Spectral analysis for gene communities in cancer cells

Ayumi Kikkawa*,

Mathematical and Theoretical Physics Unit,
Okinawa Institute of Science and Technology Graduate University,
1919-1 Tancha, Onna-son, Kunigami-gun, Okinawa, 904-0495
Japan
*Corresponding author: akikkawa@oist.jp

September 10, 2020

Abstract

We investigate gene interaction networks in various cancer cells by spectral analysis of the adjacency matrices. We observe the localization of the networks on hub genes, which have an extraordinary number of links. The eigenvector centralities take finite values only on special nodes when the hub degree exceeds the critical value $d_c \simeq 40$. The degree correlation function shows the disassortative behavior in the large degrees, and the nodes whose degrees are $d \gtrsim 40$ have a tendency to link to small degree nodes. The communities of the gene networks centered at the hub genes are extracted based on the amount of node degree discrepancies between linked nodes. We verify the Wigner-Dyson distribution of the nearest neighbor eigenvalues spacing distribution $P(s)$ in the small degree discrepancy communities (the assortative communities), and the Poisson $P(s)$ in the communities of large degree discrepancies (the disassortative communities) including the hubs. graph spectra; vector centrality; adjacency matrix; community structure; random matrix model

1 Introduction

In recent decades, much work has been done to probe network structures in large fields including the internet, social networks and biological networks. Among them, the investigations of the spectra of various matrices expressing the graphical topologies of the networks have been under enthusiastic study. A graph that represents a network consists of nodes (or vertices) and links (or edges). The adjacency matrix of size N, where N is the number of nodes in the graph, is constructed by setting its (i, j) element as $A_{ij} = 1$ if there is a link between the nodes i and j, or $A_{ij} = 0$ otherwise. In the case of directed networks, the links are replaced by arrows. In this paper, we discuss the undirected networks. In the undirected networks, the adjacency matrix A becomes a real symmetric matrix. All eigenvalues λ_i are real and, in the large-N limit, their distribution $\rho(\lambda)$ shows the famous Wigner's semi-circle independent of the graph details [1, 2].

$$\rho(\lambda) \propto \sqrt{2N - \lambda^2} \tag{1}$$

This universal behavior in the large-N limit obtained from the random matrix theory requires that the ratio L/N, where L is the number of links, is constant. The networks are considerably dense in this limit with large L.

In contrast to this universality, the spectra of adjacency matrices or other graph representative matrices of real-world networks are somewhat different. Since interaction resources are restricted, the number of links should be limited. Therefore, the large-N limit of the complex real-world networks is expected to lead to a very sparse network. On the other hand, the spectra of the sparse random matrix have been investigated, and the deviation from the Wigner's semi-circle law has been discussed [3, 4, 5, 6]. The center of the spectral band rises more sharply and long tails appear on both edges. In the scale-free networks, the power-law (long tail) behavior at the edge of the spectra is given by [7]

$$\rho(\lambda) \propto \lambda^{-\mu}, \qquad \mu = 2\gamma - 1 \tag{2}$$

where γ is the power of the distribution $P(k)$ of node degrees k in the corresponding graph,

$$P(k) \propto k^{-\gamma}. \tag{3}$$

This power-law distribution $P(k)$ is also a well-known feature of the scale-free networks [8, 9]. It is conjectured that many real-world networks including the biological networks share such scale-free features. In spite of many studies with numerically simulated model networks or real-world networks obtained from large amounts of empirical data, the theoretical basis of this universality is still unclear.

The largest eigenvalue of the adjacency matrix $\lambda_{\max}$ has a relation to the node that has the largest degree $d \equiv k_{\max}$, a so-called a hub, and sometimes forms a distinguishable narrow band above the bulk of the spectrum. In this case, the network is 'localized' around the hub, in relation to impurity bands studied in the condensed matter physics. Furthermore, several peaks at the band center are formed by many zero eigenvalues of star-graphs, in which one node of degree k is connected to other $k = 1$ nodes [10]. These peaks have close relationships to the modular structure of the network. Here we use the term ‘ modular ’ to describe the community structures of the network. In a network community, the nodes are bound together inwardly more frequently than outwardly. Also, the eigenvector in correspondence with the second largest eigenvalue is called the largest diffusive modes and has been investigated by using a random walk model on a lattice [11, 12].

Types of measures, such as degree centrality, eigenvector centrality, modularity, bipartivity and assortativity, have been used to investigate the community structures of complex networks so far. For reviews see Refs. [13, 14]. In the case of the gene interaction network, which consists of more than 20,000 nodes and several hundreds of thousands of edges, it is inevitable to divide the network into sub-networks and extract some community structures including some specific nodes. Extraction of some gene communities in cancer cells might be extremely useful for the detection of the gene groups that determine cell characteristics, the classification of pathological conditions, the selection of therapeutic methods, and so on. Also, it has been pointed out that the marginal genes that connect different communities play some important role in the cell [15].

