

A Continuous-Time Quantum Walk Kernel for Unattributed Graphs

Luca Rossi¹, Andrea Torsello¹, and Edwin R. Hancock²

¹ Department of Environmental Science, Informatics and Statistics,
Ca' Foscari University of Venice, Italy
{lurossi,torsello}@dsi.unive.it

² Department of Computer Science, University of York, YO10 5GH, UK
edwin.hancock@york.ac.uk

Abstract. Kernel methods provide a way to apply a wide range of learning techniques to complex and structured data by shifting the representational problem from one of finding an embedding of the data to that of defining a positive semidefinite kernel. In this paper, we propose a novel kernel on unattributed graphs where the structure is characterized through the evolution of a continuous-time quantum walk. More precisely, given a pair of graphs, we create a derived structure whose degree of symmetry is maximum when the original graphs are isomorphic. With this new graph to hand, we compute the density operators of the quantum systems representing the evolutions of two suitably defined quantum walks. Finally, we define the kernel between the two original graphs as the quantum Jensen-Shannon divergence between these two density operators. The experimental evaluation shows the effectiveness of the proposed approach.

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

1 Introduction

Graph-based representations have become increasingly popular due to their ability to characterize in a natural way a large number of systems which are best described in terms of their structure. Concrete examples include the use of graphs to represent shapes [1], metabolic networks [2], protein structure [3], and road maps [4]. Unfortunately, our ability to analyse this wealth of data is severely limited by the restrictions posed by standard pattern recognition techniques, which usually require the graphs to be first embedded into a vectorial space, a procedure which is far from being trivial. The reason for this is that there is no canonical ordering for the nodes in a graph and a correspondence order must be established before analysis can commence. Moreover, even if a correspondence order can be established, graphs do not necessarily map to vectors of fixed length, as the number of nodes and edges can vary.

Kernel methods [5], whose best known example is furnished by support vector machines (SVMs) [6], provide a neat way to shift the problem from that of

finding an embedding to that of defining a positive semidefinite kernel, via the well-known kernel trick. In fact, once we define a positive semidefinite kernel $k : X \times X \rightarrow \mathbb{R}$ on a set X , then we know that there exists a map $\phi : X \rightarrow H$ into a Hilbert space H , such that $k(x, y) = \phi(x)^\top \phi(y)$ for all $x, y \in X$. Thus, any algorithm that can be formulated in terms of scalar products of the $\phi(x)$ s can be applied to a set of data on which we have defined our kernel. As a consequence, we are now faced with the problem of defining a positive semidefinite kernel on graphs rather than computing an embedding. However, due to the rich expressiveness of graphs, also this task has proven to be difficult.

Many different graph kernels have been proposed in the literature [7–9]. Graph kernels are generally instances of the family of R-convolution kernels introduced by Haussler [10]. The fundamental idea is that of defining a kernel between two discrete objects by decomposing them and comparing some simpler substructures. For example, Gärtner et al. [7] propose to count the number of common random walks between two graphs, while Borgwardt and Kriegel [8] measure the similarity based on the shortest paths in the graphs. Shervashidze et al. [9], on the other hand, count the number of graphlets, i.e. subgraphs with k nodes. Note that these kernels can be defined both on unattributed and attributed graphs, although we will restrict our analysis to the simpler case of unattributed graphs, while the more general case will be the focus of future work. Another interesting approach is that of Bai and Hancock [11], where the authors investigate the possibility of defining a graph kernel based on the Jensen-Shannon kernel.

In this paper, we introduce a novel kernel on unattributed graphs where we probe the graph structure through the evolution of a continuous-time quantum walk [12, 13]. In particular, we are taking advantage of the fact that the interference effects introduced by the quantum walk seem to be enhanced by the presence of symmetrical motifs in the graph [14, 15]. To this end, we define a walk onto a new structure that is maximally symmetric when the original graphs are isomorphic. Finally, to define the kernel we make use of the quantum Jensen-Shannon divergence, a measure which has recently been introduced as a means to compute the distance between quantum states [16, 17].

The remainder of this paper is organized as follows: Section 2 provides an essential introduction to the basic terminology required to understand the proposed quantum mechanical framework. With these notions to hand, we introduce our graph kernel in Section 3. Section 4 illustrates the experimental results, while the conclusions are presented in Section 5.

