

A Quantum Jensen-Shannon Graph Kernel Using the Continuous-Time Quantum Walk

Lu Bai¹, Edwin R. Hancock^{1,*}, Andrea Torsello², and Luca Rossi²

¹ Department of Computer Science, University of York, UK
{lu,erh}@cs.york.ac.uk

² Ca' Foscari University of Venice, Italy
{torsello,lurossi}@dsi.unive.it

Abstract. In this paper, we use the quantum Jensen-Shannon divergence as a means to establish the similarity between a pair of graphs and to develop a novel graph kernel. In quantum theory, the quantum Jensen-Shannon divergence is defined as a distance measure between quantum states. In order to compute the quantum Jensen-Shannon divergence between a pair of graphs, we first need to associate a density operator with each of them. Hence, we decide to simulate the evolution of a continuous-time quantum walk on each graph and we propose a way to associate a suitable quantum state with it. With the density operator of this quantum state to hand, the graph kernel is defined as a function of the quantum Jensen-Shannon divergence between the graph density operators. We evaluate the performance of our kernel on several standard graph datasets from bioinformatics. We use the Principle Component Analysis (PCA) on the kernel matrix to embed the graphs into a feature space for classification. The experimental results demonstrate the effectiveness of the proposed approach.

Keywords: Graph Kernels, Continuous-time Quantum Walk, Quantum Jensen-Shannon Divergence.

1 Introduction

There has been an increasing interest in learning graph structures using graph kernels. A graph kernel is usually defined in terms of a similarity measure between graphs [1]. To extend the large spectrum of kernel methods from the vectorial domain to the graph domain, Haussler [2] proposed a generic way, named R-convolution, to define a kernel between two graphs by decomposing them and measuring the pairwise similarities between the resulting substructures. For example, Kashima et al. [3] proposed a graph kernel where they compute the number of matchings random walks in a pair of graphs. Borgwardt et al. [4] proposed a shortest path kernel where they enumerate the shortest paths which possess the same length. Shervashidze et al. [5] developed a subtree kernel on limited size subtrees where they iteratively update the vertex labels in a pair

* Edwin R. Hancock is supported by a Royal Society Wolfson Research Merit Award.

of graphs, and then count the numbers of matching vertex labels between pairs of subtrees in the two graphs. The main feature of these graph kernels is that they usually share the exploit the topological information on the arrangement of vertices and edges in a graph. An attractive alternative kernel measure between a pair of graphs is based on measuring the mutual information using the classical Jensen-Shannon divergence. In probability theory, the classical Jensen-Shannon divergence is a similarity measure between probability distributions, and it is symmetric, always well defined and bounded [6]. In [7], we have used the classical Jensen-Shannon divergence to define a Jensen-Shannon kernel for graphs. Here, the Jensen-Shannon kernel between a pair of graphs is defined using the classical Jensen-Shannon divergence between some suitably defined probability distributions over the vertices of the graphs. Since the entropy associated with a probability distribution of a graph can be directly computed without the need of decomposing the graph, the Jensen-Shannon kernel, unlike the aforementioned graph kernels, avoids the computational burdensome of comparing the similarities between all the pairs of substructures of the graphs.

To develop the work in [7] further, we aim to extend the classical Jensen-Shannon divergence measure of graphs into the context of quantum theory by using the quantum Jensen-Shannon divergence [6,8], and then use this as a means of defining a novel graph kernel. Unlike the classical divergence, which is defined as a similarity measure between probability distributions, the quantum Jensen-Shannon divergence is introduced in quantum theory as a distance measure between quantum states, where a quantum state is described by its density operator [8]. In order to compute the quantum Jensen-Shannon divergence between a pair of graphs, we first need to associate a quantum state with each graph. To this end, we propose to define a mixed quantum state which is based on the evolution of a continuous-time quantum walk on each graph.