Gene network data associated with cancer cells are available from several public databases. One of them, the cancer network galaxy (TCNG) database (http://tcng.hgc.jp), stores gene co-expression networks that are numerically inferred by Bayesian network algorithms called SiGN-BN NNSR [16]. Using these data, we have found the spacing distribution of the two succeeding eigenvalues $P(s)$ of the adjacency matrices follows the Wigner distribution [17]. This is the result in a dense gene network group, where the number of inferred edges is large.

In this work, the same adjacency matrices are used to examine the community structure of the gene networks. In particular, when the degree of the hub node becomes larger than a certain threshold $d > d_c$, the network localization on the hub node is observed. Furthermore, we evaluate the node degree correlation function. We show that some disassortative gene communities localized around the hub can be extracted.

2 Methods

2.1 Gene interaction networks in the TCNG database

In this study, we use numerically inferred gene interaction networks, each of which consists of 8,000 nodes downloaded from the TCNG database. These gene networks were obtained from huge calculations based on the Bayesian network model. The original expression data of human cancer cell samples used for the gene network inference were taken from the GEO database [18]. The number of expression samples used for a computational assay of the Bayesian network inference is around 100. The samples are taken mainly from various types of cancer (tumor) cells including a small portion of normal cells as controls. In the Bayesian networks, the interactions (or causal dependencies) between the nodes are represented by arrows, and the direction corresponds to the causality of the interaction. In this study we ignore the direction of the gene interactions (edges) for simplicity. We also ignore any self-interactions. Then the adjacency matrix A of the gene interaction network becomes a real symmetric matrix.

The degree k_i of node i is given by the sum of each row (or column) of the adjacency matrix A.

$$k_i = \sum_j A_{ij} \tag{4}$$

The mean degree $\bar{k}$ of all gene expression networks used in this study is $\bar{k} = 9.68$, which is within the range where the sparse random matrix predictions apply.

2.2 Eigenvector centrality and the inverse participation ratio

The eigenvector centrality v_i is one of the measures often studied in a framework of the graph theory in order to probe network properties [19, 20]. It is evaluated from the eigenvector corresponding to the largest eigenvalue λ_1 by diagonalizing the adjacency matrix A.

$$AV = \lambda V, \quad V = [v^1, v^2, \cdots, v^N], \quad \lambda = (\lambda_1, \cdots, \lambda_N), \tag{5}$$

where we label the eigenvalues λ_i in descending order $\{\lambda_1 > \lambda_2 > \cdots > \lambda_N\}$, and v^i $(i = 1 \cdots N)$ is the corresponding eigenvector. The centrality of node i is given by the i-th element of v^1.

$$v_i \equiv v_i^1 \tag{6}$$

Its variant is also known as the Google's PageRank. In localized networks, where the hub degree d becomes very large, the eigenvector centralities take finite values only on the hub node and the nodes connected to the hub. It takes the values of $O(1/N)$ on other nodes. Each of our gene networks has a different mean degree. In the original database TCNG, each edge has an edge attribute named 'edge factor', which corresponds to the likelihood of the inference. To fix the mean degree $\bar{k}$ of all networks, the edges are extracted in order from larger to smaller edge factors.

We also evaluate the inverse participation ratio Ψ, which is obtained by

$$\Psi(\lambda_i) = \sum_j (v_j^i)^4. \tag{7}$$

We note that all eigenvalues are normalized as $|v|^2 = 1$. In the localized regime, it takes finite values $\Psi \sim O(1)$ only on localized elements of λ_i.

2.3 The degree correlation fuction

By averaging several gene interaction networks obtained from the gene expression experiments in various cancers, we are able to see a coarse-grained property of the gene networks in cancer cells. We let h be the maximum node degree of all 254 gene networks, and we evaluate the degree correlation matrix E of size $h \times h$. Here, the element E_{ij} is the probability that two nodes have degrees

$k = i$ and $k = j$, respectively, at each end of an edge taken randomly from the whole network.

$$\sum_{i,j=1}^{h} E_{ij} = 1 \tag{8}$$

The sum is taken from all of the edges in 254 gene interaction networks. The degree correlation function $k_{nn}(k)$ is evaluated from the correlation matrix E by

$$k_{nn}(k) = \sum_{k'} P(k'|k) \tag{9}$$

$$P(k'|k) = \frac{E_{kk'}}{\sum_{k'} E_{kk'}}. \tag{10}$$

[21, 22, 23] When the degree correlation function $k_{nn}(k)$ is plotted against the degree k, the positive slopes are obtained in the case of social networks including the collaborative research networks. They are called the degree assortative networks. On the contrary, in the metabolic network of E. coli or in some other biological networks, the degree dissasortative behavior is obtained. In this case, the hub that has a large degree has a tendency to connect to small degree nodes, and $k_{nn}(k)$ shows the negative slope in the log-log plot. The neutral degree correlation networks are also known, for example, in a power grid network [24, 25].

