrees of separation* [20]. These networks are highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs. The neural network of the worm *Caenorhabditis elegans*, the power grid of the western United States, and the collaboration graph of film actors are shown to be small-world networks. In the following, we present an elaborate discussion on the three network models.

1.6.1 Erdős-Rényi (ER) random networks

The algorithm for the random network was given by Hungarian mathematicians Paul Erdős and Alfred Rényi.

Generation: Starting with N nodes, all the a pair of the nodes in the network is connected with a probability p , keeping the number of nodes in the network fixed.

Following are few of the properties of ER random networks:

- The total number of connections in the network is approximately $pN(N - 1)/2$ and its expected average degree ($\langle k \rangle$) is $p(N - 1)$ [9, 10].

- The degree distribution was derived by Bollobás [21] as a Binomial distribution and probability of a node having k degree is given by

$$\rho(k) = \binom{N-1}{k} p^k (1-p)^{N-k-1}$$

For large values of N , degree distribution turns out to be Poisson distribution and is given by

$$\rho(k) = \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}$$

- Diameter of ER random network is given by $\frac{\ln(N)}{\ln(pN)}$. Random networks have a small diameter, provided p is not very small.
- The clustering coefficient of ER random networks is given as $\frac{\langle k \rangle}{N}$.

Most of the properties exhibited by ER random networks do not have a close resemblance with the networks generated from real-world data, for instance, metabolic networks, protein-protein interaction networks, social networks, and so on. These networks do not exhibit Poisson degree distribution. It is increasingly found that a large number of systems exhibit behavior which is close to the power-law degree distribution.

1.6.2 Barabási-Albert (BA) scale-free networks

The degree distribution of a random graph is a Poisson distribution with a peak at $P(\langle k \rangle)$. However, in most large networks such as citation networks [22], the mobile communication networks [23], the cellular networks [24], the Internet networks [25], and the worldwide web [26], the degree distribution significantly deviates from a Poisson distribution but has a power-law tail $P(k) \sim k^{-\gamma}$. It has been sufficiently proven that the degree distribution of real-world networks is not random, most of them having a long right tail corresponding to values that are far above the mean [12]. In order to understand the evolution of the power-law degree distribution in these networks, several models have been presented. Price presented a model towards the explanation of the origin of the power-law degree distribution, prevalent in real-world networks [27] by incorporating Simons's 'rich-get-richer' phenomenon from Economics. Later A.-L. Barabási and Reka Albert proposed a

model to explain efficiently the emergence of the power-law degree distributions in networks, which is based on network evolution rather than the fixed size model of ER random networks and is the most widely accepted model.

Generation: This model generated as per BA algorithm is based on two mechanisms, (i) growth and (ii) preferential attachment. *Growth:* Starting with a very small number of nodes (say n_0), in each time step a new node is added which has ($< n_0$) edges. *Preferential attachment:* The newly added node makes connections with the already existing nodes with a probability π_i , which depends on the degree of the node i as:

$$\pi_i = \frac{k_i}{N_c}, \quad (1.6)$$

where N_c is the total number of connections in the network at time t .

Following are few of the properties of scale-free networks:

- The probability of a node having a k degree is given by the power-law degree distribution ($P(k) \sim k^{-\lambda}$) where λ is a constant and lies between 2 and 3 for sufficiently large networks [18].
- Scale-free networks exhibit a lower value diameter of about $\ln(\ln(N))$ as compared to random networks having similar connectivity and dimension [63].
- Clustering coefficient of scale-free networks is proportional to $N^{-0.75}$ [18].

The scale-free structure is known to provide robustness to the system against random attacks while makes it fragile to the targeted attack. For example, virus spreading in the Internet network becomes more harmful when the hub nodes get infected [18].

1.7 Watts-Strogatz (WS) small-world networks

Many of the real-world systems possess diameter as small as that of the random networks and clustering coefficient as large as that of the regular networks. The social psychologist Stanley Milgram, through his experiment, demonstrated that

most of the people in the United States of America have six degrees of separation between them [20]. Taking this into consideration, Watts and Strogatz proposed a model termed as the small-world network [11].

Generation: Begin with a ring lattice of N nodes in which each node is connected to its k neighbors, having $k/2$ neighbors on either side. Each edge is randomly rewired in the lattice with probability p_r , such that there are no self-connections and duplicate connections. The value of p being zero corresponds to a regular graph, and $p = 1$ corresponds to the random graph.

Following are few of the properties of small-world networks:

- On increasing p_r value on a logarithmic scale from zero, the clustering coefficient does not significantly decrease whereas the diameter drastically decreases.
- Further increase in p_r leads to a decrease in the value of clustering coefficient on a faster pace and at this value The diameter gets saturated. Networks at this p_r value are known as small-world networks and have both features of high clustering and low value of the average shortest path.

1.7.1 Configuration model

The configuration model also of having the same size and number of connections as of a given network, preserves the degree sequence of the given network, by generating a random network with a given degree sequence. We construct the configuration model network by taking the degree sequence d_1, d_2, \dots, d_N of various PPI networks as input. Each node of the corresponding configuration model is allocated stubs equal to their degree, and then these studs are paired with a uniform probability [19]. The network thus generated is a configuration model for a given degree sequence.

1.8 Network spectra

The eigenvalues of the adjacency matrix of a network are related to many basic topological invariants of networks, such as the diameter of a network [28]. In or-

der to physically understand the concept of eigenvalues and eigenvectors, a set of axes is chosen in the multidimensional space occupied by a network. The axes are rotated so that the first axis points in the direction of the greatest variability in the data and the second axis, orthogonal to the first, points in the direction of greatest remaining variability, and so on. This set of axes is a coordinate system that can be used to describe the relative positions of the set of points in the data. Most of the variability in the locations of points will be accounted for by the first few dimensions of this coordinate system. The coordinates of the points along each axis denote an eigenvector, and the length of the projection is termed as an eigenvalue [29]. The set of all eigenvalues is the spectrum of the network. Spectra of networks have been proposed to present a fingerprint of the networks, rendering the characterization of networks easier [30–32]. The rich information about the topological structure and diffusion processes can be extracted from the spectral analysis of the networks [34]. Studies of spectral properties of the complex networks may also have a general theoretical interest.

1.8.1 Spectral density

The distribution of the eigenvalues is entailed as the spectral density ($\rho(\lambda)$). The spectral density of uncorrelated random graphs is known to follow a semicircular distribution [30]. For scale-free networks, the spectral density exhibits a triangular distribution with an exponential decay around the center, followed by power-law long tails at both spectrum edges [32].

1.8.2 Degeneracies in a network spectra

Real-world networks exhibit properties that are very different from those of the corresponding model graphs [35]. One of these properties is the occurrence of degeneracy at 0, -1 , and -2 eigenvalues [36]. Few papers have related 0 and -1 eigenvalues to stars and cliques respectively [31, 32, 37]. However, graphs in the absence of stars and cliques can still show a degeneracy at 0 and -1 eigenvalues, respectively. For instance, the 0 degeneracy has been shown to arise from the complete and the partial duplications of nodes [?]. Recently, it has been reported that a

simple transformation of the network's adjacency matrix provides an understanding of the origins of the occurrence of high multiplicities in the network spectra. The eigenvectors associated with the degenerate eigenvalues shed light on the structures contributing to the degeneracy [38].

1.9 Random matrix theory and networks - What is the connection?

Random matrix theory (RMT), proposed by Wigner to explain the statistical properties of nuclear spectra, has elucidated a remarkable success in understanding complex systems, which includes disordered systems, quantum-chaotic systems, spectra of large complex atoms, etc. [39]. The random matrix theory (RMT) approach proposed by Wigner regarded the Hamiltonian of a heavy nucleus (which is very complex due to the complexity of interactions between various nucleons) as behaving like a random matrix chosen from Gaussian orthogonal ensemble (GOE) having a probability density P . Energy levels were approximated by the eigenvalues of this matrix, and their statistics were studied [40]. The functional form of P defines the type of ensemble. Later this theory was successfully applied in the study of spectra of different complex systems, including disordered systems, quantum-chaotic systems, spectra of large complex atoms, etc. [41]. RMT is also shown to be of great use while understanding the mathematical structure of the empirical cross-correlation matrices appearing in the study of multivariate time series. The classical complex systems where RMT has been successfully applied are stock market [42]; brain [43]; patterns of atmospheric variability [44], physiology and DNA-binding proteins [45] etc. Previous studies elucidate that different model networks, namely scale-free, small world, random networks, and modular networks, ensue universal GOE statistics of RMT.

1.9.1 Tools of RMT used in networks' investigations

The random matrix studies of eigenvalue spectra consider two properties: (1) global properties such as the spectral distribution of eigenvalues $\rho(\lambda)$, and (2) local properties such as eigenvalue fluctuations around $\rho(\lambda)$. Eigenvalue fluctuations are the

most popular technique in RMT and are generally obtained from the spacing distribution of eigenvalues. We denote the eigenvalues of a network by λ_i , $i = 1, \dots, N$ and $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_N$. In order to get universal properties of the fluctuations of eigenvalues, it is customary in RMT to unfold the eigenvalues by a transformation $\bar{\lambda}_i = \bar{N}(\lambda_i)$, where \bar{N} is average integrated eigenvalue density [40]. Since we do not have any analytical form for N , we numerically unfold the spectrum by polynomial curve fitting [40]. After unfolding, average spacings are unity, independent of the system. Using the unfolded spectra, spacings are calculated as $s^{(i)} = \bar{\lambda}_{i+1} - \bar{\lambda}_i$.

1.9.1.1 Nearest neighbor spacing distribution

In the case of GOE statistics, the nearest neighbor spacing distribution (NNSD) is denoted by

$$P(s_1) = \frac{\pi}{2} s_1 \exp\left(-\frac{\pi s_1^2}{4}\right). \quad (1.7)$$

For intermediate cases, the spacing distribution is described by Brody parameter [46].

$$P_\beta(s_1) = A s_1^\beta \exp\left(-\alpha s_1^{\beta+1}\right) \quad (1.8a)$$

where A and α are determined by the parameter β as follows:

$$A = (1 + \beta)\alpha, \quad \alpha = \left[\Gamma\left(\frac{\beta + 2}{\beta + 1}\right)\right]^{\beta+1} \quad (1.8b)$$

This is a semi-empirical formula characterized by parameter β . As β goes from 0 to 1, the Brody distribution smoothly changes from Poisson to GOE. We fit 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, GOE, or the intermediate of these two [46]. As connections are rewired randomly with a probability, say p_r , thereby increasing the randomness in the network, the β value increases. At rewiring probability $p = 1$, the regular lattice becomes a random network, and the NNSD follows Gaussian Orthogonal Ensemble (GOE) statistics with a value of β being one. The transition from Poisson to the GOE statistics happens at a very small value of the rewiring probability p_r and precisely at the onset of the small-world transition β attains a

value one, demonstrating that nearest neighbor eigenvalues are correlated [47].

1.9.1.2 next-Nearest neighbor spacing distribution

Apart from NNSD, the next nearest-neighbor spacing distribution (nNNSD) is also used to characterize the statistics of eigenvalue fluctuations. We calculate this distribution $P(s_2)$ of next nearest-neighbor spacing,

$$s_2(i) = (\lambda_{i+2} - \lambda_i)/2 \quad (1.9)$$

between the unfolded eigenvalues. Factor of two at the denominator is inserted to make the average of next nearest-neighbor spacing $s_2(i)$ unity. According to Ref. [40], the nNNSD of GOE matrices is identical to the NNSD of Gaussian symplectic ensemble (GSE) matrices, i.e.,

$$P(s_2) = \frac{2^{18}}{3^6 \pi^3} s_2^4 \exp\left(-\frac{64}{9\pi} s_2^2\right) \quad (1.10)$$

The NNSD and nNNSD reflect only local correlations among the eigenvalues. The spectral rigidity, measured by the Δ_3 -statistics of RMT, preserves information about the long-range correlations among eigenvalues and is a more sensitive test for studying RMT properties of the matrix under investigation. In the following, we describe the procedure to calculate this quantity.

1.9.1.3 Δ_3 statistics

The Δ_3 -statistics measures the least-square deviation of the spectral staircase function representing average integrated eigenvalue density $\bar{N}(\lambda)$ from the best fitted straight line for a finite interval of length L of the spectrum given by

$$\Delta_3(L; x) = \frac{1}{L} \min_{a,b} \int_x^{x+L} [N(\bar{\lambda}) - a\bar{\lambda} - b]^2 d\bar{\lambda} \quad (1.11)$$

where a and b are regression coefficients obtained after least square fit. Average over several choices of x gives the spectral rigidity $\Delta_3(L)$. For GOE case, $\Delta_3(L)$ depends logarithmically on L , i.e.

$$\Delta_3(L) \sim \frac{1}{\pi^2} \ln L. \quad (1.12)$$

1.9.2 Randomness: structural and spectral perspective

While studying the transition from a regular 1-d lattice to a random graph, starting from a ring lattice with n vertices and k edges per vertex, each edge is rewired at random with a certain probability, p . This construction allows to ‘tune’ the graph between regularity ($p = 0$) and disorder ($p = 1$). This rewiring procedure generates a network with some random connections, without altering the number of vertices or edges. For $p = 0$, the structure of the regular lattice or k -nearest neighbor coupled network remains the same; on the other hand, for $p = 1$, the regular lattice becomes a random network. For intermediate values of p , the graph is a small-world network. For the regular lattice ($p = 0$), NNSD follows Poisson statistics, while for $p = 1$, it follows GOE statistics, and for ($0 < p < 1$) it shows statistics intermediate of Poisson and GOE. Based on the comparison between the diameter and clustering coefficient of networks with the Brody parameter (β), a relation exists between small-world transition and GOE transition (a figure explaining this behavior is adopted from [47] with prior permission from the corresponding author). It has been shown that the NNSD changes from Poisson to GOE with a tiny increment in p , and the transition to GOE takes place exactly at the onset of small-world transition.

1.10 motivation behind this thesis

Despite a rapid advancement in the field of complex systems research in the past decades, which have uplifted our understanding of underlying complexities of systems surrounding us [18, 48, 49], unraveling the complexity in brain [50], human behavior [51], disease [52], socio-economic behavior [53], etc. still remains a challenge. All these studies were driven by the easy access of high computing facilities, availability of established techniques, and a large amount of data [54]; the past decade has witnessed a tremendous amount of work on the real-world networks [55]. As most of the works focus on analyzing various structural properties of the networks, with very few investigations on spectral properties. Moreover, those spectral studies focus on extremal eigenvalues [56] and eigenvectors [57]. In this

thesis, we have demonstrated how the network $\lambda = 0$ eigenvalues capture important properties of the underlying system individually and collectively. Further, the 0,1-embedded image of a network provided a new method to study networks that are beyond the purview of previous studies.

Recently, complex networks have also been analyzed in the RMT framework bringing them into the universality class of Gaussian orthogonal ensemble (GOE) [47, 58–60]. The universality means that universal spectral behaviors, such as statistics of nearest neighbor spacing distribution (NNSD), are not only confined to random matrices but get extended to other systems. A wide variety of complex systems fall under this class, i.e., their spectra follow GOE statistics ([41] and references therein).

In this thesis, we have shown how systematic investigations performed on the model and real-world networks establish a correlation between their structural properties and spectral properties, inspected by RMT.

1.11 Readers Guide to the thesis

We have divided our study of the network as 0,1-matrix systems or adjacency matrix into three chapters. In the first chapter, we have studied with RMT, how the networks degree-degree correlation information is stored in the spectrum of eigenvalues collectively. This information is revealed in three parts first in the short-range correlation study by the neighbor spacing distribution of the eigenvalues of the adjacency matrices in the form of Poisson to GOE distribution transition for very high degree-degree correlation with a little decrement. Second, in the long-range study by Δ_3 -statistics for rest positive correlation range. Where the correlation length increases with the decrease in the degree-degree correlation in the network and reaches a maximum value of correlation length in Δ_3 -statistics for $r = 0$. Finally, for all negative correlation in degree connections are captured in spectral density distribution plots. Here, we find that the typical density distribution plot gradually changes its shape to double-humped shape with a peak of $\lambda = 0$.

In the second chapter, here we were fascinated by the fact that in the spectrum

of real-world inspire networks, a large number of $\lambda = 0$ are observed, and we also found that negative degree-degree correlation also induces the peak at $\lambda = 0$. So, our study was here to look for more discreetly at the origins of the $\lambda = 0$ in the model and real-world networks. As easily can be guessed with the concept of rank in the linear algebra that the number of eigenvalues at $\lambda = 0$, say r , reduces the rank of $N \times N$ matrix to $N - r$. As configuration model networks with same degree-degree correlation do not have, and as high number of $\lambda = 0$ as a real-world biological network. Therefore, we found that for a network with no isolated node, it can shed light on the evolution and working of the networks, like gene duplication.

In the third chapter, we explored the adjacency matrix as the 0,1-embedded image of the network and analyzed the multifractal properties of this matrix plane image with multifractal tools. We found that our results were in agreement with all the previous results. These results provided another possibility for network study through 0,1-embedded images.

Finally, our studies have shown that the adjacency matrix provides a robust structure for network study. This structure is yet to reach its peak as eigenvalues apart from degenerate eigenvalues, and the largest eigenvalue is, however, to be explored individually and collectively. Further, the new possibility of $(0, 1)$ -embedded image structure to study of a network, where no study apart from multifractal nature is done. The beauty of the 0,1-embedded image is that it transformed the basic unit of structure to edges from nodes.

Network Structure and Eigenvalues Spectra of Networks

Chapter 2

2.1 Introduction

Last two decades have witnessed a rapid advancement in the field of complex networks [18, 19, 24]. The prime idea governing this framework is to consider a system made of interacting units. To categorize and understand real world systems based on interacting units, many models have been proposed, among which Erdős-Rényi (ER) random [9], scale-free (SF) [12] and small world [11] are the most popular ones. Further, degree-degree correlations have also been used as one of the key properties of networks characterization [19, 61–76] and is known to confer robustness to biological networks [77]. The tendency of (un)like degree nodes to stick together is termed as (dis)assortativity. Various social networks are known to be assortative, while few of the biological and technological networks have been reported to be disassortative [69–76]. Despite its importance for real networks, (dis)assortativity does not appear in any of the model networks discussed above and is driven by some other mechanism, for example, reshuffling algorithm [78, 79]. While spectral behavior of uncorrelated networks have been quite well-understood [80], despite real-world systems being highly correlated [66], such understanding for the correlated networks still needs to be developed.

Spectral graph theory is an established branch of mathematics, and eigenvalues of corresponding adjacency matrices are known as fingerprints of the underlying graphs [28, 36, 81, 82]. With recent advancement in the network theory, the spectral graph theory, traditionally used in investigations of random and regular graphs, got extended to studies of graphs motivated by real-world systems. These spectral studies, apart from presenting bounds for extremal eigenvalues, highlight their importance by relating them with the various structural as well as dynamical properties of the networks [83, 84]. The studies of networks further reveal a crucial impact of assortativity on the extremal eigenvalues [85], which has been explored in the context of disease spreading [86] and diffusion processes [87], thereby exhibiting the importance of spectral studies of networks for a more comprehensive understanding of complex systems. In this chapter, we present a systematic analysis of the impact of degree-degree correlations on the spectral properties of various networks under the random matrix theory (RMT) framework. Since its introduction in the 1960s, in the context of nuclear spectra, the theory has been successfully applied to a wide range of complex systems ranging from the quantum chaos to galaxy [39, 41]. Recently, with a spurt in the activities of a network framework, the RMT got its extension in the analysis of spectral properties of various model networks [58, 88] as well as those arising from real-world systems [89, 90].

2.2 Methods and Techniques

To quantify the degree-degree correlations of a network, we consider the Pearson (degree-degree) correlation coefficient, given as [64, 66]

$$r = \frac{[M^{-1} \sum_{i=1}^M j_i k_i] - [M^{-1} \sum_{i=1}^M \frac{1}{2}(j_i + k_i)]^2}{[M^{-1} \sum_{i=1}^M \frac{1}{2}(j_i^2 + k_i^2)] - [M^{-1} \sum_{i=1}^M \frac{1}{2}(j_i + k_i)]^2}, \quad (2.1)$$

where j_i, k_i are the degrees of nodes at both the ends of the i^{th} connection and M represents the total connections in the network.