Recently, the continuous-time quantum walk has been introduced as the natural quantum analogue of the classical random walk by Farhi and Gutmann in [9]. Similarly to the classical random walk on a graph, its state space is the set of vertices of the graph. However, unlike the classical random walk, whose state vector is real-valued and whose evolution is governed by a double stochastic matrix, the state vector of the continuous-time quantum walk is complex-valued and its evolution is governed by a time-varying unitary matrix. The continuous-time quantum walk possesses a number of interesting properties which are not exhibited by the classical random walk. For instance, the continuous-time quantum walk is reversible and non-ergodic, and does not have a limiting distribution. Hence, the continuous-time quantum walk can offer us an elegant way to design quantum algorithms on graphs which have some interesting properties. For further information on quantum walks, we refer the readers to the textbook [10].

In this paper we are interested in developing a quantum kernel for graphs using the quantum Jensen-Shannon divergence and the continuous-time quantum walk. Given a graph G , we start by evolving a continuous-time quantum walk on G . The quantum walk evolution can then be described by an ensemble of pure states each describing the state of the quantum walker at time t .

As a consequence, we can associate with G the resulting mixed state and its density operator. With the density operators of a pair of graphs to hand, the proposed graph kernel is finally defined as the quantum Jensen-Shannon divergence between their density operators. We evaluate the performance of our kernel on several standard graph datasets from bioinformatics. We use the kernel Principle Component Analysis (kPCA) on the kernel matrix to embed graphs into a feature space where we perform the classification. The experimental results demonstrate the effectiveness of the proposed framework.

2 Quantum Mechanical Background

In this section, we describe the quantum mechanical formalisms that will be used in this work. We commence by reviewing the fundamental concept of continuous-time quantum walk on a graph. We then describe how to associate a density operator with a given graph through a continuous-time quantum walk. With the density operator to hand, we finally show how to compute the von Neumann entropy of the graph.

2.1 The Continuous-Time Quantum Walk

The continuous-time quantum walk is the natural quantum analogue of the classical random walk [11,9]. Similarly to the classical random walk, the state space of the continuous-time quantum walk defined on a graph $G(V, E)$ is the set of the vertices V of $G(V, E)$. However, the evolution of the quantum walk is governed by an unitary matrix rather than a stochastic matrix.

In [9], the basis state corresponding to the walk being at vertex $u \in V$ in $G(V, E)$ is denoted, by Dirac notation, as $|u\rangle$. Here $|u\rangle$ are orthonormal vectors in a n -dimensional complex-valued Hilbert space \mathcal{H} . A general state of the walk is then a complex linear combination of the basis states, and the state of the walk at time t is defined as

$$|\psi_t\rangle = \sum_{u \in V} \alpha_u(t) |u\rangle \quad (1)$$

where the amplitude $\alpha_u(t) \in \mathbb{C}$ and $|\psi_t\rangle \in \mathbb{C}^{|V|}$ are both complex. The probability of the walk being at a particular vertex of the graph $G(V, E)$ is given by the square of the norm of the amplitude of the relative state. More formally, let X^t be a random variable giving the location of the walk at time t . Then the probability of the walk being at vertex $u \in V$ at time t is given by

$$\Pr(X^t = u) = \alpha_u(t) \alpha_u^*(t) \quad (2)$$

where $\alpha_u^*(t)$ is the complex conjugate of $\alpha_u(t)$. Moreover $\sum_{u \in V} \alpha_u(t) \alpha_u^*(t) = 1$ and $\alpha_u(t) \alpha_u^*(t) \in [0, 1]$, for all $u \in V$, $t \in \mathbb{R}^+$.

Let A is the adjacency matrix of $G(V, E)$, then the vertex degree matrix of $G(V, E)$ is a diagonal matrix D whose elements are given by $D(u, u) = d_u =$

$\sum_{v \in V} A(u, v)$. From the degree matrix D and the adjacency matrix A we can construct the Laplacian matrix $L = D - A$. Then the evolution of the walk at time t is given by Schrödinger equation, where we take the Hamiltonian of the system to be the graph Laplacian matrix L , as

$$\frac{\partial}{\partial t} |\psi_t\rangle = -iL |\psi_t\rangle \quad (3)$$

Given a initial state $|\psi_0\rangle$, Eq.(3) can be solved to calculate the state of the walk at a time, t , as