$k_{nn}(k)$ is also expected to follow a power-law, which has been observed in many studies,

$$k_{nn}(k) \propto \; k^{\eta}. \tag{11}$$

We evaluate the power η from the average over all edges of the whole gene network in the 254 cancer studies.

2.4 Nearest neighbor eigenvalues spacing distribution $P(s)$

In the random matrix theory, in addition to the eigenvalue density, the universal distribution of the nearest level spacing $P(s)$ is important [26]. In the case of the real symmetric random matrices, the Wigner distribution is

$$P(s) = \frac{\pi s}{2} \exp\left(-\frac{\pi s^2}{4}\right). \tag{12}$$

This universal behavior has been widely observed in the adjacency matrices of real-world networks including biological networks . Eq.(12) also shows that there is a correlation between the eigenvalues. On the other hand, when the eigenvalues distribute randomly with no correlation, for example, if the eigenvalues are taken from the Poisson distributed random values, their interval distribution $P(s)$ becomes

$$P(s) = \exp(-s), \tag{13}$$

where s is the spacing between two adjacent eigenvalues normalized by the mean eigenvalue spacing. Eq.(13) is called the Poisson distribution in the random matrix theory.

Before evaluating the intervals between two succeeding eigenvalues of the adjacency matrices, we 'unfold' the original eigenvalues numerically. We divide eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_N$ into segments, each of which consists of several hundreds of eigenvalues. Afterward, we rescale λ_i according to the local mean spacing $\bar{s}$ in each segment. With these unfolded eigenvalues, the local distributions of the nearest neighbor level spacing are obtained. The Kolmogorov-Smirnov test for these local distributions against the null-distributions eqs.(12) and (13) has been performed with the significance level $\alpha = 0.05$. For details about the unfolding method, see ref.[17].

3 Results

Figure 1 shows the relation between the total edge number in the inferred 254 gene interaction networks and the hub degree d. Up to $d \lesssim 100$, the hub degree grows almost linear in relation to the total number of edges in the network. However, in the region where $d \gtrsim 100$, the hub degrees are almost independent of the total edge numbers, and the average node degree $\bar{k}$ of the networks becomes smaller: $\bar{k} = 8.4$ (where $\bar{k} = 9.7$ for the whole network). In the gene networks data, the number of nodes N is fixed as 8000. It can be predicted that the edges concentrate more and more on the hub nodes in much sparser gene interaction networks where N becomes much larger.

In FIG.2(a) and (b), we plot the degree distribution $P(k)$ and the spectrum $\rho(\lambda)$ in log-log scale, where k is the node degree and λ is the eigenvalue of the adjacency matrix A_{ij}, respectively. To obtain the histogram of k in FIG.2(a), the logarithmic bins are used. The number of corresponding nodes decreases rapidly in larger bins, so the width of the n-th bin is taken in proportion to c^n, where c is some number $c > 1$. The n-th bin edge is $k^{(n)} = (c^n - 1)/(c - 1)$. Using the MATLAB function 'fit' [27], we evaluate the power γ for several values of c and over some succeeding bins. The minimum width of a 95% confidence interval of the fitted γ is obtained when $c = 1.7$ and when $n = 5$ to 10 bins, which gives the result $\gamma = 3.87 \pm 0.01$.

In FIG.2(b), we plot the tail part of $\rho(\lambda)$ for $\lambda \geq 5$. Here we take a constant bin width of 0.25. We apply the same fitting procedure for the power μ. The minimum of a 95% confidence interval is obtained in the region $7 < \lambda < 13$, which gives the power $\mu = 6.52 \pm 0.05$. Although our fitting procedure may not be precise enough, the value of μ is just slightly smaller than what was expected from eq. (2).

In Ref.[19, 20], the eigenvalue centrality v_i is investigated for a model network where the node degree k is taken as Poisson distributed random values: the Poisson random graph model. In the random graphs, the eigenvalue centralities v_i are $O(1/\sqrt{N})$ for all $i = 1, \cdots, N$. However, by adding an extra hub node of degree d, it was shown analytically that v_i takes $O(1)$ value only on the hub

