Discriminating Different Classes of Biological Networks by Analyzing the Graphs Spectra Distribution

Daniel Yasumasa Takahashi[1], João Ricardo Sato[2], Carlos Eduardo Ferreira[3], André Fujita[3]*

1 Department of Psychology and Neuroscience Institute, Princeton University, Princeton, New Jersey, United States of America, 2 Center of Mathematics, Computation, and Cognition, Federal University of ABC, Santo André, Sao Paulo, Brazil, 3 Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, Sao Paulo, Brazil

Abstract

The brain's structural and functional systems, protein-protein interaction, and gene networks are examples of biological systems that share some features of complex networks, such as highly connected nodes, modularity, and small-world topology. Recent studies indicate that some pathologies present topological network alterations relative to norms seen in the general population. Therefore, methods to discriminate the processes that generate the different classes of networks (*e.g.*, normal and disease) might be crucial for the diagnosis, prognosis, and treatment of the disease. It is known that several topological properties of a network (graph) can be described by the distribution of the spectrum of its adjacency matrix. Moreover, large networks generated by the same random process have the same spectrum distribution, allowing us to use it as a "fingerprint". Based on this relationship, we introduce and propose the entropy of a graph spectrum to measure the "uncertainty" of a random graph and the Kullback-Leibler and Jensen-Shannon divergences between graph spectra to compare networks. We also introduce general methods for model selection and network model parameter estimation, as well as a statistical procedure to test the nullity of divergence between two classes of complex networks. Finally, we demonstrate the usefulness of the proposed methods by applying them to (1) protein-protein interaction networks of different species and (2) on networks derived from children diagnosed with Attention Deficit Hyperactivity Disorder (ADHD) and typically developing children. We conclude that scale-free networks best describe all the protein-protein interactions. Also, we show that our proposed measures succeeded in the identification of topological changes in the network while other commonly used measures (number of edges, clustering coefficient, average path length) failed.

Introduction

In the last decades, attempts to understand the mechanisms that determine the topology of complex real world networks using random graphs (graphs that are generated by some random process) has gained much attention [1]. Some examples of complex networks are the World Wide Web [2], human social networks [3], protein-protein interaction networks [4], metabolic networks [5], and brain connectivity networks [6]. On studying these complex networks, some questions naturally arise. For example, how complex is a given random graph? How different are two random graphs? Given a realization of a random graph, how can one infer which random graph processes generated it? Attempts to answer some of these questions have been made on purely theoretical grounds [7], but interestingly, to the best of our knowledge, no simple and robust procedure exists to answer these questions using empirical data sets. Our aim in this work is to introduce such procedures.

Interactions are essential to understand complex systems where, to determine the behavior of the system, it is important to understand the way each component of the system interacts with others. For most classes of complex systems, interactions are neither invariant in time nor across systems from the same class. For example, neural networks in the cortex of the same individual can change in time, and synaptic organization is different among individuals. Therefore, a search for an exact common network structure seems to be unfruitful. What seem to be invariant are some statistical features that can be reproduced in classes of random graphs; therefore, the corresponding ensemble of random graphs can be used as a plausible model for an ensemble of cortical networks.

Two random graph models that are widely used to model natural phenomena are the scale-free [8] and the small-world networks [9]. The main characteristics of these random graphs are the non-trivial topological features that differ from the Erdös-Rényi random graphs [10], i.e., complex networks present heavy tail in the degree distribution, high clustering coefficient, community, hierarchical structures, and short path lengths. Usually, the scale-free network is characterized by its power-law degree distribution while the small-world network presents short path length and high clustering. However, although these characteristics are essential features of these random graphs, they are not sufficient to unambiguously identify a graph as belonging to a particular class. For example, small-world networks are highly

clustered like regular lattices and have small characteristic path lengths like Erdös-Rényi random graphs.

In this work we propose that the random graph spectrum, i.e., the ensemble average of the eigenvalues of the adjacency matrix, is a better and more general characterization of complex networks in comparison with other commonly used measures: number of edges, clustering coefficient, and average path length. For instance, it is known that several topological properties of a random graph, such as the number of walks, diameter, and cliques can be described by the spectrum of its adjacency matrix [7]. Based on this relationship between the topological properties of the random graph and its spectrum, we introduce the definition of entropy of a random graph spectrum and the Kullback-Leibler divergence between two random graph spectra. By simulation experiments, we observe that the entropy of random graph spectrum is related to the intuitive idea of amount of uncertainty of a random graph and that the Kullback-Leibler divergence between random graph spectra can discriminate two random graphs that were generated by different random process.

Statistical approaches such as model selection, parameter estimation, and hypothesis testing to discriminate two classes of random graphs are also presented. We illustrate practical use of the model selection approach in protein-protein interaction networks of eight different species. By analyzing the random graph spectrum instead of the degree distribution, we classified all the eight protein-protein interaction networks as scale-free graphs. Finally, the power of Kullback-Leibler based statistical test is illustrated by an application in networks derived from children with Attention Deficit Hyperactivity Disorder and with typical development. We succeeded in the identification of topological changes between children with typical development and ADHD patients, while standard measures such as number of edges, clustering coefficient and average path length failed.

Definition of graphs and graph spectrum

A *graph* is a pair of sets $G = (P,E)$, where P is a set of n nodes and E is a set of m edges that connect two nodes (elements of P). A random graph g is a family of graphs, where each member of the family is generated by some probability law. Among several classes of random graphs, there are three that have known importance due to their capability to model real world events, namely, Erdös-Rényi random (Figure 1A) [10], scale-free (Figure 1B) [8], and small-world graphs (Figure 1C) [9].

Erdös-Rényi random graphs are the simplest ones in terms of construction. Erdös and Rényi define a random graph as n labeled nodes in which each pair of nodes (i,j) is connected by an edge with a given probability p.

Scale-free networks, proposed by Barabási and Albert (1999), have a power-law degree distribution due to node preferential attachment. Barabási and Albert (1999) proposed the following construction of a scale-free network: start with a small number of (n_0) nodes and at every time-step, add a new node with $m_1 (\leq n_0)$ edges that link the new node to m_1 different nodes already present in the system. When choosing the nodes to which the new node connects, assume that the probability that a new node will be connected to node i is proportional to the degree of node i and the scaling exponent p_s which indicates the order of the proportionality ($p_s = 1$ linear, $p_s = 2$ quadratic and son on).

Small-world graphs, proposed by Watts and Strogatz (1998) are one-parameter models that interpolate between a regular lattice and an Erdös-Rényi random graph [11]. First, a ring lattice with n nodes is constructed, in which every node is connected to its first K neighbors ($K/2$ on either side). Then, we choose a vertex and the edge that connects it to its nearest neighbor in a clockwise sense.

With probability p_s we reconnect this edge to a vertex chosen uniformly at random over the entire ring. This process is repeated by moving clockwise around the ring, considering each vertex in turn until one lap is completed. Next, the edges that connect vertices to their second-nearest neighbors clockwise are considered. As in the previous step, each edge is randomly rewired with probability p_s; we continue this process, circulating around the ring and proceeding outward to more distant neighbors after each lap, until each edge in the original lattice has been considered once [9].

Any undirected graph G with n nodes can be represented by its adjacency matrix $A(G)$ with $n \times n$ elements A_{ij}, whose value is $A_{ij} = A_{ji} = 1$ if nodes i and j are connected, and 0 otherwise. The *spectrum* of graph G is the set of eigenvalues of its adjacency matrix $A(G)$. A graph with n nodes has n real eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Now, given a random graph g, the eigenvalues are random vectors for which we can take the expectation with respect to the probability law of the random graph. We define the *spectral density* distribution of a random graph g as

$$\rho_g(\lambda) = \lim_{n \to \infty} \langle \frac{1}{n} \sum_{j=1}^{n} \delta(\lambda - \lambda_j / \sqrt{n}) \rangle, \qquad (1)$$

where δ is the Dirac delta function and the brackets "$\langle \rangle$" indicate the expectation with respect to the probability law of the random graph. In what follows, we use the shorthand name spectrum of g to indicate ρ_g. The interest in spectral properties is related to the fact that the spectral density can be directly related to the graph's topological features [12].

