Spectral properties of random geometric graphs

C. P. Dettmann, O. Georgiou, G. Knight
University of Bristol, UK

Oxford, April 8, 2019

Outline

1. A spatial network model: the random geometric graph (RGG).
2. The spectral density of the adjacency matrix and its peaks
3. Random matrix theory and correlations in the spectrum.

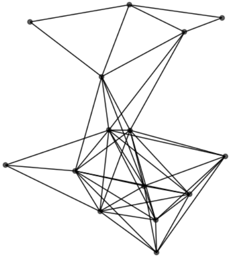
Spatially embedded networks

SPATIALLY EMBEDDED NETWORKS

- EPSRC-funded project investigating spatial networks with application to wireless communications
- Led by Justin Coon (Oxford) and CPD (Bristol).
- Find out more : www.eng.ox.ac.uk/sen/

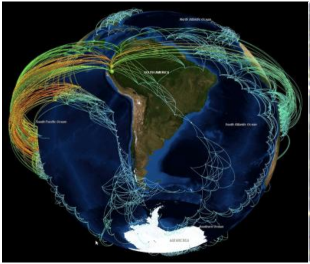
Networks

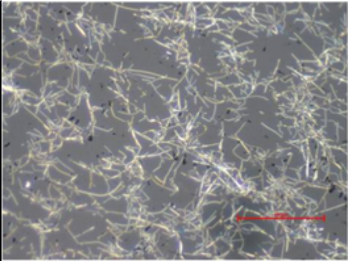
- A network consists of a set of nodes joined by edges.

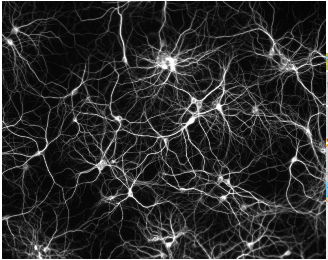


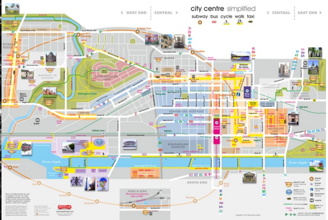
- Model for many types of *complex system*.
- Nodes: People, computers, stations, neurons...
- Edges: Relationships, contact, trips, synapses...

Spatial networks



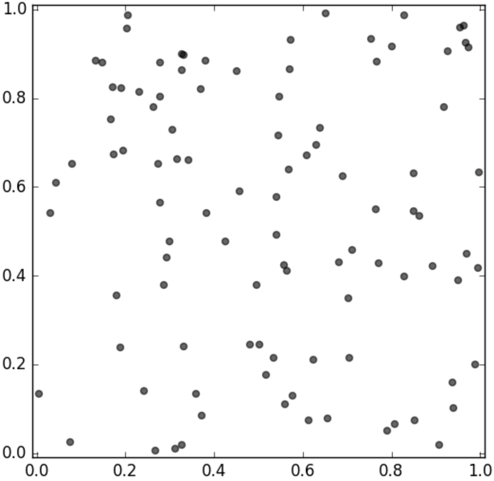






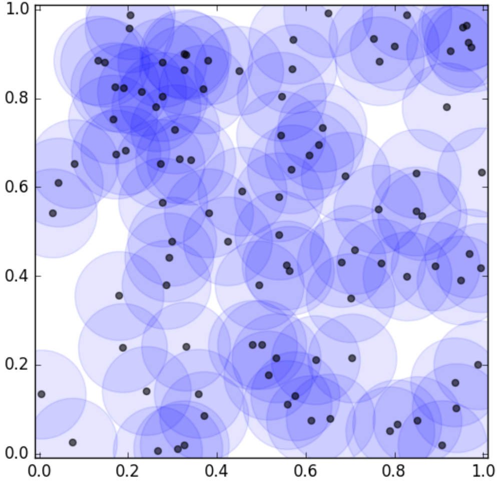
Random geometric graph (RGG)

- Nodes are distributed uniformly at random.



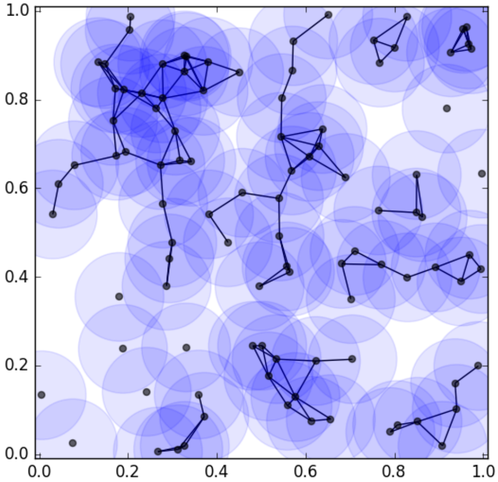
Random geometric graph

- Nodes are equipped with a connection radius.