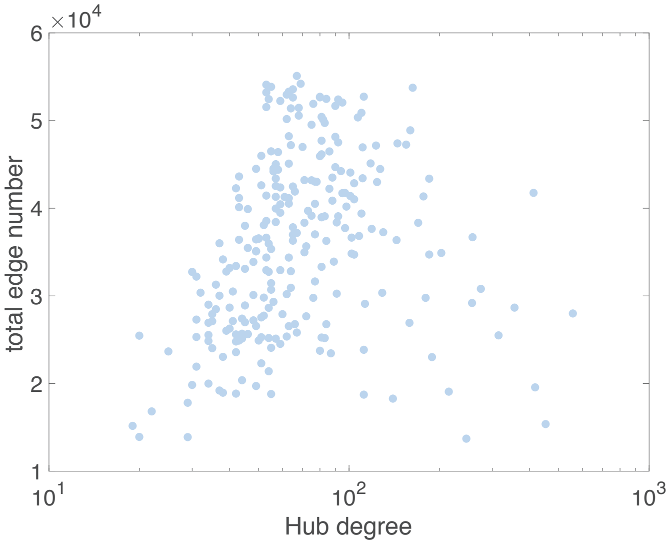

Figure 1: For all 254 gene networks, the number of edges is plotted against the degree of the hub node d. The number of nodes is fixed as $N \simeq 8000$ for all gene networks.

and its neighboring nodes, and $v_i \sim O(1/N)$ on any other nodes. The square of the eigenvector centrality in the large N limit is given by,

$$v_i^2 = \begin{cases} (\frac{1}{2}d - c)/(d - c) & \text{for } i = 1 \text{ (the hub)}, \\ (\frac{1}{2}d - c)/(d - c)^2 & \text{for } i \text{ a neighbor of the hub}, \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where c is the mean degree $\bar{k}$ and $d > 2c$. Note that the eigenvalues are enumerated in descending order.

To see the localization properties, we plot v_1^2 versus the hub degrees d of the gene interaction networks. In FIG.3, two results of different c $(= \bar{k})$ values $c = 4.7$ and $c = 6.6$ are shown. Before diagonalizing the adjacency matrix A, the edges are extracted according to the edge factors from high to low values, thus more confidential edges in the numerical inference are preferred, until the average degree attains the fixed values. We employ the curve fitting for the results by eq.(14) of $i = 1$ case, setting c as a fitting parameter. As seen from the dotted lines in FIGs.3(a) and (b), the fitting parameters $c^\dagger$ become larger than the empirical value of c in both cases, although the total behavior is well described by the theoretical result. The localization transition is expected at $d_c = 26$ and 40 from the fitting in FIG.3(a) and (b), respectively, where $d_c = 2c^\dagger$. The data points shown in the figures are evaluated with the gene interaction networks, which have edges more than $L > cN/2$.

Since the gene networks are the scale-free networks, which is shown by the power-law distribution $P(k)$ in FIG.2(a), there are correlations between node

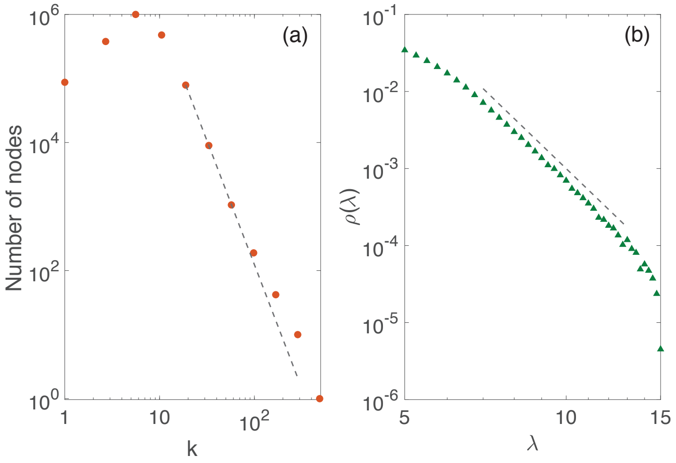

Figure 2: (a) The distribution of node degree $P(k)$ in the log-log scale. The logarithmic bins are used and the n-th bin edge is $k^{(n)} = (1.7^n - 1)/0.7$. In this figure $1 \leq n \leq 11$ bins are shown. The curve fitted by a power function for the 5th to 10th bins is performed. The result $P(k) \propto k^{-3.8}$ is plotted by the dotted line. (b) The eigenvalue density $\rho(\lambda)$ in the tail region $\lambda > 5$. The bin width is 0.25. The power-law behavior $\rho(\lambda) \propto \lambda^{-6.7}$ is also shown by the dotted line as a guide to the eye.

degrees. We note that in the study of the node disassortativity of the scale-free networks [23], transition to one highly connected graph (the giant component formation) at the value of $\bar{k} = 1$ also shifts to a larger value of $\bar{k} > 1$. The nodes of larger degrees tend to connect to lower degree nodes, thus preventing the network to form forming one fully connected graph. The larger values of d_c for the expected localization transition of the gene interaction networks might be also explained by this dissasortative scale-free network behavior. We also note that in the dense gene interaction networks of higher mean degree $c > 8$, the discrepancy from eq.(14) becomes very large and the curve fitting method does not work anymore. In such dense gene networks, the correlation of node degrees is very large, and the assumption of the uncorrelated edge degree of the Poisson random graph for the derivation of eq.(14) cannot be applied.