The random network of size N and average degree $\langle k \rangle$ is constructed using the ER model by connecting each pair of nodes with the probability $p = \langle k \rangle / N$ [9]. These networks have an assortativity coefficient (r) being close to zero or exactly zero. To generate the networks with various assortativity, we use the reshuffling

algorithm [78, 79]. In this algorithm, after selecting two pairs of nodes randomly, we sort them according to the degree. The highest degree node is then connected to the second-highest degree node with the reshuffling probability p_r , which governs the (dis)assortative mixing, i.e., we reconnect a high degree node to a (low) high degree one and low degree node to a (high) low degree one. With the probability of $1 - p_r$, we rewire them randomly. If a new connection resulting from this rewiring already exists, it is discarded, and the previous steps are performed. The process is carried out until a constant value of r is attained. For assortative networks, the k degree nodes to form a complete graph with the value of r being one, the network should have at least $(k + 1 + 2n)$ nodes, where n can be any integer starting from 0. As this condition is not satisfied for all the degrees present in the network, the network takes a value lesser than one. Similarly, the disassortative network can have a value of r more than -1.0 .

We make a further note that at high assortativity values, all the similar degree nodes being connected among themselves form groups [78, 79]. As we decrease the assortativity, the connections within the groups of similar degree nodes decrease, and the connections between different groups of identical degree nodes increase. For disassortative networks, connections between different groups of similar degree nodes exist, giving rise to a bipartite-like structure [78, 79].

As the reshuffling algorithm used for changing the degree-degree Pearson correlation function for ER random network, only require to first select four unique nodes with only two connections. This selection of four unique nodes is chosen regardless of the degree distribution they are the part of, like poisson, Power-law distributions. Therefore, this method applies to any network, regardless of the degree distribution it follows. The SF networks of size N and average degree $\langle k \rangle$ are generated using the Barabási-Albert algorithm by starting with a completely connected network seed and adding new nodes one by one which connects with existing nodes using the preferential attachment method [18].

The networks are represented in the form of adjacency matrix by defining $A_{ij} = 1$, if i , and j nodes are connected otherwise $A_{ij} = 0$. For an undirected and un-

weighted network with N nodes, the adjacency matrix is $N \times N$ symmetric square matrix entailing all real eigenvalues. We denote the eigenvalues as $\lambda_i, i = 1, 2..N$ and $\lambda_i \leq \lambda_{i+1}$ and analyse them under the RMT framework. The random matrix studies consider two properties of spectra: (1) global properties such as the spectral distribution of the eigenvalues $\rho(\lambda)$, and (2) local properties such as eigenvalue fluctuations around $\bar{\lambda}$. In RMT, calculations of spectral fluctuations are done using the unfolded eigenvalues $\bar{\lambda}_i = \bar{N}(\lambda_i)$, where $\bar{N}(\lambda) = \int_{\lambda_{min}}^{\lambda} \rho(\lambda) d\lambda$ is the average integrated eigenvalue density [40]. By using these unfolded eigenvalues, nearest neighbour spacings are calculated as $s_i = \bar{\lambda}_{i+1} - \bar{\lambda}_i$. For symmetric random matrices with the mean zero and the variance one, the nearest neighbor spacing distribution (NNSD) follows GOE statistics given as:

$$P(s) = \frac{\pi}{2} s \exp\left(-\frac{\pi s^2}{4}\right), \quad (2.2)$$

This shows a level repulsion at small spacing values with an exponential fall for larger spacings indicating that nearest neighbor eigenvalues are correlated [40]. Whereas the spacing distribution of a matrix whose diagonal elements are Gaussian distributed random numbers and rest of the elements are zero exhibit Poisson statistics ($P(s) = \exp(-s)$) indicating that eigenvalues are uncorrelated [40].

The intermediate of these two distributions can be characterized using the Brody equation [46]:

$$P_{\beta}(s) = A s^{\beta} \exp(-\alpha s^{\beta+1}), \quad (2.3)$$

where A and α are determined by the parameter β as $A = (1 + \beta)\alpha$ and $\alpha = \left[\Gamma\left(\frac{\beta+2}{\beta+1}\right)\right]^{\beta+1}$. The value of Brody parameter lies in the range ($0 \leq \beta \leq 1$). The value of β being 0, indicates the Poisson distribution, where as $\beta = 1$ corresponds to the GOE distribution. Other values of β indicates that the distribution lies intermediate to these two.

The NNSD provides a correlation measure of subsequent eigenvalues, whereas the $\Delta_3(L)$ statistic measures how the eigenvalues which are L distance apart are correlated, and can be estimated using the least-square deviation of the spectral staircase function representing average integrated eigenvalue density $\bar{N}(\lambda)$ from the best fitted straight line for a finite interval of length L of the spectrum given by

[39]:

$$\Delta_3(L; x) = \frac{1}{L} \min_{a,b} \int_x^{x+L} [N(\bar{\lambda}) - a\bar{\lambda} - b]^2 d\bar{\lambda} \quad (2.4)$$

where a and b are regression coefficients obtained after least square fit. Average over several choices of x gives the spectral rigidity, the $\Delta_3(L)$. For the GOE statistics, the $\Delta_3(L)$ depends on L in the following manner:

$$\Delta_3(L) \sim \frac{1}{\pi^2} \ln L \quad (2.5)$$

For the network spectra considered in this chapter, there is no analytical form of \bar{N} , and we perform unfolding by numerical polynomial fitting using the smooth part of the spectra by discarding eigenvalues towards both the ends as well as degenerate eigenvalues, if any [39, 41]. This renders the dimension of the unfolded eigenvalues less than the size of the network.

2.3 Results

The bulk part of the spectra of ER random networks with r value being close to zero, follow the well known semi-circular law [30, 35] (Fig. 2.1(f)). The extremal eigenvalues deviate from the random matrix predictions and indeed provide various information about structural and dynamical properties of corresponding systems [68, 83, 84, 86, 91–96]. In the following, we present results about the impact of assortativity on the spectral properties of networks. It turns out that with an increase in the assortativity, the semicircular distribution, as observed for the uncorrelated ER random networks, remains unchanged (Fig. 2.1(a)-(e)). The largest eigenvalue exhibits an increasing trend, as already discussed in [85, 86]. As the network is rewired entailing disassortativity, spectral distribution ($\rho(\lambda)$) acquires a very different structure than those of the assortative networks. The networks start exhibiting a high degeneracy at zero, with overall spectra resembling a double-humped structure (Fig. 2.1(h)), which becomes more pronounced as the disassortativity becomes higher or the value of r becomes more negative (Fig. 2.1(i)). This increase in disassortativity is also accompanied by more number of degenerate eigenvalues at zero. There could be various reasons for this high degeneracy; few of them, appropriate in the present context, are: First, as discussed that disassortativity supports bipartite-

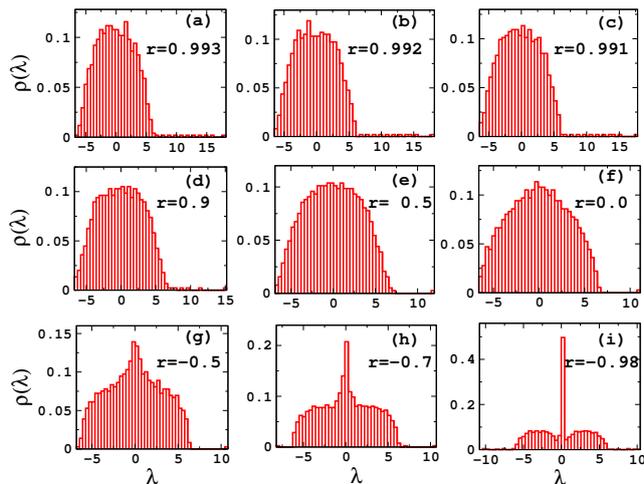


Figure 2.1: Spectral density for Erdős-Rényi random networks with different values of assortativity coefficient r . All graphs are plotted for the networks with size $N = 1000$ and connection probability $p = 0.01$, averaged over twenty different realizations of the networks.

like structure [78, 79] and a complete bipartite network has all zero eigenvalues except two. Hence bipartite-like behavior of the disassortative networks presents one of the reasons for the occurrence of high degeneracy at zero. Second, the tree-like structure has been demonstrated to yield degeneracy at zero eigenvalues [31], and disassortativity encourages tree-like structure [78, 79], which in turn indicates high degeneracy at zero.

We remark that for large N , the limiting shape of $\rho(\lambda)$ is known for various cases, which for sufficiently dense matrices, tend to follow the Wigner semicircular law typical for the Gaussian matrix ensembles [30, 35]. In contrast, an ensemble of sparse random matrices of finite size is known to yield states beyond the semicircular law in the tails of the distribution [97–99]. For sparse random graphs, i.e. matrices with 0 and 1 entries having smaller p values, while the density distribution $\rho(\lambda)$ of an ensemble exhibit singularities, with the height of the peaks being the corresponding multiplicities, the bulk is still shown to comply with random matrix predictions of Wigner’s semicircular law [100, 101]. Moreover, investigations of various model networks mimicking real-world properties have revealed that the spectra of these networks exhibit degeneracy at zero [40], as observed for the sparse random matrices. On that account, despite degeneracy at zero, the bulk of the as-

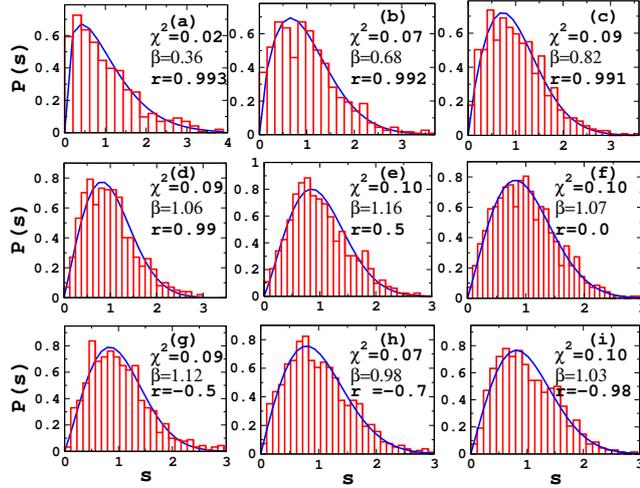


Figure 2.2: The NNSD for Erdős-Rényi random networks with different values of assortativity coefficient r . All graphs are plotted for the networks with size $N = 1000$ and connection probability $p = 0.01$. Histograms are from the data points, and the solid line is for fitting with Brody distribution (Eq. 2.3).

sortative networks following the semicircular distribution is not surprising.

As the spectral density only provides a global behavior of eigenvalues, to get insight into local fluctuations, we further analyze the short-range and long-range correlations in eigenvalues. The NNSD follows GOE statistics of RMT (Eq. 2.2) for all the values of r except for very high values corresponding to the highly assortative networks (Fig. 2.2). What is interesting that the values of r for which $\rho(\lambda)$ exhibits very similar behavior, except for a change in the value of the largest eigenvalue, the NNSD captures crucial structural changes reflected through the value of the Brody parameter. For the highest achievable amount of the assortativity coefficient for the particular network parameter for which results are presented, the value of β comes out to be close to 0.3 (Fig. 2.2(a)), and as the assortativity decreases we witness a smooth transition to the GOE statistics with value of the β turning one. Depending upon the network size, average degree, and degree sequences, the highest possible value of r for that network may be different (as discussed in Section II), which might lead to a different value of β . Fig. 2.2(a)-(d) depict that a tiny change in the value of r is capable of entailing a profound change in the statistics, in-fact it approaches from the Poisson to the GOE. Since microscopic randomness is known

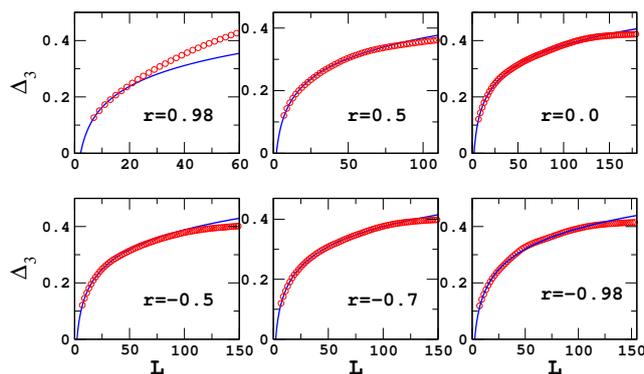


Figure 2.3: The $\Delta_3(L)$ statistic for Erdős-Rényi random networks with different values of assortativity coefficient r . All graphs are plotted for the networks with size $N = 1000$ and connection probability $p = 0.01$. The solid line is the prediction from GOE statistics (Eqs. 2.4 and 2.5) and open circle are calculated from the network.

to be enough in introducing the short-range correlation in eigenvalues [47], for a tiny deviation from the highest assortativity entails GOE statistics. Since the assortativity in a network supports the groups having similar degree nodes and as soon as assortativity, r , is decreased, these distinct groups of similar degree nodes for very high values of r get destroyed, leading to a transition from the Poisson to the GOE statistics. That is, as soon as the value of r is decreased, and sufficient random connections among the groups of similar degree nodes are induced, the value of Brody parameter β becomes one, and no further signature of structural changes on the value of β is found with a further decrease in the assortativity.

For disassortative networks which are characterized with negative values of r , what is remarkable is that despite these networks displaying distinguishable spectral distributions than those of the assortative networks, the NNSD yields the value of the Brody parameter ($\beta = 1$) bringing them into the universality class of GOE. This is not surprising as NNSD is analyzed by taking the non-degenerate part of the spectra, and high degeneracy at a particular value, for instance at zero, does not account for any effect in the NNSD. As long as the underlying network has some random connections, the NNSD displays the GOE statistics [47]. We remark that all the networks considered here form a single connected cluster as for disconnected networks, even though each sub-network follows GOE statistics, the spectra were

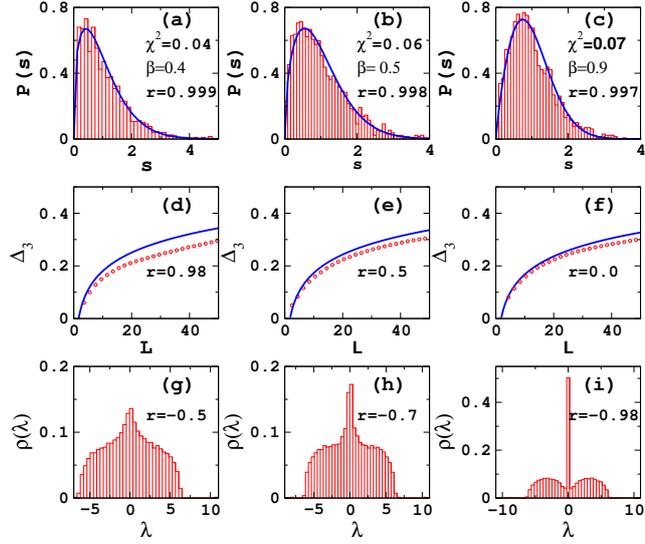


Figure 2.4: (a)-(c) represent the NNSD, (d)-(f) present the $\Delta_3(L)$ statistic and (g)-(i) depict the spectral density distribution of the networks ER random networks. All graphs are plotted for the networks with size $N = 2000$, $\langle k \rangle = 10$ and for average over twenty different realizations of the network.

taken together may lead to a different spacing statistics [88].

To get further insight into the structural changes arising due to the changes in r values, we probe for the long-range correlations in eigenvalues for those sets which yield the β value one. We find that for all these values of r , the long range correlations, measured using the $\Delta_3(L)$ statistic (Eq. 2.4) follow the universal GOE statistics as given by Eq. 2.5 for a certain value of L (denoted as L_0) and deviates from this universality afterwards (Fig. 2.3).

Note that a regular network, for instance, 1-d lattice with a periodic boundary condition, follows a Poisson distribution. As connections are rewired, thereby increasing the randomness in the network. Therefore, the value of the Brody parameter rises with an increase in the rewiring probability. It becomes one at the onset of the small-world transition, demonstrating that nearest neighbor eigenvalues are correlated [47]. For such a small change in the network structure, there is no visible change in the density distribution, but the Brody distribution detects even such a slight change in the number of random connections and hence has been proposed to be used as a measure of randomness at a fine-scale [47]. After the Brody parameter attains a value one, the $\Delta_3(L)$ the statistic has been shown to measure

the randomness (in terms of L_0 in this chapter), in the underlying network [102]. As rewiring probability increases further, the value of L_0 for which $\Delta_3(L)$ statistic follows RMT predictions increases, demonstrating that the eigenvalues are L_0 distance apart are also correlated. Since L_0 provides a measure of randomness in a network [102], for the networks under investigation in the present work, it turns out that the highest assortative network is least random, as the value of L_0 is least for that particular r value (Fig. 2.3(a)). As the assortativity of the network is decreased, the randomness of the network increases, reflected in the higher value of L_0 . This increase in the size of L_0 continues up to r being zero, supporting the fact that the network reaches to the maximum randomness. The value of L_0 then remains steady for a further decrease in the value of assortativity to the minimum possible value of r , i.e., to the maximum disassortativity (Fig. 2.3(c)). As most of the real-world networks have been reported to possess a certain level of disassortativity [66], based on the $\Delta_3(L)$ results, we can argue that real-world systems attempt to have more randomness, thereby leading to being disassortative. What follows that as the value of r decreases, by keeping network size and average degree the same, the value of L_0 for which $\Delta_3(L)$ statistic follows RMT predictions increases, indicating an increased amount of randomness in the underlying network. Fig. 2.4 demonstrates that the behavior of various spectral properties remains unchanged as network size increases. Figs. 2.4(a)-(c) indicate that the value of Brody parameter β becomes one with a very small decrease in the value of r . With a further decrease in the value of r , the value of L_0 for which the $\Delta_3(L)$ statistic follows GOE statistic increases, indicating an increase in the randomness as discussed earlier. With a further decrease in the value of r in the disassortativity regime, there occurs a peak at zero eigenvalues, which becomes more pronounced as the network becomes more dissociative, which is also accompanied by the deviation from the semicircular distribution at a meager value of r .

Further, to demonstrate the robustness of the universal RMT predictions against changes in the network architecture, we present results for the SF networks for various values of r . For r being close to zero, the density distribution of SF networks

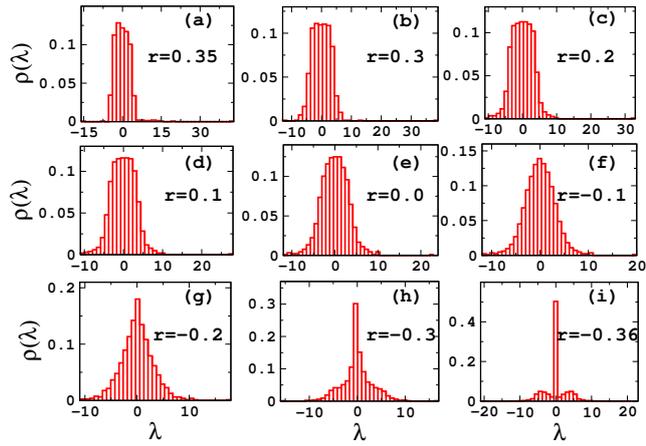


Figure 2.5: Spectral density for scale-free networks with different values of assortativity coefficient r . All graphs are plotted for the networks with size $N = 1000$, $\langle k \rangle = 10$, for twenty different realizations.

exhibits the triangular shape [35], which, with an increase in the assortativity, tends to display flattening of the peak. The range of the distribution also shrinks as the assortativity increases (Fig. 2.5(a)-(e)). On the other hand, as we decrease assortativity, i.e., make the network more disassortative, the shape of density distribution starts changing from its signature triangular distribution, with the peak at zero eigenvalues being more pronounced (Fig. 2.5(f)-(g)). As we further increase the disassortativity, the eigenvalues distribute themselves symmetrically and adopt a double-hump shape for highly disassortative networks, clearly visible in Fig. 2.5(i) which is accompanied by a high peak at the zero eigenvalues similar to that of the ER random networks. It is noteworthy that for highly disassortative networks, the spectral density of ER and SF model networks behave similarly, deviating from their respective signature distributions. Further, the β value exhibits a transition from the Poisson to the GOE statistics with a decrease in the r value. Despite the overall spectral density being different from that of the ER networks, the NNSD and $\Delta_3(L)$ statistic display similarity in behaviour, which is in line of the argument that the eigenvalues fluctuations are calculated from the smooth, homogeneous part of the spectra by not taking degeneracy into account, and density is not known to be a real test of GOE statistics [103].

