

fMRI Activation Network Analysis Using Bose-Einstein Entropy

Jianjia Wang^(✉), Richard C. Wilson, and Edwin R. Hancock

Department of Computer Science, University of York, York YO10 5DD, UK
jw1157@york.ac.uk

Abstract. In this paper, we present a novel method for characterizing networks using the entropy associated with bosonic particles in thermal equilibrium with a heat-bath. According to this analogy, the normalized Laplacian plays the role of Hamiltonian operator, and the associated energy states are populated according to Bose-Einstein statistics. This model is subject to thermal agitation by the heat reservoir. The physics of the system can be captured by using a partition function defined over the normalized Laplacian eigenvalues. Various global thermodynamic characterizations of the network including its entropy and energy then can be computed from the derivative of corresponding partition function with respect to temperature. We explore whether the resulting entropy can be used to construct an effective information theoretic graph-kernel for the purposes of classifying different types of graph or network structure. To this end, we construct a Jensen-Shannon kernel using the Bose-Einstein entropy for a sample of networks, and then apply kernel principle components analysis (kPCA) to map graphs into low dimensional feature space. We apply the resulting method to classify fMRI activation networks from patients with suspected Alzheimer disease.

Keywords: Bose-Einstein statistics · Network entropy · Jensen-Shannon divergence

1 Introduction

Graphs are powerful tools for representing complex patterns of interaction in high dimensional data [1]. For instance, in the application later considered in this paper, graphs provide the activation patterns in fMRI data, which can be indicative of the early onset of Alzheimer's disease. On the other hand, kernel-methods on graphs provide emerging and powerful set of tools to determine the class-structure of different graphs. There are many examples in the literature where graph kernels have successfully exploited topological information, and these include the heat diffusion kernel [2], the random walk kernel [3], and the shortest path kernel [4]. Once a graph kernel is to hand, it provides a convenient starting point from which machine learning techniques can be applied to learn potentially complex class-structure [5]. Despite the success of existing graph

kernels, one of the main challenges that remains open is to capture the variations present in different classes of graph in a probabilistic manner [6].

Recently, statistical mechanics and information theory have been used to understand more deeply variations in network structure. One of the successes here has been to use quantum spin statistics to describe the geometries of complex networks [7]. For example, using a physical analogy based on a Bose gas, the phenomenon of Bose-Einstein condensation has been applied to study the salient aspects network structure [8]. This has been extended to understand processes such as supersymmetry in networks [9]. Although these types of analogy are useful and provide powerful tools for network analysis, they are not easily accommodated into the kernel-based approach to machine learning.

The aim in this paper is to bridge this gap in the literature. Our aim is to develop a link between statistical mechanics and kernel methods. To do that we define information theoretic kernels in terms of network entropy. The Jensen-Shannon kernel is computed from the Jensen-Shannon divergence, which is an entropic measure of the similarity between two probability distributions. To compute the divergence, the distribution of data must be characterized in terms of a probability distribution and the associated entropy must be to hand. For graph or network structures, both the probability distribution and the associated entropy can be elusive. To solve this problem, in prior work, they have computed the required entropy using von Neumann entropy (essentially the Shannon entropy of the normalized Laplacian eigenvalues) [5]. Here we aim to explore whether the physical heat bath analogy and Bose-Einstein statistics can be used to furnish the required entropy, and implicitly the underlying probability distribution.

We proceed as follows. We commence from a physical analogy in which the normalized Laplacian plays the role as Hamiltonian (energy operator) and the normalized Laplacian eigenvalues are energy states. These states are occupied by bosonic particles (assumed the 0 spin) and the resulting system is in thermodynamic equilibrium with a heat-bath, which is characterized by temperature. The bosons are indistinguishable, and each energy level can accommodate an unlimited number of particles. The effect of the heat bath is to thermalise or randomise the population of energy levels. The occupation of the energy states is therefore governed by Bose-Einstein statistics, and can be characterized using an appropriate partition function. The quantities can be derived from the partition function over the energy states in the network when the system of particles is in thermodynamic equilibrium with the heat bath. From the partition function, we can compute the entropy of the system of particles, and hence compute the Jensen-Shannon kernel. Once the kernel matrix is to hand, we use kernel principle components analysis (kPCA) [10] to embed the graphs into a low dimensional feature space where classification is performed.