The inverse participation ratio $\Psi(\lambda_i)$ is plotted in log-scale in FIG.4 for the gene networks that have $\lambda_{\max}^2 > 200$, where $\lambda_{\max} \equiv \lambda_1$ is the largest eigenvalue of A. It is also given by

$$\lambda_{\max} \simeq \sqrt{d} \tag{15}$$

where d is the hub degree [19]. The scatter plot in FIG.4 shows Ψ for the strongly localized gene interaction networks. Ψ takes $O(1)$ values only on the localized nodes, which consist of the hub node and its neighbors. The complicated structures of Ψ are observed due to multiple heavily localized hub nodes of large degrees. We also see several peaks around the center $\lambda = 0$, which may

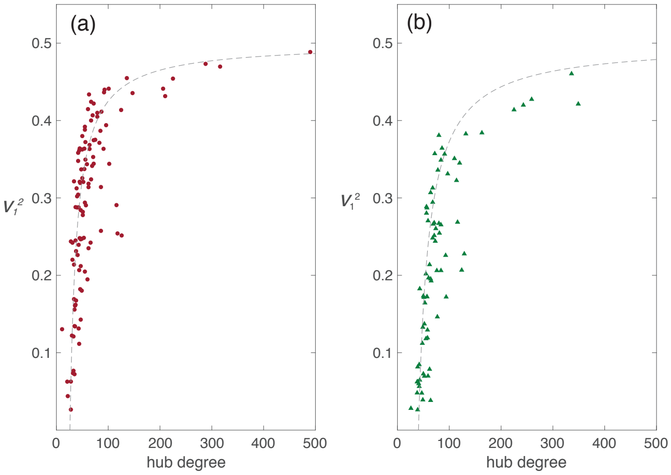

Figure 3: (a) v_1^2 v.s. the hub degrees are plotted for 112 gene interaction networks where the mean degree $c = 4.7$. The dashed line is the result of a curve fitting by the function in eq.(14). The fitting parameter is $c^\dagger = 13$. The localization transition at $d_c = 2c^\dagger = 26$ is larger than the theoretical prediction $d_c = 9.4$. (b) The same plot for the networks where $c = 6.6$ and we obtain $d_c = 40$ from the fitting (the dashed line). The number of networks plotted in this figure is 76.

be related to the eigenvalues of the star-graphs [10, 28].

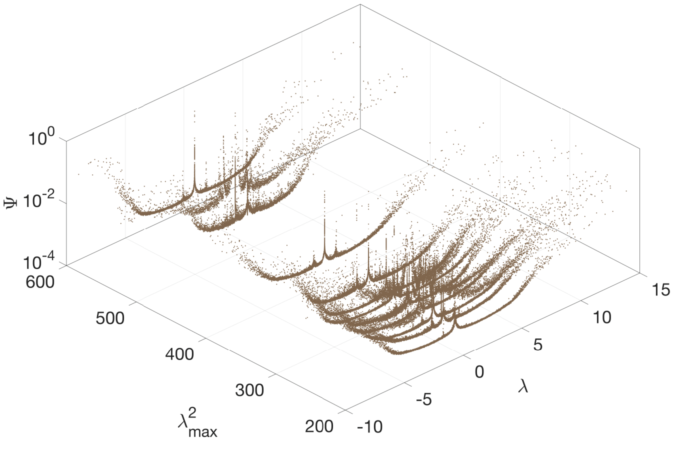

Figure 4: The scatter plot of the inverse participation ratio $\Psi(\lambda_i)$ for the strongly localized gene interaction networks, which have the largest eigenvalue $\lambda^2_{\max} > 200$.

To study the topology of the gene interaction networks with multiple localized hubs of degree $d \gtrsim 40$, we show a histogram of degree discrepancies Δ between two nodes on both ends of an edge in FIG.5.

$$\Delta = |k_i - k_j|, \tag{16}$$

where k_i and k_j are the degrees of the connected i and j nodes, respectively. Figure 5 is the result of the gene interaction network ID:101 obtained from the gene expression experiments on cells from lung cancer. The total number of edges is $24,654$ in this network and there are two heavily localized nodes whose degrees are $d_1 = 112$ and $d_2 = 69$. From FIG.5 we see two distinct peaks near $\Delta \simeq 110$ and 70, which should correspond to the edges in the sub-networks centered on the two hubs d_1 and d_2, respectively. We find the naive histogram of Δ also shows the disassortative feature of the hub nodes, where $\Delta \simeq d$ for $\Delta \gtrsim 40$.