We would like to remark here on the impact and reliability of network size con-

Table 2.1: Comparison of number of zero eigenvalues of PPI networks of different species and their corresponding configuration models. r_0 denotes the value of the assortativity coefficient for the PPI networks. $N_0(PPI)$ denotes the number of zero eigenvalues in the spectra of the PPI networks. $N_0(r = 0)$ stands for degeneracy at zero for configuration model with $r = 0$, whereas $N_0(r = r_0)$ denotes the same for the configuration models taking r values equal to the corresponding PPI network.

PPI networks	N	r_0	$N_0(PPI)$	$N_0(r = 0)$	$N_0(r = r_0)$
<i>H.pylori</i>	709	-0.243	317	115	152
<i>C.elegans</i>	2386	-0.183	1354	465	1124
<i>S.cerevisiae</i>	5019	-0.088	976	717	1149
<i>H.sapiens</i>	2138	-0.084	864	423	643
<i>D.melanogaster</i>	7321	-0.083	2311	1389	1975
<i>E.coli</i>	2209	-0.012	487	487	497

sidered in the present investigation. In RMT, different quantities are calculated by averaging an ensemble of matrices. However, for real systems, calculations are made as running averages over part of the whole spectrum. The random matrix predictions can be applied to real-world systems if the above two are equivalent, a property known as ergodicity. More explicitly, it means that all members of the ensemble, except for a set of measure zero, satisfies the above equivalence [104, 105]. Due to the ergodicity, one can construct matrix ensembles in different ways: (a) large dimensional random matrices with less number of realizations or (b) smaller dimensional matrices with a large number of realizations. We consider an ensemble of twenty network realizations with a large dimension, which is already shown to be good enough to study various structural properties of networks, such as degree distributions, clustering coefficients, etc. [11]. Moreover, individual entities of each ensemble follow RMT predictions for NNSD with a reasonable accuracy, characterized by χ^2 values. As we increase the realizations, accuracy increases (Fig. 2.2 and Appendix). Consideration of an ensemble consisting of many more number of network realizations would not lead to significant betterment or difference in the following properties of the network spectra: (1) the Brody parameter smoothly turning one with a decrease in the value of r at a very fine scale; (2) a further reduction in the amounts of r leading to an increase in the value of L for which spectra follow

GOE statistics; and (3) increasing height of the peak at zero eigenvalues with an increase in the disassortativity, owing to the bipartite-like structure of the network.

Next, to investigate if the degree-degree correlations in a real-world system have different spectral behavior than those of the model networks discussed above, we consider the protein-protein interaction (PPI) networks of six different species. These networks have already been shown to follow universal RMT predictions of GOE statistics [106]. We concentrate here on the occurrence of high degeneracy at zero. The assortativity coefficient and fraction of degenerate eigenvalues are tabulated in Table I. As all the PPI networks possess a negative value of r as well as have a high degeneracy at zero, we expect disassortativity to be one of the factors governing the degeneracy in the real-world networks. To probe more into the correlation between disassortativity and degeneracy at zero, we compare the corresponding configuration model for all PPI networks presented above (Table I). It is clearly indicated that as soon as the value of r takes a negative value (close to the corresponding PPI network) while keeping all other parameters of the system the same, there is an increase in the degeneracy at zero eigenvalues.

2.4 Discussion and Conclusions

The density distribution of the random networks for r value being zero follows the Wigner semicircular distribution. Even with change in the assortativity ($0 \leq r < 1$), the bulk part of the spectra keeps displaying semi-circular distribution (Fig. 2.1(a)-(f)), Whereas an increase in the disassortativity ($-1 \leq r < 0$) leads to the double hump, which is symmetrically distributed around a peak at zero eigenvalues (Fig. 2.1(i)). The height of the peak increases with the increase in the disassortativity of the network.

The NNSD of the networks with the various (dis)assortativity values ($1 < r < -1$), reveal that there is a smooth transition in the β -value around the very high assortativity regime. For very high assortativity values, β values lie close to zero, and as the network becomes less assortative β progresses to one. It might be due to the reason that the networks with the highest assortativity have groups of similar degree

nodes that get perturbed as r decreases by making random connections among these different groups. For the rest of the assortativity values, the β remains fixed at 1, which corresponds to the universal GOE distribution as r value goes to a negative end.

Further, the property of the Brody parameter being able to detect changes in network structure at a fine scale and the increase in L_0 of $\Delta_3(L)$ statistic after attainment of β value one have several implications. One of them that concerns the present work is that the value of β distinguishes two networks based on random connections present, while the other is that more assortativity in the network corresponds to less randomness. Decreasing the assortativity leads to an increase in randomness, which continues up to the value of $r = 0$, for which the network is most random (L_0 value being maximum). Then the value of L for which $\Delta_3(L)$ statistic follows GOE prediction starts decreasing and remains steady for a further decrease in the amount of assortativity up to the minimum possible value of r (i.e., up to the maximum disassortativity case). The SF networks also exhibit the similar statistics of eigenvalue fluctuations as for ER random networks, where density distribution for $r = 0$ and for lower $|r|$ values show triangular distribution instead of semicircular. Both the networks, however, exhibit high degeneracy at zero for the disassortative networks. By considering different PPI networks, we further demonstrate the role of disassortativity governing the appearance of degeneracy at zero eigenvalues.

In spectral graph theory, most of the works concentrate on extremal eigenvalues [36]. In contrast, the RMT research focuses on the distribution of various spectral properties of random matrices with an extension to the random graphs, mainly ignoring many graphs properties existing in real-world systems. The analysis carried here is a step towards bridging this gap by considering the two most popular tools of random matrix theory, i.e., density and spacing distributions, to understand the impact of one of the essential properties of graphs, i.e., assortativity. This property has been increasingly realized as a characteristic of a system [107–109]. Our analysis is another demonstration of the importance of spacing analysis in understanding

impact of degree-degree correlation on the network detected through the spectra as for very minute changes in r , there are no visible changes in the spectral density, but this leads to a very drastic changes in the eigenvalues fluctuations demonstrating the impact of r values on randomness in a network.

Furthermore, $\Delta_3(L)$ statistic provides an insight into the reason why social networks tend to be assortative, while biological and technological networks tend to be disassortative. As randomness, measured in terms of L_0 for which the $\Delta_3(L)$ -statistic follows RMT prediction, increases with a decrease in r . A direct implication of this result can be witnessed in case of social networks where entities are known to be associated in an ordered fashion (people with similar age or educational profile are often more connected) [110], thus providing a probable reason as to why social networks tend to assume an assortative topology. On the other hand, it has been reported that most of the biological and technological networks possess a certain level of disassortativity [64, 66, 76]. Also, the biological networks, for instance, the PPI networks, exhibit varying amounts of randomness in their underlying networks detected through different values of L_0 for which the $\Delta_3(L)$ statistic follows GOE statistics [106]. This randomness has been attributed to mutations occurring in the course of evolution [111]. Relating the disassortative nature of the PPI networks and the randomness they possess, with the results obtained from our analysis of the model networks, suggests that biological networks tend to become more disassortative to comply with their underlying randomness.

To conclude, we present a systematic analysis of the spectral properties of the networks with varying (dis)assortativity. We find that assortativity has a profound impact on the spectral properties of the underlying networks. At a very high assortativity regime, even with a slight decrease in the value of r , the Brody parameter smoothly turns one. A further decrease in the values of r leads to an increase in the value of L_0 of $\Delta_3(L)$ -statistic for which the spectra follows GOE statistics. As Brody parameter β captures the changes in assortativity coefficient at a fine-scale [47] and L_0 at large scale [102], which further suggest that when r decreases, randomness increases. With a further decrease in r , at around $r = 0$, the density

distribution starts exhibiting a peak at zero eigenvalues, which becomes more pronounced as r decreases further. Interestingly, most of the studies on network spectra report that the bulk part of the spectra of the networks having Gaussian and scale-free degree distribution follows semicircular and triangular distributions [30, 35] respectively. Still, for highly disassortative networks, the spectral density of both the degree distributions can have entirely different behavior.

Recently, the realm of assortativity has been realized in understanding adaptive synchronization [108], which, combined with our results of the varying amount of randomness for various values of r can be explored further to understand dynamical processes on networks. Further, Table I indicates that disassortativity is one of the factors contributing to degeneracy at zero. The prevalence of zero degeneracy has been implicated in terms of gene duplication [112]. This, along with the impact of change in the topology of a network, brought upon by assortativity, leading to a profound shift in the spectral density, provides a direction to explore the evolutionary origin of real-world systems [113, 114]. Lastly, since randomness or random connections in a network have already been emphasized for the proper functioning of corresponding systems [115], the profound role of assortativity, parameter revealed through the sophisticated random matrix the technique is not only crucial for network community attempting to model complex systems but is interesting for random matrix communities at the fundamental level as well.

2.5 Appendix

Numerical calculations about assortative mixing, eigenvalues calculations, and $\Delta_3(L)$ statistic are done using FORTRAN code written by the Authors. The eigenvalues are calculated by calling LAPACK (Linear Algebra PACKage) subroutines into the FORTRAN code. The calculation of spacings and polynomial fittings are done using MATLAB.

We present the χ^2 values as a measure of goodness of fit of the model to data, a lower of χ^2 indicating a better fitting. As depicted from Fig. 2.6, the χ^2 values consistently decrease with an increase in the number of network realizations in the

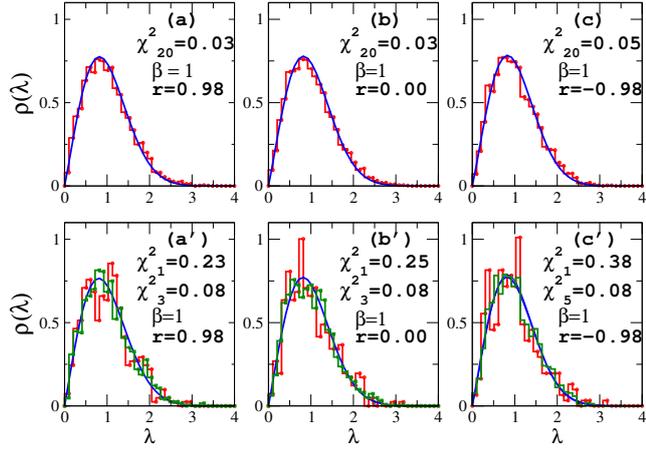


Figure 2.6: (a), (b) and (c) plot average NNSD for an ensemble of twenty realizations for different values of r , whereas (a'), (b') and (c') plot the ensemble having different number of the network realizations. The histogram is drawn using the data from the networks, and the solid line is the fitted Brody distribution. For all the graphs $N = 1000$ and $\langle k \rangle = 10$.

ensemble, implicating an increase in the accuracy reaching to the amount of χ^2 being less than one lying in the acceptable range [116]. For assortative networks, as less as three realizations in the individual ensemble are good enough to bring χ^2 within the acceptable range, whereas for r taking negative values, the number of realizations in the ensemble increases little bit more (five as depicted in Fig. 2.6(c')) in order to bring χ^2 within the acceptable range. This happens as for disassortative networks, there is high degeneracy at zero eigenvalues leading to the less effective dimension of the unfolded spectra (refer discussions in the Methods and Techniques section), and hence several realizations of the networks are required.

Network Structure and Zero-Eigenvalue of Networks

Chapter 3

quotation Network theory provides a framework to understand the dynamical behavior and evolution of complex systems. Particularly, spectral graph theory has been very successful in predicting the synchronization and diffusion properties of dynamical units in underlying networks. Hence, a considerable amount of work is done in recent years on the spectra of the model and real-world networks. These analyses have indicated that while the bulk of the spectra match closely to that of the corresponding random network models, few spectral properties differ considerably. In this work, we explore the origin and implications of one such property, which is degeneracy at zero eigenvalues in the spectra. quotation

3.1 Introduction

The last two decades have witnessed tremendous growth in the studies of complex systems under the graph theory framework [117]. This framework describing a complex system in terms of its interacting units has not only enabled us to understand the properties of large complex systems but has also shed light on the dynamics of evolution of these structural properties [18, 19, 68, 118]. Though spectral graph theory is a well-established domain [28, 36, 82], most of the studies in complex systems research pertain to understanding the properties and behavior of

a system by analysis of various structural measures. In contrast, studies on spectral analyses of graphs generated for real-world systems are comparatively limited [30, 90, 106, 119]. The spectral investigations indicate that the patterns of density distributions are distinguishing features of different classes of model networks [32]. Further, extremal eigenvalues have been shown to contain useful information about the structure of the graphs [68]. Although the bulk of real-world networks bear reasonable similarities with the model networks [28], some properties differ significantly. One such property is the degeneracy at zero eigenvalues. Almost all biological and technological networks exhibit high degeneracy at the zero eigenvalues [31, 35, 106], gene duplication is one of the suggested reasons behind the occurrence of high degeneracy at zero eigenvalues in biological systems [37]. Many biological systems are known to follow gene duplication as the primary mechanism behind their growth [112]. From a straightforward matrix algebra calculation, we know that node duplication leads to a lowering of the rank of the corresponding matrix, hence contributing one additional zero eigenvalues in the spectra. Although node duplication provides a clue to the origin of zero degeneracy [120], it fails to provide a quantitative measure of actual degeneracy observed in real-world networks [31], indicating the contribution from other factors. Scale-free behavior or sparseness of real-world networks have been argued out to be other reasons responsible for degeneracy at the zero eigenvalues [31, 35, 106]. In this work, we explore the origin of zero eigenvalues in various model networks. We substantiate the results by considering various real-world networks.

3.2 Methods and Techniques

A network can be represented in terms of adjacency matrix which is defined as,

$$A_{ij} = \begin{cases} 1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases} .$$

The eigenvalues of the adjacency matrix are denoted by λ_i , $i = 1, 2, \dots, N$ such that $\lambda_1 < \lambda_2 < \lambda_3 < \dots < \lambda_N$. A theorem [121] relating the degeneracy at zero eigenvalues with the properties of the matrix states that for an adjacency matrix

Table 3.1: In order to demonstrate one to one relation between number of duplicates and zero eigenvalues, each time one new add is added in a network of size N in such a manner it satisfies either complete duplication criteria (ii) or the partial duplication criteria (iii). D_c represents number of complete duplicate node groups, D_p represents number of partial duplicate node groups and λ_0 indicates number of zero eigenvalues. Seed network of size $N = 100$ and average degree $\langle k \rangle = 10$.

N	100	101	102	102	103	104	104	105	106	106	107	108	108	109	110	110
D_c	0	1	2	0	3	4	0	5	6	0	7	8	0	9	10	0
D_p	0	0	0	1	0	0	2	0	0	3	0	0	4	0	0	5
λ_0	0	1	2	1	3	4	2	5	6	3	7	8	4	9	10	5
Condition	-	a	a	b	a	a	b	a	a	b	a	a	b	a	a	b

of size N and rank r , there will be exactly $N - r$ zero eigenvalues. Therefore, if we know the rank of an adjacency matrix, we can find out the degenerate zero eigenvalues. Factors responsible for lowering of the rank of an adjacency matrix are enlisted in the following:

(a) When two rows (columns) have the same entries, it is termed as complete row (column) duplication:

$$R_1 = R_2 \quad (3.1)$$

Subtracting one such row from the other yields one of the rows to attain all zero values, thus reducing the rank of the matrix by one.

(b) When two or more rows (columns) added together have exactly same entries as some other row or column, we call it partial row (column) duplication. For example;

$$R_1 = R_2 + R_3, \text{ or } R_1 + R_2 = R_3 + R_4 + R_5 \quad (3.2)$$

(c) An isolated node in the network leads to all zero entries in the corresponding row and column, thus lowering the rank of the matrix by one.

Conditions (a) and (b) lead to the linear dependence of row (column), reducing the rank of the matrix. Note that we consider a connected network in order to rule out the trivial possibility (c) of occurrence of zero eigenvalues. Further there are $N(N - 1)/2$ possible ways in which condition (a) of complete duplication can be realized, while for the partial duplication (b) among ' x ' number of nodes with ' y ' number of nodes, there can be $\frac{1}{2} \frac{N!}{(N-x-y)!x!y!}$ possibilities. Hence, for a given net-

work, checking the existence of condition (b) becomes computationally exhaustive as with increase in network size the number of possibilities becomes very large.

- **The possible number of ways in which duplication can happen.** We look for the duplication of x nodes by y nodes in an adjacency matrix of network of N nodes. Here, we have to look for all the possibilities of corresponding to x and y , where $1 < x, y < (N - 1)$.

Now, we can choose x nodes out of N nodes in $\binom{N}{x}$ number of ways.

Similarly, we can choose y nodes out of these $(N - x)$ nodes in $\binom{N-x}{y}$ number of ways. Therefore, total number of ways in which we can select these (x, y)

pair is

$$\binom{N}{x} \binom{N-x}{y} = \frac{N!}{(N-x-y)!x!y!}$$

As the number of nodes involved are very few, we can neglect $x!$ and $y!$ being order of ~ 1 and, Since all these selections are symmetric i.e. $AB = BA$, therefore in total

$$\sum_{x,y=1}^{N-1} \frac{1}{2} \frac{N!}{(N-x-y)!}$$

$2 \leq (x + y) \leq N$

Since N is very large ($\sim 10^3$) quantity, finding partial duplicate nodes is computationally exhaustive process.

- **The probability of two nodes of k degree being duplicated nodes.** To explain this, we consider number of ways in which two k -degree nodes can select their neighbours excluding themselves is $\frac{[(N-2)(N-3)\dots(N-3-k)]^2}{k!} = \binom{N-2}{k}^2$

and number of ways in which these two k -degree nodes select same neighbours

$$\binom{N-2}{k} \binom{k}{k} = \frac{[(N-2)(N-3)\dots(N-3-k)]}{k!}$$

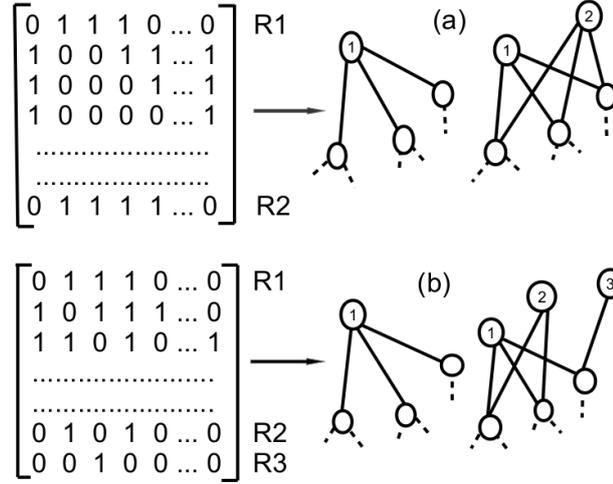


Figure 3.1: Schematic diagram representing (a) complete node duplication (Eq. 3.1) and (b) partial node duplication (Eq. 3.2) in networks.

Therefore, for two nodes of k -degree to be duplicated is given by

$$\frac{\binom{N-2}{k} \binom{k}{k}}{\binom{N-2}{k} \binom{N-2}{k}} = \frac{k!}{(N-2)(N-3)\dots(N-3-k)}$$

If $k \ll N$, then $(N-2) \approx N$; $(N-3) \approx N$; \dots ; $(N-3-k) \approx N$

$$\frac{\binom{N-2}{k} \binom{k}{k}}{\binom{N-2}{k} \binom{N-2}{k}} = \frac{k!}{N^k}$$

Hence, the probability of two nodes of k degree being duplicated nodes is $\frac{k!}{N^k}$.