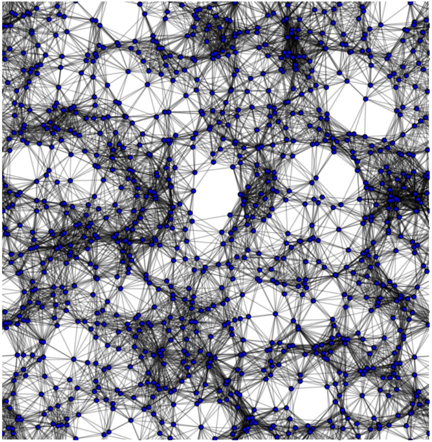


Random geometric graph

- Edges are made when nodes are within connection range.



RGG

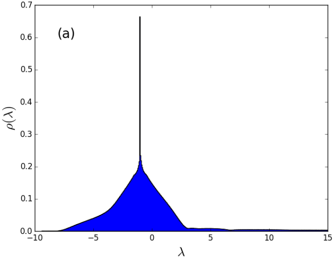


- We study RGGs with periodic boundary conditions.
- RGG with $N = 10^3$ and $r = 0.09375$ on the torus.

Adjacency matrix

- The zero-one, $N \times N$ adjacency matrix $\mathbf{A}$ has entries $a_{ij} = 1$ if there is a connection between nodes i and j, zero otherwise.
- $\mathbf{A}$ real and symmetric, its spectrum consists of real eigenvalues $\lambda_i, i = 1, .., N$ with $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_N$.

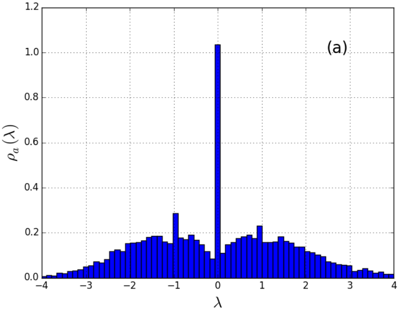
Ensemble-averaged adjacency spectral density.



- $N = 10^3$. (a): $r = 0.09375$, mean degree 28 and connected. (b): $r = 0.3$, mean degree 283.
- Clear peak at -1 for both r values.

Peaks in adjacency spectrum

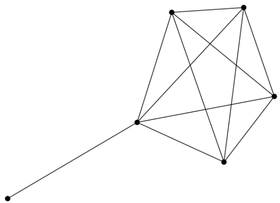
- Peaks are found in many *real-world* networks.
- Spectral density of the adjacency matrix of the Western States Power Grid of the United States.
- Peaks not common in other random graph models.



Symmetric motifs

- A network *motif* containing symmetric nodes, gives rise to eigenvalue multiplicities.
- Subgraph whose vertices are invariant under permutation.
- When the vertices are connected *Type-I orbits*.
- When disconnected *Type-II orbits*
- *Network redundancy*, nodes with identical roles.
- Eigenvectors localise on these symmetric nodes.

Type-I



Type-II

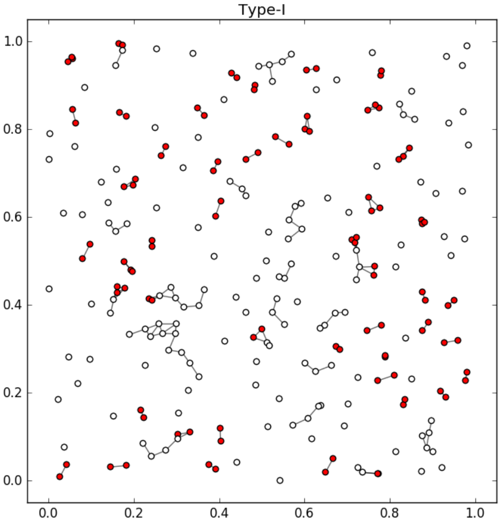
Symmetry in the adjacency matrix

- Consider two symmetric nodes n_1 and n_2 connected (type-I) with adjacency matrix