In application, a closed form for the spectral density is rarely available, so we have to rely on some statistical estimators $\hat{\rho}_g$. In order to estimate the spectral densities, first the eigenvalues are computed, and then Gaussian kernel regression using the Nadaraya-Watson estimator [13] is applied for the regularization of the estimator. Finally, the density is normalized to obtain the integral below the curve equal to one. The bandwidth of the kernel can be chosen by (max(eigenvalues) - min(eigenvalues))/number of bins [14], where the number of bins can be selected by using any objective criterion. In this work, we used the Sturges' criterion [15].

It is worth mentioning that the study of spectral density distribution of complex networks is still an active area of research [7,16], but the aim has been in general to obtain the exact or approximate properties of spectrum distribution for a given model. In this article, we are instead concerned with their statistical properties and applications to crucial biological systems.

Results

First we will present the definitions of entropy and divergence for graphs spectra, along with statistical methods for estimation and significance testing. Then, the performance of each method is evaluated by simulations and finally applied to actual data for illustration.

Entropy of graph spectrum

Let ρ_g be the spectrum of a random graph g. We define the *spectral entropy* $H(\rho_g)$ as

$$H(\rho_g) = - \int_{-\infty}^{+\infty} \rho_g(\lambda) \log \rho_g(\lambda) d\lambda, \qquad (2)$$
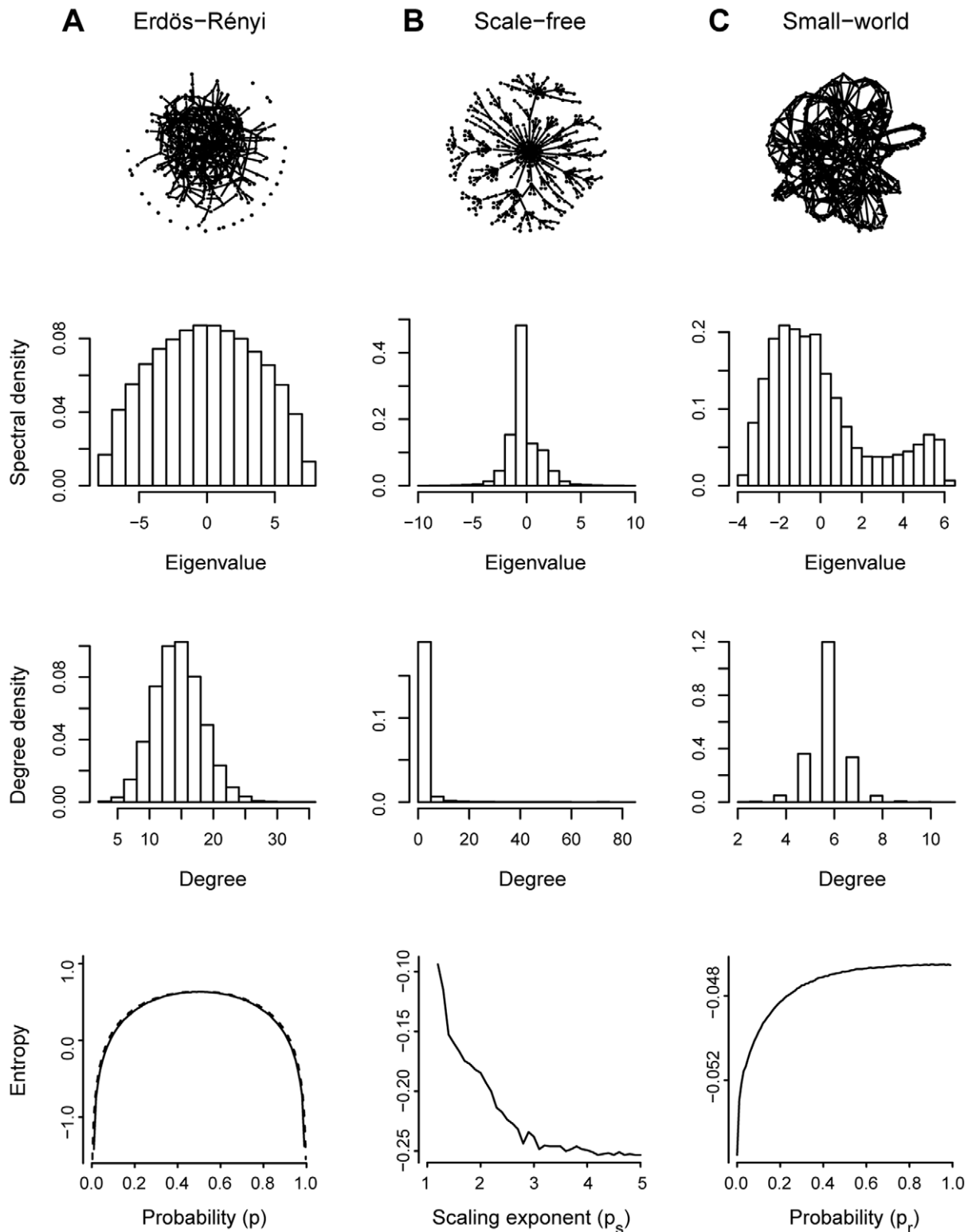
Figure 1. Illustrative figure of the three different complex network models. (**A**) Erdös-Rényi; (**B**) Scale-free; and (**C**) Small-world and their respective spectra, degree distributions, and entropies, in this order from top to bottom. The estimated entropies are computed for the respective graph type for the respective parameters (probability p for the Erdös-Rényi, scaling exponent p_s for the scale-free, and probability p_r for the small-world random graphs). In (**A**) the entropy values estimated from the simulation data is depicted by a solid line and the theoretical value of the entropy computed using equation 4 is indicated by a dashed line.
doi:10.1371/journal.pone.0049949.g001

where, as usual, we assume $0 \log 0 = 0$. Observe that the entropy defined above is also known as differential entropy [17] and can assume negative values, in contrast to the entropy defined for discrete distributions.

Since the spectral density of an adjacency matrix of a random graph has a tight relationship with the random graph structure and can be considered a fingerprint of the random graph [7], we propose that the corresponding spectral entropy also describes important characteristics of the random graph. More specifically, we propose that the spectral entropy measures a form of "uncertainty" associated to the random graph. To gain some intuition, we can compute the approximate spectral entropy for the Erdös-Rényi random graph g with parameter p as follows. For large n, we have

$$\rho_g(\lambda) \sim \frac{\sqrt{4p(1-p) - \lambda^2}}{2\pi p(1-p)} \qquad (3)$$

for $0 < |\lambda| < 2\sqrt{(p(1-p))}$ and 0 otherwise [18,19]. Using the above approximation, we have that

$$H(\rho_g) \sim \frac{1}{2}\ln(4\pi^2 p(1-p)) - \frac{1}{2}. \qquad (4)$$

This formula shows that the maximum spectral entropy for the Erdös-Rényi graph is achieved for $p = 0.5$, which is in accordance to the intuition that this is the model with the largest uncertainty. To confirm our point, the Erdös-Rényi random graph spectral entropy was calculated for many different values of probability p (bottom panel Figure 1A, dashed line). For the Erdös-Rényi graphs, not surprisingly, the entropy achieved its maximum value on $p = 0.5$, and the minimum values on $p = 0$ and $p = 1$, which is the situation where there is only one possible graph, i.e., the empty and complete graphs, respectively (Figure 1A). Furthermore, it is important to point out that the entropy function is symmetric due to the symmetry of the spectrum function, i.e., the spectral density of the Erdös-Rényi graph generated with parameter p is equal to the spectral density of the Erdös-Rényi graph generated with parameter $1-p$.