The remainder of paper is organised as follows. Section 2 briefly reviews the preliminaries of graph representation in the statistical mechanical domain. Sections 3 and 4 respectively describe the underpinning concepts of how entropy is computed from the Bose-Einstein partition function, and how the Jensen-Shannon kernel is constructed from the resulting entropy. Section 5 presents our experimental evaluation. Finally, Sect. 6 provides conclusions and directions for future work.

2 Graph in Quantum Representation

In this section, we give the preliminaries on the graph representation in quantum domain. We specify the density matrix as the normalized Laplacian matrix. We then introduce the idea of a Hamiltonian operator on a graph and its relationship with the normalized Laplacian, which associated von Neumann entropy.

2.1 Density Matrix

The density matrix, in quantum mechanics, is used to describe a system whose state is an ensemble of pure quantum states $|\psi_i\rangle$, each with probability p_i . It is defined as $\rho = \sum_{i=1}^V p_i |\psi_i\rangle \langle \psi_i|$.

With this notation, Passerini and Severini [11] have extended this idea to the graph domain. Specifically, they show that a density matrix for a graph or network can be obtained by scaling the normalized discrete Laplacian matrix by the reciprocal of the number of nodes in the graph.

When defined in this way the density matrix is Hermitian i.e. $\rho = \rho^\dagger$ and $\rho \geq 0$, $\text{Tr} \rho = 1$. It plays an important role in the quantum observation process, which can be used to calculate the expectation value of measurable quantity.

2.2 Hamiltonian Operator

In quantum mechanics, the Hamiltonian operator is the sum of the kinetic energy and potential energy of all the particles in the system, and it dictates the Schrödinger equation for the relevant system. The Hamiltonian is given by

$$\hat{H} = -\nabla^2 + U(r, t) \quad (1)$$

Here we specify the kinetic energy operator to be the negative of the normalized adjacency matrix, and the potential energy to be the identity matrix. Then, the normalized form of the graph Laplacian can be viewed as the Hamiltonian operator $\hat{H} = \hat{L}$.

In this case, the eigenvalues of the Laplacian matrix can be viewed as the energy eigenstates, and these determine the Hamiltonian and hence the relevant Schrödinger equation which govern a system of particles. The graph as a thermodynamic system specified by N particles with energy states given by the network Hamiltonian and immersed in a heat bath at temperature T .

2.3 von Neumann Entropy

The interpretation of the scaled normalized Laplacian as a density matrix opens up the possibility of characterizing a graph using the von Neumann entropy from quantum information theory.

The von Neumann entropy is defined as the entropy of the density matrix associated with the state vector of a system. As noted above, the von Neumann entropy can be computed by scaling the normalized discrete Laplacian matrix

for a network [11]. As a result, it is given in terms of the eigenvalues $\lambda_1, \dots, \lambda_V$ of the density matrix ρ ,

$$S = -\text{Tr}(\rho \log \rho) = -\sum_{i=1}^V \frac{\hat{\lambda}_i}{V} \log \frac{\hat{\lambda}_i}{V} \quad (2)$$

Since the normalized Laplacian spectrum has been proved to provide an effective characterization for networks or graphs [12], the von Neumann entropy derived from the spectrum may also be anticipated to be an effective tool for network characterization.