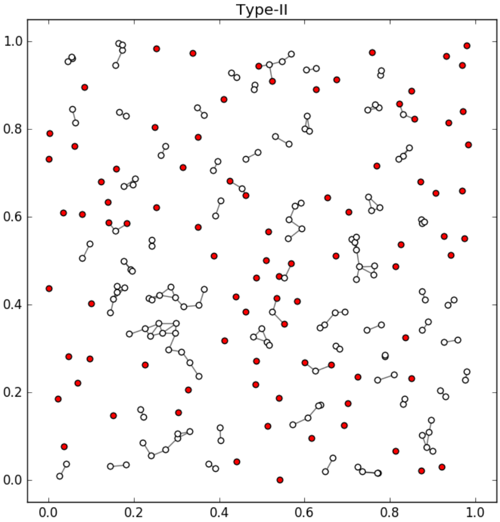
$$\left(\begin{array}{cc|c} 0 & 1 & \dots \\ 1 & 0 & \dots \\ \hline 1 & 1 & \\ 0 & 0 & \\ \vdots & \vdots & \end{array}\right) \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ \vdots \end{pmatrix} = -1 \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

- If n_1, n_2 are not connected by an edge (type-II), get eigenvalue 0.

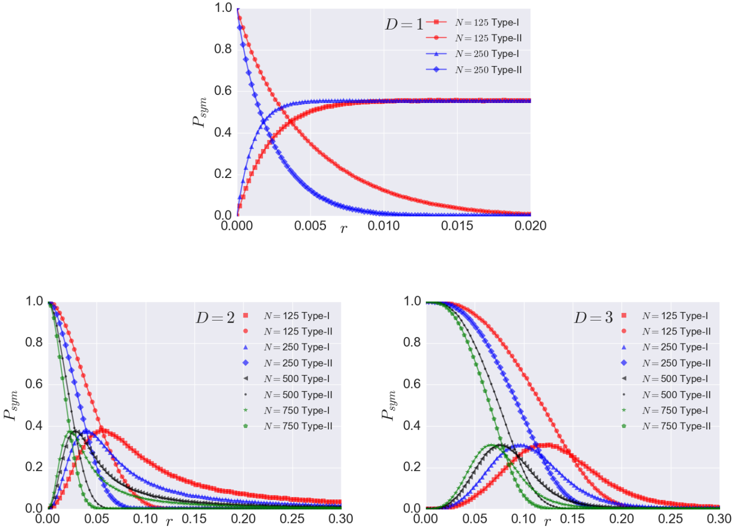
Type-I symmetry



Type-II symmetry



Symmetry probabilities by dimension



Minimal distance of two nodes

There are of order N^2 internode distances. We can show using the Chen-Stein method that the smallest distance $s_{min} \sim C_D N^{-2/D}$ for some constant C_D.

We can then find the probability that the region around this closest pair, $\mathcal{N}_{ex}$, is empty, leading to a Type-I symmetry.

Limits

$N \to \infty$ such that...

r is constant: **Intensive limit**

$r = CN^{-1/D}$ and mean degree constant: **Thermodynamic limit**

D=1

$\mathbb{P}(N(\mathcal{N}_{ex}) = 0) = \left(1 - \frac{2C_1}{N^2}\right)^{N-2} \to 1$

This holds in either the intensive limit (r const) or the thermodynamic limit: Lots of symmetric motifs.

D=2

$$\mathbb{P}(N(\mathcal{N}_{ex})=0)=\left(1-4r^2\sin^{-1}\left(\frac{C_2}{2rN}\right)-\frac{2C_2}{N}\sqrt{r^2-\frac{C_2^2}{4N^2}}\right)^{N-2}$$

Intensive limit: $\mathbb{P}(N(\mathcal{N}_{ex})=0)\to e^{-4C_2 r}$

Thermodynamic limit: $\mathbb{P}(N(\mathcal{N}_{ex})=0)\to 1$

D=3

$$\mathbb{P}(N(\mathcal{N}_{ex}) = 0) = \left(1 - 2\pi r^2 C_3 N^{-\frac{2}{3}} + \frac{\pi}{6} C_3 N^{-2}\right)^{N-2}$$

Intensive limit: $\mathbb{P}(N(\mathcal{N}_{ex}) = 0) \to 0$

Thermodynamic limit: $\mathbb{P}(N(\mathcal{N}_{ex}) = 0) \to 1$

Spectral density

- What about the rest of the spectral density?
- How does it compare with non-spatial random networks?