For the scale-free and small-world networks, an exact formula for the spectral entropy is not known, therefore, we estimated the entropy for different parameters of the models. A straightforward way to obtain an estimator $\hat{H}(\rho_g)$ for the spectral entropy is to first obtain an estimator $\hat{\rho}(\lambda)$ of $\rho(\lambda)$ and plug in to the equation (2). This is the procedure adopted in this work. To verify the accuracy of our estimator we compared the average estimated entropy values for 100 Erdös-Rényi random graphs with 500 nodes (bottom panel Figure 1A, solid line) and the theoretical value in equation 4 (bottom panel Figure 1A, dashed line). A visual inspection shows that the estimator is very accurate. The average bias for this example was -0.015, i.e., a small negative bias.

For the scale-free graphs we observe (Figure 1B) that the estimated entropy is higher in low scaling exponents (p_s) because it becomes similar to an Erdös-Rényi random graph, whereas when the scaling exponent goes to infinity it becomes closer to a complete bipartite graph resulting in a lower entropy. Finally, for small-world graphs (Figure 1C), the entropy is higher when the randomness of the graph (probability p_r) increases. Notice that when $p_r = 1$, the small-world graph becomes an Erdös-Rényi graph, whereas when $p_r = 0$ the graph is a ring [9], therefore presenting lower entropy. For both scale-free and small-world graphs, the number of nodes and edges were set to 500 and 600,

respectively, and for each scaling exponent (p_s) or probability (p_r), an average entropy of 100 graphs were calculated.

Kullback-Leibler divergence between graphs

Once the spectral entropy is defined, one may introduce a measure of similarity between two spectral densities, which is also a measure of similarity between two random graphs. It is clear that if two spectral densities are different, then the respective graphs should be different, although the converse is not always true (i.e., there are non-isomorphic graphs which are isospectral).

We define the *Kullback-Leibler divergence* (for sake of brevity we call it KL divergence) between two spectral densities ρ_{g_1} and ρ_{g_2} as

$$KL(\rho_{g_1}|\rho_{g_2}) = \int_{-\infty}^{+\infty} \rho_{g_1}(\lambda) \log \frac{\rho_{g_1}(\lambda)}{\rho_{g_2}(\lambda)} d\lambda, \qquad (5)$$

if the support of ρ_{g_2} contains the support of ρ_{g_1}. Otherwise, $KL(\rho_{g_1}|\rho_{g_2}) = +\infty$. As usual, we assume $0 \log \frac{0}{0} = 0$.

For the above equation, ρ_{g_2} is called the *reference measure*. This divergence is asymmetric and non-negative. It is also zero if and only if ρ_{g_1} and ρ_{g_2} are equal.

The KL divergence can be interpreted as a measure of discrepancy between two random graphs, and thus can be used to build an estimator for the parameter of a model given an observation. Specifically, let g be a random graph with spectral density ρ_g. Also let $\{\rho_\theta\}$ be a parametric family of spectral distributions indexed by a real vector θ. Assume that there exists a value of the parameter θ, which we denote θ^*, that minimizes $KL(\rho_g|\rho_\theta)$. An estimator $\hat{\theta}$ of θ^* is given by

$$\hat{\theta} = \arg \min_\theta KL(\hat{\rho}_g|\rho_\theta). \qquad (6)$$

The idea is that among all possible choices of models in a parametric class of random graphs ρ_θ, we choose the one for which the corresponding spectral density minimizes the divergence with the non-parametrically estimated spectral density. This is in the same spirit as nonparametric likelihood estimators of which the Whittle estimator is an example [20].

To show the performance of our estimator, different complex network models (Erdös-Rényi, scale-free, and small-world) with sizes equal to 50, 100, 200, and 300 nodes were simulated. The parameters to be estimated for each random graph model are: the probability p of connecting two nodes for Erdös-Rényi graphs, the scaling exponent of the preferential attachment p_s for scale-free graphs, and the rewiring probability p_r for small-world graphs. The estimated parameters were averaged values calculated for 50 repetitions, and the results are shown in Table 1. Brackets indicate one standard deviation. From the results in Table 1, we conclude that the estimator is reasonable and it can recover the correct parameter with relatively small bias and variance, i.e., one or two order of magnitudes smaller than the value of the estimated parameter. We observe from Table 1 and further simulations not shown here that the direction of the bias depend on the specific parameter of the model and size of the graph, and therefore no systematic bias direction seems to exist. The performance of the estimator is further discussed in Section.

Another use of the KL is to build a model selection criterion to select good models among a set of candidate random graphs. More specifically, given a graph, it is important to decide if the graph was more likely to have been generated by an Erdös-Rényi, scale-

Table 1. Average parameters estimated by minimum distance estimator based on KL divergence for Erdös-Rényi random, scale-free, and small-world graphs.

	Random (p)	Scale-free (p_s)	Small-world (p_r)
Number of nodes/true parameters	**0.50**	**1.50**	**0.30**
50	0.51 (0.04)	1.53 (0.06)	0.33 (0.05)
100	0.50 (0.03)	1.53 (0.05)	0.33 (0.03)
200	0.50 (0.03)	1.56 (0.03)	0.34 (0.03)
300	0.50 (0.03)	1.55 (0.05)	0.34 (0.03)
500	0.50 (0.02)	1.54 (0.04)	0.33 (0.03)

One standard deviation is indicated between brackets. Calculations were carried out for 50 repetitions. The parameters to be estimated for each graph model are: the probability p of connecting two nodes for Erdös-Rényi graphs, the power of the preferential attachment p_s for scale-free graphs, and the rewiring probability p_r for small-world graphs.
doi:10.1371/journal.pone.0049949.t001

free, or small-world network. The KL divergence between the given graph spectrum and the spectrum of different classes of graphs can be interpreted as the quality of fitting the graph to the model.

Given a graph g and its spectrum ρ_g, several candidate graph models may be ranked according to their KL divergence values and the models with smaller KL divergence values should be considered as good candidates to explain the data. Thus, KL divergence provides an objective comparison among models, i.e., a tool for model selection. Specifically, let $\hat{\rho}_g$ be the empirical spectral distribution and $\{\rho_{\theta_1}\}, \ldots, \{\rho_{\theta_m}\}$ be m different parametric families of spectral distributions. Let $\hat{\theta}_i$ for $i = 1, \ldots, m$ be the estimators given in equation 6. We denote by $\#(\theta_i)$ the dimension of θ_i. The best candidate model $\hat{\theta}_j$ is chosen by

$$j = \arg\min_i 2KL(\hat{\rho}|\rho_{\hat{\theta}_i}) + 2\#(\hat{\theta}_i) \qquad (7)$$

The motivation for this criterion is the AIC (Akaike Information Criterion) [21] model selection criterion. Informally, the model that minimizes equation 7 is the one that has the most similar spectral distribution when compared to the spectral distribution of the data. The penalization term $2\#(\hat{\theta}_i)$ is added to avoid overfitting. The three random graph models analyzed here have the same number of parameters; therefore, the penalization term is not strictly required here, but may be necessary in more general settings.

Simulations were carried out in order to verify the accuracy of the proposed model selection approach. Ten thousand graphs of each class were generated and classified as Erdös-Rényi, scale-free, or small-world by the model selection approach. The graph size varied from 10 to 120 nodes. Figure 2 illustrates the performance of the model selection method. For all graph class (Erdös-Rényi (Figure 2A), scale-free (Figure 2B) or small-world (Figure 2C)), when the number of nodes increases, the correct proportion of hits also increases, demonstrating that the method is consistent and improves with the graph size.

Usually, in real applications, complex networks are composed of hundreds to thousands of nodes. In Figure 2, we observe that the accuracy is high even for graphs smaller than 100 nodes. Indeed, this implies that the proposed model selection method should be useful for applications in data set with realistic data size.

Interestingly, the performance to identify small-world graphs is very high, close to 100% even when the graph is very small (10 nodes). This is probably due to the specific algorithm used to

construct such a graph. Remember that the construction of a small-world graph based on Watts-Strogatz algorithm starts with a deterministic step, i.e., a ring lattice with n nodes which every node is connected to its first K neighbors ($K/2$ on either side). It is likely that this first step results in a more deterministic spectrum, in comparison to Erdös-Rényi or scale-free graphs that are totally non-deterministic.