3.3 Results

In order to demonstrate the effect of duplication on zero degeneracy, we construct an Erdős-Renýi (ER) random network for size N and connection probability p using ER model [18], such that it has no duplicates and no zero eigenvalues (row 1 of Table 3.1). Next, we add a node to the existing network in a way that satisfies the complete node duplication criteria, i.e., condition (a). This leads to precisely one zero eigenvalue corresponding to one duplicate node. The addition of one more node mimicking the previous node leads to two zero eigenvalues (row 2 and 3 of Table 3.1). This demonstrates how complete node duplication leads to zero eigen-

values (Fig. 3.1 (a)). Further, we consider another situation where we devise our algorithm such that two new nodes are added to the existing random network in a way that in coalition they mimic the neighbors of an existing node (condition (b)), i.e., they duplicate an existing node (row 4 of Table 3.1) as demonstrated in Fig. 3.1 (b). The impact of duplications (by both conditions (a) and (b)) on the zero eigenvalues are presented in the subsequent rows of Table 3.1. Thus, we observe that with the entry of every new node in a network, satisfying condition (a) or (b) of complete or partial duplication, there is an addition of exactly one zero eigenvalues in the spectra. The number of duplicates (complete or partial) equals the number of zero eigenvalues. The density distribution at a low average degree yields a peak at zero eigenvalues. With an increase in $\langle k \rangle$, the peak of the density distribution flattens (Fig. 3.2 (a)).

To study the impact of network architecture on the duplication phenomenon, we present results for an ensemble average of the scale-free (SF) networks as they are known to have high degeneracy at zero eigenvalues. We generate the SF network using the preferential attachment mechanism [12], where each new node gets attached to the existing nodes with the probability proportional to their respective degrees. This phenomenon gives rise to a power-law degree distribution. Here at each time step, a new node enters, which is most likely to connect with the highest degree nodes owing to the preferential attachment algorithm. The next entry also tends to attach with the highest degree nodes. From the power-law degree distribution of SF networks, it is evident that there are very few high degree nodes, which are known as the hubs of the network and a large number of low degree nodes. At low values of $\langle k \rangle$, there is a high degeneracy at zero eigenvalues indicating high duplication. This is because, at a low average degree, most of the low degree nodes attain very few connections. Under preferential attachment property, these low degree nodes have the highest probability to connect with the hubs of the network, which increases the likelihood of any two nodes to have the same neighbors, leading to a pair of duplicate nodes. Although even with an increase in average degree, the density distribution remains triangular, there is flattening of the peak (Fig. 3.2 (b)). This might

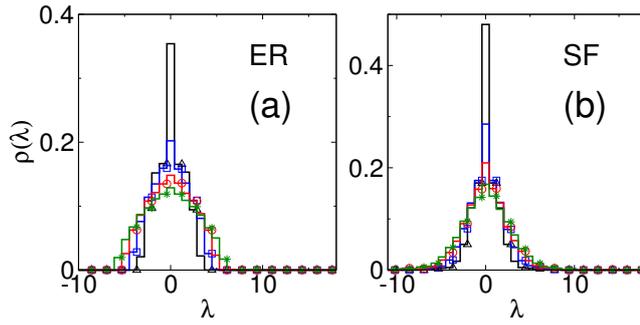


Figure 3.2: The density distribution of Erdős-Renýi (ER) random networks and scale-free (SF) networks for different average degrees and $N = 1000$. Δ , \square , \circ and $*$ represent the data points of density distribution for $\langle k \rangle = 2, 4, 6$ and 8 , respectively. All values are averaged over ten realizations of the networks.

be because the low degree nodes also tend to acquire connections with nodes other than the hubs. All these findings indicate that a low average degree favors duplication. The explanation behind this can be given in terms of the possible number of ways of duplication, which is the ratio of the possible number of combinations of duplication possessed by two k -degree nodes to the possible number of combinations of random connections of those nodes. This is given as $\frac{k!}{N^k}$, Where N is the total number of nodes. As k increases, the possible number of ways of duplication drastically decreases, thus explaining why low degree supports duplication.

Since networks with power-law degree distribution, i.e., SF networks, lead to high degeneracy in the spectra, we further explore other model networks with power-law degree sequence. We construct the configuration model network by taking the degree sequence of the connected SF network as input. Each node of the corresponding configuration model is allotted stubs equal to their degree; then, these stubs are paired with uniform probability [122–124]. This generates a configuration model for a given degree sequence. Only connected networks are carried-further, rest all are discarded. In spite of having randomly assigned connections, they display a much higher zero degeneracy as compared to the ER random networks (Fig. 3.3 (a)). The configuration model networks are not generated using the preferential attachment property as of the SF networks. So this possibility is ruled out as a reason behind the higher degeneracy of configuration model networks as

compared to ER random networks. The particular (the power-law) degree sequence emerges as a probable reason behind high degeneracy at zero eigenvalues in the configuration model. Due to the power-law behavior, there exists a large proportion of nodes with a low degree, which for $\langle k \rangle = 2$ are peripheral nodes. Only a few nodes had high degree act as the hubs. A large number of low degree nodes get randomly attached to high degree nodes, i.e., the peripheral nodes attach with the hubs only, leading to the complete duplication of these nodes. While in the case of ER networks, duplication is less likely as all the nodes have their degrees are fluctuating around the average degree. The configuration model networks exhibit a lower peak at zero eigenvalues as compared to SF networks (Fig. 3.3) as it is a randomized version of the SF network. This indicates that apart from the preferential attachment phenomenon, the particular degree sequence is also responsible for high degeneracy at zero eigenvalues.

Table 3.2: Properties of the six PPI networks and their comparison with ER random networks, SF networks generated using BA algorithm and corresponding configuration model networks (of the same degree sequence as of the PPI networks). D_c , D_c^{ER} , D_c^{BA} , and D_c^{conf} denote the number of complete duplicates in the PPI networks, ER random networks, SF networks and configuration model networks, respectively. λ_0 , λ_0^{ER} , λ_0^{BA} and λ_0^{conf} represent the number of degenerate zero eigenvalues in the PPI networks, ER random networks, SF networks and configuration model networks, respectively. γ denotes the values of the exponent of the degree distribution of the PPI networks. The γ values of the SF networks turn out to be ~ 3 . The ER and SF networks are generated for an average of over ten different realizations of the networks by keeping N and average degree same.

Species	N	$\langle k \rangle$	λ_0	D_c	γ	λ_0^{ER}	D_c^{ER}	λ_0^{BA}	D_c^{BA}	λ_0^{conf}	D_c^{conf}
<i>H. pylori</i>	709	3.935	317	146	1.94	0	0	155	17	163	79
<i>H. sapiens</i>	2138	2.872	976	662	3.03	0	0	469	33	512	309
<i>E. coli</i>	2209	9.895	487	323	1.7	0	0	0	0	569	200
<i>C. elegans</i>	2386	3.206	1354	940	2.08	0	0	528	29	818	569
<i>S. cerevisiae</i>	5019	8.803	864	491	2.00	0	0	0	0	950	314
<i>D. melanogaster</i>	7321	6.159	2311	1046	2.26	0	0	110	0	1621	687

So far, we have discussed the impact of network architecture on the duplication and the zero degeneracy. In the following, we evaluate the impact of average degree and size on the same. For a fixed network size, when the average degree of the

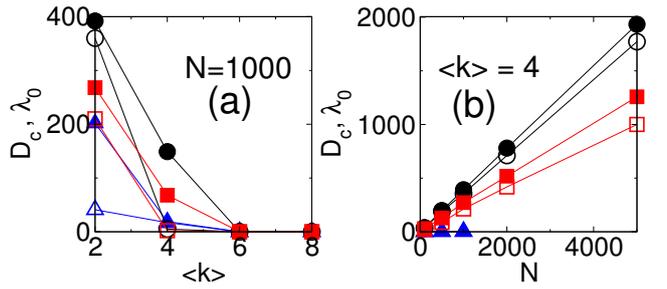


Figure 3.3: (Effect of the change in the network parameters, namely, size (N) and the average degree ($\langle k \rangle$) on the number of the duplicates and zero eigenvalues in different model networks. All values are averaged over 10 random realizations of the networks. Note that for ER random networks, connected component could not be obtained at $\langle k \rangle = 4$ for N above 1000. Here the Δ , \circ and \square represent the D_c of the ER, SF and configuration model networks, respectively. The solid Δ , \circ and \square represent the λ_0 of the ER, SF and configuration model networks, respectively.

ER random network increases, there is an increase in the number of connections. The probability that any node has the same set of neighbors as any other node is given by $\frac{\langle k \rangle^{2k}}{e^{2\langle k \rangle} k! N^k}$. With an increase in $\langle k \rangle$, this probability diminishes exponentially. This implicates a reduction in node's duplication. Fig. 3.3 (a) exhibits that at a low average degree, the number of complete duplicates is much less as compared to the number of zero eigenvalues, indicating that the contribution to the occurrence of the zero eigenvalues comes mainly from the partial duplicates. With an increase in the average degree, the number of duplicate nodes, as well as the number of zero eigenvalues, decreases. In order to further explore the impact of duplicates on zero degeneracy, we consider model networks other than the ER random networks generated using different algorithms, some of which might support node duplication. In case of an SF network of the same size and the average degree, a much higher number of duplicates and zero eigenvalues are exhibited as compared to the ER random network (Fig. 3.3 (a)), but both the counts decrease with an increase in the average degree. The configuration model networks display less complete duplicates and zero eigenvalues than those of the SF networks, as apart from the degree sequence, there does not exist any other preference for the association of nodes. With an increase in the average degree, the number of complete duplicates and the zero eigenvalues decrease. On a further increase in the average degree, the number of

complete duplicates and zero eigenvalues coincide with negligible values for all the three networks. At a fixed average degree, with an increase in the size of the networks, both numbers of complete duplicates and zero eigenvalues increase in the case of SF and configuration model networks, with zero degeneracy being higher as compared to complete duplication. The number of duplicates and zero eigenvalues, however, remain negligible in the case of ER random networks even with an increase in size (Fig. 3.3 (b)).

3.4 Discussion and Conclusions

Keeping in view the high zero degeneracy prevalent in real-world systems [125], in the following, we attempt to analyze how our investigation on model networks sheds light on the reasons behind high degeneracy at zero in real-world systems. We analyze the protein-protein interaction (PPI) networks of six different species, namely *H. Pylori*, *H. sapiens*, *D. melanogaster*, *S. cerevisiae*, *C. elegans* and *E. coli*. As depicted in Table 3.2, the number of zero eigenvalues are more than the number of complete duplicates indicating the existence of partial duplicates in the underlying networks. We generate ER random networks of the same size and average degree as of the six PPI networks. Table 3.2 reveals that the generated ER random networks have no degeneracy at zero eigenvalues and no duplicates. As the PPI networks are SF in nature [106], we generate SF networks of the same size and average degree as of the PPI networks using the BA algorithm. We find that though corresponding SF networks lead to a high degeneracy at zero, as expected, the number of zero eigenvalues and complete duplicates are much less than those of the corresponding PPI networks. It may be because the SF networks so generated display power-law behavior but need not have the same degree sequence as of the PPI networks. We further construct corresponding randomized models of the real systems, i.e., configuration models having the same degree sequence as of the six PPI networks. We observe that the configuration model networks have much higher zero degeneracy and complete duplication as compared to the SF networks generated using the BA algorithm. This observation is quite intriguing as it has been

demonstrated that the SF networks generated using the BA algorithm have higher duplication and degeneracy as compared to their corresponding configuration models (Fig. 3.2 and 3.3). But it would be noteworthy to mention that the configuration models generated in the case of model networks were the ones having the same degree sequence as of the BA-algorithm generated SF networks. While in the case of the PPI networks, the configuration models preserve the degree sequence of the PPI networks. This indicates that not only the power-law behavior of the networks accounts for duplication but the very nature by which the real world PPI networks have evolved and acquired a degree sequence that favors duplication and leads to degeneracy at zero also contributes to it.

Further, the results presented for the configuration model in Table 3.2 uses the same degree sequence as of the real PPI networks of the various species and consequently has the same power-law exponent values as of the latter. It may be possible that the BA algorithm leads to a different exponent for the same values of N and $\langle k \rangle$ taken from the PPI networks. As evident from columns 6 and 11 of Table 3.2, the BA networks for *H. sapiens* have the closest values of zero degeneracy, and complete duplicate nodes as of the PPI networks, which might be occurring as both the networks have the power-law exponents being almost equal. The values of exponent deviating more from those of the real PPI networks lead to more deviation of the values of zero degeneracy and complete duplication (Table 3.2) indicating that the exponent, as well as particular degree sequence collectively, contribute to the occurrence of the zero degeneracy.

To conclude, all the real-world networks show more zero degeneracy as compared to the corresponding random models, indicating that the equivalent number of nodes are completely duplicated (condition (a)) or have partial duplications (condition (b)), as depicted by Table 3.2. Gene duplication mechanism has been emphasized in evolutionary biology to be a driving force for creating new genes in a genome [112, 126–129]. As duplicated genes are known to acquire mutations faster than other genes resulting in divergence of functions [130], understanding the PPI networks exhibiting the prevalence of duplicates is quite interesting as well as

intriguing.

Most of the real-world networks are scale-free, which renders few nodes connected with almost all other nodes in the network, leaving a lot of nodes having as less as one connection with the hub only. This naturally leads to a lot of complete duplicate nodes (condition (a)), which in turn leads to a high degeneracy at zero. Scale-free networks generated through preferential attachment algorithm yield a good number of duplicated nodes, owing to the very nature of the algorithm, in turn leading to equivalent degeneracy at the zero eigenvalues. In the preferential attachment, a new entry in the network has more probability of connecting with the existing hubs, making it more probable that it connects with the same set of nodes which gained connections with the previous entry. Since such kind of bias does not exist in the configuration model, as expected, it exhibits fewer duplicates and hence less degeneracy at zero. With an increase in the average degree, it becomes difficult for pair(s) of nodes to satisfy condition (a) or (b), as more connections will lead to more probability of destroying duplication, leading to a constant decrease in the zero eigenvalues.

We explore the origin of zero degeneracy in the network spectra. Our analysis sheds light on the mechanisms which collectively lead to zero degeneracy in the real networks. Further, we correlate the occurrence of zero degeneracy with the evolutionary origin of a network. Comparison of the number of duplicates and zero degeneracy in the PPI networks of six different species with their corresponding configuration models reveals that in addition to the power-law behavior of the real networks, other factors also contribute to node duplication leading to comparatively high zero degeneracy. Duplicated gene pairs have been emphasized to confer evolutionary stability to many biological systems [131, 132]. The analysis carried out in this work combined with the occurrence of an exceptionally high peak at zero degeneracy in real-world networks can be extended to understand other complex systems as well as to build up robust technological networks.

Network Structure and (0,1)- Embedded Matrix Image of Net- works

Chapter 4

4.1 Introduction

The fractal geometry of nature was first described by Mandelbrot already in 1967 [133], and the approach has been extensively used, since then, to gain insights into the underlying scaling of a variety of visually complex structures, like fracture surfaces of metals [134], strange attractors [135, 136], diffusion [137] and medical imaging [138] patterns, galaxies [139], and atomic spectra [140], just to quote a few examples. In complex networks, so far, the focus has been investigating self-similarity [141, 142] and fractal structure of skeleton [143], as well as growth models which capture the observed fractal behavior[48, 144].

A fundamental issue is identifying whether a system is a mono- or a multi-fractal, i.e., whether or not a unique fractal scaling spans all the different system's parts, regions, or components. Multi-fractal analysis requires considering a physical measure: the number of nodes within a box of size, say l , has been used so far to analyze how the distribution of the such number of nodes scales in a network, as the box size increases [143]. In this Letter, instead, we quarter the focus on the scaling of the number of edges in a partition-box of the network's adjacency matrix \mathcal{A} . Precisely, we consider a *spatial distribution* of the network's edges (instead of

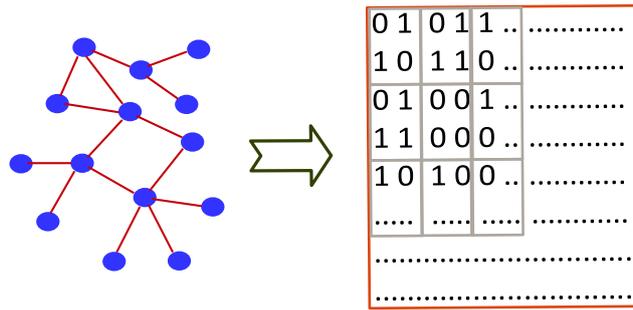


Figure 4.1: (Color Online). Schematic representation of the used box counting method. Left: original network. Right: partition of the adjacency matrix \mathcal{A} .

the network's nodes), this way making edges (the entries of the \mathcal{A} 's plane) as the basic units for the evaluation of fractal dimensions.

The aim is to introduce a practical and computationally non-demanding procedure, able to distinguish mono- from multi-fractal characters in complex networks. To do so, the distribution of 1's in square boxes of length ϵ in \mathcal{A} is analyzed by the use of the box-counting method (BCM) [145–147], which considers only non overlapping boxes, thus preventing specific parts of the matrix from getting over-weighted due to systematic over-counting. Furthermore, the robustness of the results is pledged by reshuffling the nodes, i.e., by avoiding any form of biasness generated by a specific node indexing. Precisely, the supplementary material contains evidence that each shuffling step provides a new ensemble of adjacency matrices [148].

4.2 Methods and Techniques

A schematic representation of the procedure is sketched in Fig. 4.1. We start with partitioning the adjacency matrix into ϵ -size boxes, and counting the number of boxes $n(\epsilon)$ with at least one non-zero entry (edge), with ϵ varying from 2 to $N/2$. $n(\epsilon)$ exhibits typically the scaling $n(\epsilon) \sim \epsilon^{D_0}$, with D_0 giving the dimension of the network. If D_0 is a non-integer number, the network is said to be *fractal*. D_0 , however, gives no information whatsoever on the system's multi-fractality, which should be accounted for, instead, by a multi-fractal approach [145]. Let therefore digitalize \mathcal{A} into ϵ -size boxes, and let $n_i(\epsilon)$ be now the number of '1' entries in

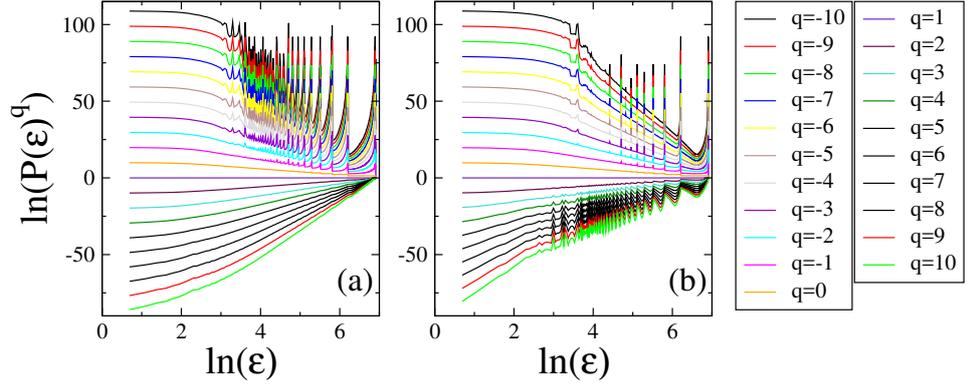


Figure 4.2: (Color Online). P_ϵ^q vs. ϵ in a double logarithmic scale with $q \in (-10, 10)$ for (a) ER and (b) SF networks. $N = 1,000$ and $\langle k \rangle = 20$.

the i^{th} element of the partition. The occurrence probability of 1's in the i^{th} box of size ϵ , denoted by $p_i(\epsilon)$, ranges therefore between $1/2N_c$ and $\epsilon^2/2N_c$, where N_c is total number of connection in network. At each ϵ value, the q^{th} moment (for all $-\infty < q < +\infty$ real numbers) of this probability is given by

$$P_\epsilon^q = \sum_{i=1}^{n(\epsilon)} [p_i(\epsilon)]^q. \quad (4.1)$$

The scaling exponent $\tau(q)$ is given by

$$\tau(q) = \lim_{\epsilon \rightarrow 0} \frac{\ln P_\epsilon^q}{\ln \epsilon}, \quad (4.2)$$

and is obtained from the slope of $\ln P_\epsilon^q$ vs. $\ln \epsilon$ at all q -value. Notice that $\tau(q)$ is here calculated *always* as ensemble average over the P_ϵ^q values obtained by shuffling the node indices. The q^{th} -moment dimension D_q is then given by

$$D_q = \frac{\tau(q)}{q-1}. \quad (4.3)$$

For each box of size ϵ and occupation probability $p_i(\epsilon)$, the singularity strength α_i is given by $p_i(\epsilon) = \epsilon^{\alpha_i}$, and at every q , α is evaluated as $1/n(\epsilon) \sum_{i=1}^{n(\epsilon)} \alpha_i$. The singularity spectrum, $f(\alpha)$, is related with $\tau(q)$ by means of the Legendre transform

$$f(\alpha) = q\alpha - \tau(q), \quad (4.4)$$

with $\alpha = d\tau(q)/dq$ being the Hölder exponent, and $f(\alpha)$ indicating the dimension of the subset scaling with α .