$$|\psi_t\rangle = e^{-iLt} |\psi_0\rangle \quad (4)$$

where $U_t = e^{-iLt}$ is the unitary matrix. To implement the simulation of the quantum walk evolution, we re-write Eq.(4) in terms of the spectral decomposition $L = \Phi^\top \Lambda \Phi$ of the Laplacian matrix L , where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{|V|})$ is a diagonal matrix with the ordered eigenvalues as elements ($0 = \lambda_1 < \lambda_2 < \dots < \lambda_{|V|}$) and $\Phi = (\phi_1 | \phi_2 | \dots | \phi_{|V|})$ is a matrix with the corresponding ordered orthonormal eigenvectors as columns. Hence, Eq.(4) can also be re-written as

$$|\psi_t\rangle = \Phi^\top e^{-i\Lambda t} \Phi |\psi_0\rangle \quad (5)$$

In this work, we define the initial state $|\psi_0\rangle$ of $G(V, E)$ as

$$|\psi_0\rangle = \sum_{u \in V} \frac{d_u}{\sqrt{\sum_{u \in V} d_u d_u^*}} |u\rangle \quad (6)$$

where $\frac{d_u}{\sqrt{\sum_{u \in V} d_u d_u^*}}$ is the initial amplitude on vertex u . In other words, the initial probability distribution induced by $|\psi_0\rangle$ is equal to the steady state of random walk on $G(V, E)$.

2.2 A Density Operator for Graphs

In quantum mechanics, the density operator ρ is a matrix that describes an ensemble of pure states, i.e. a mixed state. A pure state is a quantum state that can be described by a single state vector $|\psi\rangle$ and its density operator ρ can be written as $|\psi\rangle \langle \psi|$. On the other hand, we can think of a mixed quantum state as an ensemble of pure states described by a density operator ρ . More formally, consider a quantum system that can be found in a number of pure states $\{(|\psi_n\rangle, n) | (n = 1, 2, \dots, N)\}$ each with a probability p_n . The density operator (i.e. density matrix) of the system is then defined as

$$\rho = \sum_n p_n |\psi_n\rangle \langle \psi_n| \quad (7)$$

Now we proceed to show how to associate a density operator with a graph through a continuous-time quantum walk. Consider a graph $G(V, E)$ and continuous-time quantum walk $|\psi_t\rangle$ on G . We can see $|\psi_t\rangle$ as a pure state describing the state of the walk at time t . If we associate with each of these pure states

a probability $p_t = 1/T$, we obtain a mixed state $\{(|\psi_t\rangle, t) | (t = 1, 2, \dots, T)\}$ describing the quantum walk evolution on $G(V, E)$. Hence, the density operator ρ_G of $G(V, E)$ of this mixed state is defined as

$$\rho_G = \sum_{t=1}^T p_t |\psi_t\rangle \langle \psi_t| \tag{8}$$

2.3 The von Neumann Entropy of A Graph

In quantum mechanics, the von Neumann entropy is an extension of the classical Shannon entropy, and it is defined on a density matrix (i.e. density operator) ρ as $H_N(\rho) = -Tr(\rho \log \rho)$ [12]. Note that if ρ is the density matrix associated with a pure state, then the von Neumann entropy of ρ vanishes. Consider a graph $G(V, E)$ and its density operator ρ_G defined as in Eq.(8), its von Neumann entropy is

$$H_N(\rho_G) = -Tr(\rho_G \log \rho_G) \tag{9}$$

Note that computing $H_N(\rho_G)$ is a rather complex operation, since it involves taking the logarithm of the density operator matrix ρ_G . In practice, it is more convenient to firstly determine the spectral decomposition of $\rho_G = \Phi_{\rho;G}^\top \Lambda_{\rho;G} \Phi_{\rho;G}$, and then Eq.(9) can be re-written as

$$H_N(\rho_G) = - \sum_j^{|\mathcal{V}|} \lambda_{j;\rho;G} \log \lambda_{j;\rho;G} \tag{10}$$

where $\lambda_{1;\rho;G}, \dots, \lambda_{j;\rho;G}, \dots, \lambda_{|\mathcal{V}|\rho;G}$ is the ordered eigenvalues of ρ_G .

