This is a repository copy of *Transitive state alignment for the quantum jensen-shannon kernel*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/85368/

Version: Submitted Version

Proceedings Paper:
Torsello, Andrea, Gasparetto, Andrea, Rossi, Luca et al. (2 more authors) (2014) Transitive state alignment for the quantum jensen-shannon kernel. In: Fränti,, Pasi, Brown, Gavin, Loog, Marco, Escolano, Francisco and Pelillo, Marcello, (eds.) Structural, Syntactic, and Statistical Pattern Recognition:Joint IAPR International Workshop, S+SSPR 2014, Joensuu, Finland, August 20-22, 2014. Proceedings. Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition, S+SSPR 2014, 20-22 Aug 2014 Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) . Springer , GBR , pp. 22-31.

https://doi.org/10.1007/978-3-662-44415-3_3

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Transitive State Alignment for the Quantum Jensen-Shannon Kernel

Andrea Torsello[1], Andrea Gasparetto[1], Luca Rossi[2], Lu Bai[3], and Edwin R. Hancock[3]

1 Università Ca' Foscari Venezia, Italy
2 University of Birmingham, UK
3 University of York, UK
torsello@dais.unive.it,andrea.gasparetto@unive.it,l.rossi@cs.bham.ac.uk, lu@cs.york.ac.uk,erh@cs.york.ac.uk

Abstract. Kernel methods provide a convenient 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. One problem with the most widely used kernels is that they neglect the locational information within the structures, resulting in less discrimination. Correspondence-based kernels, on the other hand, are in general more discriminating, at the cost of sacrificing positive-definiteness due to their inability to guarantee transitivity of the correspondences between multiple graphs. In this paper we generalize a recent structural kernel based on the Jensen-Shannon divergence between quantum walks over the structures by introducing a novel alignment step which rather than permuting the nodes of the structures, aligns the quantum states of their walks. This results in a novel kernel that maintains localization within the structures, but still guarantees positive definiteness. Experimental evaluation validates the effectiveness of the kernel for several structural classification tasks.

1 Introduction

Structural representations have become increasingly popular due to their representational power. However, the descriptiveness comes at the cost of an increased difficulty in applying standard machine learning and pattern recognition techniques to them, as these usually require data that reside in a vector space. The famous kernel trick allows the focus to be shifted from the vectorial representation of data, which now becomes implicit, to a similarity representation. This allows standard learning techniques to be applied to structural data for which no obvious vectorial representation exists.

One of the most influential works on structural kernels was the generic R-convolution kernel proposed by Haussler [6]. Here graph kernels are computed by comparing the similarity of the basic elements for a given decomposition of the two graphs. Depending on the decomposition chosen, we obtain different

kernels. Most R-convolution kernels simply count the number of isomorphic substructures in the two graphs. For example, Kashima et al. [8] compute the kernel by decomposing the graph into random walks, while Borgwardt et al. [3] have proposed a kernel based on shortest paths. Here, the similarity is determined by counting the numbers of pairs of shortest paths of the same length in a pair of graphs. Shervashidze et al. [16] have developed a subtree kernel on subtrees of limited size, where the number of subtrees common between two graphs is computed efficiently using the Weisfeiler-Lehman graph invariant.

One drawback of these kernels is that they neglect the locational information for the substructures in a graph. In other words, the similarity does not depend on the relationships between substructures. As a consequence, these kernels cannot establish reliable structural correspondences between the substructures. This limits the precision of the resulting similarity measure. To overcome this problem, Fröhlich et al. [5] introduced alternative optimal assignment kernels. Here each pair of structures is aligned before comparison. However, the introduction of the alignment step results in a kernel that is not positive definite in general [19]. The problem results from the fact that alignments are not in general transitive. In other words, if σ is the vertex-alignment between graph A and graph B, and π is the alignment between graph B and graph C, in general we cannot guarantee that the alignment between graph A and graph C is $\pi \circ \sigma$. On the other hand, when the alignments are transitive, there is a common simultaneous alignment of all the graphs. Under this alignment, the optimal assignment kernel is simply the sum over all the vertex/edge kernels, which is positive definite since it is the sum of separate positive definite kernels. While lacking positive definiteness the optimal assignment kernels cannot be guaranteed to represent an implicit embedding into a Hilbert space, they have nonetheless been proven to be very effective in classifying structures.