2 Quantum Mechanical Background

Quantum walks are the quantum analogue of classical random walks [13]. In this paper we consider only continuous-time quantum walks, as first introduced by Farhi and Gutmann in [12]. Given a graph $G = (V, E)$, the state space of the continuous-time quantum walk defined on G is the set of the vertices V of the graph. Unlike the classical case, where the evolution of the walk is governed by a stochastic matrix (i.e. a matrix whose columns sum to unity), in

the quantum case the dynamics of the walker is governed by a complex unitary matrix i.e., a matrix that multiplied by its conjugate transpose yields the identity matrix. Hence, the evolution of the quantum walk is reversible, which implies that quantum walks are non-ergodic and do not possess a limiting distribution. Using Dirac notation, we denote the basis state corresponding to the walk being at vertex $u \in V$ as $|u\rangle$. A general state of the walk is a complex linear combination of the basis states, such that 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.

At each point in time the probability of the walker being at a particular vertex of the graph 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 walker at time t . Then the probability of the walker being at the vertex u 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 $\alpha_u(t) \alpha_u^*(t) \in [0, 1]$, for all $u \in V$, $t \in \mathbb{R}^+$, and in a closed system $\sum_{u \in V} \alpha_u(t) \alpha_u^*(t) = 1$.

Recall that the adjacency matrix of the graph G has elements

$$A_{uv} = \begin{cases} 1 & \text{if } (u, v) \in E \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

The evolution of the walk is governed by Schrödinger equation, where we take the Hamiltonian of the system to be the graph adjacency matrix, which yields

$$\frac{d}{dt} |\psi_t\rangle = -iA |\psi_t\rangle \quad (4)$$

Given an initial state $|\psi_0\rangle$, we can solve Equation 4 to determine the state vector at time t

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

where $A = \Phi \Lambda \Phi^\top$ is the spectral decomposition of the adjacency matrix.

2.1 Quantum Jensen-Shannon Divergence

A pure state is defined as a state that can be described by a ket vector $|\psi_i\rangle$. Consider a quantum system that can be in a number of states $|\psi_i\rangle$ each with probability p_i . The system is said to be in the ensemble of pure states $\{|\psi_i\rangle, p_i\}$. The density operator (or density matrix) of such a system is defined as

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i| \quad (6)$$

The Von Neumann entropy [18] of a density operator ρ is

$$H_N(\rho) = -Tr(\rho \log \rho) = - \sum_j \lambda_j \log \lambda_j, \tag{7}$$

where the λ_j s are the eigenvalues of ρ . With the Von Neumann entropy to hand, we can define the quantum Jensen-Shannon divergence between two density operators ρ and σ as

$$D_{JS}(\rho, \sigma) = H_N\left(\frac{\rho + \sigma}{2}\right) - \frac{1}{2}H_N(\rho) - \frac{1}{2}H_N(\sigma) \tag{8}$$

This quantity is always well defined, symmetric and negative definite [19]. It can also be shown that $D_{JS}(\rho, \sigma)$ is bounded, i.e., $0 \leq D_{JS}(\rho, \sigma) \leq 1$. Let $\rho = \sum_i p_i \rho_i$ be a mixture of quantum states ρ_i , with $p_i \in \mathbb{R}^+$ such that $\sum_i p_i = 1$, then we can prove that

$$H_N\left(\sum_i p_i \rho_i\right) \leq H_S(p_i) + \sum_i p_i H_N(\rho_i) \tag{9}$$

where the equality is attained if and only if the states ρ_i have support on orthogonal subspaces. By setting $p_1 = p_2 = 0.5$, we see that

$$D_{JS}(\rho, \sigma) = H_N\left(\frac{\rho + \sigma}{2}\right) - \frac{1}{2}H_N(\rho) - \frac{1}{2}H_N(\sigma) \leq 1 \tag{10}$$

Hence D_{JS} is always less or equal than 1, and the equality is attained only if ρ and σ have support on orthogonal subspaces.