3 A Quantum Jensen-Shannon Graph Kernel

In this section, we use the quantum Jensen-Shannon divergence to develop a novel kernel for graphs. We commence by reviewing the concept of quantum Jensen-Shannon divergence, which can be seen as an extension of the classical Jensen-Shannon divergence to the quantum realm. The quantum Jensen-Shannon divergence between a pair of density operators is defined as a function of the von Neumann entropy associated with the operators. With the density operators to hand, we show that the quantum kernel between the pair of graphs can be computed as a function of the quantum Jensen-Shannon divergence between their density operators.

3.1 Classical and Quantum Jensen-Shannon Divergence

The classical Jensen-Shannon divergence is a non-extensive mutual information measure defined between probability distributions over structured data, and it is related to the Shannon entropy [13]. Assume there are two (discrete) probability

distributions $P = (p_1, p_2, \dots, p_X)$ and $Q = (q_1, q_2, \dots, q_Y)$, the classical Jensen-Shannon divergence between P and Q is defined as

$$D_{JS}(P, Q) = H_S\left(\frac{P+Q}{2}\right) - \frac{1}{2}H_S(P) - \frac{1}{2}H_S(Q) \quad (11)$$

where $H_S(P) = \sum_x p_x \log p_x$ is the Shannon entropy. The classical Jensen-Shannon divergence is always well defined, symmetric, negative definite and bounded, i.e., $0 \leq D_{JS} \leq 1$. By replacing the Shannon entropy H_S of a probability distribution with the von Neumann entropy of a density operator as defined in Eq.(10), in [6] the classical Jensen-Shannon divergence has been extended to the quantum realm to define the quantum Jensen-Shannon divergence between quantum states

$$D_{QJS}(\rho, \sigma) = H_N\left(\frac{\rho+\sigma}{2}\right) - \frac{1}{2}H_N(\rho) - \frac{1}{2}H_N(\sigma) \quad (12)$$

where ρ and σ are two density operators describing the corresponding quantum states, and $H_N(\cdot)$ is the von Neumann entropy of a density operator. The quantum Jensen-Shannon divergence is always well defined, symmetric, positive definite and bounded, i.e., $0 \leq D_{QJS} \leq 1$ [6].

3.2 A Quantum Jensen-Shannon Kernel for Graphs

We propose a novel quantum kernel for graphs by using the quantum Jensen-Shannon divergence between the density operators associated with the graphs. To this end, we let a continuous-time quantum walk evolve on a pair of graphs $G_a(V_a, E_a)$ and $G_b(V_b, E_b)$ with time t ($t = 1, \dots, T$). Then the density operators $\rho_{G;a}$ and $\sigma_{G;b}$ of $G_a(V_a, E_a)$ and $G_b(V_b, E_b)$ can be computed from their mixed states using Eq.(8). With the density operators $\rho_{G;a}$ and $\sigma_{G;b}$, and the quantum Jensen-Shannon divergence $D_{QJS}(\rho_{G;a}, \sigma_{G;b})$ between $\rho_{G;a}$ and $\sigma_{G;b}$ computed using Eq.(12) to hand, the quantum Jensen-Shannon divergence $D_{QJS}(G_a, G_b)$ between the pair of graphs $G_a(V_a, E_a)$ and $G_b(V_b, E_b)$ is

$$D_{QJS}(G_a, G_b) = H_N\left(\frac{\rho_{G;a} + \sigma_{G;b}}{2}\right) - \frac{1}{2}H_N(\rho_{G;a}) - \frac{1}{2}H_N(\sigma_{G;a}) \quad (13)$$

Then, we define the quantum Jensen-Shannon kernel $k_{QJS}(G_a, G_b)$ between $G_a(V_a, E_a)$ and $G_b(V_b, E_b)$ as

$$\begin{aligned} k_{QJS}(G_a, G_b) &= \exp(\lambda D_{QJS}(G_a, G_b)) \\ &= \exp\left(\lambda H_N\left(\frac{\rho_{G;a} + \sigma_{G;b}}{2}\right) - \frac{1}{2}\lambda H_N(\rho_{G;a}) - \frac{1}{2}\lambda H_N(\sigma_{G;a})\right) \end{aligned} \quad (14)$$

where λ is a decay factor which satisfies $0 < \lambda < 1$, and $H_N(\cdot)$ is the von Neumann entropy of the density operator associated to the graph. Here λ is used to ensure that the large values do not tend to dominant the kernel value. In particular, in this work we use $\lambda = 0.1$.