Jensen-Shannon divergence

Given two random graphs g_1 and g_2, now we would like to define a notion of distance between them based on entropy. In other words, we are interested in identifying graphs that are generated by the same random process instead of isomorphism in graphs (an isomorphism of graphs g_1 and g_2 is a bijection f from the vertex sets of g_1 to the vertex sets of g_2 such that any two vertices u and v of g_1 are adjacent if and only if $f(u)$ and $f(v)$ are adjacent in g_2)

The KL divergence is suited for the purpose of parameter estimation and model selection as explained in previous section. Nevertheless, it is not symmetric, i.e., in general $KL(\rho_1|\rho_2) \neq KL(\rho_2|\rho_1)$. For this reason, KL divergence is not suited when it is not clear which distribution is the reference distribution. This is indeed the case for statistical test comparing two graphs spectra ρ_1 and ρ_2. We would like to avoid inconsistency in the results when considering $KL(\rho_1|\rho_2)$ or $KL(\rho_2|\rho_1)$.

Therefore, we introduce the *Jensen-Shannon divergence* (JS) between two spectral densities ρ_{g_1} and ρ_{g_2} defined as

$$JS(\rho_{g_1}, \rho_{g_2}) = \frac{1}{2}KL(\rho_{g_1}|\rho_M) + \frac{1}{2}KL(\rho_{g_2}|\rho_M) \qquad (8)$$

where $\rho_M = \frac{1}{2}(\rho_{g_1} + \rho_{g_2})$.

This divergence is symmetric and non-negative. It is also zero if and only if ρ_{g_1} and ρ_{g_2} are equal. Moreover, the square root of the JS divergence satisfies the triangle inequality.

It is natural to ask if the JS divergence between two distributions is zero or not. Therefore, we set the statistical test for JS divergence between two sets of graphs spectra ρ_{g_1} and ρ_{g_2} as ($H_0 : JS(\rho_{g_1}, \rho_{g_2}) = 0$ versus $H_1 : JS(\rho_{g_1}, \rho_{g_2}) > 0$). Details of the respective bootstrap-based test are provided in the Materials and Methods section.

When a statistical test is proposed, at least two properties must be shown: the power of the test under the alternative hypothesis
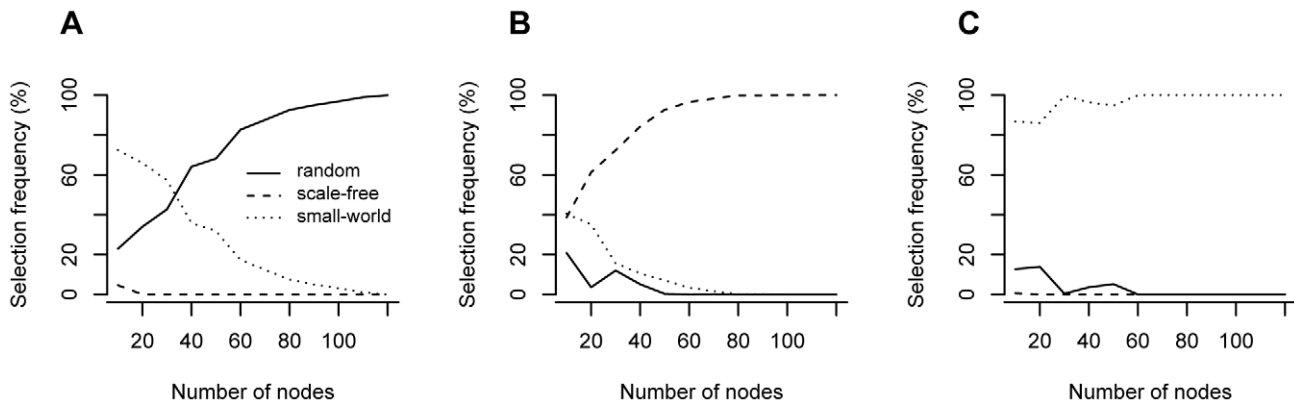
Figure 2. Figure illustrating the performance of the model selection approach as a function of number of nodes. Given a graph belonging to (**A**) Erdös-Rényi with parameter $p = 0.3$, (**B**) scale-free with parameter $p_s = 1$, and (**C**) small-world with parameter $p_r = 0.3$, the solid, dashed, and dotted lines represent the proportion of graphs classified as Erdös-Rényi, scale-free, and small-world, respectively. Notice that the larger is the graph, the higher is the proportion of correct hits, showing that the model selection approach is consistent. For each graph size (10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120 nodes), 1,000 repetitions were carried out.
doi:10.1371/journal.pone.0049949.g002

(H_1) and the control of the rate of false positives under the null hypothesis (H_0).

In order to check the power of the statistical test, i.e., if the method based on the spectral distribution actually discriminates between two sets of graphs characterized by slightly different parameters (details in the Materials and Methods section), receiver operating characteristic (ROC) curves were constructed and compared to the test based on the degree distribution. The ROC curve is useful in evaluating the power of the test and it consists in a bidimensional plot of sensitivity (y-axis) versus 1 - specificity (x-axis), where sensitivity = number of true positives/ (number of true positives+number of false negatives) and specificity = number of true negatives/(number of true negatives+number of false positives). The area below the ROC curve is a quantitative summary of the power of the test. In other words, an area closer to one (a curve above the diagonal line) denotes high power while an area close to 0.5 (a curve close to the diagonal line) is equivalent to random decisions. The top panels in Figure 3 illustrate the ROC curves with 10,000 repetitions for each class (Erdös-Rényi, scale-free, and small-world). The solid and dashed lines represent the test based on the spectral and degree distributions, respectively. Despite the small differences between the two conditions (parameters $p_1 = 0.10$ versus $p_2 = 0.11$ for Erdös-Rényi graphs; the scaling exponent $p_{s1} = 1.0$ versus $p_{s2} = 1.1$ for scale-free networks; and $p_{r1} = 0.30$ versus $p_{r2} = 0.31$ for small-world graphs) and relatively small sizes (100 nodes), our statistical test based on the spectra was able to identify the graphs that were generated by different sets of parameters with high accuracy as can be observed by the ROC curves clearly above the diagonal line. On the other hand, the statistical test based on the degree distribution had comparable power to the spectra-based test only when the evaluated networks were Erdös-Rényi graphs. When the degree-based test was applied to scale-free and small-world graphs, the discriminative power was not much better than by chance, i.e., the ROC curves were close to the diagonal. This probably occurred because the degree distribution is closely related to the number of edges while the spectrum is related to the whole structure of the graph. Notice that the parameter p of the Erdös-Rényi graph is associated to the number of edges, while the parameters p_s of the scale-free network and p_r of the small-world network are associated to the structure of the graph.

It is also necessary to verify if the bootstrap-based test is actually controlling the rate of false positives under the null hypothesis, i.e., when both sets of graphs are generated by the same model and same set of parameters. By simulating two random graphs g_1 and g_2, each one generated by the same model and parameters (see Materials and Methods section), and testing $H_0 : JS(\rho_{g_1}, \rho_{g_2}) = 0$ versus $H_1 : JS(\rho_{g_1}, \rho_{g_2}) > 0$, the p-value distribution should be a uniform distribution. The uniform distribution of p-values illustrates that the rate of false positives is actually controlled by our bootstrap procedure under any p-value threshold. Notice that for a p-value threshold set to 1%, it is expected to obtain 1% of false positives, for a threshold of 5%, 5% are expected to be false positive and so on and so forth. The bottom panels in Figure 3 show the p-value distributions (x-axis represents the p-values while the y-axis is the frequency or density of the respective p-value in 10,000 repetitions under the null hypothesis), one for each class (Erdös-Rényi, scale-free, and small-world), indicating that all of them are very similar to uniform distributions on [0,1] under the null hypothesis. In other words, the bootstrap test is controlling the rate of false positives, as expected.