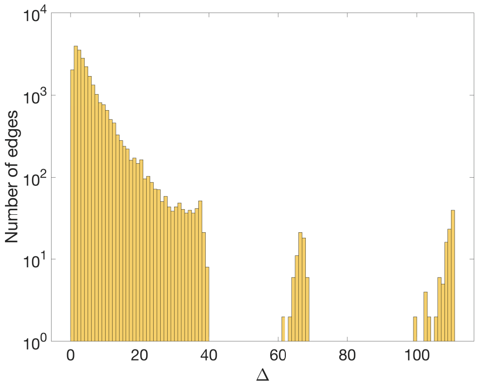

Figure 5: Histogram of the degree discrepancies Δ between two nodes linked by an edge in the ID:101 gene interaction network, which was inferred from expression data from lung cancer cells. The bin width is 1. The total number of edges is 24654.

In FIG.6, we plot the degree correlation function $k_{nn}(k)$ obtained from 254 gene interaction networks in log-log scale. The averaging has been done for all edges in these networks. We use the logarithmic bins again, where the n-th bin edge is given by $k = (1.5^n - 1)/0.5$. In the large k regime, we see the disassortative behavior, which is apparent from the negative slope of $k_{nn}(k)$ for $k \gtrsim 40$. We also apply a curve fitting for the power η of the slope, by which we get $k_{nn}(k) \propto k^{-0.37}$. The fitting result is also shown in FIG.6 with a dashed line. In the medium k regime, where $6 \lesssim k \lesssim 40$, we obtain the neutral or probably the assortative behavior of the correlation of the node degrees.

From these observations, we expect that the whole cancer gene interaction network contains multiple disassortative sub-networks centered on the hub whose degree is $d > 40$.

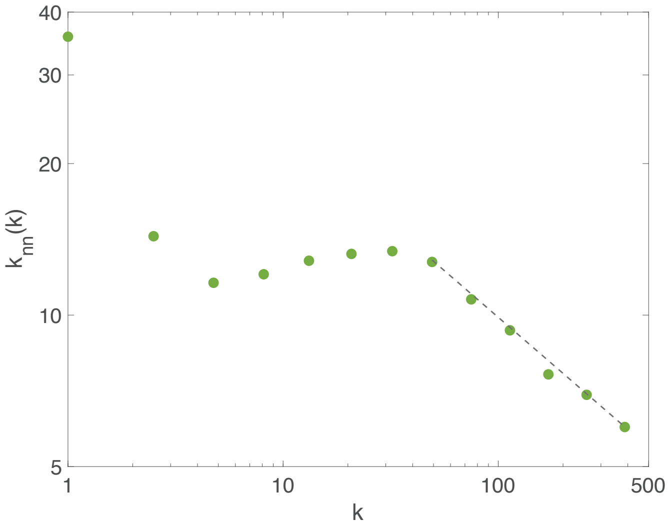

Figure 6: The correlation function of the node degrees averaged over all the edges in the 254 gene interaction networks. The dashed line is the result of the fitting. The negative power $k_{nn}(k) \propto k^{-0.37}$ describes the disassortative degree correlation that exists in the sub-networks centered on the hub nodes whose degrees are $k \gtrsim 40$.

The gene interaction networks, which have a localized super-hub of a degree larger than $d > 125$, are selected for the calculations of the distribution of the nearest neighbor eigenvalues spacing $P(s)$. In each of such 27 strongly localized networks with huge hubs, we extract the edges according to the value of Δ (the degree discrepancies of linked nodes). We divide each network into two sub-networks, which consists of edges with (a) $\Delta < 25$ (the assortative community) and (b) $\Delta \geq 25$ (the disassortative communitiy), respectively. FIG.7(a) shows $P(s)$ obtained from the sub-network (a). The unfolding procedure has been done for each of the 27 networks separately, then we took the average of local $P(s)$. The coincidence with the Wigner-Dyson distribution is very nice, which has also been tested by the one sample Kormogorov-Smirnov test. We have tested 906 segments of unfolded eignevalues from 27 networks, and 93% of them have p-values larger than the significance level $\alpha = 0.05$.

We follow the same procedure for the sub-networks (b), which consist of the edges $\Delta \geq 25$, and show $P(s)$ in FIG.7(b). The Poisson distribution of $P(s)$ is obtained in 76% of the total 502 eigenvalue segments over 27 networks at the significance level $\alpha = 0.05$, which also suggests the modular behavior of the strongly localized sub-networks. We note that the number of edges that belong to the sub-network (b) is 4315 (31% of total), for example in the ID:191 gene interaction network. The number of contributing nodes is 3825 (48% of total) in the same sub-network. In Table 1, we show the statistics of the sub-networks

(b) in each of the 27 gene interaction networks. The average of the ratio of the edges belonging to the hub centered communities $\Delta > 25$ is 19%.