Lemma. *The quantum Jensen-Shannon graph kernel is positive definite **pd**.*

Proof. This follows the definitions in [6,8,14]. In [14], a diffusion kernel $k_s = \exp(\lambda s(G_p, G_q))$ associated with any symmetric similarity measure $s(G_p, G_q)$ has been proven to be **pd**. Since the quantum Jensen-Shannon divergence between a pair of density operators is symmetric [6,8], then the proposed quantum Jensen-Shannon graph kernel is **pd**.

Note that, a positive definite graph kernel is often called a *valid kernel*. Clearly, the positive definiteness constraint restricts the class of valid graph kernels which can be defined by different similarity measures on graphs. However, the property of positive definiteness is crucial for the definition of kernel machines and turns out to be sufficiently strong to implicate a considerable number of theoretical properties associated with graph kernels [15].

3.3 Algorithmic Complexity

The computational complexity of the proposed quantum Jensen-Shannon kernel of graphs depends on several factors. These factors include 1) the computation of the initial state of the continuous-time quantum walk, 2) the simulation of the continuous-time quantum walk evolution at each time step $t = 1, 2, \dots, T$, 3) the construction of the density operator associated with the graph, and 4) the computation of the von Neumann entropy of the density operator. It is clear that these operations are all dependent on the graph size, and that the complexity is highly influenced by the termination time T of the continuous-time quantum walk evolution. For example, we can easily show that the cost of computing the von Neumann entropy of the density operator is $O(|V|^2)$, where $|V|$ is the number of vertices of the graph. However, it's easy to see that the overall complexity is dominated by that of computing the eigendecomposition of the graph Laplacian, which is cubic in the number of vertices of the graph, i.e. $O(|V|^3)$. This algorithmic complexity analysis reveals that our quantum Jensen-Shannon kernel between graphs can be computed in polynomial time.

4 Experimental Evaluations

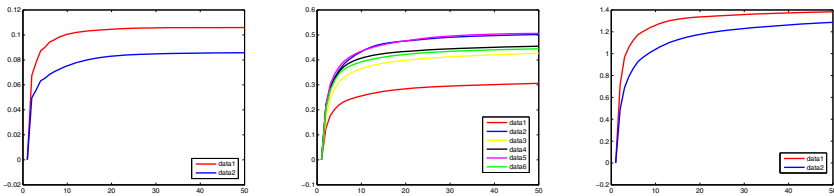
In this section, we demonstrate the performance of our quantum Jensen-Shannon kernel, and then compare it to several state of the art graph based learning methods on three standard graph datasets abstracted from bioinformatics. These datasets include: MUTAG, ENZYMES and PPIs. The MUTAG dataset contains graphs representing 188 chemical compounds to predict mutagenicity. The ENZYMES dataset contains graphs representing protein tertiary structures consisting of 600 enzymes from the BRENDA enzyme database. The PPIs dataset consists of protein-protein interaction networks (PPIs). Here we select Proteobacteria40 PPIs and Acidobacteria46 PPIs as the testing graphs. Details about the datasets are shown in Table.1 and [5,16].

Table 1. Information of the Graph based Datasets

| Datasets | MUTAG | ENZYMES | PPIs |
|--------------------|-------|---------|--------|
| Maximum # vertices | 28 | 126 | 232 |
| Minimum # vertices | 10 | 2 | 3 |
| Average # vertices | 17.93 | 32.63 | 109.60 |
| # testing graphs | 188 | 600 | 86 |

4.1 Von Neumann Entropy Evaluation

We commence by exploring the relationship between the von Neumann entropy of a graph and its corresponding increasing time t . In our experiments, we utilize the testing graphs in the MUTAG, ENZYMES and PPIs datasets. For each graph, we let the continuous-time quantum walk evolve until a maximum time t , where we vary t from 1 to 50. For each time t we compute the density operator associated with the graph using Eq.(8). Then the von Neumann entropy of the graph at each time t can be computed from its corresponding density operator. The experimental results are shown in Fig.1. The left, middle and right subfigures of Fig.1 show the results of the evaluation on the MUTAG, Enzymes and PPIs datasets separately. The x-axis shows the time t which is from 1 to 50, and the y-axis shows the mean value of the von Neumann entropies of graphs belonging to the same class at each time t . Here the different lines represent the entropies of different classes of graphs separately. This evaluation suggests that the von Neumann entropies of different classes of graphs can be divided well, and tend to be stable with increasing time t .