A multi-fractal structure is indicated by the following marks [145]: *i*) multiple slopes of $\tau(q)$ vs. q ; *ii*) a non constant value of D_q vs. q , and *iii*) $f(\alpha)$ vs. α covering

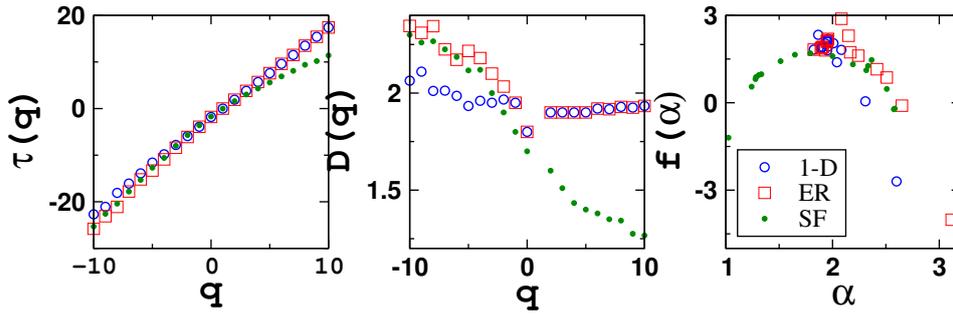


Figure 4.3: $\tau(q)$ (a) and D_q (b) vs. q . (c) $f(\alpha)$ vs. α . In all plots results are reported for 1-d lattices, ER and SF networks with $N = 1,000$ and $\langle k \rangle = 10$. All quantities are calculated for 50 random realizations of the networks generated by reshuffling indices 100,000 times for each realization. The left and right region slopes of $\tau(q)$ for the 1-d lattices, ER and SF networks are, respectively, 2.1 and 1.9; 2.4 and 1.9; 2.3 and 1.2.

a broad range instead of being accumulated at nearby non-integer values of α .

We start by briefly discussing the effect of different reshuffling of the nodes' labeling, in particular (i) random reshuffling, and (ii) degree-based reshuffling. For an ϵ -size box (located at the (i, j) coordinate in the adjacency matrix plane and denoted as b_{ij}) the probability of occurrence of 1's is $p_\epsilon^{b_{ij}} \propto \langle k \rangle$ in case i) (detailed derivations are provided in Supplementary Material), and is therefore independent of the network architecture. Case ii) leads instead to interesting behavior, as the probability of occurrence of 1's depends on the heterogeneity present in the networks. For 1-d lattices, where no degree heterogeneity is present, random and degree-based reshuffling provides the same results [148]. At variance, Erdős-Rényi (ER) and scale-free (SF) networks, which display varying levels of degree heterogeneity, exhibit different behaviors for the two cases. For the degree-based reshuffling one has $p_\epsilon^{b_{ij}} \propto \frac{\epsilon^2}{N^4} \frac{k^2}{\rho(k)^2}$, where k is the degree of a node inside the box b_{ij} and $\rho(k)$ is the probability of occurrence of k -degree nodes in the network. Note that $p_\epsilon^{b_{ij}}$ depends upon the degree as well as the probability density of the degree of all the nodes inside the box [148]. In the following, we first sort the nodes of a network in decreasing order of their degrees and assign their indices. As the usual case is that many nodes have the same degree, in order to avoid the possibility of the preference being conferred on a particular node in a set of the same degree nodes, we carry

out several realizations of the shuffling of indices among the nodes with the same degree.

4.3 Results

We now present the results for 1-d lattices, ER and SF networks, and real-world protein-protein interaction (PPI) networks of six different species. With a range of ϵ values spanning from 2 to $N/2$, P_ϵ^q is evaluated for q values ranging between -10 to $+10$. The slope of P_ϵ^q (calculated using Eq. 4.1) versus ϵ (on a double logarithmic scale) provides the estimation of the value of the scaling exponent at each q , denoted as $\tau(q)$ (Eq. 4.2). For all networks, the value of the scaling exponent is zero at $q = 1$ and non-zero at other q values (see Fig. 4.2). The q range can be divided, then into the left ($q \leq 0$) and the right ($q > 1$) region. We do not consider $q = 1$, as this is the fractal dimension of the system. The same value of the slope of $\tau(q)$ in both regions indicates a mono-fractal nature of the network, while different values are a signature of a multi-fractal structure. For a regular lattice, all nodes have the same degree. This kind of uniformity results in a single value of the scaling exponent on both left and right regions (see Fig. 4.3(a)). While the slopes of $\tau(q)$ are indicative measures of the multi-, mono- or non-fractality of the system, the range of D_q values gives the dimension(s) of the corresponding networks. For 1-d lattices, the D_q values shrink to a very narrow range about 2, confirming again the mono-fractal character of the graph, while the singularity spectrum, f_α (calculated using Eq. 4.4) shows that, for a specific range of α , most of the f_α values are accumulated in a nearby non-integer region, with only a few points deviating from this behavior.

The case of ER networks is far different. There, by construction, a new edge does not have a preference to connect with specific nodes [27], leading to approximately similar degrees of all the nodes, and to an adjacency matrix where the 1 entries are dispersed in the entire plane, without clusters or patterns. The associated homogeneity gives nearly close values of scaling exponents on the left and right regions on performing multi-fractal analysis (see Fig. 4.3(a) and consult the Supplementary Material for the analysis of the impact of degree homogeneity on

multi-fractal dimensions). Further, on plotting D_q vs. q , one finds that the right region exhibits the same behavior as that of the 1-d lattice, whereas the left region deviates significantly from a mono-fractal pattern (see Fig. 4.3(b)). The conclusion is that ER networks can be said to exhibit a multi-fractal nature in the left region of the spectrum, and a quasi-mono-fractal character for $q > 1$. When SF networks are taken into consideration, the growth mechanism producing them has a bias towards high degree nodes getting more and more edges (as preferential attachment induces new nodes to get linked to existing high degree nodes [27]), and gives a large spread in the degrees of the network's units. In the adjacency matrix, this is reflected by a few rows and columns (corresponding to high degree nodes) with a vast number of entries, with the majority of rows and columns having only very few 1's. Two large structures in the adjacency matrix then arise: one is a strip structure with a dense accumulation of 1's, and another is the sparse dispersion of entries in the rest of the space. The mixing of these two patterns yields different scaling exponents, and is indicative of a multi-fractal behavior (see Fig. 4.3(a)).

The probability of occurrence of 1's in the i^{th} box of size ϵ , $p_i(\epsilon)$, ranges between $p_{min}(\epsilon) = 1/2N_c$ and $p_{max}(\epsilon) = \epsilon^2/2N_c$. On considering the q^{th} moments of these probabilities (Eq. 4.1), one finds that in the positive q region the $p_{max}^q(\epsilon)$ values dominate over the $p_{min}^q(\epsilon)$, while in the negative q region the $p_{min}^q(\epsilon)$ values exhibit dominance over $p_{max}^q(\epsilon)$ values with increasing magnitude of q . This means that the behavior of the densely packed boxes of entries in the adjacency matrix plane is reflected in the right q region, while the sparsely packed boxes are determining the behavior of the left q region. The conclusion is that the relevant characteristics of the system (in terms of the high density of edges) are described by the right region, while generally small fluctuations in the system's fractal properties are accounted for the left region (as tiny changes in the q^{th} moments of the probabilities get magnified in the negative q region).

In 1-d lattices, all nodes have the same degree, thus rendering uniform the distribution of 1's across the adjacency matrix planes. As a consequence, there are only two probabilities of occurrence of 1's: for all the boxes in the off-diagonal region,

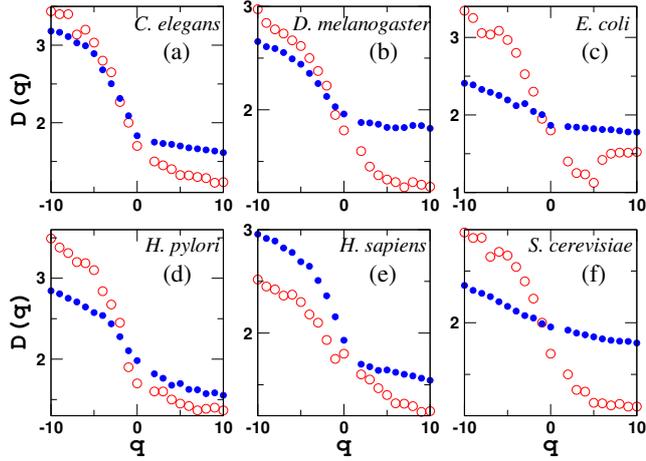


Figure 4.4: D_q vs. q for PPI networks (open circles) of six species, along with their corresponding ER random networks (closed circles) having the same size and average degree. The results for each of the corresponding ER random networks are presented for 100,000 node reshufflings, and for 50 realizations.

the probability is $\frac{2N_c\epsilon^2}{(N^2-N)}$, while for all the boxes lying in the diagonal region, the probability is $\frac{2N_c(\epsilon^2-\epsilon)}{(N^2-N)}$. The slight kink in the D_q values in the left region can be therefore attributed to the probabilities corresponding to the diagonal regions.

The degree distribution of an ER random network follows Poisson statistics. Though most of the nodes have a degree close to $\langle k \rangle$, nodes have higher and lower degrees as well. The probability of having 1s in a box depends on its degree distribution and deviates from that of a 1-d lattice. The deviation δ in the homogeneity of node degrees is reflected in the lower and upper bounds for the probability of occurrence of 1's: for all the boxes in the off-diagonal region, the probability is $(\frac{2N_c}{(N^2-N)} \pm \delta)\epsilon^2$, whereas for those in the diagonal region, the probability is $(\frac{2N_c}{(N^2-N)} \pm \delta)(\epsilon^2 - \epsilon)$. Though the distribution of 1's in the adjacency matrix plane of the ER random networks appears very similar to that of a 1-d lattice (see Supplementary Material), the deviation in the homogeneity of node degrees results in significant deviations within the left region of D_q , as small changes are actually magnified in this region.

For SF networks, two types of dominant structures (patterns) exist in the adjacency matrix plane. The first structure is made up of densely packed rows and columns of 1's, while the sparse population of 1's forms the second structure. The

mixing of these two structures gives rise to a wide variety of probabilities $p_i(\epsilon)$'s. For positive (negative) q values, the principal contribution to the multi-fractal nature of the graph comes from the first (the second) structure. Thus, for negative q values, SF networks exhibit the same behavior as ER random networks, while their behavior sharply differs from that of ER graphs for positive q values.

Finally, we apply our measure for shedding light on the structure and nature of associations in real-world protein-protein interaction (PPI) networks. Namely, we consider the PPI networks of six different species: the *C. elegans*, the *D. melanogaster*, the *E. coli*, the *H. pylori*, the *H. sapiens* and the *S. cerevisiae*. In order to draw a fair comparison among the different PPI networks, we construct corresponding ER random networks having the same size and average degree as the real networks, and perform comparative multi-fractal analysis for 100,000 rounds of node reshufflings, and 50 realizations. On plotting D_q vs. q for the PPI networks and their corresponding random controls, one finds that for *E. coli* and *S. cerevisiae* both left and right regions deviate significantly from their random counterparts (see Fig. 4.4(c) and (f)), indicating a genuine multi-fractal behavior of the underlying networks, which is quite expected as all these species are known to display scale-free properties in the degree distribution [106]. In the cases of *D. melanogaster* and *H. pylori*, the extent of deviation of the PPI networks from their corresponding ER graphs is less evident (see Fig. 4.4(b) and (d)). Interestingly, for *C. elegans*, the PPI network very closely resembles the behavior of its corresponding random network (see Fig. 4.4(a)).

In the case of *H. sapiens*, the behavior is very different as compared to all other PPI networks. Indeed, in contrast to the other five graphs, the *H. sapiens* PPI network exhibits a fractal behavior intermediate between ER random networks and 1-d lattices in the left region (see Fig. 4.4(e) and 4.3(b)), indicating some kind of structural homogeneity in the underlying system, as also revealed recently through an analysis based on random matrix theory [106]. The multi-fractal analysis supports and corroborates the hypotheses that real-world biological systems are formed based on different designing principles, which are then reflected in the difference in their dimensionality.

4.4 Discussion and Conclusions

In conclusion, this study provides a practical and computationally not demanding method for inspecting the complexity of real-world systems by analyzing the scaling behavior of edges in the underlying networks. Our study reveals that networks of the same size and average degree can have a completely different fractal character and nature, depending upon how edges are distributed in the networks. Further, we demonstrated how homogeneity in node degrees influences the scaling behavior of the underlying networks, and a deviation from degree homogeneity leads to significant changes in the graph's dimensionality. Our results point out the importance of edge distribution in determining the behavior of the system and have diverse applications. For instance, different biological networks such as protein-protein interaction networks, metabolic networks, transcription regulatory networks, neural networks (which are based on different basic design and functional principles) may display similar fractal behavior depending upon similarity in their edge distributions. It is indeed vital to remark that the underlying biological systems have evolved and survived over a long period without any significant changes in their properties [149]. Additional examples demonstrating the importance of edge distribution are the edge-server network, which facilitates the search on the internet by Google [150, 151], and the identification of influential nodes in social networks which take into account the arrangement of the edges in the graph [152]. Therefore the secure method for the multi-fractal analysis presented here can be of interest and use for a better understanding and description of complexity in these systems. It could be furthermore interesting to investigate multifractality in networks constructed based on a given ordering of nodes, which would result in an adjacency matrix with a more pronounced block structure (specifically more dense diagonal blocks). Looking at the differences between a random permutation of the nodes and the ordering of the nodes that give the best community structure can be one of the dimensions of future investigation.

Conclusions and Scope for Future Work

Chapter 5

Conclusions and Scope for Future Work

5.1 Conclusions and Discussions

In network science, we study any complex system by mapping it in the node-edge form, and this information can be represented by pictorially, list, or as a matrix. Here, the matrix representation is a potent tool because the matrix itself works as a picture of the network in the 2D-matrix plane with orientation being all the number of ways can be indexed usually in the order of $(N!)$. Simultaneously, information stored in the matrix also can be used to study the structural properties of networks as a list do. However, the matrix provides another extra set of information, namely eigenspectrum.

Any N -nodes network can completely describe by $(N^2 + N)$ -set of variables, N for indexing of N -nodes, and N^2 for their pair-wise interaction, including the self-loops. Similarly, eigenspectrum consists of N -eigenvalues, and for each eigenvalue, we get the corresponding eigenvector with N -entries, i.e., in total N^2 -eigenvectors entries.

Surprisingly, These new $(N + N^2)$ -variable set gives us some information not

easily accessible from N for indexing of N -nodes and N^2 for their pair-wise interaction. Like, which network will synchronize first in two or more given networks.

Similar to the degree distribution, spectral density distribution has unique information on degree distribution. Also, as our results demonstrated, eigenvalues collectively store degree-degree correlation data of network stored. Further, zero-eigenvalues degeneracy has structural as well as evolution information of networks.

Further, with the help of zero eigenvalue degeneracy, other eigenvalue degeneracies were explained and also gave rise to many new structural constitutional information about the network.

All these studies have shown that network representation as a matrix is advantageous and provides three authoritative paths to study networks of which structural study is well established while the spectral study is produced by many for new insight in networks. However, the third, matrix image approach was pursued by us and is new. These results provide a new scope to explore network studies.

In this thesis, we covered three significant results for the adjacency matrix of an independent or single-layer network. First, with RMT in chapter 2, how the networks degree-degree correlation information reveal from the spectrum of eigenvalues. This information has three parts early in the short-range correlation study by the neighbor spacing distribution of the eigenvalues of the adjacency matrices in the form of Poisson to GOE distribution transition for very high degree-degree correlation with a little decrement. Second, in the long-range study by Δ_3 -statistics for rest positive correlation range. Where the correlation length increases with the decrease in the degree-degree correlation in the network and reaches a maximum value of correlation length in Δ_3 -statistics for $r = 0$. Finally, for all negative correlation in degree connections are captured in spectral density distribution plots. Here, we find that the typical density distribution plot gradually changes its shape to double-humped shape with a peak of $\lambda = 0$ [2].

For the second result in chapter 3, we explored why in the spectrum of real-world inspire networks, a large number of $\lambda = 0$ are observed and also with the previous finding that negative degree-degree correlation also induces the peak at

$\lambda = 0$. We look for discreetly at the origins of the $\lambda = 0$ in the model and real-world networks. Also, it found that the standard model networks do not have a large amount of $\lambda = 0$ as real-world networks. Further, even the configuration models based on the same degree-sequences of real-world networks failed to mimic this observation. This inability of the configuration model to have nearly the same zero-degeneracy points to the fact that the system evolves very differently than the configuration model where a pair of nodes are connected randomly. Also, mathematically, the concept of rank in the linear algebra that the number of eigenvalues at $\lambda = 0$, say r , reduces the rank of $N \times N$ matrix to $N - r$. As configuration model networks with same degree-degree correlation do not have, and as high number of $\lambda = 0$ as a real-world network. Therefore, we found that for a network with no isolated node, it can shed light on the evolution and working of the networks, like gene duplication [3].

Furthermore, for the final result in chapter 4, we explored that adjacency matrix as the $(0, 1)$ -embedded image of the network and analyzed the multifractal properties of this matrix plane image with multifractal tools. We found that our results were in agreement with all the previous results. These results provided another possibility for network study through $(0, 1)$ -embedded images [4].

All these three results have shown that the adjacency matrix provides a robust structure for network study. The adjacency matrix approach is with vast possibilities. Much of the information about eigenvalues apart from the degenerate eigenvalues and the largest eigenvalue is not clear individually and collectively. The new possibility of $(0, 1)$ -embedded image structure for the study of a network has lots of potentials, as no study apart from multifractal nature is done. The beauty of the $(0, 1)$ -embedded image is that it transformed the basic unit of structure to edges from nodes.

All these results presented here are for the adjacency matrix of single-layer networks and needs for research multilayer networks case. Recently, network scientists have realized that multilayer is a more realistic abstraction of complex real-world systems. With the addition of intralayer connection information in the adjacency

matrix, we expect to predict behavior about complex systems more accurately. Also, what information is there in the eigenvalues and eigenvector spectra of networks?. For example, the impact of specific structures or properties in the network on eigen-system of the adjacency matrix.

Appendix

A.1 Random reshuffling and degree-based reshuffling of the nodes

In our work, two different methods of reshuffling the nodes have been considered.

1. The labeling of the nodes is reshuffled randomly. Usually saturation of measurements occurs for a number of reshuffling being 10 times larger with respect to the number of connections.
2. A degree based reshuffling, where nodes are first sequenced based on their degree, and reshuffling operations are made only among those nodes which have the same degree. For this case, the saturation of measurements is reached much earlier.

A.1.1 Random reshuffling

For random reshuffling, the probability of occurrence of 1's in boxes of the adjacency matrix plane is proportional to $\langle k \rangle$. In the following, we analytically derive the differences.

For a network of size N , upon shuffling randomly the nodes' indices, a node of degree d_i (say) can have an index j (say) without any relation between i and j . Therefore, the probability for any node (say u_i) to have index j is given by $p_{(u_i)(j)} = 1/N$. Let us now represent each box by its position in the adjacency

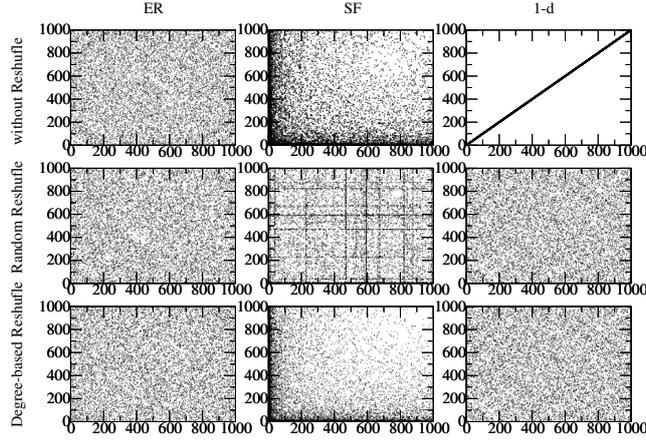


Figure A.1: The effect of random reshuffling and degree-based reshuffling on the adjacency matrix. (Left) ER random networks, (middle) SF networks and (right) 1-d lattices. In each panel, each dot is the position of an entry in the adjacency matrix plane.

matrix plane. b_{ij} represents a box which starts from (i, j) in the adjacency matrix plane. The number of 1's in a box b_{ij} is given as

$$x_{ij} = \sum_{l_1=i+1}^{i+\epsilon} \sum_{l_2=j+1}^{j+\epsilon} a_{l_1 l_2}, \quad (\text{A.1})$$

where l_1 and l_2 are the adjacency matrix coordinates inside a box.