In fact, Han et al. [13] have shown how to approximate the calculation of von Neumann entropy in terms of simple degree statistics. Their approximation allows the cubic complexity of computing the von Neumann entropy from the Laplacian spectrum, to be reduced to one of quadratic complexity using simple edge degree statistics, i.e.

$$S = 1 - \frac{1}{V} - \frac{1}{V^2} \sum_{(u,v) \in E} \frac{1}{d_u d_v} \quad (3)$$

This expression for the von Neumann entropy allows the approximate entropy of the network to be efficiently computed and has been shown to be an effective tool for characterizing structural property of networks, with extremal values for cycle and fully connected graphs.

3 Quantum Statistics and Network Entropy

In this section, we describe the statistical mechanics for particles with quantum spin statistics. We then explain how the Laplacian eigenstates are occupied by a system of bosonic particles in equilibrium with a heat bath. This can be characterized using the Bose-Einstein partition function. From the partition function, various thermodynamic quantities, including entropy, can then be computed.

3.1 Thermal Quantities

Based on the heat bath analogy, particles occupy the energy states of the Hamiltonian subject to thermal agitation. The number of particles in each energy state is determined by (a) the temperature, (b) the assumed model of occupation statistics, (c) the relevant chemical potential.

When specified in this way, the various thermodynamic characterizations of the network can be computed from the partition function $Z(\beta, N)$, where $\beta = 1/k_B T$ and k_B is the Boltzmann constant. For instance, the Helmholtz free energy of the system is

$$F(\beta, N) = -\frac{1}{\beta} \log Z(\beta, N) = -k_B T \log Z(\beta, N) \quad (4)$$

and the thermodynamic entropy is given by

$$S = k_B \left[\frac{\partial}{\partial T} T \log Z \right]_N = - \left(\frac{\partial F}{\partial T} \right)_N \tag{5}$$

The incremental change in Helmholtz free energy is related to the incremental change in β and N ,

$$dF = \left(\frac{\partial F}{\partial \beta} \right)_N d\beta + \left(\frac{\partial F}{\partial N} \right)_T dN = \frac{S}{k_B} \frac{1}{\beta^2} d\beta + \mu dN \tag{6}$$

where μ is the chemical potential, given by

$$\mu = \left(\frac{\partial F}{\partial N} \right)_T = -k_B T \left[\frac{\partial}{\partial N} \log Z \right]_\beta \tag{7}$$

3.2 Bose-Einstein Statistics

Specified by the network Hamiltonian, each energy state can accommodate an unlimited number of integral spin particles. These particles are indistinguishable and subject to Bose-Einstein statistics. In other words, they do not obey the Pauli exclusion principle, and can aggregate in the same energy state without interacting.

Base on the thermodynamic system specified by N particles with the energy states, the Bose-Einstein partition function is defined as the product of all sets of occupation number in each energy state, and the matrix form is given as

$$Z_{BE} = \det \left(I - e^{\beta\mu} \exp[-\beta\tilde{L}] \right)^{-1} = \prod_{i=1}^V \left(\frac{1}{1 - e^{\beta(\mu - \varepsilon_i)}} \right) \tag{8}$$

where the chemical potential μ is defined by Eq. (7), indicating the varying number of particles in the network. From Eq. (5), the corresponding entropy is

$$\begin{aligned} S_{BE} &= \log Z + \beta \langle U \rangle = -\text{Tr} \left\{ \log [I - e^{\beta\mu} \exp(-\beta\tilde{L})] \right\} \\ &\quad - \text{Tr} \left\{ \beta [I - e^{\beta\mu} \exp(-\beta\tilde{L})]^{-1} (\mu I - \tilde{L}) e^{\beta\mu} \exp(-\beta\tilde{L}) \right\} \\ &= - \sum_{i=1}^V \log \left(1 - e^{\beta(\mu - \varepsilon_i)} \right) - \beta \sum_{i=1}^V \frac{(\mu - \varepsilon_i) e^{\beta(\mu - \varepsilon_i)}}{1 - e^{\beta(\mu - \varepsilon_i)}} \end{aligned} \tag{9}$$