Random matrix theory of complex networks

PHYSICAL REVIEW E **76**, 046107 (2007)

Random matrix analysis of complex networks

Sarika Jalan* and Jayendra N. Bandyopadhyay†
Max-Planck Institute for the Physics of Complex Systems, Nöthnitzerstrasse 38, D-01187 Dresden, Germany
(Received 18 March 2007; published 12 October 2007)

We study complex networks under random matrix theory (RMT) framework. Using nearest-neighbor and next-nearest-neighbor spacing distributions we analyze the eigenvalues of the adjacency matrix of various model networks, namely, random, scale-free, and small-world networks. These distributions follow the Gaussian orthogonal ensemble statistic of RMT. To probe long-range correlations in the eigenvalues we study spectral rigidity via the Δ_3 statistic of RMT as well. It follows RMT prediction of linear behavior in semi-logarithmic scale with the slope being $\sim 1/\pi^2$. Random and scale-free networks follow RMT prediction for very large scale. A small-world network follows it for sufficiently large scale, but much less than the random and scale-free networks.

DOI: 10.1103/PhysRevE.76.046107 PACS number(s): 89.75.Hc, 64.60.Cn, 89.20.−a

- Use ideas from RMT to study complex networks.
- Random networks (ER)$N = 2000, p_{edge} = 0.01, \langle d \rangle = 20$.
- Scale-free (BA) $N = 2000, \langle d \rangle = 20$.
- Small-world (WS) $N = 2000, \langle d \rangle = 20, p_{rewire} = 0.005$
- Find *universality* in the statistics (Bandyopadhyay, Jalan '07, Mendez-Bermudez et al '15). What about RGGs?

Random matrix theory background

- Gaussian Orthogonal Ensemble (GOE): Real, symmetric random matrices whose elements are Gaussian distributed rvs.
- For GOE the nearest neighbour spacing distribution (NNSD) $P(s)$ is given by the Wigner-Dyson formula

$$P(s) \approx \frac{\pi}{2} s e^{-\frac{\pi s^2}{4}}$$

- No correlation $P(s)$ is Poisson distribution

$$P(s) = e^{-s}$$

- Interpolating between these is the (empirical) Brody distribution.

$$P_\beta(s) = (\beta + 1)\alpha s^\beta e^{-\alpha s^{\beta+1}}$$

$$\alpha = \left(\frac{\Gamma(\beta + 2)}{\Gamma(\beta + 1)} \right)^{\beta+1}$$

- $\Gamma()$ Gamma function. $\beta = 0$ Poisson, $\beta = 1$ Wigner-Dyson.

Complex networks.

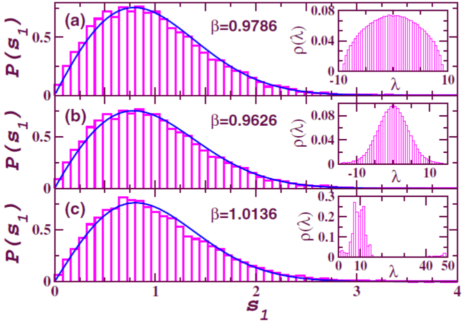


FIG. 1. (Color online) Nearest-neighbor spacings distribution (NNSD) $P(s_1)$ of the adjacency matrices of different networks [(a) random network, (b) scale-free network, and (c) small-world network]. All follow GOE statistics. The histograms are numerical results and the solid lines represent the fitted Brody distribution [Eq. (3a)]. All networks have $N=2000$ nodes and an average degree $k=20$ per node. Figures are plotted for the average over ten random realizations of the networks. Insets show respective spectral densities $\rho(\lambda)$.

Analytical results

- L. Erdős, A. Knowles, H.-T. Yau, J. Yin (2012,2013)
- Local semi-circle law, was proven for E-R graphs under the restriction $pN \to \infty$ (with at least logarithmic speed in N)
- This used to prove the presence of GOE statistics in the level spacings of E-R graphs.

Applications

- A. Rai, A. V. Menon, and S. Jalan (2014)
- RMT framework used to differentiate between cancerous and healthy protein networks.
- Nodes are proteins, edges are interactions.