$a_{l_1 l_2} = 1$, if l_1 is connected to l_2 and 0 otherwise, meaning that the contribution of l_1 row to x_{ij} is proportional to the degree of l_1 -index node, as the l_2 -column is randomly allotted for a fixed l_1 . This condition is arising due to the fact that for a fixed l_1 , the probability that l_2 is connected to l_1 node is proportional to the degree of l_1 node.

Therefore, the probability of any $a_{l_1, i} \neq 0 \forall i$ is given as

$$p_{a_{l_1, i}} = \frac{1}{N} d_{l_1}, \quad (\text{A.2})$$

and the probability of any $a_{l_1, i} \neq 0 \forall i$ in a box is

$$p_{a_{l_1, i}}^{b_{ij}} = \frac{1}{N} d_{l_1} \epsilon. \quad (\text{A.3})$$

x_{ij} is given by

$$x_{ij} = \sum_{l_1=i+1}^{i+\epsilon} \sum_{l_2=j+1}^{j+\epsilon} a_{ij} \neq 0$$

Let $p_\epsilon^{b_{ij}}$ be the probability that a given box b_{ij} of size ϵ has at least one entry 1.

Then one has $p_\epsilon^{b_{ij}} = \sum_{l_1} p_{a_{l_1, i}}^{b_{ij}} \times \text{Probability of } d_{l_1}\text{-degree node in the box of size } \epsilon.$

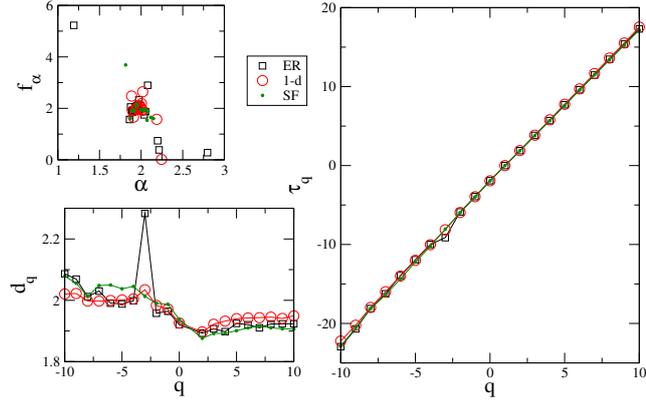


Figure A.2: Various multi-fractal measures plotted for the case of random reshuffling.

One then has

$$p_\epsilon^{b_{ij}} = \frac{\epsilon}{N} \sum_{l_1} d_{l_1} \frac{\epsilon}{N} = \frac{\epsilon^2}{N^2} \sum_{l_1} d_{l_1}$$

$$d_{l_1} = \sum_{j=1}^N P_{l_1}(d_j) \cdot d_j$$

where $P_{l_1}(d_j)$ is the probability of d_j node being allotted l_1 index.

In random reshuffling method, all the nodes in a network have equal probability of occurring at a particular index. Hence, $P_{l_1}(d_j) = \frac{1}{N}$. Therefore,

$$d_{l_1} = \frac{1}{N} \sum_{j=1}^N d_j = \frac{\langle k \rangle N}{N} = \langle k \rangle$$

Since l_1 index runs within a box having size ϵ , hence

$$\sum_{l_1} d_{l_1} = \epsilon \langle k \rangle \quad (\text{A.4})$$

Therefore,

$$p_\epsilon^{b_{ij}} = \frac{\epsilon^3 \langle k \rangle}{N^2} \quad (\text{A.5})$$

This indicates that on considering random reshuffling of the nodes, the probability of occurrence of 1's in a box of size ϵ is directly proportional to the average degree of the underlying network. Further, to capture minor changes in this probability, with respect to various degree distributions, we consider 1-d, ER and SF networks.

In the 1-d lattices with circular boundary condition, all the nodes have exactly the same degree which is equal to the average degree of the network. Thus each

box exhibits exactly the same behaviour. For the ER random networks, the degree of nodes are closely distributed about the average degree of the underlying network. Therefore, probability of occurrence of 1's in each box has small deviations from the average behaviour. In the case of SF networks, a very high degree heterogeneity is present in the networks, which leads to significant local deviations from the average behaviour, depending on the type of the node (high degree node or low degree node) occupying the boxes (see Fig. S1 and S2).

To further assess the importance of the reshuffling method on the probability of occurrence of 1's in a box, in the following we adopt a different method of reshuffling considering degree homogeneity.

A.1.2 Degree-based reshuffling

In the degree-based reshuffling method, we arrange the nodes in a decreasing order of their degrees and assign them indices. To avoid any preferential allocation of indices between the same degree nodes, we reshuffle the indices between the same degree nodes. For a network of size N , $\sum \rho(k) = 1$ and $\sum N\rho(k) = N$, where $\rho(k)$ is the probability of occurrence of k -degree node and $N\rho(k)$ is number of nodes with degree k . Given that $0 \leq N\rho(k) \leq N$,

- If $N\rho(k) = 1$, any k -degree node having index i is given as $i = \sum_{k=k_{max}}^k N\rho(k)$.
- If $N\rho(k) > 1$, any k -degree node having index i is given as

$$i \in \left(\sum_{k=k_{max}}^{k+1} N\rho(k) + 1, \sum_{k=k_{max}}^k N\rho(k) \right) \quad (\text{A.6})$$

For instance, let us consider a degree sequence viz. [200, 136, 80, 60, 60, 60, 30, 30, 22, 22, 22, 19]. We want to find the range of indices which can have a node with degree 60. From Eq. A.6, the upper and lower bound of indices that can have a node with degree 60 is given as $(\sum_{k=200}^{80} N\rho(k) + 1, \sum_{k=200}^{60} N\rho(k)) = ((1 + 1 + 1) + 1, (1 + 1 + 1 + 3)) = (4, 6)$.

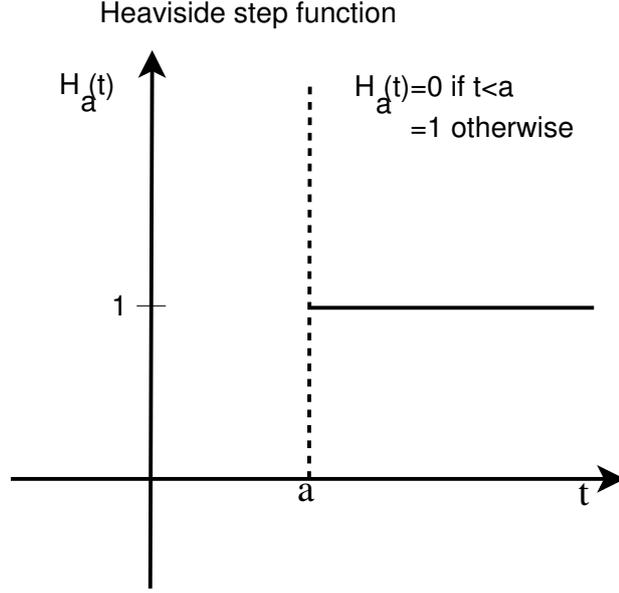


Figure A.3: Graph representation of Heaviside step function used in Eq. A.7.

Since in the degree-based reshuffling, the nodes having similar degrees get reshuffled, the indices accessible to a particular degree (say k_x) are given by the probability

$$p_i^{k_x} = \frac{H_{[\sum_{k=k_{max}}^{k_x+1} N\rho(k)]}(i) - H_{[\sum_{k=k_{max}}^{k_x} N\rho(k)]}(i)}{N\rho(k_x)}, \quad (\text{A.7})$$

where $H_a(i)$ is the Heaviside step function (Fig. S3).

Assuming there is no preference of the nodes being connected based on their degrees, we say that probability that any given k -degree node to be connected to any other node is equal to $\frac{k}{N}$.

Therefore, a box of ϵ size (b_{ij}) starting from a point in the adjacency matrix plane (i, j) will have x_{ij} number of 1's, given as;

$$x_{ij} = \sum_{l_1, l_2=i+1, j+1}^{i+\epsilon, j+\epsilon} a_{l_1 l_2}, \quad (\text{A.8})$$

where $a_{l_1 l_2} = 1$, if l_1 and l_2 are connected and 0, otherwise.

The probability of any i^{th} node being connected to the j^{th} node is given by

$$\begin{aligned}
p_{(a_{l_1 l_2}=1)} &= \sum_{k_1 k_2} p_{l_1}^{k_1} p_{l_2}^{k_2} \frac{k_1}{N} \frac{k_2}{N} \\
&= \sum_{k_1 k_2} \frac{k_1 k_2}{N^2} \frac{H_{[\sum_{k=k_{max}}^{k_1+1} N\rho(k)]}(l_1) - H_{[\sum_{k=k_{max}}^{k_1} N\rho(k)]}(l_1)}{N\rho(k_1)} \frac{H_{[\sum_{k=k_{max}}^{k_2+1} N\rho(k)]}(l_2) - H_{[\sum_{k=k_{max}}^{k_2} N\rho(k)]}(l_2)}{N\rho(k_2)} \\
&= \sum_{k_1 k_2} \frac{k_1 k_2}{N^4 \rho(k_1) \rho(k_2)} (H_{[\sum_{k=k_{max}}^{k_1+1} N\rho(k)]}(l_1) - H_{[\sum_{k=k_{max}}^{k_1} N\rho(k)]}(l_1)) (H_{[\sum_{k=k_{max}}^{k_2+1} N\rho(k)]}(l_2) - H_{[\sum_{k=k_{max}}^{k_2} N\rho(k)]}(l_2)) \\
&\quad - H_{[\sum_{k=k_{max}}^{k_2} N\rho(k)]}(l_2).
\end{aligned} \tag{A.9}$$

The probability of getting 1's in b_{ij} box of ϵ -size is given by

$$p_{(b_{ij})}^\epsilon = \sum_{l_1=i+1}^{i+\epsilon} \sum_{l_2=j+1}^{j+\epsilon} p_{(a_{l_1 l_2}=1)} \tag{A.10}$$

Now, according to the degree sequence of the network and the position (i, j) , $p_{(b_{ij})}^\epsilon$ can assume various values which are discussed in the following subsections.

A.1.2.1 Case 1: *The nodes have only one type of degree (say k) in the box b_{ij} .*

The indices of ϵ -size box, i.e. $(i + 1$ to $i + \epsilon, j + 1$ to $j + \epsilon)$, which lies within the range given by Eq. A.6 can have only k -degree nodes. Therefore, the probability of occurrence of k -degree node within the box is $\frac{1}{N\rho(k)}$, while the probability of occurrence of other degree node is zero. The probability of having 1's at (l_1, l_2) position within the given box is drawn from Eq. A.9 as

$$p_{(a_{l_1 l_2}=1)} = \frac{k^2}{N^4(\rho(k))^2} \tag{A.11}$$

Since $p_{(a_{l_1 l_2}=1)}$ is the same for all $a_{l_1 l_2}$ within the box, the consequence is that

$$p_{b_{ij}}^\epsilon = \epsilon^2 p_{(a_{l_1 l_2}=1)} = \frac{k^2 \epsilon^2}{N^4 \rho(k)^2} \tag{A.12}$$

A.1.2.2 Case 2: *The nodes have two types of degrees (say k and $k + 1$) in the b_{ij} box.*

The indices of a ϵ -size box, i.e. $(i + 1$ to $i + \epsilon, j + 1$ to $j + \epsilon)$ lies within the range given by Eq. A.6 for k and $k + 1$ degree nodes. Therefore, the probability of any k -degree node allotted within the box is $p_i^k = \frac{H_{[\sum_{k=k_{max}}^{k+1} N\rho(k)]}(i) - H_{[\sum_{k=k_{max}}^k N\rho(k)]}(i)}{N\rho(k)}$

and the probability of any $k + 1$ -degree node allotted within the box is $p_i^{k+1} = \frac{H_{[\sum_{k=k_{max}}^{k+1+1} N\rho(k)]}(i) - H_{[\sum_{k=k_{max}}^{k+1} N\rho(k)]}(i)}{N\rho(k+1)}$. The probability of occurrence of any other de-

gree nodes is zero. Therefore, the probability of any $a_{l_1 l_2}$ being '1' within the box b_{ij} is given by

$$\begin{aligned}
p_{(a_{l_1 l_2}=1)} &= \frac{(k+1)^2}{N^4 \rho(k+1)^2} (H_{[\sum_{k=k_{max}}^{(k+1)+1} N\rho(k)]}(l_1) - H_{[\sum_{k=k_{max}}^{(k+1)} N\rho(k)]}(l_1)) \\
&\quad \times (H_{[\sum_{k=k_{max}}^{(k+1)+1} N\rho(k)]}(l_2) - H_{[\sum_{k=k_{max}}^{(k+1)} N\rho(k)]}(l_2)) \\
&\quad + \frac{k^2}{N^4 \rho(k)^2} (H_{[\sum_{k=k_{max}}^{(k)+1} N\rho(k)]}(l_1) - H_{[\sum_{k=k_{max}}^{(k)} N\rho(k)]}(l_1)) \\
&\quad \times (H_{[\sum_{k=k_{max}}^{(k)+1} N\rho(k)]}(l_2) - H_{[\sum_{k=k_{max}}^{(k)} N\rho(k)]}(l_2)) \\
&\quad + \frac{k(k+1)}{N^4 \rho(k)\rho(k+1)} [(H_{[\sum_{k=k_{max}}^{(k)+1} N\rho(k)]}(l_1) - H_{[\sum_{k=k_{max}}^{(k)} N\rho(k)]}(l_1)) \\
&\quad \times (H_{[\sum_{k=k_{max}}^{(k+1)+1} N\rho(k)]}(l_2) - H_{[\sum_{k=k_{max}}^{(k+1)} N\rho(k)]}(l_2)) \\
&\quad + (H_{[\sum_{k=k_{max}}^{(k+1)+1} N\rho(k)]}(l_1) - H_{[\sum_{k=k_{max}}^{(k+1)} N\rho(k)]}(l_1)) \\
&\quad \times (H_{[\sum_{k=k_{max}}^{(k)+1} N\rho(k)]}(l_2) - H_{[\sum_{k=k_{max}}^{(k)} N\rho(k)]}(l_2))].
\end{aligned}$$

Therefore, the probability of having 1's inside the box is

$$\begin{aligned}
p_\epsilon^{b_{ij}} &= \sum_{l_1=i+1}^{i+\epsilon} \sum_{l_2=j+1}^{j+\epsilon} p_{(a_{l_1 l_2}=1)} \\
&= \sum_{l_1=i+1}^{i+\epsilon_1} \sum_{l_2=j+1}^{j+\epsilon_2} p_{(a_{l_1 l_2}=1)}^{(k+1)^2} + \sum_{l_1=i+1}^{i+\epsilon_1} \sum_{l_2=j+\epsilon_2+1}^{j+\epsilon} p_{(a_{l_1 l_2}=1)}^{k(k+1)} \\
&\quad + \sum_{l_1=i+\epsilon_1+1}^{i+\epsilon} \sum_{l_2=j+1}^{j+\epsilon_2} p_{(a_{l_1 l_2}=1)}^{k(k+1)} + \sum_{l_1=i+\epsilon_1+1}^{i+\epsilon} \sum_{l_2=j+\epsilon_2+1}^{j+\epsilon} p_{(a_{l_1 l_2}=1)}^{k^2}.
\end{aligned} \tag{A.13}$$

On substituting Eq. A.1.2.2 into Eq. A.13, one has

$$p_\epsilon^{b_{ij}} = \frac{1}{N^4} [\epsilon_1 \epsilon_2 \left(\frac{k+1}{\rho(k+1)} - \frac{k}{\rho(k)} \right)^2 + \frac{\epsilon^2 k^2}{\rho(k)^2} + \frac{\epsilon(\epsilon_1 + \epsilon_2)k}{\rho(k)} \left(\frac{k+1}{\rho(k+1)} - \frac{k}{\rho(k)} \right)]. \tag{A.14}$$

In comparison to the random reshuffling, the degree-based reshuffling highlights following differences for both the case. For the case of degree-based reshuffling, the probability of occurrence of 1's within a box depends on the position of the box as well as the coordinates within the box (for instance, $(l_1, l_2) \in (i+1, j+1) \dots \dots (i+\epsilon, j+\epsilon)$). More importantly, the probability of occurrence of 1's in the degree-based reshuffling method depends on $\frac{k}{\rho(k)}$, i.e., network's degree distribution. Whereas, in the case of random reshuffling, the probability of occurrence of 1's within a

box depends only on the average degree of the underlying network. Therefore, the degree-based reshuffling leads to different behaviours for 1-d lattices, ER random networks and SF networks of the same average degree and size (see Fig. S1).

After discussing impact of reshuffling on the probability of occurrence of 1's in a particular box, in the following we derive expressions for probability distribution as well as q^{th} moment of fractal dimension for various different types of the edge distributions.

A.2 Edge distribution of different networks on adjacency matrix plane

For the edge distribution of random networks and 1-d lattices, we present three simplified cases: (a) a strictly homogeneous distribution, (b) an approximately homogeneous distribution and (c) a strictly linear distribution.

A.2.1 Case A: Strictly Homogeneous distribution

Any ϵ -size box in the adjacency matrix plane will have $\frac{\langle k \rangle \epsilon^2}{N}$ number of 1's. Here, we consider large size network to avoid any boundary effect while assigning the boxes. Therefore, the probability of occurrence of 1's in a i^{th} box of size ϵ is given as

$$p_i(\epsilon) = \frac{\epsilon^2}{N^2}.$$

Summing up these probabilities gives

$$P_\epsilon = \sum_{i=1}^{i=\frac{N^2}{\epsilon^2}} p_i(\epsilon) = \sum_{i=1}^{i=\frac{N^2}{\epsilon^2}} \frac{\epsilon^2}{N^2} = 1.$$

The q^{th} -moment of the probability of occurrence of 1's for a ϵ size box is given as

$$P_\epsilon^q = \sum_{i=1}^{i=\frac{N^2}{\epsilon^2}} [p_i(\epsilon)]^q = \sum_{i=1}^{i=\frac{N^2}{\epsilon^2}} \left(\frac{\epsilon^2}{N^2}\right)^q = \frac{\epsilon^{2(q-1)}}{N^{2(q-1)}}.$$

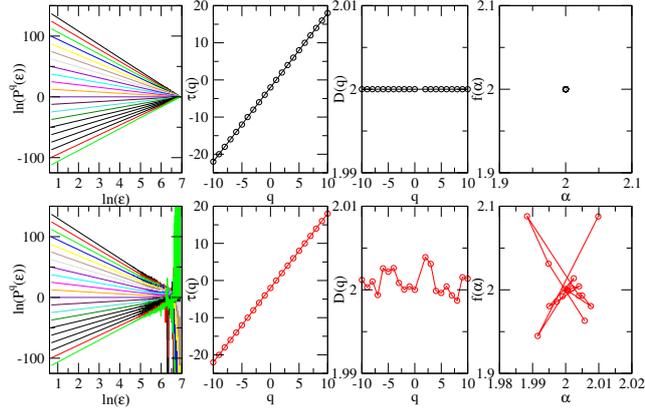


Figure A.4: (Color Online) Comparison between, respectively, $\ln(P_i^q(\epsilon))$ vs $\ln(\epsilon)$ plot, $\tau(q)$ vs q plot, $D(q)$ vs q plot and $f(\alpha)$ vs α plots for strictly (top panel) and approximately (bottom panel) homogeneous distribution of edges.