3 QJSD Kernel

Given two graphs $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ we build a new graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = V_1 \cup V_2$, $\mathcal{E} = E_1 \cup E_2 \cup E_{12}$, and $(u, v) \in E_{12}$ only if $u \in V_1$ and $v \in V_2$. With this new structure to hand, we define two continuous-time quantum walks $|\psi_t^-\rangle = \sum_{u \in V} \psi_{0u}^- |u\rangle$ and $|\psi_t^+\rangle = \sum_{u \in V} \psi_{0u}^+ |u\rangle$ on \mathcal{G} with starting states

$$\psi_{0u}^- = \begin{cases} +\frac{d_u}{C} & \text{if } u \in G_1 \\ -\frac{d_u}{C} & \text{if } u \in G_2 \end{cases} \quad \psi_{0u}^+ = \begin{cases} +\frac{d_u}{C} & \text{if } u \in G_1 \\ +\frac{d_u}{C} & \text{if } u \in G_2 \end{cases} \tag{11}$$

where d_u is the degree of the node u and C is the normalisation constant such that the probabilities sum to one.

We let the two quantum walks evolve until a time T and we define the average density operators ρ_T and σ_T over this time as

$$\rho_T = \frac{1}{T} \int_0^T |\psi_t^-\rangle \langle \psi_t^-| dt \quad \sigma_T = \frac{1}{T} \int_0^T |\psi_t^+\rangle \langle \psi_t^+| dt \tag{12}$$

In other words, we defined two mixed systems with equal probability of being in any of the pure states defined by the quantum walks evolutions.

Then, given two unattributed graphs G_1 and G_2 , we define the quantum Jensen-Shannon kernel $k_T(G_1, G_2)$ between them as

$$k_T(G_1, G_2) = D_{JS}(\rho_T, \sigma_T) \quad (13)$$

where ρ_T and σ_T are the density operators defined as in Eq. 12. Note that this kernel is parametrised by the time T . As it is not clear how we should set this parameter, in this paper we propose to let $T \rightarrow \infty$. However, in Section 4 we will show that a proper choice of T can yield an increased average accuracy in an SVM classification task.

We now proceed to show some interesting properties of our kernel. First, however, we need to prove the following

Lemma 1. *If G_1 and G_2 are two isomorphic graphs, then ρ_T and σ_T have support on orthogonal subspaces.*

Proof. We need to prove that

$$(\rho_T)^\dagger \sigma_T = \frac{1}{T^2} \int_0^T \rho_{t_1} dt_1 \int_0^T \sigma_{t_2} dt_2 = \mathbf{0} \quad (14)$$

where $\mathbf{0}$ is the matrix of all zeros, $\rho_t = |\psi_t^-\rangle \langle \psi_t^-|$ and $\sigma_t = |\psi_t^+\rangle \langle \psi_t^+|$. Note that if $\rho_{t_1}^\dagger \sigma_{t_2} = \mathbf{0}$ for every t_1 and t_2 , then $(\rho_T)^\dagger \sigma_T = \mathbf{0}$. We now prove that if G_1 is isomorphic to G_2 then $\langle \psi_{t_1}^- | \psi_{t_2}^+ \rangle = 0$ for every t_1 and t_2 .

Let $U = e^{-iAt}$ be the unitary evolution operator of the quantum walk. If $t_1 = t_2 = t$, then $\langle \psi_0^- | (U^t)^\dagger U^t | \psi_0^+ \rangle = 0$ since $(U^t)^\dagger U^t$ is the identity matrix and the initial states are orthogonal by construction. On the other hand, if $t_1 \neq t_2$, we have $\langle \psi_0^- | U^{\Delta t} | \psi_0^+ \rangle = 0$ where $\Delta t = t_2 - t_1$. To conclude the proof we rewrite the previous equation as

$$\begin{aligned} \langle \psi_0^- | U^{\Delta t} | \psi_0^+ \rangle &= \sum_k \psi_{k0}^+ \sum_l \psi_{l0}^+ U_{lk}^{\Delta t} \\ &= \sum_{k_1} \psi_{k_1 0}^+ \sum_l \psi_{l0}^+ U_{lk_1}^{\Delta t} - \sum_{k_2} \psi_{k_2 0}^+ \sum_l \psi_{l0}^+ U_{lk_2}^{\Delta t} \\ &= \sum_l \psi_{l0}^+ \left(\sum_{k_1} \psi_{k_1 0}^+ U_{lk_1}^{\Delta t} - \sum_{k_2} \psi_{k_2 0}^+ U_{lk_2}^{\Delta t} \right) = 0 \end{aligned} \quad (15)$$