Given the temperature $T = 1/\beta$, the average number of particles at the energy level indexed i with energy ε_i is

$$n_i = -\frac{1}{\beta} \left(\frac{\partial \log Z}{\partial \varepsilon_i} \right) = \frac{1}{\exp[\beta(\varepsilon_i - \mu)] - 1} \tag{10}$$

and as a result the total number of particles in the network is

$$N = \sum_{i=1}^V n_i = \sum_{i=1}^V \frac{1}{\exp[\beta(\varepsilon_i - \mu)] - 1} \quad (11)$$

The matrix form is

$$N = \text{Tr} \left[\frac{1}{\exp(-\beta\mu) \exp[\beta\tilde{L}] - I} \right] \quad (12)$$

where I is the identity matrix.

In order for the number of particles in each energy state to be non-negative, the chemical potential must be less than the minimum energy level, i.e. $\mu < \min \varepsilon_i$. Thus, the entropy derived from Bose-Einstein statistics is related to the temperature, energy states and chemical potential.

4 Graph Kernel Construction

In this section, we show how to compute the Jensen-Shannon divergence between a pair of graphs using the network entropy derived from the heat bath analogy and Bose-Einstein statistics. Once the combined the idea from graph embedding, we establish graph kernel associated with kernel principle component analysis (kPCA) to classify graphs.

4.1 Jensen-Shannon Divergence

The Jensen-Shannon kernel [14] is given by

$$k_{JS}(G_i, G_j) = \log 2 - D_{JS}(G_i, G_j) \quad (13)$$

where $D_{JS}(G_i, G_j)$ is the Jensen-Shannon divergence between the probability distributions defined on graphs G_i and G_j .

We now apply the kernel method of Jensen-Shannon divergence to construct a graph kernel between pairs of graphs. Suppose the graphs are represented by a set $G = G_i, i = 1, \dots, n$, where $G_i = (V_i, E_i)$. V_i is the set of nodes on graph G_i and $E_i \subset V_i \times V_i$ is the set of edges.

For a pair of graphs G_i and G_j , the union graph is defined as $G_i \oplus G_j$, which has the nodes on both graphs G_i and G_j , that is $G_U = V_i \cup V_j$. It also contains the edge sets between pairs of nodes, such that if $(k, l) \in E_i$, and $(k, l) \in E_j$, then $((k, l), (k, l)) \in E_U$, which means the union graph contains the combined edges of two graphs. With the union graph to hand, the Jensen-Shannon divergence for a couple of graphs G_i and G_j is

$$D_{JS}(G_i, G_j) = H(G_i \oplus G_j) - \frac{H(G_i) + H(G_j)}{2} \quad (14)$$

where $H(G_i)$ is the entropy associated with the graph G_i , and $H(G_i \oplus G_j)$ is the entropy associated with the corresponding union graph G_U .

Using the Bose-Einstein entropy in Eq. (9) for the graphs G_i , G_j and their union $G_i \oplus G_j$, we compute the Jensen-Shannon divergence for the pair of graphs and hence the Jensen-Shannon kernel matrix.

With the graph kernel to hand, we apply kernel-PCA to embed the graphs into a vector space. To compute the embedding, we commence by computing the eigen decomposition of the kernel matrix, which will reproduce the Hilbert space with a non-linear mapping. In such a case, graph features can be mapped to low dimensional feature space with linear separation. So the graph kernel can be decomposed as

$$k_{JS} = U\Lambda U^T \quad (15)$$

where Λ is the diagonal eigenvalue matrix and U is the matrix with eigenvectors as columns. To recover the matrix X with embedding co-ordinate vectors as columns, we write the kernel matrix in Gram-form, where each element is an inner product of embedding co-ordinate vectors

$$k_{JS} = XX^T \quad (16)$$

and as a result $X = \sqrt{\Lambda}U^T$.

5 Experiments

In this section, we present experiments on fMRI data. Our aim is to explore whether we can classify the subjects on the basis of similarity of the activation networks from the fMRI scans. To do this we embed the network similarity data into a vector-space by applying kernel-PCA to the Jensen-Shannon kernel. To simplify the calculation, the Boltzmann constant is set to unity through the experiment.