By taking log of both the sides, one has

$$\ln(P_\epsilon^q) = 2(q-1)\ln(\epsilon) - 2(q-1)\ln(N)$$

This is an equation for a straight line with a slope $2(q-1)$, which we define as the scaling exponent $\tau(q)$.

The q^{th} -moment dimension of the system $D(q)$ defined as $\frac{\tau(q)}{q-1}$ gives a single dimension i.e. 2.

The singularity spectrum, $f(\alpha)$ is related with $\tau(q)$ by means of Legendre transform

$$f(\alpha) = q\alpha - \tau(q).$$

$$\text{where } \alpha = \frac{d\tau(q)}{dq}.$$

$$\text{Substituting the value of } \tau(q), \alpha = 2 \text{ and } f(\alpha) = 2q - 2(q-1) = 2.$$

This indicates that a homogeneous distribution over a large system (where boundary effect can be neglected) leads to a non-fractal structure with dimension 2.

A.2.2 Case B : Approximately Homogeneous distribution

In this case, we assume that the number of 1's in the i^{th} box deviates from the homogeneous distribution of 1's by a small amount r_i generated randomly for each box with alternately changing sign in such a way that sum over all the multi-fractal measures is normalized. This gives the following expression for the multi-fractal measure,

$$P_i(\epsilon) = \sum_{i=1}^{\frac{N^2}{\epsilon^2}} \left(\frac{\epsilon^2}{N^2} (1 + (-1)^{i+1} r_i) \right) = 1,$$

which further gives us the following restriction on the random numbers (r_i) to be generated,

$$\sum_{i=1}^{\frac{N^2}{\epsilon^2}} \left(\frac{\epsilon^2}{N^2} (-1)^{i+1} r_i \right) = 0$$

Now, q^{th} moment of $P_i(\epsilon)$ can be written as,

$$P_i^q(\epsilon) = \sum_{i=1}^{\frac{N^2}{\epsilon^2}} \left(\frac{\epsilon^2}{N^2} (1 + (-1)^{i+1} r_i) \right)^q$$

For small r_i values, the q -th moment can be approximated to,

$$P_i^q(\epsilon) = \sum_{i=1}^{\frac{N^2}{\epsilon^2}} \left(\frac{\epsilon}{N} \right)^{2q} (1 + (-1)^{i+1} q r_i)$$

$$P_i^q(\epsilon) = \left(\frac{\epsilon}{N} \right)^{2(q-1)} + \sum_{i=1}^{\frac{N^2}{\epsilon^2}} \left(\frac{\epsilon}{N} \right)^{2q} (-1)^{i+1} q r_i$$

$$P_i^q(\epsilon) = \left(\frac{\epsilon}{N} \right)^{2(q-1)} \left(1 + \left(\frac{q\epsilon^2}{N^2} \right) \sum_{i=1}^{\frac{N^2}{\epsilon^2}} (-1)^{i+1} r_i \right)$$

Taking log on both the side, we get,

$$\ln(P_i^q(\epsilon)) = 2(q-1)\ln(\epsilon) - 2(q-1)\ln(N) + \ln\left(1 + \left(\frac{q\epsilon^2}{N^2}\right) \sum_{i=1}^{\frac{N^2}{\epsilon^2}} (-1)^{i+1} r_i\right)$$

By using $\ln(1+x)$ approximation for small x , we get,

$$\ln(P_i^q(\epsilon)) = 2(q-1)\ln(\epsilon) - 2(q-1)\ln(N) + \left(\frac{q\epsilon^2}{N^2}\right) \sum_{i=1}^{N^2} (-1)^{i+1} r_i - \frac{1}{2} \left(\frac{q\epsilon^2}{N^2}\right) \sum_{i=1}^{N^2} (-1)^{i+1} r_i^2 + \frac{1}{3} \left(\frac{q\epsilon^2}{N^2}\right) \sum_{i=1}^{N^2} (-1)^{i+1} r_i^3 - \dots$$

Also, $\tau(q) = \lim_{\epsilon \rightarrow 0} \frac{\ln(P_i^q(\epsilon))}{\ln(\epsilon)}$ gives

$$\tau(q) = 2(q-1)$$

$$D(q) = 2$$

One can see that theoretically this approximation does not affect values of $\tau(q)$ and $D(q)$ from those of case A. However, numerical evaluation shows significant deviations of these values from those of the case A. For these two cases of strictly homogeneous and approximately homogeneous distribution, on plotting $\ln(P_i^q(\epsilon))$ vs $\ln(\epsilon)$ we find that the strictly homogeneous case has a straight line behaviour with existence of various slopes indicating non-fractal structure with dimension 2. Whereas the approximately homogeneous case manifests a deviation leading to non-fractal dimensions distributed very closely about 2 which is in fact a multi-fractal structure (see Fig. S4). This indicates that one of the reasons behind $\ln(P_i^q(\epsilon))$ vs $\ln(\epsilon)$ plot deviating from the linear behaviour can be attributed to the non-homogeneity in the degree distribution.

A.3 τ_q versus q for PPI Networks

On plotting $\tau(q)$ versus q for the PPI networks and their corresponding random controls, for *E. coli* and *S. cerevisiae*, we find a significant deviation exhibited by the PPI networks from their corresponding random controls (see Fig. S5(c) and (f)), indicating multi-fractal behaviour of the underlying networks. This is quite expected owing to their scale-free nature of degree distribution [18]. The extent of deviation between the PPI networks of *D.melanogaster* and *H.pylori* and their corresponding ER random networks diminishes (Fig. S5(b) and (d)), while notice-

ably for *C.elegans* the PPI network very closely resembles the behaviour of its corresponding random network (Fig. S4(a)). Interestingly, *H.sapiens* exhibits a completely different behaviour from rest of the model organisms. The left region of *H.sapiens* PPI network exhibits a slope which is lower than that of its corresponding ER random network (see Fig. S5(e)), thus indicating that though the underlying system maintains an universality in large scale properties with the other PPI networks, minor fluctuations in fractal behaviour captured by the left region indicates a homogeneity in the underlying structure. These features of the PPI networks are magnified and better understood through the behaviour of D_q versus q , discussed in length in the chapter 4.

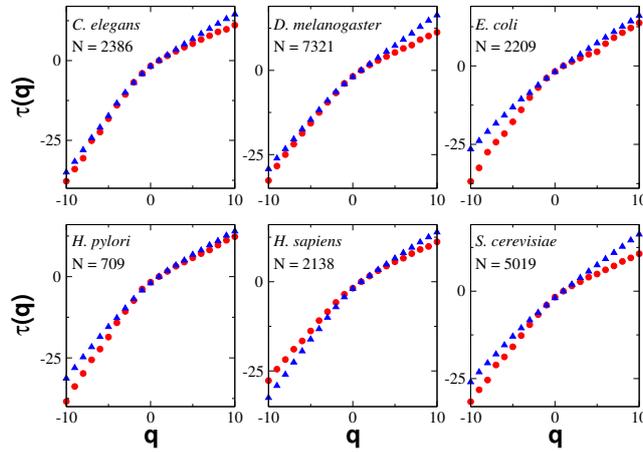


Figure A.5: (Color Online) Plots of τ_q as a function of q for PPI networks (circles) of six species along with their corresponding ER random networks (triangles) having the same size and average degree. The results for each of the corresponding ER random networks are presented for 100,000 node reshuffling done for 50 realizations.

Bibliography

- [1] Sarkar, C., Jalan, S., Spectral properties of complex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28, 10:102101(2018).
- [2] Jalan, S., Yadav, A., Assortative and disassortative mixing investigated using the spectra of graphs, *Physical Review E* 91, 1:012813 (2015).
- [3] Yadav, A., Jalan, S., Origin and implications of zero degeneracy in networks spectra, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25, 4:043110 (2015).
- [4] Jalan, S., Yadav, A., Sarkar, C., Boccaletti, S., Unveiling the multi-fractal structure of complex networks, *Chaos, Solitons and Fractals* 97:11 (2017).
- [5] Biggs, N., Lloyd, E. K., Wilson, R. J. (1986). *Graph Theory, 1736-1936*. Oxford University Press.
(ISBN: 0198539169, 9780198539162)
- [6] M. van Steen (2010). *Graph Theory and Complex Networks: An Introduction*. Maarten van Steen,
(DOI:9dba/e30f8253791138e6c1031c5b7e4c7b321185.pdf).
- [7] Travers, J., Milgram, S. (1967). The small world problem. *Psychology Today*, 1(1), 61-67.
(DOI:<https://www.jstor.org/stable/pdf/2786545>).
- [8] Travers, J., Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, 32(4):425–443.
(DOI:<https://doi.org/10.1016/B978-0-12-442450-0.50018-3>)
- [9] Erdős, P., Rényi, A. (1959). On random graphs I. *Publ. Math. Debrecen*, 6, 290-297.

- (DOI:<http://www.leonidzhukov.net/hse/2016/networks/papers/erdos-1959-11.pdf>)
- [10] Erdős P., Rényi A. (1960), On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.*, 5 (1), 17-60.
(DOI:[4201/73e087bca0bdb31985e28ff69c60a129c8ef.pdf](https://doi.org/10.4201/73e087bca0bdb31985e28ff69c60a129c8ef.pdf))
- [11] Watts D.J., Strogatz S.H. (1998), Collective dynamics of small-world networks, *Nature*, 393, 440-442
(DOI: [10.1038/30918](https://doi.org/10.1038/30918)).
- [12] Barabási A.-L., Albert R. (1999), Emergence of Scaling in Random Networks, *Science*, 286, 509-512 (DOI: [10.1126/science.286.5439.509](https://doi.org/10.1126/science.286.5439.509)).
- [13] Hüchel, E. (2000). On the quantum theory of the double bond. In *Quantum Chemistry: Classic Scientific Papers* (pp. 452-478).
(DOI:https://doi.org/10.1142/9789812795762_0026).
- [14] Hüchel, E. (1931). Quantentheoretische beiträge zum benzolproblem. *Zeitschrift für Physik*, 70(3-4), 204-286.
(DOI:<https://doi.org/10.1007/BF01339530>).
- [15] Hüchel, E. (1931). Quantentheoretische beiträge zum benzolproblem. *Zeitschrift für Physik*, 72(5-6), 310-337.
(DOI:<https://doi.org/10.1007/BF01341953>).
- [16] Hüchel, E. (1932). Quantentheoretische Beiträge zum Problem der aromatischen und ungesättigten Verbindungen. III. *Zeitschrift für Physik*, 76(9-10), 628-648.
(DOI:<https://doi.org/10.1007/BF01341936>).
- [17] Hüchel, E. (1933). Die freien Radikale der organischen Chemie. *Zeitschrift für Physik*, 83(9-10), 632-668.
(DOI:<https://doi.org/10.1007/BF01330865>).
- [18] Albert, R., Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), 47.
(DOI:<https://doi.org/10.1103/RevModPhys.74.47>)
- [19] Boccaletti, S., Latora, V., Moreno, Y., *et al.* (2006). Complex networks: Structure and dynamics. *Physics reports*, 424(4-5), 175-308.
(DOI:<https://doi.org/10.1016/j.physrep.2005.10.009>)

- [20] Guare, J. (1990) *Six Degrees of Separation: A Play*, Vintage Books, New York (ISBN: 0679734813).
- [21] Bollobás, B. (1981). Degree sequences of random graphs. *Discrete Mathematics*, 33(1), 1-19.
(DOI: 10.1016/0012-365X(81)90253-3).
- [22] Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 4(2), 131-134.
(DOI: 10.1007/s100510050359).
- [23] Onnela, J. P., Saramäki, J., Hyvönen, J., *et al.*(2007). Structure and tie strengths in mobile communication networks. *Proceedings of the national academy of sciences*, 104(18), 7332-7336.
(DOI: 10.1073/pnas.0610245104).
- [24] Barabasi, A. L., Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2), 101.
(DOI: 10.1038/nrg1272).
- [25] Motter, A. E., Lai, Y. C. (2002). Cascade-based attacks on complex networks. *Physical Review E*, 66(6), 065102.
(DOI: 10.1103/PhysRevE.66.065102).
- [26] Lawrence S., Giles C.L. (1998), Searching the world wide web, *Science*, 280 (5360), 98-100.
(DOI: 10.1126/science.280.5360.98).
- [27] Newman, M.E.J., (2010) *Networks: an introduction*, Oxford University Press, Oxford (ISBN: 0199206651).
- [28] Chung, F.R.K., (1997) *Spectral Graph Theory*, American Mathematical Society (ISBN: 0821803158).
- [29] Seary A.J., Richards W.D. (2003) *Spectral Methods for Analyzing and Visualizing Networks: An Introduction*. In: Breiger R.L., Carley K., Pattison P. (ed) *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, National Academies Press, pp. 209-228 (ISBN: 0309089522).
- [30] Farkas, I. J., Derényi, I., Barabási, A. L., *et al.*(2001). Spectra of “real-world” graphs: Beyond the semicircle law. *Physical Review E*, 64(2), 026704.
(DOI: 10.1103/PhysRevE.64.026704).

- [31] Dorogovtsev, S. N., Goltsev, A. V., Mendes, J. F., *et al.*(2003). Spectra of complex networks. *Physical Review E*, 68(4), 046109.
(DOI: 10.1103/PhysRevE.68.046109).
- [32] Goh, K. I., Kahng, B., Kim, D. (2001). Spectra and eigenvectors of scale-free networks. *Physical Review E*, 64(5), 051903.
(DOI: 10.1103/PhysRevE.64.051903).
- [33] Chauhan, S., Girvan, M., Ott, E. (2009). Spectral properties of networks with community structure. *Physical Review E*, 80(5), 056114.
(DOI:https://doi.org/10.1103/PhysRevE.80.056114).
- [34] Eriksen, K. A., Simonsen, I., Maslov, S., *et al.*(2003). Modularity and extreme edges of the Internet. *Physical review letters*, 90(14), 148701.
(DOI: 10.1103/PhysRevLett.90.148701).
- [35] de Aguiar, M. A. M., Bar-Yam, Y. (2005). Spectral analysis and the dynamic response of complex networks. *Physical Review E*, 71(1), 016106.
(DOI: 10.1103/PhysRevE.71.016106).
- [36] Mieghem, P.Van., (2011) *Graph spectra for complex networks*, first ed. Cambridge University Press, New York, pp. 113-145 (ISBN: 052119458X).
- [37] Kamp, C., Christensen, K. (2005). Spectral analysis of protein-protein interactions in *Drosophila melanogaster*. *Physical Review E*, 71(4), 041911.
(DOI: 10.1103/PhysRevE.71.041911).
- [38] Marrec, L., Jalan, S. (2017). Analysing degeneracies in networks spectra. *EPL (Europhysics Letters)*, 117(4), 48001.
(DOI: 10.1209/0295-5075/117/48001).
- [39] Papenbrock, T., Weidenmüller, H. A. (2007). Colloquium: Random matrices and chaos in nuclear spectra. *Reviews of Modern Physics*, 79(3), 997.
(DOI: 10.1103/RevModPhys.79.997).
- [40] Mehta, M.L., (1991) *Random Matrices*, second ed. Academic Press, New York (ISBN: 047177796X).
- [41] Guhr, T., Müller–Groeling, A., Weidenmüller, H. A. (1998). Random-matrix theories in quantum physics: common concepts. *Physics Reports*, 299(4-6), 189-425.
(DOI: 10.1016/S0370-1573(97)00088-4).

- [42] Plerou, V., Gopikrishnan, P., Rosenow, B., *et al.*(1999). Universal and nonuniversal properties of cross correlations in financial time series. *Physical review letters*, 83(7), 1471.
(DOI: 10.1103/PhysRevLett.83.1471).
- [43] Šeba P. (2003), Random matrix analysis of human EEG data. *Physical review letters*, 91(19), 198104.
(DOI: 10.1103/PhysRevLett.91.198104).
- [44] Santhanam, M. S., Patra, P. K. (2001). Statistics of atmospheric correlations. *Physical Review E*, 64(1), 016102.
(DOI: 10.1103/PhysRevE.64.016102).
- [45] Podobnik, B., Wang, D., Horvatic, D., *et al.*(2010). Time-lag cross-correlations in collective phenomena. *EPL (Europhysics Letters)*, 90(6), 68001.
(DOI: 10.1209/0295-5075/90/68001).
- [46] Brody, T. A. (1973). A statistical measure for the repulsion of energy levels. *Lettere al Nuovo Cimento (1971-1985)*, 7(12), 482-484.
(DOI: 10.1007/BF02727859).
- [47] Bandyopadhyay, J. N., Jalan, S. (2007). Universality in complex networks: Random matrix analysis. *Physical Review E*, 76(2), 026109.
(DOI: 10.1103/PhysRevE.76.026109).
- [48] Palla, G., Lovász, L., Vicsek, T. (2010). Multifractal network generator. *Proceedings of the National Academy of Sciences of the United States of America*, 107(17), 7640-7645.
(DOI: 10.1073/pnas.0912983107).
- [49] Milo, R., Itzkovitz, S., Kashtan, N., *et al.*(2004). Superfamilies of evolved and designed networks. *Science*, 303(5663), 1538-1542.
(DOI: 10.1126/science.1089167).
- [50] Yeo, G. S., Heisler, L. K. (2012). Unraveling the brain regulation of appetite: lessons from genetics. *Nature neuroscience*, 15(10), 1343.
(DOI: 10.1038/nn.3211).
- [51] Gracia-Lázaro, C., Cuesta, J. A., Sánchez, A., *et al.*(2012). Human behavior in Prisoner's Dilemma experiments suppresses network reciprocity. *Scientific*

- reports, 2, 325.
(DOI: 10.1038/srep00325).
- [52] Sethi, N., Kang, Y. (2011). Unravelling the complexity of metastasis—molecular understanding and targeted therapies. *Nature Reviews Cancer*, 11(10), 735.
(DOI: 10.1038/nrc3125).
- [53] Cao, H., Li, Y. (2014). Unraveling chaotic attractors by complex networks and measurements of stock market complexity. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24(1), 013134.
(DOI: 10.1063/1.4868258).
- [54] Manovich, L. (2011). Trending: The promises and the challenges of big social data. *Debates in the digital humanities*, 2, 460-475.
(DOI: 10.5749/minnesota/9780816677948.003.0047).
- [55] Borgatti, S. P., Mehra, A., Brass, D. J., *et al.*(2009). Network analysis in the social sciences. *science*, 323(5916), 892-895.
(DOI: 10.1126/science.1165821).
- [56] Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23), 8577-8582.
(DOI: 10.1073/pnas.0601602103).
- [57] Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social networks*, 29(4), 555-564.
(DOI: 10.1016/j.socnet.2007.04.002).
- [58] Jalan, S., Bandyopadhyay, J. N. (2007). Random matrix analysis of complex networks. *Physical Review E*, 76(4), 046107.
(DOI: 10.1103/PhysRevE.76.046107).
- [59] Jalan, S., Bandyopadhyay, J. N. (2008). Random matrix analysis of network Laplacians. *Physica A: Statistical Mechanics and its Applications*, 387(2-3), 667-674.
(DOI: 10.1016/j.physa.2007.09.026).
- [60] Jalan, S., Solymosi, N., Vattay, G., *et al.*(2010). Random matrix analysis of localization properties of gene coexpression network. *Physical Review E*, 81(4), 046118.
(DOI: 10.1103/PhysRevE.81.046118).

- [61] Colizza, V., Flammini, A., Serrano, M. A., *et al.* (2006). Detecting rich-club ordering in complex networks. *Nature physics*, 2(2), 110.
(DOI: <https://doi.org/10.1038/nphys209>).
- [62] Soffer, S. N., Vazquez, A. (2005). Network clustering coefficient without degree-correlation biases. *Physical Review E*, 71(5), 057101.
(DOI: <https://doi.org/10.1103/PhysRevE.71.057101>).
- [63] Cohen, R., Havlin, S. (2003). Scale-free networks are ultrasmall. *Physical review letters*, 90(5), 058701.
(DOI: <https://doi.org/10.1103/PhysRevLett.90.058701>).
- [64] Newman, M. E. (2002). Assortative mixing in networks. *Physical review letters*, 89(20), 208701.
(DOI: <https://doi.org/10.1103/PhysRevLett.89.208701>)
- [65] Rivera, M. T., Soderstrom, S. B., Uzzi, B. (2010). Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *annual Review of Sociology*, 36, 91-115
(DOI: <https://doi.org/10.1146/annurev.soc.34.040507.134743>)
- [66] Newman, M. E. (2003). Mixing patterns in networks. *Physical Review E*, 67(2), 026126.
(DOI: <https://doi.org/10.1103/PhysRevE.67.026126>).
- [67] Barthélemy, M. (2003). Crossover from scale-free to spatial networks. *EPL (Europhysics Letters)*, 63(6), 915.
(DOI: <https://doi.org/10.1209/epl/i2003-00600-6>).
- [68] Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2), 167-256.
(DOI: <https://doi.org/10.1137/S003614450342480>).
- [69] Goh, K. I., Oh, E., Kahng, B., *et al.* (2003). Betweenness centrality correlation in social networks. *Physical Review E*, 67(1), 017101.
(DOI: <https://doi.org/10.1103/PhysRevE.67.017101>).
- [70] Catanzaro, M., Caldarelli, G., Pietronero, L. (2004). Assortative model for social networks. *Physical Review E*, 70(3), 037101.
(DOI: <https://doi.org/10.1103/PhysRevE.70.037101>).