where the indices l, k, k_1 and k_2 run over the nodes of \mathcal{G} , G_1 and G_2 respectively. To see that Eq. 15 holds, note that U is a symmetric matrix and it is invariant to graph symmetries, i.e., if u and v are symmetric then $U_{uu}^{\Delta t} = U_{vv}^{\Delta t}$, and that if G_1 and G_2 are isomorphic, then $k_1 = k_2$ and $\psi_{1:k_1 0}^+ = \psi_{k_1+1:k_2 0}^+$.

Corollary 1. *Given a pair of graphs G_1 and G_2 , the kernel satisfies the following properties: 1) $0 \leq k_T(G_1, G_2) \leq 1$ and 2) if G_1 and G_2 are isomorphic, then $k_T(G_1, G_2) = 1$.*

Proof. The first property is trivially proved by noting that, according to Eq. 13, the kernel between G_1 and G_2 is defined as the quantum Jensen-Shannon divergence between two density operators, and then recalling that the value of quantum Jensen-Shannon divergence is bounded to lie between 0 and 1.

The second property follows again from Eq. 13 and Theorem 1. It is sufficient to note that the quantum Jensen-Shannon divergence reaches its maximum value if and only if the density operators have support on orthogonal spaces.

Unfortunately we cannot prove that our kernel is positive semidefinite, but both empirical evidence and the fact that the Jensen-Shannon Divergence is negative semidefinite on pure quantum states [19] while our graph similarity is maximal on orthogonal states suggest that it might be.

3.1 Kernel Computation

We conclude this section with a few remarks on the computational complexity of our kernel. Recall that $|\psi_t\rangle = e^{-iAt} |\psi_0\rangle$, then we rewrite Eq. 12 as

$$\rho_T = \frac{1}{T} \int_0^T e^{-iAt} |\psi_0\rangle \langle \psi_0| e^{iAt} dt \quad (16)$$

Since $e^{-iAt} = \Phi e^{-i\Lambda t} \Phi^\top$, we can rewrite the previous equation in terms of the spectral decomposition of the adjacency matrix,

$$\rho_T = \frac{1}{T} \int_0^T \Phi e^{-i\Lambda t} \Phi^\top |\psi_0\rangle \langle \psi_0| \Phi e^{i\Lambda t} \Phi^\top dt \quad (17)$$

The (r, c) element of ρ_T can be computed as

$$\rho_T(r, c) = \frac{1}{T} \int_0^T \left(\sum_k \sum_l \phi_{rk} e^{-i\lambda_k t} \phi_{lk} \psi_{0l}^- \right) \left(\sum_m \sum_n \psi_{0m}^\dagger \phi_{mn} e^{i\lambda_n t} \phi_{cn} \right) dt \quad (18)$$

Let $\bar{\psi}_k = \sum_l \phi_{lk} \psi_{0l}$ and $\bar{\psi}_n = \sum_m \phi_{mn} \psi_{0m}^\dagger$, then

$$\rho_T(r, c) = \frac{1}{T} \int_0^T \left(\sum_k \phi_{rk} e^{-i\lambda_k t} \bar{\psi}_k \sum_n \phi_{cn} e^{i\lambda_n t} \bar{\psi}_n \right) dt \quad (19)$$

which can be finally rewritten as

$$\rho_T(r, c) = \sum_k \sum_n \phi_{rk} \phi_{cn} \bar{\psi}_k \bar{\psi}_n \frac{1}{T} \int_0^T e^{i(\lambda_n - \lambda_k)t} dt \quad (20)$$

If we let $T \rightarrow \infty$, Eq. 20 further simplifies to

$$\rho_T(r, c) = \sum_{\lambda_k \in \tilde{\Lambda}} \sum_m \sum_n \phi(\lambda_k)_{r,m} \phi(\lambda_k)_{c,n} \bar{\psi}_m \bar{\psi}_n \quad (21)$$

where $\tilde{\Lambda}$ is the set of unique eigenvalues of A and $\phi(\lambda_k)$ is the matrix whose columns are the eigenvectors associated with λ_k . As a consequence, we see that the complexity of computing the QJSD kernel is upper bounded by that of computing the eigendecomposition of \mathcal{G} , i.e. $O(|\mathcal{V}|^3)$.