- To analyse the spectrum we need to *unfold* the eigenvalues.
- Unfolding removes effects due to spectral density.
- Spectral function which for a given *energy* E is defined as

$$S(E) = \sum_{i=1}^{N} \delta(E - \lambda_i)$$

Unfolding Eigenvalues

- Cumulative spectral function counts how many $\leq E$

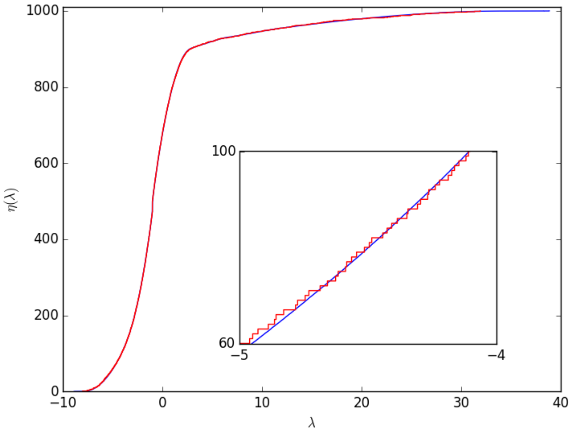
$$\eta(E) = \int_{-\infty}^{E} S(x)dx = \sum_{i=1}^{N} \Theta(E - \lambda_i)$$

- Unfolding defined via cumulative mean spectral function

$$\overline{\lambda}_i = \langle \eta(E) \rangle |_{E=\lambda_i}$$

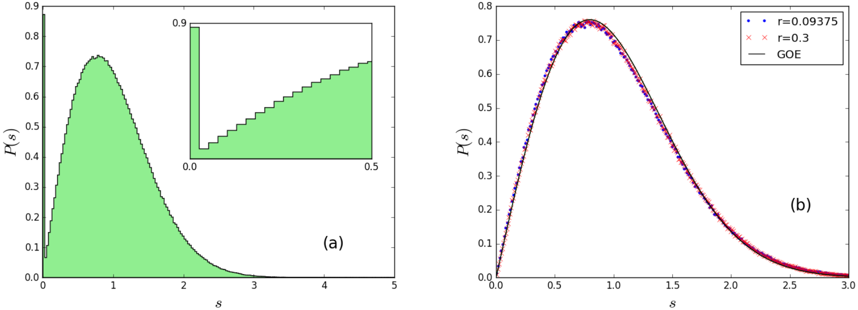
- $s_i = \overline{\lambda}_{i+1} - \overline{\lambda}_i$. $P(s)$ is the distribution of the s_i. Nearest neighbour spacing distribution NNSD. $\langle s_i \rangle = 1$.

Unfolding Eigenvalues



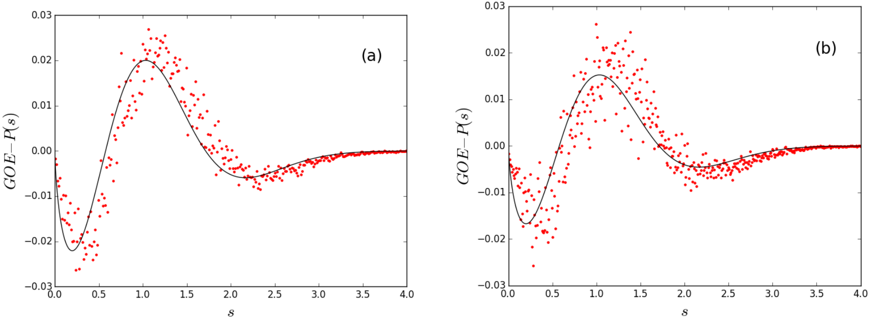
- Cumulative mean spectral function (blue). Ensemble of $N = 10^3$ RGGs, $r = 0.09375$. Cumulative spectral density of single RGG (red).

NNSD



- $P(s)$ from ensemble of $N = 10^3$ RGGs $r = 0.09375$ (a). Note the peak at zero.In (b) we compare with GOE statistics.

Brody distribution



- The difference between NNSD and GOE for $r = 0.09375$ (a) and $r = 0.3$ (b) (red dots). Also difference between GOE and the Brody distribution. Fit value of $\beta = 0.941$ (a) and $\beta = 0.955$ (b) (black lines).

nNNSD

- *Next* nearest neighbour spacing distribution.
- $s_2^i = (\overline{\lambda}_{i+2} - \overline{\lambda}_i)/2$, $P(s_2)$ their distribution.
- nNNSD of GOE is given by the NNSD of *Gaussian symplectic ensemble* matrices (GSE)

$$P(s_2) \approx \frac{2^{18}}{3^6 \pi^3} s_2^4 e^{-\frac{64}{9\pi} s_2^2}$$

- Result: As for NNSD, very close to GOE for non-spatial random networks and RGG.