Application to protein-protein interaction network

In order to illustrate the model selection application in actual data, protein-protein interaction data were downloaded from the DIP (Database of Interacting Proteins [22]) on June 29th, 2011. The DIP database is composed of eight species namely, *H. pylori* (bacterium), *R. norvegicus* (rat), *M. musculus* (mouse), *E. coli* (bacterium), *C. elegans* (worm), *S. cerevisiae* (yeast), *H. sapiens* (human), *D. melanogaster* (fruit fly). All of them present different number of nodes, edges, average degree, diameter, clustering coefficient and average path length as can be visualized in Table 2. The adjacency matrices of graphs were constructed for each species and the set of eigenvalues with the corresponding multiplicities were calculated. The frequency plot for the eigenvalues of the adjacency matrix for the eight species are displayed in Figure 4.

We evaluate how successful our algorithm based on the graph spectrum and KL divergence is by analyzing those protein-protein interaction networks that have already been classified as scale-free graphs by considering the degree distribution [23].

Remarkably, all the eight species were classified as scale-free networks by our model selection approach based on the graph
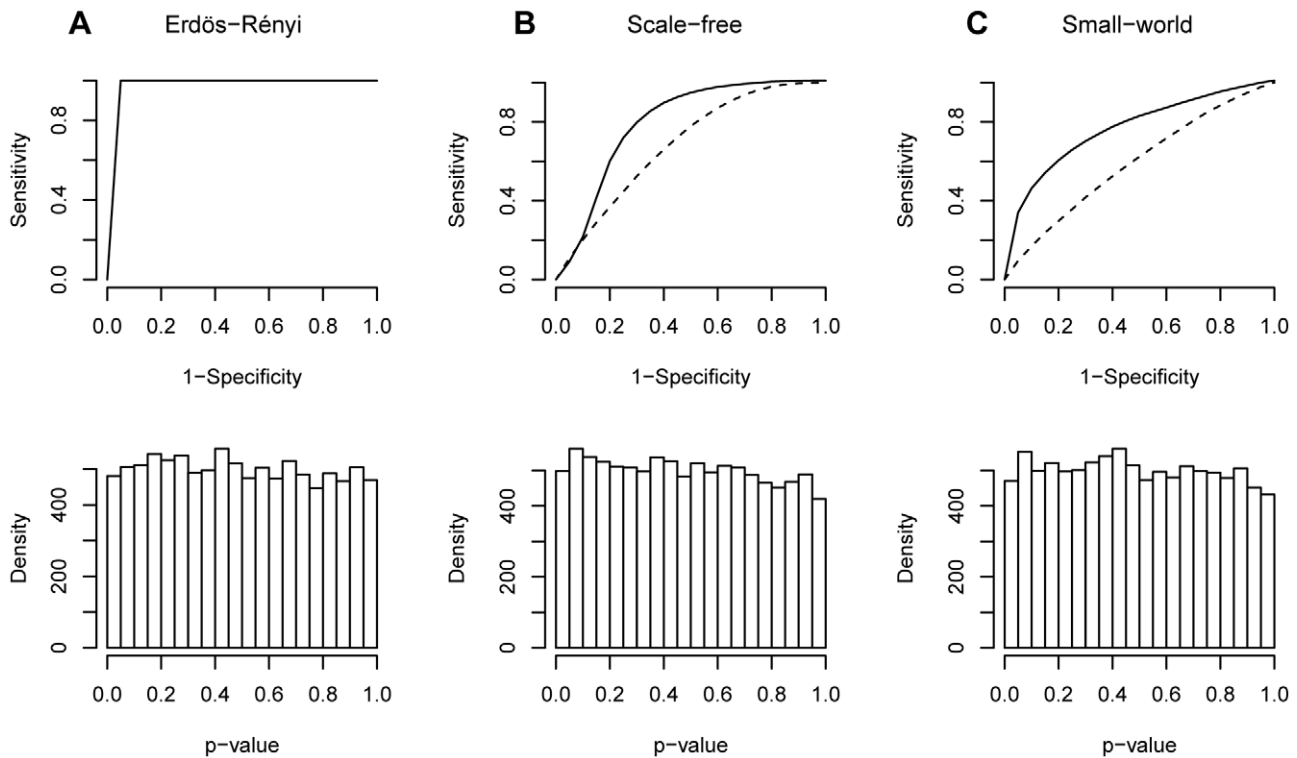
Figure 3. ROC curve under the alternative hypothesis and p-value distribution under the null hypothesis. (**A**) Erdös-Rényi graphs; (**B**) scale-free graphs, and (**C**) small-world graphs. For the ROC curves, the x-axis represents the 1-specificity and the y-axis the sensitivity. Both ROC curves and p-value distributions were constructed by analyzing 10,000 experiments. Solid and dashed lines represent the test based on the spectral and degree distributions, respectively.
doi:10.1371/journal.pone.0049949.g003

spectrum analysis (instead of the degree distribution) (Table 3) demonstrating that not only the degree distribution, but also the spectrum contains information for classification.

Application to neuroscience data

Application of JS divergence measure ("distance" between graphs) and its respective statistical test is illustrated in fMRI data of children diagnosed with Attention Deficit Hyperactivity Disorder (ADHD) and children with typical development. ADHD is a developmental disorder that affects at least 5–10% of children and is associated with difficulty on staying focused, on paying attention, difficulty controlling behavior, and hyperactivity [24]. Despite several efforts, there is no comprehensive model of this pathophysiology and the treatment is usually focused on medication that reduces the symptoms and improves functioning [25]. In order to provide new insights for this disease by using our proposed methodology, pre-processed functional magnetic resonance imaging (fMRI) data, from normal individuals and subjects diagnosed with ADHD, was downloaded from The Neuro Bureau as well as the ADHD-200 consortium [26]. The data is based on monitoring the BOLD (blood oxygenation level dependent) at different brain regions, which can be considered as an indirect measure of local neuronal activity [28]. The data was acquired under a resting state protocol, which is associated with the observation of brain spontaneous activity [27].

Pairwise Spearman correlation was calculated among 351 mean signals at different regions (using CC400 Atlas, only regions larger than five voxels) and a threshold of p-value = 0.05 (after FDR correction [29]) was set to determine the existence of an edge. The correlation between these regions describes the functional

connectivity of spontaneous activity at these areas. In other words, an adjacency matrix for each subject was constructed by considering a p-value < 0.05 as 1 and 0 otherwise. Network topological comparisons were carried out between the 478 children with typical development against 158 with combined type of ADHD (hyperactive-impulsive and inattentive).

Differences in the topology between children with typical development and with ADHD were estimated by our approach based on graph spectral distribution and four robust and often used measures, namely number of edges, clustering coefficient, average path length, and degree distribution. The Wilcoxon test was carried out in order to test differences in the number of edges, clustering coefficient, and the average path length. For the degree distribution, we applied the JS based test, similar to the one applied to test differences in the spectra. Table 4 shows that no statistical evidences to discriminate the two groups of children were identified by the number of edges (p-value = 0.82), clustering coefficient (p-value = 0.85), and average path length (p-value = 0.87). However, by analyzing the degree and spectral distributions (Figure 5), significant statistical differences were found (p-value = 0.031 for degree distribution and p-value = 0.024 for spectral distribution).