**Fig. 1.** Evaluations on the von Neumann Entropy for increasing time t

4.2 Experiments on Standard Graph Datasets from Bioinformatics

Experimental Setup. We evaluate the performance of our quantum Jensen-Shannon kernel on the graph datasets abstracted from bioinformatics databases, and then compare it with several alternative state of the art graph learning methods. These methods include 1) the Weisfeiler-Lehman subtree kernel [5], 2) the shortest path graph kernel [4], 3) the Shannon entropy associated with the information functionals FV and FP [17], and 4) the Ihara zeta function on

graphs [18]. Based on the evaluation in Section 4.2, here we set time t from 1 to 50. For the kernel methods, we compute the kernel matrix of each graph kernel on each dataset, we then apply the PCA [19] on the kernel matrix to embed the graphs into principle component space as feature vectors. For other methods, we compute the characteristics values of graphs on each dataset. We perform 10-fold cross-validation using the Support Vector Machine (SVM) Classification associated with the Sequential Minimal Optimization (SMO) [20] on the graph feature vectors or characteristics values to evaluate the performance of our kernel and that of the alternative methods. We use nine samples for training and one for testing. All the SMO-SVMs and their parameters were performed and optimized on a Weka workbench [21]. We report the average classification accuracies of each method in Table 2.

Results. a) On the MUTAG dataset, the accuracies of all the methods are similar, our kernel overcomes or is competitive to the alternatives. b) On the ENZYMES dataset, the accuracy of our kernel is obviously lower than that of the Weisfeiler-Lehman subtree kernel, but is competitive to that of the shortest path graph kernel and outperforms that of others. c) On the PPIs dataset, the accuracy of our kernel is obviously higher than that of the alternatives. As a whole, our kernel outperforms or is competitive to the alternatives, only the Weisfeiler-Lehman subtree kernel and the shortest path graph kernel are competitive to our kernel on the ENZYMES dataset.

Table 2. Accuracies Comparisons on Graph Datasets abstracted from Bioninformatics

| Datasets | MUTAG | ENZYMES | PPIs |
|--------------------------------------|--------|---------|--------|
| quantum Jensen-Shannon kernel | 84.04% | 32.16% | 76.20% |
| Weisfeiler-Lehman subtree kernel | 84.57% | 38.50% | 76.16% |
| shortest path graph kernel | 85.29% | 31.16% | 78.45% |
| Shannon entropy with FV | 84.57% | 24.17% | 70.93% |
| Shannon entropy with FP | 84.57% | 24.17% | 70.93% |
| Ihara zeta function on graphs | 80.85% | 32.00% | 70.93% |

5 Conclusion

In this paper, we developed a novel graph kernel by using the quantum Jensen-Shannon divergence and the continuous-time quantum walk on graphs. Given a graph, we evolved a continuous-time quantum walk on its structure and we showed how to associate a mixed quantum state to the graph and how to compute the von Neumann entropy of the corresponding density operator. With the von Neumann entropies to hand, the quantum Jensen-Shannon kernel between a pair of graphs was defined as a function of the quantum Jensen-Shannon divergence between the corresponding density operators. Finally, we used the Principle Component Analysis (PCA) on the kernel matrix to embed the graphs into a feature space where we performed the classification task. Experiments on several standard datasets demonstrate the effectiveness of the proposed graph

kernel. Our future work is to extend the quantum graph kernel to a quantum hypergraph kernel. In [22], we have developed a hypergraph kernel by using the classical Jensen-Shannon divergence. In [23], we have explored the use of discrete-time quantum walks on a directed line graph which can be generated by transforming a hypergraph. It would thus be interesting to extend these works by using the quantum Jensen-Shannon divergence to compare the quantum walks of hypergraphs based on their directed line graphs.