- [71] Newman, M. E., Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E*, 68(3), 036122.
(DOI:<https://doi.org/10.1103/PhysRevE.68.036122>).
- [72] Lusseau, D., Newman, M. E. (2004). Identifying the role that animals play in their social networks. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(Suppl 6), S477-S481.
(DOI:<https://doi.org/10.1098/rsbl.2004.0225>).
- [73] Croft, D. P., James, R., Ward, A. J. W., *et al.* (2005). Assortative interactions and social networks in fish. *Oecologia*, 143(2), 211-219.
(DOI:<https://doi.org/10.1007/s00442-004-1796-8>).
- [74] Bollen, J., Gonçalves, B., Ruan, G., *et al.* (2011). Happiness is assortative in online social networks. *Artificial life*, 17(3), 237-251.
(DOI:https://doi.org/10.1162/artl_a_00034).
- [75] Bagler, G., Sinha, S. (2007). Assortative mixing in protein contact networks and protein folding kinetics. *Bioinformatics*, 23(14), 1760-1767.
(DOI:<https://doi.org/10.1093/bioinformatics/btm257>).
- [76] Bassett, D. S., Bullmore, E., Verchinski, B. A., *et al.* (2008). Hierarchical organization of human cortical networks in health and schizophrenia. *Journal of Neuroscience*, 28(37), 9239-9248.
(DOI:<https://doi.org/10.1523/JNEUROSCI.1929-08.2008>).
- [77] Pechenick, D. A., Payne, J. L., Moore, J. H. (2014). Phenotypic robustness and the assortativity signature of human transcription factor networks. *PLoS computational biology*, 10(8), e1003780.
(DOI:<https://doi.org/10.1371/journal.pcbi.1003780>).
- [78] Xulvi-Brunet, R., Sokolov, I. M. (2004). Reshuffling scale-free networks: From random to assortative. *Physical Review E*, 70(6), 066102.
(DOI:<https://doi.org/10.1103/PhysRevE.70.066102>).
- [79] Xulvi-Brunet, R., Sokolov, I. M. (2005). Changing correlations in networks: assortativity and disassortativity. *Acta Physica Polonica B*, 36(5), 1431.
(DOI:<https://www.actaphys.uj.edu.pl/fulltext?series=Reg&vol=36&page=1431>).
- [80] Berg, J., Lässig, M. (2002). Correlated random networks. *Physical review letters*, 89(22), 228701.
(DOI:<https://doi.org/10.1103/PhysRevLett.89.228701>).

- [81] Akemann, G., Baik, J., Di Francesco, P. (2011). The Oxford handbook of random matrix theory. Oxford University Press.(ISBN:9780199574001).
- [82] Cvetkovic, D., Simic, S., Rowlinson, P. (2009). An introduction to the theory of graph spectra. Cambridge University Press.(ISBN 9780521118392).
- [83] Van Mieghem, P., Omic, J., Kooij, R. (2009). Virus spread in networks. IEEE/ACM Transactions on Networking (TON), 17(1), 1-14. (DOI:10.1109/TNET.2008.925623).
- [84] Restrepo, J. G., Ott, E., Hunt, B. R. (2005). Onset of synchronization in large networks of coupled oscillators. Physical Review E, 71(3), 036151. (DOI:<https://doi.org/10.1103/PhysRevE.71.036151>)
- [85] Van Mieghem, P., Wang, H., Ge, X., *et al.*(2010). Influence of assortativity and degree-preserving rewiring on the spectra of networks. The European Physical Journal B, 76(4), 643-652. (DOI: <https://doi.org/10.1140/epjb/e2010-00219-x>)
- [86] Goltsev, A. V., Dorogovtsev, S. N., Oliveira, J. G.,*et al.*(2012). Localization and spreading of diseases in complex networks. Physical review letters, 109(12), 128702. (DOI:<https://doi.org/10.1103/PhysRevLett.109.128702>)
- [87] D'Agostino, G., Scala, A., Zlatić, V.,*et al.*(2012). Robustness and assortativity for diffusion-like processes in scale-free networks. EPL (Europhysics Letters), 97(6), 68006. (DOI:<https://doi.org/10.1209/0295-5075/97/68006>).
- [88] Palla, G., Vattay, G. (2006). Spectral transitions in networks. New Journal of Physics, 8(12), 307. (DOI:<https://doi.org/10.1088/1367-2630/8/12/307>).
- [89] Potestio, R., Caccioli, F., Vivo, P. (2009). Random matrix approach to collective behavior and bulk universality in protein dynamics. Physical review letters, 103(26), 268101. (DOI:<https://doi.org/10.1103/PhysRevLett.103.268101>).
- [90] Jalan, S., Sarkar, C., Madhusudanan, A.,*et al.* (2014). Uncovering randomness and success in society. PloS one, 9(2), e88249. (DOI: <https://doi.org/10.1371/journal.pone.0088249>).

- [91] Sompolinsky, H., Crisanti, A., Sommers, H. J. (1988). Chaos in random neural networks. *Physical review letters*, 61(3), 259.
(DOI:<https://doi.org/10.1103/PhysRevLett.61.259>).
- [92] Rajan, K., Abbott, L. F. (2006). Eigenvalue spectra of random matrices for neural networks. *Physical review letters*, 97(18), 188104.
(DOI:<https://doi.org/10.1103/PhysRevLett.97.188104>).
- [93] May, R. M. (1972). Will a large complex system be stable?. *Nature*, 238(5364), 413.
(DOI:<https://doi.org/10.1038/238413a0>).
- [94] Chauhan, S., Girvan, M., Ott, E. (2009). Spectral properties of networks with community structure. *Physical Review E*, 80(5), 056114.
(DOI:<https://doi.org/10.1103/PhysRevE.80.056114>).
- [95] Watanabe, T., Masuda, N. (2010). Enhancing the spectral gap of networks by node removal. *Physical Review E*, 82(4), 046102.
(DOI:<https://doi.org/10.1103/PhysRevE.82.046102>).
- [96] Peixoto, T. P. (2013). Eigenvalue spectra of modular networks. *Physical review letters*, 111(9), 098701.
(DOI:<https://doi.org/10.1103/PhysRevLett.111.098701>).
- [97] Rodgers, G. J., Bray, A. J. (1988). Density of states of a sparse random matrix. *Physical Review B*, 37(7), 3557.
(DOI:<https://doi.org/10.1103/PhysRevB.37.3557>).
- [98] Fyodorov, Y. V., Mirlin, A. D. (1991). On the density of states of sparse random matrices. *Journal of Physics A: Mathematical and General*, 24(9), 2219.
(DOI:<https://doi.org/10.1088/0305-4470/24/9/027>).
- [99] Evangelou, S. N. (1992). A numerical study of sparse random matrices. *Journal of statistical physics*, 69(1-2), 361-383.
(DOI:<https://doi.org/10.1007/BF01053797>).
- [100] Rogers, T., Castillo, I. P., Kühn, R., *et al.*(2008). Cavity approach to the spectral density of sparse symmetric random matrices. *Physical Review E*, 78(3), 031116.
(DOI:<https://doi.org/10.1103/PhysRevE.78.031116>).

- [101] Dumitriu, I., Pal, S. (2012). Sparse regular random graphs: spectral density and eigenvectors. *The Annals of Probability*, 40(5), 2197-2235.
(DOI:10.1214/11-AOP673).
- [102] Jalan, S., Bandyopadhyay, J. N. (2009). Randomness of random networks: A random matrix analysis. *EPL (Europhysics Letters)*, 87(4), 48010.
(DOI:https://doi.org/10.1209/0295-5075/87/48010).
- [103] Pandey, A., Mehta, M. L. (1983). Gaussian ensembles of random Hermitian matrices intermediate between orthogonal and unitary ones. *Communications in Mathematical Physics*, 87(4), 449-468.
(DOI:https://doi.org/10.1007/BF01208259).
- [104] Bohigas, O., Giannoni, M. J. (1975). Level density fluctuations and random matrix theory. *Annals of Physics*, 89(2), 393-422.
(DOI:https://doi.org/10.1016/0003-4916(75)90187-6).
- [105] French, J. B., Mello, P. A., Pandey, A. (1978). Ergodic behavior in the statistical theory of nuclear reactions. *Physics Letters B*, 80(1-2), 17-19.
(DOI:https://doi.org/10.1016/0370-2693(78)90294-0).
- [106] Agrawal, A., Sarkar, C., Dwivedi, S. K. *et al.* (2014). Quantifying randomness in protein-protein interaction networks of different species: A random matrix approach. *Physica A: Statistical Mechanics and its Applications*, 404, 359-367.
(DOI: https://doi.org/10.1016/j.physa.2013.12.005).
- [107] Tenenbaum, J. N., Havlin, S., Stanley, H. E. (2012). Earthquake networks based on similar activity patterns. *Physical Review E*, 86(4), 046107.
(DOI:https://doi.org/10.1103/PhysRevE.86.046107)
- [108] Avalos-Gaytán, V., Almendral, J. A., Papo, D., *et al.* (2012). Assortative and modular networks are shaped by adaptive synchronization processes. *Physical Review E*, 86(1), 015101.
(DOI:https://doi.org/10.1103/PhysRevE.86.015101)
- [109] Papadopoulos, F., Kitsak, M., Serrano, M. Á., *et al.* (2012). Popularity versus similarity in growing networks. *Nature*, 489(7417), 537.
(DOI:https://doi.org/10.1038/nature11459)
- [110] Smith, J. A., McPherson, M., Smith-Lovin, L. (2014). Social distance in the United States: Sex, race, religion, age, and education homophily among

- confidants, 1985 to 2004. *American Sociological Review*, 79(3), 432-456.
(DOI:<https://doi.org/10.1177/0003122414531776>)
- [111] Petrov, D. A. (2002). DNA loss and evolution of genome size in *Drosophila*. *Genetica*, 115(1), 81-91.
(DOI:<https://doi.org/10.1023/A:1016076215168>)
- [112] Teichmann, S. A., Babu, M. M. (2004). Gene regulatory network growth by duplication. *Nature genetics*, 36(5), 492.
(DOI:<https://doi.org/10.1038/ng1340>)
- [113] Barclay, V. C., Smieszek, T., He, J., *et al.* (2014). Positive network assortativity of influenza vaccination at a high school: implications for outbreak risk and herd immunity. *PloS one*, 9(2), e87042.
(DOI: <https://doi.org/10.1371/journal.pone.0087042>)
- [114] Alstott, J., Madnick, S., Velu, C. (2014). Homophily and the speed of social mobilization: the effect of acquired and ascribed traits. *PloS one*, 9(4), e95140.
(DOI:<https://doi.org/10.1371/journal.pone.0095140>)
- [115] Longo, G., Montévil, M. (2012). Randomness increases order in biological evolution. In *Computation, Physics and Beyond*. Springer, Berlin, Heidelberg, pp. 289-308.
(DOI:https://doi.org/10.1007/978-3-642-27654-5_22)
- [116] Seiler, M. C., Seiler, F. A. (1989). Numerical recipes in C: the art of scientific computing. *Risk Analysis*, 9(3), 415-416.
(DOI:<https://doi.org/10.1111/j.1539-6924.1989.tb01007.x>).
- [117] Barrat, A., Barthelemy, M., Vespignani, A. (2008). *Dynamical processes on complex networks*. Cambridge university press.(ISBN:0521879507 9780521879507)
- [118] Strogatz, S. H. (2001). Exploring complex networks. *nature*, 410(6825), 268.
(DOI:<https://doi.org/10.1038/35065725>).
- [119] Nadakuditi, R. R., Newman, M. E. (2012). Graph spectra and the detectability of community structure in networks. *Physical review letters*, 108(18), 188701.
(DOI:<https://doi.org/10.1103/PhysRevLett.108.188701>)

- [120] Banerjee, A., Jost, J. (2008). On the spectrum of the normalized graph Laplacian. *Linear algebra and its applications*, 428(11-12), 3015-3022.
(DOI:<https://doi.org/10.1016/j.laa.2008.01.029>)
- [121] Poole, D. (2014). *Linear algebra: A modern introduction*. Cengage Learning.(ISBN:1285463242, 9781285463247)
- [122] Molloy, M., Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3), 161-180.
(DOI:<https://doi.org/10.1002/rsa.3240060204>)
- [123] Newman, M. E., Strogatz, S. H., Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical review E*, 64(2), 026118.
(DOI:<https://doi.org/10.1103/PhysRevE.64.026118>)
- [124] Del Genio, C. I., Kim, H., Toroczkai, Z., Bassler, K. E. (2010). Efficient and exact sampling of simple graphs with given arbitrary degree sequence. *PloS one*, 5(4), e10012.
(DOI:<https://doi.org/10.1371/journal.pone.0010012>)
- [125] Tononi, G., Sporns, O., Edelman, G. M. (1999). Measures of degeneracy and redundancy in biological networks. *Proceedings of the National Academy of Sciences*, 96(6), 3257-3262.
(DOI:<https://doi.org/10.1073/pnas.96.6.3257>)
- [126] Hurles, M. (2004). Gene duplication: the genomic trade in spare parts. *PLoS biology*, 2(7), e206.
(DOI:<https://doi.org/10.1371/journal.pbio.0020206>)
- [127] Brenner, S. E., Hubbard, T., Murzin, A., Chothia, C. (1995). Gene duplications in *H. influenzae*. *Nature*, 378, 140.
(DOI:10.1038/378140a0).
- [128] Teichmann, S. A., Park, J., Chothia, C. (1998). Structural assignments to the *Mycoplasma genitalium* proteins show extensive gene duplications and domain rearrangements. *Proceedings of the National Academy of Sciences*, 95(25), 14658-14663.
(DOI:<https://www.pnas.org/content/95/25/14658>).
- [129] Gough, J., Karplus, K., Hughey, R., *et al.* (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all

- proteins of known structure1. *Journal of molecular biology*, 313(4), 903-919.
(DOI:<https://doi.org/10.1006/jmbi.2001.5080>)
- [130] Gu, Z., Steinmetz, L. M., Gu, X., *et. al.*(2003). Role of duplicate genes in genetic robustness against null mutations. *Nature*, 421(6918), 63.
(DOI:<https://doi.org/10.1038/nature01198>)
- [131] Kitano, H. (2004). Biological robustness. *Nature Reviews Genetics*, 5(11), 826.
(DOI:<https://doi.org/10.1038/nrg1471>)
- [132] Nowak, M. A., Boerlijst, M. C., Cooke, J., *et. al.*(1997). Evolution of genetic redundancy. *Nature*, 388(6638), 167.
(DOI:<https://doi.org/10.1038/40618>)
- [133] Mandelbrot, B. B. (1982). *The fractal geometry of nature (Vol. 1)*. New York: WH freeman.(ISBN-10:0716711869)
- [134] Mandelbrot, B. B., Passoja, D. E., Paullay, A. J. (1984). Fractal character of fracture surfaces of metals. *Nature*, 308(5961), 721.
(DOI:<https://doi.org/10.1038/308721a0>)
- [135] Hentschel, H. G. E., Procaccia, I. (1983). The infinite number of generalized dimensions of fractals and strange attractors. *Physica D: Nonlinear Phenomena*, 8(3), 435-444.
(DOI:[https://doi.org/10.1016/0167-2789\(83\)90235-X](https://doi.org/10.1016/0167-2789(83)90235-X))
- [136] Egolf, D. A., Greenside, H. S. (1994). Relation between fractal dimension and spatial correlation length for extensive chaos. *Nature*, 369(6476), 129.
(DOI:<https://doi.org/10.1038/369129a0>).
- [137] Mandelbrot, B. B. (1974). Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics*, 62(2), 331-358.
(DOI:<https://doi.org/10.1017/S0022112074000711>)
- [138] Lopes, R., Betrouni, N. (2009). Fractal and multifractal analysis: a review. *Medical image analysis*, 13(4), 634-649.
(DOI:<https://doi.org/10.1016/j.media.2009.05.003>)
- [139] Wu, K. K., Lahav, O., Rees, M. J. (1999). The large-scale smoothness of the Universe. *Nature*, 397(6716), 225.
(DOI: <https://doi.org/10.1038/16637>)

- [140] Sáiz, A., Martínez, V. J. (1996). Multifractal analysis of the atomic spectral line series. *Physical Review E*, 54(3), 2431.
(DOI:<https://doi.org/10.1103/PhysRevE.54.2431>)
- [141] Song, C., Havlin, S., Makse, H. A. (2005). Self-similarity of complex networks. *Nature*, 433(7024), 392.
(DOI: <https://doi.org/10.1038/nature03248>).
- [142] Song, C., Gallos, L. K., Havlin, S., *et. al.* (2007). How to calculate the fractal dimension of a complex network: the box covering algorithm. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(03), P03006.
(DOI:<https://doi.org/10.1088/1742-5468/2007/03/P03006>).
- [143] Kim, J. S., Goh, K. I., Salvi, G., *et. al.* (2007). Fractality in complex networks: critical and supercritical skeletons. *Physical Review E*, 75(1), 016110.
(DOI:<https://doi.org/10.1103/PhysRevE.75.016110>)
- [144] Jung, S., Kim, S., Kahng, B. (2002). Geometric fractal growth model for scale-free networks. *Physical Review E*, 65(5), 056101.
(DOI:<https://doi.org/10.1103/PhysRevE.65.056101>)
- [145] Grassberger, P., Procaccia, I. (1983). Characterization of strange attractors. *Physical review letters*, 50(5), 346.
(DOI:<https://doi.org/10.1103/PhysRevLett.50.346>)
- [146] Barabási, A. L., Szèpfalusy, P., Vicsek, T. (1991). Multifractal spectra of multi-affine functions. *Physica A: Statistical Mechanics and its Applications*, 178(1), 17-28. (DOI : [https://doi.org/10.1016/0378-4371\(91\)90072-K](https://doi.org/10.1016/0378-4371(91)90072-K))
- [147] Kantelhardt, J. W., Roman, H. E., Greiner, M. (1995). Discrete wavelet approach to multifractality. *Physica A: Statistical Mechanics and its Applications*, 220(3-4), 219-238. (DOI:[https://doi.org/10.1016/0378-4371\(95\)00267-B](https://doi.org/10.1016/0378-4371(95)00267-B))
- [148] the reader is addressed to the Supplementary Material for the detailed derivations of the probability distributions (and the numerical calculations of fractal dimensions) for a random reshuffling case, as well as matrix diagrams for various types of nodes' index reshufflings.
- [149] Albert, R. (2005). Scale-free networks in cell biology. *Journal of cell science*, 118(21), 4947-4957. (DOI: [10.1242/jcs.02714](https://doi.org/10.1242/jcs.02714))

- [150] Jain, S., Kumar, A., Mandal, S., *et. al.*. (2013, August). B4: Experience with a globally-deployed software defined WAN. In ACM SIGCOMM Computer Communication Review, Vol. 43, No. 4, pp. 3-14. ACM. (DOI: 10.1145/2534169.2486019)
- [151] Nygren, E., Sitaraman, R. K., Sun, J. (2010). The akamai network: a platform for high-performance internet applications. ACM SIGOPS Operating Systems Review, 44(3), 2-19. (DOI:10.1145/1842733.1842736)
- [152] Kempe, D., Kleinberg, J., É. Tardos.(2015). Bibliography with links to cited articles. Theory of Computing, 11(4), 105-147. (DOI:10.4086/toc.2015.v011a004).