There has recently been an increasing interest in quantum computing because of the potential speed-ups over classical algorithms. Recently Bai et al. [1] introduced a graph kernel based on a Quantum analogue of the Jensen-Shannon divergence between average states of continuous-time quantum walks over the structures to be analyzed. Being based on the divergence which is conjectured to be negative definite [4], the kernel is thought to be positive definite. However it lacks permutational invariance, thus different permutations of the same graphs result in different values of the kernel. This fact, while mitigated by the long range interactions reinforced by the interference patterns in quantum walks, is a rather undesirable property for a structural kernel. For this reason in this paper we modify the kernel by adding a novel alignment step that rather than permuting the nodes of the structures, aligns the quantum states of the walks. This results in a novel kernel that is permutationally invariant and maintains similar localization property of the alignment kernels [5, **?**]. Further, we prove that the alignment transformations between multiple structures are transitive and that, for this particular alignment, the kernel is always positive definite.

2 Quantum Mechanical Background

Quantum walks are the quantum analogue of classical random walks. 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, 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. As a consequence, the evolution of the quantum walk is reversible, which implies that quantum walks are non-ergodic and do not possess a limiting distribution. See [9] for an overview of the properties of quantum walks. 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 $|\psi_t\rangle = \sum_{u \in V} \alpha_u(t) |u\rangle$ 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 $\Pr(X^t = u) = \alpha_u(t)\alpha_u^*(t)$ where $\alpha_u^*(t)$ is the complex conjugate of $\alpha_u(t)$. Moreover, in a closed system $\sum_{u \in V} \alpha_u(t)\alpha_u^*(t) = 1$.

The evolution of the walk over graph $G = (V, E)$ is governed by Schrödinger equation, where we take the Hamiltonian of the system to be the graph Laplacian L, which, eliminating scaling constants, yields

$$\frac{d}{dt} |\psi_t\rangle = -iL |\psi_t\rangle \tag{1}$$

Given an initial state $|\psi_0\rangle$, we can solve Equation 1 to determine the state vector at time t $|\psi_t\rangle = e^{-iLt} |\psi_0\rangle = \Phi e^{-i\Lambda t}\Phi^\top |\psi_0\rangle$, where $L = \Phi\Lambda\Phi^\dagger$ is the spectral decomposition of the Laplacian matrix.

While a pure state can be naturally described using a single ket vector, in general a quantum system can be in a *mixed state*, i.e., a statistical ensemble of pure quantum states $|\psi_i\rangle$, each with probability p_i. The *density operator* (or *density matrix*) of such a system is defined as

$$\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i| . \tag{2}$$

Density operators are positive unit-trace matrices directly linked with the observables of the (mixed) quantum system. Let O be an observable, i.e., an Hermitian operator acting on the quantum states and providing a measurement. The expected value of the measurement O over a mixed state can be calculated from the density matrix ρ: $\langle O \rangle = \mathrm{tr}\,(\rho O)$, where tr is the trace operator.

The Von Neumann entropy of a density operator ρ is

$$H_N(\rho) = -Tr(\rho \log \rho) = -\sum_j \lambda_j \log \lambda_j , \tag{3}$$

where the λ_js 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

$$D_{\mathrm{JS}}(\rho,\sigma) = H_N\Big(\frac{\rho+\sigma}{2}\Big) - \frac{1}{2}H_N(\rho) - \frac{1}{2}H_N(\sigma) \tag{4}$$

This quantity is symmetric, bounded between 0 and 1, and negative definite for pure states and is conjectured with ample experimental evidence to be negative definite for all states [4].

Finally, for a graph $G(V,E)$, let $|\psi_t\rangle$ denote the state corresponding to a continuous-time quantum walk that has evolved from time $t=0$ to time $t=T$. We define the time-averaged density matrix $\rho_G^{\dagger}$ for $G(V,E)$

$$\rho_G^{\dagger} = \frac{1}{T}\int_0^T |\psi_t\rangle\langle\psi_t|\,\mathrm{d}t\,. \tag{5}$$

Let ϕ_{xy} denote the (xy)th element of the matrix of eigenvectors Φ of the Laplacian. Following [14], we compute the (r,c)th element of ρ_T as follows:

$$\rho_G^{\dagger}(r,c) = \sum_{k=1}^{n}\sum_{y=1}^{n}\phi_{rk}\phi_{cy}\bar{\psi}_k\bar{\psi}_y\frac{1}{T}\int_0^{\dagger} e^{i(\lambda_y-\lambda_k)t}\,\mathrm{d}t. \tag{6}$$

If we let $T\to\infty$, Eq.(6) further simplifies to

$$\rho_G^{\infty} = \sum_{\lambda\in\tilde{\Lambda}} P_\lambda\rho_0 P_\lambda^{\top} \tag{7}$$

where $\tilde{\Lambda}$ is the set of distinct eigenvalues of the Laplacian matrix L and P_λ is the orthogonal projector onto the eigenspace associated with λ.