References

1. Schölkopf, B., Smola, A.: *Learning with Kernels*. MIT Press (2002)
2. Haussler, D.: Convolution kernels on discrete structures. Technical Report UCS-CRL-99-10, Santa Cruz, CA, USA (1999)
3. Kashima, H., Tsuda, K., Inokuchi, A.: Marginalized kernels between labeled graphs. In: *Proceedings of International Conference on Machine Learning*, pp. 321–328 (2003)
4. Borgwardt, K.M., Kriegel, H.P.: Shortest-path kernels on graphs. In: *Proceedings of the IEEE International Conference on Data Mining*, pp. 74–81 (2005)
5. Shervashidze, N., Schweitzer, P., van Leeuwen, E.J., Mehlhorn, K., Borgwardt, K.M.: Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research* 1, 1–48 (2010)
6. Lamberti, P., Majtey, A., Borrás, A., Casas, M., Plastino, A.: Metric character of the quantum jensen-shannon divergence. *Physical Review A* 77, 052311 (2008)
7. Bai, L., Hancock, E.R.: Graph kernels from the jensen-shannon divergence. *Journal of Mathematical Imaging and Vision* (to appear)
8. Majtey, A., Lamberti, P., Prato, D.: Jensen-shannon divergence as a measure of distinguishability between mixed quantum states. *Physical Review A* 72, 052310 (2005)
9. Farhi, E., Gutmann, S.: Quantum computation and decision trees. *Physical Review A* 58, 915 (1998)
10. Dirac, P.: *The Principles of Quantum Mechanics*, 4th edn. Oxford Science Publications (1958)
11. Kempe, J.: Quantum random walks: an introductory overview. *Contemporary Physics* 44, 307–327 (2003)
12. Nielsen, M., Chuang, I.: *Quantum computation and quantum information*. Cambridge university press (2010)
13. Martins, A.F., Smith, N.A., Xing, E.P., Aguiar, P.M., Figueiredo, M.A.: Nonextensive information theoretic kernels on measures. *Journal of Machine Learning Research* 10, 935–975 (2009)
14. Konder, R., Lafferty, J.: Diffusion kernels on graphs and other discrete input spaces. In: *Proceedings of International Conference on Machine Learning*, pp. 315–322 (2002)
15. Neuhaus, M., Bunke, H.: *Bridging the gap between graph edit distance and kernel machines*. World Scientific (2007)
16. Escolano, F., Hancock, E.R., Lozano, M.A.: Heat diffusion: Thermodynamic depth complexity of networks. *Physical Review E* 85, 036206 (2012)
17. Dehmer, M.: Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation* 201, 82–94 (2008)

18. Ren, P., Wilson, R.C., Hancock, E.R.: Graph characterization via ihara coefficients. *IEEE Transactions on Neural Networks* 22, 233–245 (2011)
19. Schölkopf, B., Smola, A.J., Müller, K.R.: Kernel principal component analysis. In: *Proceedings of International Conference on Artificial Neural Networks*, pp. 583–588 (1997)
20. Platt, J.C.: Fast training of support vector machines using sequential minimal optimization. In: Schölkopf, B., Burges, C.J.C., Smola, A.J. (eds.) *Advances in Kernel Methods*, pp. 185–208 (1999)
21. Witten, I.H., Frank, E., Hall, M.A.: *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann (2011)
22. Bai, L., Hancock, E.R., Ren, P.: A jensen-shannon kernel for hypergraphs. In: Gimel'farb, G., Hancock, E., Imiya, A., Kuijper, A., Kudo, M., Omachi, S., Windeatt, T., Yamada, K. (eds.) *SSPR & SPR 2012. LNCS*, vol. 7626, pp. 181–189. Springer, Heidelberg (2012)
23. Ren, P., Aleksic, T., Emms, D., Wilson, R., Hancock, E.: Quantum walks, ihara zeta functions and cospectrality in regular graphs. *Quantum Information Processing* 10, 405–417 (2011)