5.1 Dataset

The fMRI data comes from the ADNI initiative [15]. Each patient lies in the MRI scanner with eyes open. BOLD (BOLD: Blood-Oxygenation-Level-Dependent) fMRI image volumes are acquired every two seconds. The fMRI signal at each time point is measured in volumetric pixels over the whole brain. The voxels here have been aggregated into larger regions of interest (ROIs), and the blood oxygenation time-series in each region has been averaged, yielding a mean time-series for each ROI. The correlation between the average time series in different ROIs represents the functional connectivity between regions.

We construct the graph using the cross-correlation coefficients for the average time serial pairs of ROIs. To do this we create an undirected edge between two ROI's if the cross-correlation co-efficient between the time series is in the top 40 % of values. The threshold is fixed for all the available data, which provides an optimistic bias for constructing graphs. Those ROIs which have missing time series data are discarded. Subjects fall into four categories according to their degree of disease severity, namely full Alzheimer's (AD), Late Mild Cognitive

Impairment (LMCI), Early Mild Cognitive Impairment (EMCI) and Normal. The LMCI subjects are more severely affected and close to full Alzheimer’s, while the EMCI subjects are closer to the healthy control group (Normal). There are 30 subjects in the AD group, 34 in LMCI, 47 in EMCI and 38 in the Normal group.

5.2 Experimental Results

Now we describe the application of the above methods to investigate the structural dissimilarity of the fMRI activation networks, which is used to distinguish different groups of patients. We compute the Jensen-Shannon kernel matrix using the Bose-Einstein entropy and compare the performance with that obtained from von Neumann entropy. Given the spectra of a graph and the total number of particles, the chemical potential can be derived from Eq. (11), which is used to calculate the entropy. Fig. 1 shows the results of mapping the graphs into a 3-dimensional feature space obtained by kernel principal components analysis (kPCA). We use first three eigenvectors to show the cluster of each group. The common feature is that both the Bose-Einstein and von Neumann entropies separate the four groups of subjects. In the case of Bose-Einstein statistics, the clusters are better separated than those obtained with von Neumann entropy.

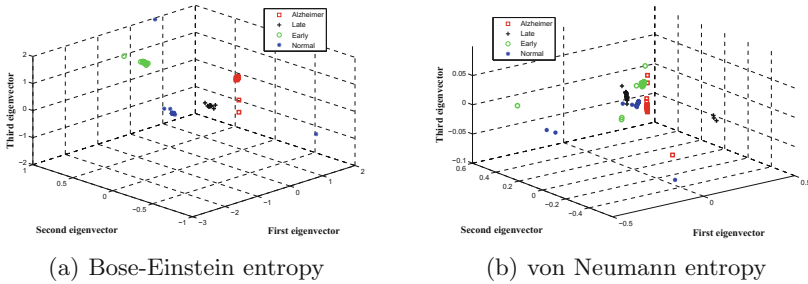


Fig. 1. Kernel PCA performance of Jensen-Shannon divergence in Bose-Einstein entropy (Fig. 1(a)) and von Neumann entropy (Fig. 1(b)). Temperature $\beta = 10$ and particle number $N = 1$.

To place our analysis on a more quantitative footing, we apply Fisher’s linear discriminant analysis to classify graphs with the kernel features, and compute the classification accuracy. Since the number of sampling in the dataset is small, we apply the leave-one-out cross-validation and use all the graphs as the testing data. Table 1 summarizes the results of classification accuracy obtained by Jensen-Shannon kernels computed from the two entropies. Compared to the accuracy with von Neumann entropy, that obtained with Bose-Einstein entropy exhibits a higher classification accuracy. The Maxwell-Boltzmann entropy outperforms the von Neumann entropy on three classes of data presented by a margin of

Table 1. Classification accuracy for entropy from Bose-Einstein statistics and von Neumann entropy

Classification accuracy	Alzheimer	LMCI	EMCI	Normal
Bose-Einstein statistics	93.33 % (28/30)	100 % (34/34)	89.36 % (42/47)	92.11 % (35/38)
von-Neumann entropy	93.33 % (28/30)	88.24 % (30/34)	82.98 % (39/47)	86.84 % (33/38)

about 10 %. This reveals that the proposed graph kernel computed with Jensen-Shannon Divergence and Bose-Einstein entropy improve the performance for the fMRI data classification.