3 State-Aligned QJSD Kernel

In [1] the Bai et al. defined a kernel based on the Quantum Jensen Shannon divergence between two continuous-time quantum walks between the graphs. The QJSD kernel was defined as

$$K_{\mathrm{QJSD}}(G_1,G_2) = \exp\left(-\beta D_{\mathrm{JS}}(\rho_1,\rho_2)\right) \tag{8}$$

where ρ_1 and ρ_2 are the time-averaged density matrices associated with the quantum walks over G_1 and G_2 respectively, and β is a decay parameter of the kernel. The walks are initialized in the starting state

$$|\Psi_0\rangle = \sum_{u\in V}\sqrt{\frac{d_u}{\sum_{v\in V} d_v}}\,|u\rangle\ . \tag{9}$$

The kernel is positive definite under the conjecture that the quantum Jensen-Shannon divergence is negative definite for all states, and exhibited good performance on several graph classification tasks, but its value is dependent on the order under which the nodes are presented due to the mixing term $\frac{\rho+\sigma}{2}$ in the definition of the divergence.

In this paper we solve the permutational invariance problem of the QJSD kernel by adding an alignment step to the computation of the kernel. In contrast to alternative alignment kernels such as [5] and [**?**], the alignment is not performed over the node permutations Σ_n of the graphs. Rather it is performed over the quantum basis under which the walker can be observed. In classical random walks the nodes of the graph provide a preferred basis for observation as the walker cannot be simultaneously on multiple nodes, thus the only available degree of freedom is in the choice of an order within the basis vectors, i.e., the observation basis is fully defined up to a permutation $\pi \in \Sigma_n$. This is in stark contrast with quantum mechanics where, due to quantum superposition, prior to observation a quantum walker can be simultaneously at multiple nodes, and the observation itself can be performed under any quantum superposition of states. This means that any orthogonal basis is valid for observation and, thus, the basis is defined up to a unitary transformation $O \in U(n)$, where $U(n)$ is the Unitary group over $\mathbb{C}^n$.

Following this property, we define a *State-aligned* QJSD kernel as

$$\begin{aligned} K_{\mathrm{SAQJSD}}(G_1, G_2) &= \max_{O \in U(n)} \exp\left(-\beta D_{\mathrm{JS}}(\rho_1, O\rho_2 O^\dagger)\right) \\ &= \exp\left(-\beta \min_{O \in U(n)} D_{\mathrm{JS}}(\rho_1, O\rho_2 O^\dagger)\right) \end{aligned} \tag{10}$$

In the following we will prove some important properties of the state-aligned kernel. Namely we will give a closed form solution to the alignment, prove that the optimal transformation are transitive, and prove that the resulting kernel is positive definite without making use of the negative-definiteness conjecture for the quantum Jensen-Shannon divergence.

3.1 Properties of the State-Aligned QJSD Kernel

We start by enunciating a theorem relating the optimal state-alignment to the eigenvectors of the density matrices. For a proof of this result see [18].

Theorem 1. *Let $\rho 1 = \Phi_1 \Lambda_1 \Phi_1^\dagger$ and $\rho 2 = \Phi_2 \Lambda_2 \Phi_2^\dagger$ be the singular value decompositions of ρ_1 and ρ_2 respectively, with the eigenvalues in descending order in both Λ_1 and Λ_2, then the global minimum of $\bar{H}_N(O)$ is attained by $O^* = \Phi_1 \Phi_2^\dagger$.*