In order to check whether the differences in the spectral distributions are not due to numerical fluctuation, the control of the rate of false positives in biological data was verified. The set of 478 children with typical development was split randomly into two subsets, and the JS divergence test in graphs spectra was applied between these subsets. This procedure was repeated 10,000 times. The proportion of falsely rejected hypothesis for p-values equal to 0.1, 1, 5, and 10% were 0.16, 1.04, 5.55, and 11.05%,
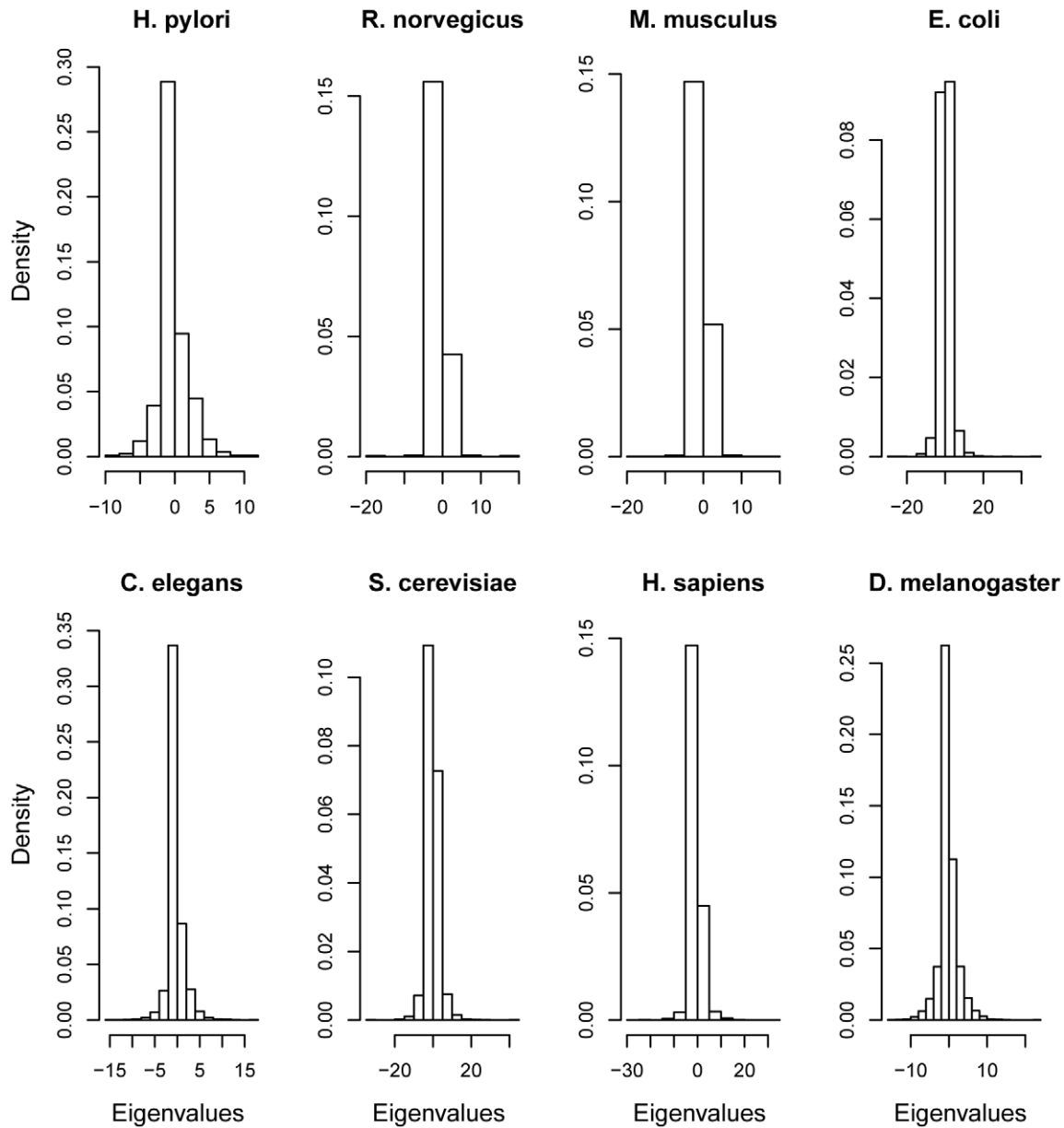
Figure 4. The frequency plots. The frequency plot for the eigenvalues of the eight species (*H. pylori, R. norvegicus, M. musculus, E. coli, C. elegans, S. cerevisiae, H. sapiens, D. melanogaster*).
doi:10.1371/journal.pone.0049949.g004

Table 2. The general characteristics of eight protein-protein interaction networks. For each network we indicate the number of nodes, the number of edges, the average degree, the diameter, the clustering coefficient and the average path length.

Species	Number of nodes	Number of edges	Average degree	Diameter	Clustering coefficient	Average path length
H. pylori	714	1,393	3.90	9	0.016	4.139
R. norvegicus	758	691	1.82	9	0.001	3.651
M. musculus	1,868	1,895	2.03	20	0.006	6.280
E. coli	2,997	12,348	8.24	12	0.115	3.986
C. elegans	3,183	5,068	3.18	13	0.012	4.803
S. cerevisiae	5,213	25,073	9.62	10	0.058	3.860
H. sapiens	5,940	14,144	4.76	17	0.017	4.755
D. melanogaster	7,931	23,386	5.90	12	0.012	4.468

doi:10.1371/journal.pone.0049949.t002

Table 3. The estimated Kullback-Leibler divergence between the eight species and the three random graph models. In bold are the lowest KL divergence values.

Species	Erdös-Rényi	Scale-free	Small-world
H. pylori	15.07	**1.46**	11.36
R. norvegicus	134.67	**100.47**	118.67
M. musculus	14.10	**6.93**	24.51
E. coli	21.15	**1.91**	17.90
C. elegans	30.48	**2.66**	30.23
S. cerevisiae	24.21	**0.87**	18.25
H. sapiens	47.10	**11.31**	44.04
D. melanogaster	17.40	**0.39**	18.06

doi:10.1371/journal.pone.0049949.t003

respectively, confirming that the type I error is effectively controlled in this biological data. Moreover, in order to verify the site effect, the JS based test on the spectra was carried out among laboratories. The tests were carried out under the null hypothesis, i.e., in typical development children datasets of different laboratories. Table 5 shows the p-values after Bonferroni correction for multiple tests. Notice that since no null hypothesis was rejected (significance level of 0.05), there are no statistical evidences of site effect that may significantly affect our results. These results suggest that the differences between children with typical development and with ADHD graphs spectra are statistically significant.

Discussion

The topology of the network represents the set of interactions between the nodes of the network. The topology affects the system's dynamics and carries information about the functional needs of the system, its evolution and the role of each individual unit [30]. Therefore, network analyses comparing control cases and disease cases is becoming a reference in the medical area [31]. Findings of significant differences when doing this comparison will

possibly lead to the improvement of diagnostic, prognostic, and therapy.

Most of the network analyses are based on algorithms that identify punctual changes (presence or absence of a certain edge) in their node connectivity. However, in Systems Biology, different subjects with the same disease may display topologically different molecular networks or brain networks due to genetic variability rather than disease variability. Therefore, a single graph will probably not be representative of the network; instead, a class of graphs generated by a random mechanism seems to be more appropriate.

This situation requires statistical procedures to analyze graphs. The difficulty is then to understand which parameter is representative of the class of graphs. The spectral distribution of a graph gives characteristics for ensemble of graphs generated by the random graphs, and the entropy of a spectrum and Kullback-Leibler divergence between spectra are natural information theoretical quantities to be studied.

Parameter estimation

For some classes of graphs, the parameters of the model can be easily estimated. For example, the parameter p of an Erdös-Rényi graph can be estimated by counting the edges and dividing it by the total number of possible edges of the graph ($n^2 - n$). However, for more complex models such as the small-world graph proposed by Watts and Strogatz, it is not trivial to estimate the probability p_r of edge permutation. Here, we demonstrated that the estimator based on the KL minimum distance (equation 6) is a general and straightforward method that can be successfully applied to estimate parameters of diverse complex networks.

One may argue whether the application of KL minimum distance estimator could not be applied to degree distribution instead of the graph spectrum. Notice in Figure 3 that the degree distribution showed a lower power to discriminate graphs generated by different parameters than the spectra. Therefore, the spectrum might be a better feature to be analyzed than the degree in order to estimate the parameters.