5.3 Discussion

The main parameters of the Bose-Einstein entropy are the temperature and number of particles in the system. In this section, we discuss the effects of the temperature on the energy level occupation statistics and hence upon the entropic kernel performance at low and high temperatures. We first focus on the average number of particles given the temperature β at each energy level ε_i from Eq. (10). In Fig. 2(a), we plot the occupation number for the different normalised Laplacian energy states with different values of temperature.

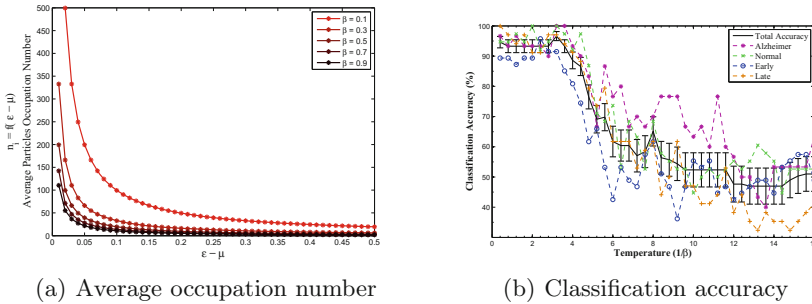


Fig. 2. Average occupation number for energy state set different temperature for Bose-Einstein statistics (Fig. 2(a)). Classification accuracy changes with temperature in Jensen-Shannon divergence with entropy from Bose-Einstein statistics (Fig. 2(b)).

As shown in this figure, with fixed temperature and increasing energy, the number of particles in each energy level decreases. As a result the lower energy levels are occupied with the largest number of particles. Furthermore, as the temperature decreasing, the number of particles in each energy state decreases. From Eq. (10), it should be noted that the number of particles in each state is determined by two factors, namely (a) the Bose-Einstein occupation statistics, and (b) the number of particles as determined by the chemical potential.

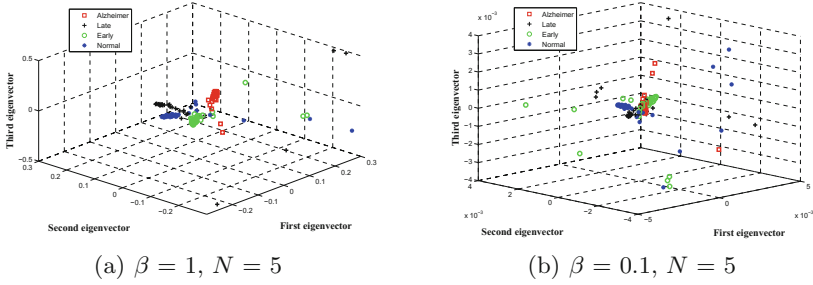


Fig. 3. Kernel PCA performance of Jensen-Shannon divergence with entropy from Bose-Einstein statistics at different values of temperature.

In order to evaluate how temperature effects the performance of the Jensen-Shannon kernel, we compare its behaviour at low and high temperature. For the fMRI brain activation data, we set $\beta = 1$ and $\beta = 0.1$, leaving the total particle number $N = 5$ unchanged. Compared to the low temperature case ($\beta = 10$) in Fig. 1, increasing temperature makes the four classes of graphs more densely clustered in feature space, shown in Fig. 3(a) and (b). This is term which reduces the performance of kernel PCA. Figure 2(b) shows the performance of classification changes with temperature. As the temperature increasing, the occupation number at each energy level increases and particles become to propagate at the high energy states. The probabilities of energy states in the system become identical to each other, reaching the uniform distribution as the temperature approaching to infinite. So all the groups with the same number of states, in statistical perspective, are rather similar to each other at high temperature. This reduces the performance of classification accuracy.