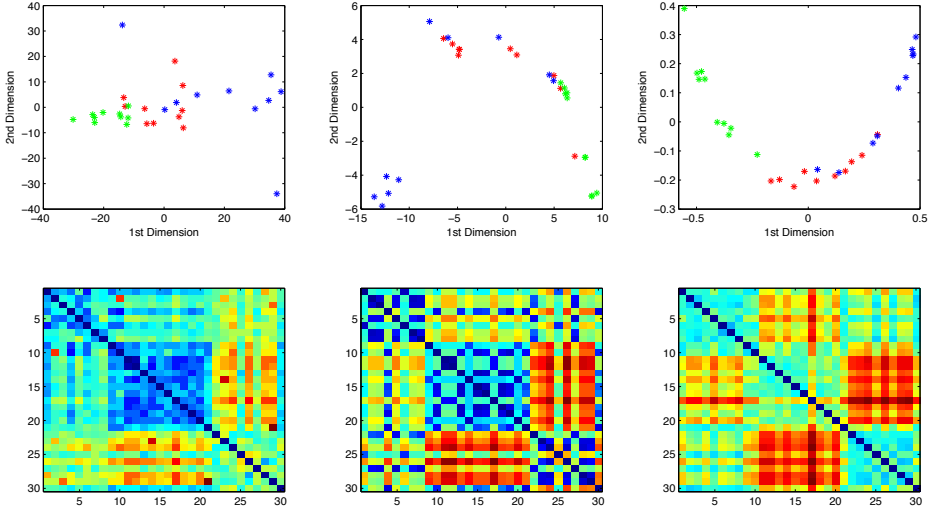


Fig. 1. Two-dimensional MDS embeddings of the synthetic data (top row) on different distance matrices (bottom row). From left to right, the distance is computed as the edit distance between the graphs, the distance between the graph spectra and the distance associated with the QJSD kernel.

4 Experimental Results

In this section, we evaluate the performance of our kernel and we compare it with a number of well-known alternative graph kernels, namely the classic random walk kernel [7], the shortest-path kernel [8] and a set of graphlet kernels [9]. We test different variants of the graphlet kernel, where we vary the graphlet sizes $k \in \{3, 4\}$ and the type of graphlets (all possible size k graphlets vs only those which are fully connected).

The experiments are performed on three different standard dataset, namely MUTAG, Enzymes and PPI. Table 1 reports some statistics about these datasets. MUTAG is a dataset of 188 mutagenic aromatic and heteroaromatic compounds labeled according to whether or not they have a mutagenic effect on the Gram-negative bacterium *Salmonella typhimurium*. Enzymes is a dataset of graphs representing protein tertiary structures that consists of 600 enzymes from the BRENDA enzyme database. Finally, the PPI dataset consists of protein-protein interaction (PPIs) networks related to histidine kinase from two different groups: 40 PPIs from *Acidovorax avenae* and 46 PPIs from *Acidobacteria*. To these three datasets, we add a fourth set of 30 synthetically generated graphs, 10 for each class. The graphs belonging to each class were sampled from a generative model with size 12,14 and 16 respectively. Details about the generative model can be found in [20].

We first evaluate the Multidimensional Scaling embedding of the synthetic graphs for three different distance matrices, namely the edit distance, the distance between the graph spectra and the distance corresponding to our kernel function. The distance between the graph spectra is computed as follows. For each graph G with adjacency matrix A , we compute the column vector s_G of the ordered eigenvalues of A . As the graphs are of different sizes and thus their spectra are of different lengths, the vectors are all made to be the same length by padding zeros to the end of the shorter vector. The (i, j) th element of the distance matrix is then $d_{ij} = \|s_i - s_j\|$. Figure 1 shows the MDS embeddings and the graph distance matrices. It is clear that the distance matrix associated with our kernel has a well-defined block structure which is reflected in the MDS embedding, where the three classes seem to be easily separable.