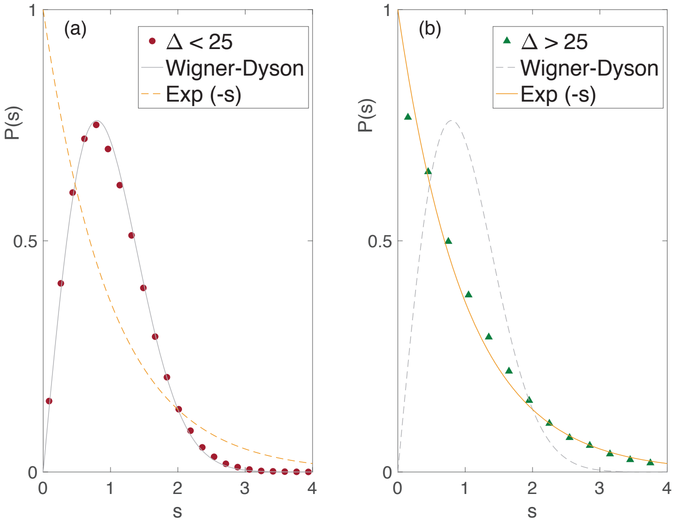

Figure 7: (a) The distribution of the nearest neighbor eigenvalues spacing $P(s)$ of the adjacency matrices of sub-networks, which consists of edges that have $\Delta < 25$ (the assortative or neutral degree correlation network). s is the distance between (unfolded) λ_i and λ_{i+1} normalized by the mean level spacing. The average has been taken over the 27 networks listed in Table.1. (b) The $P(s)$ obtained from the sub-networks of edges $\Delta \geq 25$ (the disassortative degree correlation sub-networks localized on the hub nodes).

In NDEx [29], we show a network graph of sub-network (b) of the ID:230 gene interaction network in a colon tumor. The URL accessible to the graph is given in the Supplementary file. It can be seen that the hubs are connected via mediator nodes of relatively small degree, including the genes RECK, BCAS1, RNF168, etc.. A detailed study of the nature of these mediator genes is important for further understanding of the disease.

4 Conclusions

254 networks have been analyzed to obtain the average behavior of gene networks in human cancer cells. More than 9×10^6 edges are included in the calculation of the degree distribution $P(k)$. The power-law behavior has been observed where $P(k) \propto k^{-3.8}$.

The cancer gene networks are strongly localized on the hub gene when the degree of the hub $d_c > 26$ for the networks where the mean degree $\bar{k} = 4.7$ and $d_c > 40$ in the networks of $\bar{k} = 6.6$. The eigenvector centralities on the hub fit well on the theoretical line, which was obtained for the model networks of random graphs with the hub. However, the critical values of the hub degree d_c become larger than the theoretical result $d_c = 2\bar{k}$.

The localized cancer gene interaction networks are the disassortative (hubs avoiding hubs) networks, and the degree correlation function $k_{nn}(k)$ has a negative power in the region $k \gtrsim 40$. This result is also consistent with the evaluation from the eigenvector centrality for the networks where $\bar{k} = 6.6$.

The localized hub communities in which the edges have large degree discrepancies $\Delta > 25$ have been extracted. They show the modular behavior, which was also confirmed from the Poisson distribution of the nearest neighbor eigenvalues spacing $P(s)$. The sub-network in which $\Delta < 25$ form a giant component (a hair ball), and we obtain the Wigner-Dyson distribution of $P(s)$ in this assortative sub-network.

In the localized hub-centered community, the hubs are linked by the mediator genes that have small degrees. An investigation of the nature of these mediator genes might be helpful for the diagnosis of and medical treatment decisions for human cancers.

References

[1] Mehta, M. L. (1991) *RANDOM MATRICES.* Academic Press, Inc.

[2] Guhr, T., Mueller-Groeling, A. & Weidenmueller, H. A. (1998) Random Matrix Theories in Quantum Physics: Common Concepts. *Phys. Rep.*, **299**, 189–425.

[3] Kuhn, R. (2008) Spectra of sparse random matrices. *J. Phys. A Math. Theor.* **41**, 295002.

[4] Mirlin, A. D. & Fyodorov, Y. V. (1991) Universality of Level Correlation-Function of Sparse Random Matrices. *J. Phys. A. Math. Gen.* **24**, 2273–2286.

[5] Rodgers, G. J. & Bray, A. J. (1988) Density of states of a sparse random matrix. *Phys. Rev. B* **37**, 3557–3562.

[6] Nagao, T. & Rodgers, G. J. (2008) Spectral density of complex networks with a finite mean degree. *J. Phys. A Math. Theor.* **41**, 265002.

[7] Dorogovtsev, S. N., Goltsev, A. V., Mendes, J. F. F. & Samukhin, A. N. (2003) Spectra of complex networks. *Phys. Rev. E* **68**, 046109.