Model selection

Jeong and others [23] were the first group to classify protein-protein interaction networks as scale-free graphs by analyzing the
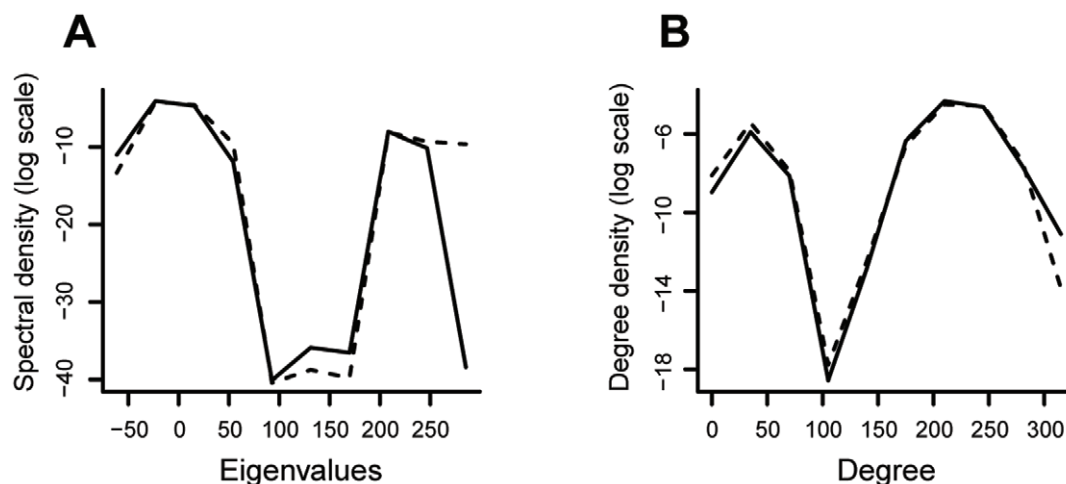

Figure 5. (A) Spectral and (B) degree distributions in the log-scale. Solid line represents the children with typical development. Dashed line represents children with combined type of ADHD (hyperactive-impulsive and inattentive).
doi:10.1371/journal.pone.0049949.g005

Table 4. Different metrics to measure graph discrepancy between children with typical development and children with combined type of ADHD (hyperactive-impulsive and inattentive) and their respective p-values.

| | Number of | Clustering | Average path | Degree | Spectrum |
	edges	coefficient	length	Distribution	
normal vs ADHD	0.82	0.85	0.87	0.031	0.024

For number of edges, clustering coefficient and average path length, the Wilcoxon test was carried out. For degree and spectral distributions, the JS divergence with the bootstrap test was calculated.

degree distribution. Later, several other groups re-analyzed the degree distribution of protein-protein interaction networks and came to differing conclusions regarding whether it was appropriate to refer to these graphs as scale-free [32,33]. One difficulty was the lack of an objective statistical procedure to decide which random graph model fits better the data set.

By applying our model selection approach it is possible to choose objectively, from a choice of candidate graph models, which model best fits the data. By our graph spectrum analysis, all the eight protein-protein interaction networks were classified as scale-free networks among Erdös-Rényi, scale-free, and small-world models. We note that, in the simulation study, our model selection approach has correctly classified 100% of the graphs with 120 nodes and the protein-protein interaction networks analyzed here are larger than 700 nodes, which adds to the evidence that among these three candidate networks, the scale-free network seems to fit better.

Despite these results, it is important to notice that the model selection approach is an objective criterion to select the model that best fits the data *among candidate models*. Therefore, by analyzing the graph spectrum instead of the degree distribution, this study only provides one more piece of evidence that scale-free graphs fit better to protein-protein interaction networks than ER and small-world networks. If another complex network model is proposed, one may use this approach to verify which one best fits the given graph.

Another point to be analyzed is the fact that, since only part of the protein-protein network is available, it is always possible that the observed sample is not representative of the entire network, consequently, resulting in a sampling artifact problem [34]. Unfortunately, this is a problem about the original data set that should be addressed when the data is collected or by introduction

Table 5. P-values obtained by testing the Jensen-Shannon divergence in the spectra distributions among different laboratories.

Labs.	#2	#3	#4	#5	#6	#7
#1	0.9	1.0	0.2	1.0	1.0	1.0
#2		1.0	1.0	1.0	1.0	1.0
#3			1.0	1.0	1.0	1.0
#4				1.0	1.0	1.0
#5					1.0	1.0
#6						1.0

The tests were carried out under the null hypothesis, i.e., in typical development children datasets of different laboratories. The laboratories were numbered from one to seven and the p-values are after Bonferroni correction for multiple tests.
doi:10.1371/journal.pone.0049949.t005

of *a priori* model of the network. The analysis proposed here is conditioned to the quality of the data sets.

Materials and Methods

We present below the details of the computational experiments. The statistical analyses were done using custom made programs in R [35] (language and environment for statistical computing and graphics). The R library *igraph* was used to generate the random graphs.

Parameter estimation

The performance of the parameter estimator based on minimization of the KL divergence was evaluated on different complex network models namely Erdös-Rényi random graph, scale-free, and small-world, with sizes varying from 50 to 300 nodes. The parameters to be estimated are the probability $p = 0.50$, the scaling exponent of the preferential attachment $p_s = 1.50$ and the rewiring probability $p_r = 0.30$ for Erdös-Rényi, scale-free, and small-world networks, respectively. The spectral densities (ρ_g) of each graph were estimated by a Gaussian kernel regression using the Nadaraya-Watson estimator. Since the theoretical spectrum distribution (ρ_θ) is unknown for scale-free and small-world networks, the spectrum distribution was estimated by simulating 50 graphs and calculating the average spectra distribution ($\hat{\rho}_\theta$) as an approximation for the theoretical distribution (ρ_θ). A grid search was carried out in order to determine the argument θ that minimizes $KL(\hat{\rho}_g|\hat{\rho}_\theta)$.

Model selection

In order to evaluate the performance of the proposed model selection approach, one random graph G is generated (among Erdös-Rényi, scale-free, and small-world) with parameters $p = 0.3$ for Erdös-Rényi graph, $p_s = 1$ for scale-free graphs and $p_r = 0.3$ for small-world graphs, with sizes varying from 10 to 120 nodes. Then, the spectrum of G is estimated. In order to search the optimum set of parameters for each graph model (the set of parameters that minimizes the KL divergence), a grid search was carried out. Fifty graphs for each class ($g_1 =$ Erdös-Rényi random; $g_2 =$ scale-free; and $g_3 =$ small-world) are generated. The KL divergence is estimated between the spectrum of G and the average spectrum of the 50 graphs of each graph type (g_1, g_2, g_3). The graph model g_i ($i = 1, 2, 3$) which has the minimum KL divergence value between G and the three models (g_1, g_2, g_3) is the one which best fits G. This experiment was repeated 1,000 times for each graph type (Erdös-Rényi, scale-free, or small-world) and each graph size (10 to 120 nodes).

Statistical test for JS divergence between graph spectra

Given two sets of graphs g_1 and g_2, the test consists of verifying if the JS divergence between the average graph spectrum of set g_1

and the average graph spectrum of g_2 is zero or not. Formally, we test $H_0 : JS(\rho_{g_1}, \rho_{g_2}) = 0$ versus $H_1 : JS(\rho_{g_1}, \rho_{g_2}) > 0$.

One alternative to perform the test is to use a bootstrap procedure. The *bootstrap* was introduced in 1979 as a computer-based method for estimating the standard error of the statistic or to construct confidential intervals that could be used to provide a significance level for a hypothesis test [36].

Let $\#g_1$ and $\#g_2$ be the quantity of graphs contained in sets g_1 and g_2, respectively. The bootstrap implementation of this test is as follows:

1. Create a set of graphs spectra $\tilde{g}_1$ (the bootstrap sample) by resampling with replacement, $\#g_1$ spectra distributions from $g_1 \bigcup g_2$.

2. Create a set of graphs spectra $\tilde{g}_2$ (the bootstrap sample) by resampling with replacement, $\#g_2$ spectra distributions from $g_1 \bigcup g_2$.