A second experiment uses a binary C-SVM to test the efficacy of our kernel for classification. We perform 10-fold cross validation, where for each sample we independently tune the value of C , the SVM regularizer constant, by considering the training data from that sample. The process is averaged over 100 random partitions of the data. Given this setting, we first investigate the effect of the time parameter in the classification accuracy. Fig. 2 shows the value of the average accuracy (\pm standard error) on the synthetic dataset as the time parameter T varies. Here the red horizontal line shows the mean accuracy for $T \rightarrow \infty$. The plot shows that the choice of the time greatly influences the performance of our kernel, as we can clearly see that the average accuracy reaches a maximum before stabilizing around the asymptotic value. This should be compared with the average accuracy that we achieve for $T \rightarrow \infty$, which, although not optimal, is not too far from the maximum. However, a more detailed study of the time parameter is beyond the scope of this paper and will thus be the subject of future work.

Finally, Table 2 reports the average classification accuracies (\pm standard error) of the different kernels. As we can see, the proposed kernel achieves the best result on three out of four datasets. The poor accuracy on the Enzymes dataset is likely to be linked to the presence of disjoint graphs, as this will affect the way in which the walk spreads through the graph. Note, however, that this is a particularly hard dataset where the structures of the graphs provide limited information about the underlying class structure. In fact, all kernels based only on graph structure perform only marginally better than random guess, and node and edge attributes need to be taken into account too.

Table 1. Statistics on the graph datasets

datasets	# graphs	# classes	avg # nodes	disjoint
Synth	30	3 (10 each)	13.77	N
MUTAG	188	2 (125 vs. 63)	17.93	N
Enzymes	600	6 (100 each)	32.63	Y
PPI	86	2 (40 vs. 46)	109.60	N

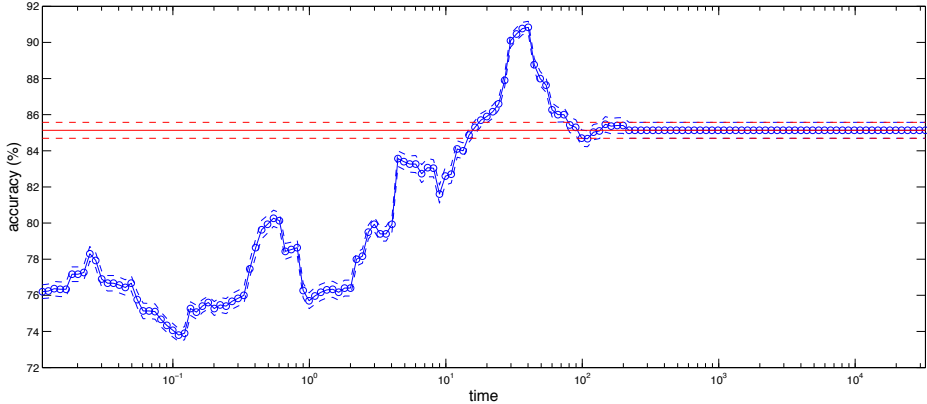


Fig. 2. The mean accuracy (\pm standard error) of the QJSD kernel as the time parameter T varies. The red horizontal line shows the mean accuracy for $T \rightarrow \infty$.

Table 2. Classification accuracy (\pm standard error) on unattributed graph datasets. QJSD is the proposed kernel, SP is the shortest-path kernel [8], RW is the random walk kernel [7], while G_k (CG_k) denotes the graphlet kernel computed using all graphlets (all the connected graphlets, respectively) of size k [9].