This theorem tells us how to efficiently compute the state alignment. Further, this transformation aligns the eignevectors resulting in a mixed density matrix $\frac{1}{2}(\rho_1 + O^* \rho_2 O^{*\dagger})$ with eigenvalues $\frac{1}{2}(\lambda_1 + \mu_i)$ where $\lambda_1, \ldots, \lambda_n$ and $\mu_1, \ldots, \mu_n$ are the eigenvalues of ρ_1 and ρ_2 respectively taken in descending order with their

multiplicity. This means that the aligned Jensen Shannon divergence only needs the eigenvalues of ρ_1 and ρ_2 to be computed, in fact:

$$\min_{O\in U_n} D_{\mathrm{JS}}(\rho_1, O\rho_2 O^\dagger) = \sum_{j=1}^{n} -\frac{\lambda_j+\mu_j}{2}\log\left(\frac{\lambda_j+\mu_j}{2}\right) + \frac{\lambda_j\log(\lambda_j)+\mu_j\log(\mu_j)}{2}\,. \tag{11}$$

This reduces the computational complexity of computing the kernel for all times at which the mixed density matrix is computed, as we do not need to perform the eignedecomposition of the mixed matrix $\frac{1}{2}(\rho_1+\rho_2)$ for each pair of graphs in the kernel. Rather, we only need to compute the eigenvalues (not the eigenvectors) of all the density matrices beforehand. The resulting complexity for the whole kernel computation is $O(Nn^3+N^2n)$ where N is the number of graphs and n their (maximum) size. In contrast, the QJSD kernel has complexity $O(Nn^3+N^2n^2)$ due to the eigenvalue computation for each pair of graphs.

Further, in the case of the infinite-time mixing matrix, we can significantly reduce the computational burden of computing the eigenvalues of the density matrix, by using a result presented in [14]. There it was proven that the infinite-time mixing matrix commuted with the graph Laplacian. As a consequence, ρ^∞ expressed in the eigenbasis of the Laplacian, is a block diagonal matrix where blocks correspond to eigenspaces associated with a single eigenvalue. Let $L=\Phi\Lambda\Phi^\dagger$, be the spectral decomposition of the graph Laplacian, we denote with Φ_j the matrix formed with the columns of Φ corresponding to the eigenvectors associated with the j-th distinct eigenvalue. The j-th diagonal block of ρ^∞ expressed in the eigenbasis Φ is $\Phi_j^\dagger\rho^\infty\Phi_j$. using Eq. (7) and recalling that $P_j=\Phi_j^\dagger\Phi_j$, we have

$$\Phi_j^\dagger\rho^\infty\Phi_j = \Phi_j^\dagger\rho^\infty\Phi_j = \Phi_j^\dagger\rho^0\Phi_j = \Phi_j^\dagger\left|\psi_0\right\rangle\left\langle\psi_0\right|\Phi_j = \left|\Phi_j^\dagger\psi_0\right\rangle\left\langle\Phi_j^\dagger\psi_0\right| \tag{12}$$

which is a rank 1 matrix with a single non-zero eigenvalue $\lambda_j=\|\Phi_j^\dagger\psi_0\|^2$. Hence, once the singular value decomposition of the graph's Laplacian is to hand, we can compute the eigenvalues of the infinite-time mixing matrix directly, without the need for an additional decomposition. This makes the infinite-time kernel particularly efficient to compute.

It is worth noting that as the graph Laplacian has eigenvalues with higher multiplicity the infinite-time mixing matrix has more zero eigenvalues resulting in a lower Von Neumann entropy. This is particularly interesting since higher multiplicities of the eigenvalues is associated with the presence of symmetries in the graph [12] which, in turn, have been used to characterize the entropy of the structure [11].

We can now prove the following properties for the state-aligned kernel

Theorem 2. *The Unitary transformations minimizing the quantum Jensen Shannon divergence between pairs of density matrices in a set are transitive, i.e. let*

$$O_{i,j} = \operatorname*{argmin}_{O\in U(n)} D_{\mathrm{JS}}(\rho_i, O\rho_j O^\dagger)$$

with $i, j \in \{1, 2, 3\}$*, then*

$$D_{\mathrm{JS}}\big(\rho_1, O_{1,2}O_{2,3}\rho_3 O_{2,3}^{\dagger}O_{1,2}^{\dagger}\big) = D_{\mathrm{JS}}\big(\rho_1, O_{1,3}\rho_3 O_{1,3}^{\dagger}\big)$$

Proof The optimal transformation between two density matrices is completely determined by the relation $O_{1,2}^{*} = \Phi_1\Phi_2^{\dagger}$ up to a change of sign of the eigenvalue and a change of base for each eignespace associated with an eigenvalue with multiplicity greater than one. In any case these changes do not affect the value of the divergence. However,

$$O_{1,2}^{*}O_{2,3}^{*} = \Phi_1\Phi_2^{\dagger}\Phi_2\Phi_3^{\dagger} = \Phi_1\Phi_3^{\dagger} = O_{1,3}^{*} \tag{13}$$

QED.