Spectral rigidity

- Δ_3 statistic,(Dyson Mehta 1963), which measures long range correlation over distance L.
- $\Delta_3(L,x)$ measures the least-square deviation of the unfolded spectral staircase function $\overline{\eta}$ to the line of best fit over the interval $[x, x+L]$.

$$\Delta_3(L,x) = \frac{1}{L} \min_{A,B} \int_x^{x+L} \left(\overline{\eta}(\overline{\lambda}) - A\overline{\lambda} - B\right)^2 d\overline{\lambda}.$$

- $\overline{\eta}$ counts how many unfolded eigenvalues there are less than or equal to a given value

$$\overline{\eta}(E) = \sum_{i=1}^{N} \Theta(E - \overline{\lambda}_i).$$

- The average over non-intersecting intervals of length L $\langle ... \rangle_x$ is then the spectral rigidity $\Delta_3(L)$.

$$\langle \Delta_3(L,x) \rangle_x = \Delta_3(L).$$

Spectral rigidity

- Full correlation, equal spacings, *picket fence* spectrum, no L dependence.

$$\Delta_3(L) = \frac{1}{12}.$$

- Uncorrelated, Poisson statistics, linear dependence on L

$$\Delta_3(L) = \frac{L}{15}.$$

- GOE statistics, logarithmic dependence on L. For large L

$$\Delta_3(L) \simeq \frac{1}{\pi^2}\left(\ln(2\pi L) + \gamma - \frac{5}{4} - \frac{\pi^2}{8}\right),$$

to order $1/L$, γ is Euler's constant.

Evaluating Δ_3

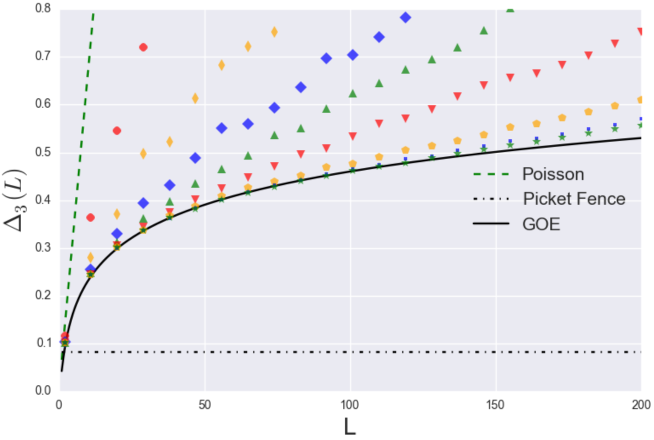
- Analytically evaluate $\Delta_3(L,x)$ (Bohigas Giannoni 1975) for experimentally obtained sequence.
- Centre interval $[x, x+L]$ at the origin. Transform n unfolded eigenvalues in the interval $\overline{\lambda}_i, \overline{\lambda}_{i+1}, ..., \overline{\lambda}_{i+n-1}$

$$\hat{\lambda}_j = \overline{\lambda}_{i-1+j} - \left(x + \frac{L}{2}\right),$$

- After transformation we can use

$$\Delta_3(L,x) = \frac{n^2}{16} - \frac{1}{L^2}\left(\sum_{j=1}^{n}\hat{\lambda}_j\right)^2 + \frac{3n}{2L^2}\left(\sum_{j=1}^{n}\hat{\lambda}_j^2\right) - \frac{3}{L^4}\left(\sum_{j=1}^{n}\hat{\lambda}_j^2\right)^2 + \frac{1}{L}\left(\sum_{j=1}^{n}(n-2j+1)\hat{\lambda}_j\right).$$

Spectral rigidity



- Spectral rigidity of 10^3 node RGGs.

Interpretation

- RGGs follow GOE statistics up to some value L_0 and then deviate towards Poisson statistics.
- L_0 has been related to community structure (Jalan 2009) and randomness of connections (Jalan and Bandyopadhyay 2009), for example rewiring probability in regular networks.

Summary

- We used RMT statistics to study the adjacency spectrum in RGGs.
- Short range correlations: same universality class (GOE) as non-spatial complex networks.
- Long range correlations: deviations towards Poisson.
- Future: Different connection functions, continuum limit

Further reading

- C. P. Dettmann, O. Georgiou and G. Knight, *Spectral statistics of random geometric graphs*, EPL **118**, 18003 (2017).
- C. P. Dettmann and G. Knight, *Symmetric motifs in random geometric graphs*, J. Complex Networks, **6**, 95-105 (2018).