Kernel	Synth	MUTAG	Enzymes	PPI
QJSD	85.20 ± 0.47	86.55 ± 0.15	24.20 ± 0.38	78.43 ± 0.30
SP	74.90 ± 0.33	85.02 ± 0.17	28.55 ± 0.42	66.14 ± 0.40
RW	78.53 ± 0.43	77.87 ± 0.21	22.15 ± 0.37	69.70 ± 0.30
G_3	79.33 ± 0.39	82.04 ± 0.14	24.87 ± 0.22	51.95 ± 0.44
G_4	83.60 ± 0.48	81.89 ± 0.13	28.60 ± 0.21	73.14 ± 0.37
CG_3	56.57 ± 0.47	66.43 ± 0.08	19.92 ± 0.27	52.89 ± 0.50
CG_4	81.57 ± 0.54	69.08 ± 0.15	23.05 ± 0.06	61.56 ± 0.41

5 Conclusions

In this paper, we have introduced a novel kernel on unattributed graphs where we probe the graph structure using the time evolution of a continuous-time quantum walk. More precisely, given a pair of graphs we computed the quantum Jensen-Shannon divergence between the evolutions of two quantum walks on a suitably defined union of the original graphs. With the quantum Jensen-Shannon divergence to hand, we established our graph kernel. We performed an extensive experimental evaluation and we demonstrated the effectiveness of the proposed approach. Future work will focus on incorporating node and edge labels information, as well as studying the role of the time parameter more in depth.

Acknowledgments. Edwin Hancock was supported by a Royal Society Wolfson Research Merit Award.

References

1. Siddiqi, K., Shokoufandeh, A., Dickinson, S., Zucker, S.: Shock graphs and shape matching. *International Journal of Computer Vision* 35, 13–32 (1999)
2. Jeong, H., Tombor, B., Albert, R., Oltvai, Z., Barabási, A.: The large-scale organization of metabolic networks. *Nature* 407, 651–654 (2000)
3. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y.: A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences* 98, 4569 (2001)
4. Kalapala, V., Sanwalani, V., Moore, C.: The structure of the united states road network. University of New Mexico (2003) (preprint)
5. Schölkopf, B., Smola, A.J.: *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press (2001)
6. Vapnik, V.: *Statistical learning theory* (1998)
7. Gärtner, T., Flach, P.A., Wrobel, S.: On graph kernels: Hardness results and efficient alternatives. In: Schölkopf, B., Warmuth, M.K. (eds.) *COLT/Kernel 2003*. LNCS (LNAI), vol. 2777, pp. 129–143. Springer, Heidelberg (2003)
8. Borgwardt, K., Kriegel, H.: Shortest-path kernels on graphs. In: *Fifth IEEE International Conference on Data Mining*, p. 8. IEEE (2005)
9. Shervashidze, N., Vishwanathan, S., Petri, T., Mehlhorn, K., Borgwardt, K.: Efficient graphlet kernels for large graph comparison. In: *Proceedings of the International Workshop on Artificial Intelligence and Statistics*. Society for Artificial Intelligence and Statistics (2009)
10. Haussler, D.: Convolution kernels on discrete structures. Technical report, UC Santa Cruz (1999)
11. Bai, L., Hancock, E.: Graph kernels from the Jensen-Shannon divergence. *Journal of Mathematical Imaging and Vision*, 1–10 (2012)
12. Farhi, E., Gutmann, S.: Quantum computation and decision trees. *Physical Review A* 58, 915 (1998)
13. Kempe, J.: Quantum random walks: an introductory overview. *Contemporary Physics* 44, 307–327 (2003)
14. Emms, D., Wilson, R., Hancock, E.: Graph embedding using quantum commute times. *Graph-Based Representations in Pattern Recognition*, 371–382 (2007)
15. Rossi, L., Torsello, A., Hancock, E.R.: Approximate axial symmetries from continuous time quantum walks. 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. 144–152. Springer, Heidelberg (2012)
16. Majtey, A., Lamberti, P., Prato, D.: Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states. *Physical Review A* 72, 052310 (2005)
17. Lamberti, P., Majtey, A., Borrás, A., Casas, M., Plastino, A.: Metric character of the quantum Jensen-Shannon divergence. *Physical Review A* 77, 052311 (2008)
18. Nielsen, M., Chuang, I.: *Quantum computation and quantum information*. Cambridge University Press (2010)
19. Briët, J., Harremoës, P.: Properties of classical and quantum jensen-shannon divergence. *Physical review A* 79, 52311 (2009)
20. Torsello, A., Rossi, L.: Supervised learning of graph structure. *Similarity-Based Pattern Recognition*, 117–132 (2011)