3. Let $\rho_{\tilde{g}_1^i}$ is the i-th spectra distribution of $\tilde{g}_1$ and $\rho_{\tilde{g}_2^i}$ is the i-th spectra distribution of $\tilde{g}_2$. Calculate the average spectra distributions $\rho_{g_1^*}$, i.e., $\rho_{g_1^*}(\lambda) = \frac{\sum_{i=1}^{\#g_1} \rho_{\tilde{g}_1^i}(\lambda)}{\#g_1}$, and $\rho_{g_2^*}$, i.e. $\rho_{g_2^*}(\lambda) = \frac{\sum_{i=1}^{\#g_2} \rho_{\tilde{g}_2^i}(\lambda)}{\#g_2}$, of $\tilde{g}_1$ and $\tilde{g}_2$, respectively.

4. Calculate $\hat{J}S(\rho_{g_1^*}|\rho_{g_2^*})$ (the bootstrap replication).

5. Repeat steps 1 to 5 until obtaining the desired number of bootstrap replications.

6. Test if $\hat{J}S(\rho_{g_1}|\rho_{g_2}) = 0$ using the empirical distribution obtained in steps 1 to 5. Gather the information from the empirical distribution of $\hat{J}S(\rho_{g_1^*}|\rho_{g_2^*})$ to obtain a p-value for $\hat{J}S(\rho_{g_1}|\rho_{g_2}) = 0$, by analyzing the probability of obtaining values equal or greater than $\hat{J}S(\rho_{g_1^*}|\rho_{g_2^*})$.

The purpose of steps 1 and 2 is to construct new sets $\tilde{g}_1$ and $\tilde{g}_2$ that are under the null hypothesis. This is exactly done by sampling graphs spectra distributions from $g_1 \bigcup g_2$. In order to verify whether the bootstrap based statistical test is actually controlling the rate of false positives, p-value histograms under the null hypothesis were constructed. For each class of graph (Erdös-Rényi random, scale-free, and small-world), 100 graphs with 100 nodes with the same set of parameters ($p = 0.5$ for Erdös-Rényi graphs; $p_s = 1$ for scale-free graphs and $p_r = 0.3$ for small-world graphs) were constructed. The 100 graphs of each class were split into two sets of 50 graphs and the statistical test performed with 1,000 bootstrap resampling. These experiments were repeated 10,000 times in order to construct the p-value distributions.

We were concerned in evaluating the power of the proposed test, therefore the parameters of the 50 graphs of one group and the 50 graphs of the other were set with small differences. The parameters are set as follows: $p_1 = 0.50$ versus $p_2 = 0.52$ for Erdös-Rényi graphs; the scaling exponent $p_{s1} = 1.0$ versus $p_{s2} = 1.1$ for scale-free networks and $p_{r1} = 0.30$ versus $p_{r2} = 0.31$ for small-world graphs. The parameters p_1 and p_2 for Erdös-Rényi graphs represent the probability of a pair of nodes be connected by an edge. The parameters p_{s1} and p_{s2} represent the degree of proportionality (scaling exponent) that a new node in the scale-free graph will be connected to node i. For example, $p_s = 1$ means that the new node attaches to node i linearly proportional to the degree of node i. $p_s = 2$ means that the new node attaches to node i quadratic proportional to the degree of node i and so on and so forth. The parameters p_{r1} and p_{r2} represent the probability of rewiring (permuting the edges) in the small-world graph. All other parameters (number of nodes for Erdös-Rényi graphs and number of nodes and edges for scale-free and small-world graphs) were maintained equal between the two groups.

Conclusions and Future Applications

Our findings indicate that there are significant differences in the graph spectra of brain networks between children with and without ADHD. We anticipate that future studies in the field of graph spectra may illuminate the topological significance of these features, and consequently help in the investigation of the relationship of these differences with brain function.

The proposed approaches are flexible enough to allow generalizations to other arbitrarily sophisticated families of graphs. Here, we limited the analysis to three well-known classes of random graphs, but the analysis can be extended to other graphs without restriction and it is applicable to many areas where network data is a source of concern.

Acknowledgments

Author Contributions

Conceived and designed the experiments: DYT JRS CEF AF. Performed the experiments: DYT JRS AF. Analyzed the data: DYT AF. Contributed reagents/materials/analysis tools: DYT JRS AF. Wrote the paper: DYT AF.

References

1. Bollobás B, Riordan O (2004) The diameter of a scale-free random graph. Combinatorica 24: 5–34.

2. Huberman BA, Adamic LA (1999) Growth dynamics of the World-Wide-Web. Nature 401: 131.

3. Wasserman S, Faust K (1994) Social network analysis. Cambridge University Press, Cambridge.

4. Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. Science 296: 910–913.

5. Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. Nature 402: 47–52.

6. Eguíluz VM, Chialvo DR, Cecchi GA, Baliki M, Apkarian V (2005) Scale-free brain functional networks. Physical Review Letters 94: 018102.

7. Mieghem PV (2011) Graph spectra for complex networks. Cambridge University Press.

8. Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science 286: 509–512.

9. Watts DJ, Strogatz SH (1998) Collective dynamics of small-world networks. Nature 393: 440–442.

10. Erdös P, Rényi A (1959) On random graphs. Publicationes Mathematicae (Debrecen) 6: 290–297.

11. Newman MEJ, Watts DJ (1999) Renormalization group analysis of the small-world network model. Physics Letters A 263: 341–346.

12. Albert R, Barabasi A-L (2002) Statistical mechanics of complex networks. Reviews of Modern Physics 74: 47–97.

13. Nadaraya EA (1964) On estimating regression. Theory of probability and its applications 10: 186–190.

14. Sain SR, Scott W (1996) On locally adaptive density estimation. Journal of the American Statistical Association 91: 1925–1934.

15. Sturges HA (1926) The choice of a class interval. Journal of the American Statistical Association 21: 65–66.

16. Dorogovtsev SN, Goltsev AV, Mendes JFF, Samukhin AN (2003) Spectra of complex networks. Physical Review E E68: 046109.

17. Cover TM, Thomas JA (2006) Elements of information theory 2nd edition, Willey-Interscience: NJ.

18. Wigner E (1955) Characteristic vectors of bordered matrices with in_nite dimensions. Ann Math 62: 548–564.

19. Wigner E (1958) On the distribution of the roots of certain symmetric matrices. Ann Math 67: 325–328.

20. Whittle P (1953) Estimation and information in stationary time series. Ark Math 2: 423–434Z.

21. Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19: 716–723.

22. Database of Interacting Proteins. Available: http://dip.doe-mbi.ucla.edu/dip/. Accessed 2011 June 29.

23. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabsi A-L (2000) The large-scale organization of metabolic networks. Nature 407: 651–654.

24. American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders (4th edition) American Psychiatric Association, Washington, DC.

25. Singh I (2008) Beyond polemics: science and ethics of ADHD. Nature Reviews Neuroscience 9: 957–964.

26. ADHD-200 Preprocessed Data. Available: http://neurobureau.projects.nitrc.org/ADHD200/Introduction.html. Accessed 2011 August 20.

27. Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, et al. (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. PNAS 102: 9673–8.

28. Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investication of the basis of the fMRI signal. Nature 412: 150–157.

29. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 57: 289–300.

30. Guimerà R, Sales-Pardo M (2009) Missing and spurious interactions and the reconstruction of complex networks. PNAS 106: 22073–22078.

31. Barabási A-L, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nature Reviews Genetics 12: 56–68.

32. Khanin R, Wit E (2006) How scale-free are biological networks. Journal of Computational Biology 13: 810–818.

33. Lima-Mendez G, van Helden J (2009) The powerful law of the power law and other myths in network biology. Molecular BioSystems 5: 14821493.

34. Han J-D D, Dupuy D, Bertin N, Cusik ME, Vidal M (2005) Effect of sampling on topology predictions of protein-protein interaction networks. Nature Biotechnology 23: 839–844.

35. R Development Core Team (2011) R: A Language and Environment for Statistical Computing, ISBN 3-900051-07-0.

36. Efron B, Tibshirani RJ (1993) An introduction to the bootstrap, Chapman & Hall/CRC.