[8] Barábasi, A.-L. (2002) Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97.

[9] Barábasi, A.-L. (2016) *NETWORK SCIENCE.* Cambridge University Press.

Table 1: The statisitics of the localized sub-networks where $\Delta \geq 25$. Δ is the difference of degrees of linked nodes.

Network ID	Total edges	Edges of $\Delta \geq 25$	Consisting nodes	Ratio of edges	Ratio of nodes
ID:191	13787	4315	3825	0.319	0.478
ID:248	15961	3478	2786	0.217	0.348
ID:117	19267	1393	1224	0.072	0.153
ID:247	19270	9254	6652	0.480	0.831
ID:153	19720	12957	6867	0.657	0.927
ID:243	23300	11456	7073	0.491	0.885
ID:251	26742	2715	2322	0.101	0.290
ID:111	27525	4014	3608	0.145	0.451
ID:86	28671	4309	3919	0.150	0.489
ID:148	29215	6535	5237	0.223	0.654
ID:57	29882	4252	3840	0.142	0.480
ID:143	30714	6879	5210	0.223	0.651
ID:146	31424	5459	4542	0.173	0.567
ID:125	32213	1900	1767	0.058	0.220
ID:162	35648	8204	5689	0.230	0.711
ID:203	35965	7589	5450	0.211	0.681
ID:147	37468	5246	4175	0.140	0.521
ID:109	38016	3804	3282	0.100	0.410
ID:112	39315	2259	2125	0.057	0.265
ID:119	39754	3594	2988	0.090	0.373
ID:24	43079	4325	3486	0.100	0.435
ID:210	43184	2965	2623	0.068	0.327
ID:114	44674	5735	4233	0.128	0.529
ID:179	46849	5943	4246	0.126	0.530
ID:1	49622	4966	3600	0.100	0.450
ID:221	50594	5896	4217	0.116	0.527
ID:26	51994	4091	3048	0.078	0.381

[10] Bauer, M. & Golinelli, O. (2001) Random incidence matrices: Moments of the spectral density. *J. Stat. Phys.* **103**, 301–337.

[11] Eriksen, K. A., Simonsen, I., Maslov, S. & Sneppen, K. (2003) Modularity and Extreme Edges of the Internet. *Phys. Rev. Lett.* **90**, 148701.

[12] Simonsen, I., Astrup Eriksen, K., Maslov, S. & Sneppen, K. (2004) Diffusion on complex networks: a way to probe their large-scale topological structures. *Phys. A Stat. Mech. its Appl.* **336**, 163–173.

[13] Newman, M. E. J. (2004) Detecting community structure in networks. *Eur. Phys. J. B* **38**, 321–330.

[14] Estrada, E. & Higham, D. J. (2010) Network Properties Revealed through Matrix Functions. *SIAM Rev.* **52**, 696–714.

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

[16] Tamada, Y. *et al.* (2011) Estimating genome-wide gene networks using nonparametric bayesian network models on massively parallel computers. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **8**, 683–697.

[17] Kikkawa, A. (2018) Random Matrix Analysis for Gene Interaction Networks in Cancer Cells. *Sci. Rep.* **8**, 10607.

[18] Barrett, T. *et al.* (2012) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995.

[19] Nadakuditi, R. R. & Newman, M. E. J. (2013) Spectra of random graphs with arbitrary expected degrees. *Phys. Rev. E* **87**, 012803.

[20] Martin, T., Zhang, X. & Newman, M. E. J. (2014) Localization and centrality in networks. *Phys. Rev. E* **90**, 052808.

[21] Pastor-Satorras, R., Vázquez, A. & Vespignani, A. (2001) Dynamical and correlation properties of the internet. *Phys. Rev. Lett.* **87**, 258701.

[22] Costa, L. da F., Rodrigues, F. A., Travieso, G. & Boas, P. R. V. (2007) Characterization of complex networks: A survey of measurements. *Adv. Phys.* **56**, 167–242.

[23] Newman, M. E. J. (2002) Assortative Mixing in Networks. *Phys. Rev. Lett.* **89**, 208701.

[24] Newman, M. E. J. (2003) Mixing patterns in networks. *Phys. Rev. E* **67**, 026126.

[25] Jalan, S. & Yadav, A. (2015) Assortative and disassortative mixing investigated using the spectra of graphs. *Phys. Rev. E* **91**, 012813.

[26] *The Oxford handbook of random matrix theory.* (2011) Oxford University Press.

[27] *MATLAB 2018a* (The MathWorks, Inc., Natick, Massachusetts, U. S.)

[28] Evangelou, S. N. (1992) A numerical study of sparse random matrices. *J. Stat. Phys.* **69**, 361–383.

[29] Pratt, D. *et al.* (2015) NDEx, the Network Data Exchange. *Cell Syst.* **1**, 302–305.