6 Conclusion

In this paper, we show how to compute an information theoretic graph-kernel using Bose-Einstein entropy and the Jensen-Shannon divergence. This method is based on quantum statistics associated with bosonic population of the normalized Laplacian eigenstates.

By applying kernel PCA to the Jensen-Shannon kernel matrix, we embed sets of graphs into a low dimensional space. In order to evaluate the performance of thermal entropies, we use discriminant classifier analysis to assign the graphs to different groups. Experimental results reveal that the method improves the classification performance for graphs extracted from fMRI data. The kernel method combined Bose-Einstein entropy and the Jensen-Shannon divergence provides an effective and efficient method in fMRI network analysis. Further work maybe focus on investigating the confusion matrix to evaluate the performance of classification.

References

1. Kang, U., Tong, H.H., Sun, J.M.: Fast random walk graph kernel. In: Proceedings of SIAM International Conference on Data Mining (2012)
2. Kondor, R.I., Lafferty, J.D.: Diffusion kernels on graphs and other discrete structures. In: Proceedings of 19th International Conference on Machine Learning (2002)
3. Kashima, H., Tsuda, K., Inokuchi, A.: Marginalized kernels between labeled graphs. In: International Conference on Machine Learning (2003)
4. Borgwardt, K.M., Kriegel, H.P.: Shortest-path kernels on graphs. In: Proceedings of IEEE International Conference on Data Mining (2005)
5. Bai, L., Hancock, E.R.: Graph kernels from the Jensen-Shannon divergence. *J. Math. Imaging Vis.* **47**, 60–69 (2012)
6. Torsello, A., Rossi, L.: Supervised learning of graph structure. In: Pelillo, M., Hancock, E.R. (eds.) SIMBAD 2011. LNCS, vol. 7005, pp. 117–132. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-24471-1_9](https://doi.org/10.1007/978-3-642-24471-1_9)
7. Bianconi, G., Rahmede, C., Wu, Z.: Complex Quantum Network Geometries: Evolution and Phase Transitions (2015). [arxiv:1503.04739v2](https://arxiv.org/abs/1503.04739v2)
8. Bianconi, G., Barabasi, A.L.: Bose-Einstein condensation in complex network. *Phys. Rev. Lett.* **86**, 5632–5635 (2001)
9. Bianconi, G.: Supersymmetric multiplex networks described by coupled Bose and Fermi statistics. *Phys. Rev. E* **91**, 012810 (2015)
10. Schölkopf, B., Smola, A., Müller, K.-R.: Kernel principal component analysis. In: Gerstner, W., Germond, A., Hasler, M., Nicoud, J.-D. (eds.) ICANN 1997. LNCS, vol. 1327, pp. 583–588. Springer, Heidelberg (1997). doi:[10.1007/BFb0020217](https://doi.org/10.1007/BFb0020217)
11. Passerini, F., Severini, S.: The von Neumann entropy of networks. *Int. J. Agent Technol. Syst.* **1**, 58–67 (2008)
12. Chung, F.: Spectral graph theory. In: CBMS Regional Conference Series in Mathematics, no. 92 (1997)
13. Han, L., Hancock, E., Wilson, R.: Characterizing graphs using approximate von Neumann entropy. *Pattern Recogn. Lett.* **33**, 1958–1967 (2012)
14. Martins, A.F.T., Smith, N.A., Xing, E.P., Aguiar, P.M.Q., Figueiredo, M.A.T.: Nonextensive information theoretic kernels on measures. *J. Mach. Learn. Res.* **10**, 935–975 (2009)
15. fMRI Image Data Alzheimer’s Disease Neuroimaging Institute (ADNI). <http://adni.loni.usc.edu/>