Theorem 3. *The quantum aligned QJSD kernel is positive definite.*

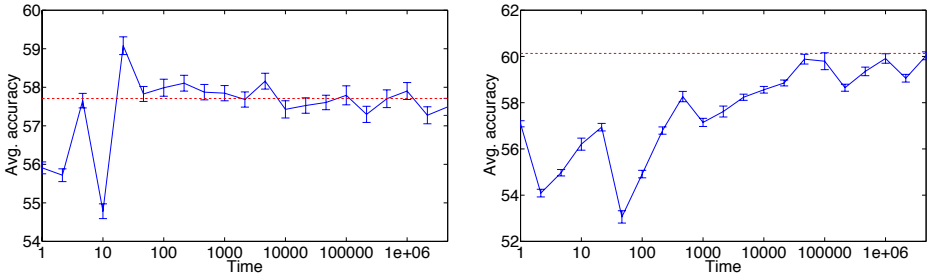
Proof As a consequence of the previous theorems, the value of the quantum Jensen Shannon divergence of the optimally aligned density matrices is equal to the normal Jensen Shannon divergence of the sorted eigenvalues of the density matrices (taken as probability distributions). Since the Jensen Shannon divergence is proven to be negative definite [4] the state-aligned QJSD kernel, being an exponentiation of a negative definite kernel is positive definite [10]. QED.

4 Experimental Results

We now evaluate the performance of the State-Aligned (SA) QJSD kernel and we compare it with a number of well-known alternative graph kernels. More specifically, we compare our kernel with the unaligned QJSD kernel [1], the Weisfeiler-Lehman kernel [16], the graphlet kernel [17], the shortest-path kernel [3], and the random walk kernel [8]. Note that for the Weisfeiler-Lehman we set the number of iterations $h = 3$ and we attribute each node with its degree.

We run our experiments on the following datasets: 1) The **PPI** dataset, which consists of protein-protein interaction (PPIs) networks related to histidine kinase [7] (40 PPIs from *Acidovorax avenae* and 46 PPIs from *Acidobacteria*). 2) The **PTC** (The Predictive Toxicology Challenge) dataset, which records the carcinogenicity of several hundred chemical compounds for male rats (MR), female rats (FR), male mice (MM) and female mice (FM) [15] (here we use the 344 graphs in the MR class). 3) The **COIL** dataset, which consists of 5 objects from [13], each with 72 views obtained from equally spaced viewing directions, where for each view a graph was built by triangulating the extracted Harris corner points. 4) The **Reeb** dataset, which consists of a set of adjacency matrices associated to the computation of reeb graphs of 3D shapes [2]. Further information on the four selected datasets is shown in Table **??**.

We use a binary C-SVM to test the efficacy of the kernels. 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



(a) Sub-optimal β (b) Optimal β

Fig. 1. The average classification accuracy as the time parameter of the continuous-time quantum walk varies, for an optimal (left) and sub-optimal value of the decay factor β.

sample. The process is averaged over 100 random partitions of the data, and the results are reported in terms of average accuracy $\pm$ standard error.

Fig. 1 shows the value of the average classification accuracy ($\pm$ standard error) on the PTC dataset as we let the time parameter of the continuous-time quantum walk vary. Here the red horizontal line denotes the average accuracy for $T \to \infty$. Note that in Fig. 1(a) we set the decay parameter β of the kernel to a sub-optimal value, while in Fig. 1(b) we set it to its optimal value, i.e., the value that results in the best classification accuracy. The plot shows that when β is sub-optimal the choice of the time parameter has a clear influence on the performance of our kernel. In fact, we see that the average accuracy reaches a maximum before stabilizing around the asymptotic value. On the other hand, when β is optimized the best classification performance is achieved when $T \to \infty$. Moreover, in the latter case the average classification accuracy is higher than that recorded for smaller values of T and a sub-optimal β.

Table 1. Classification accuracy ($\pm$ standard error) on unattributed graph datasets. SA QJSD and QJSD denote the proposed kernel and its original unaligned version, respectively, WL is the Weisfeiler-Lehman kernel [16], GR denotes the graphlet kernel computed using all graphlets of size 3 [17], SP is the shortest-path kernel [3], and RW is the random walk kernel [8]. For each kernel and dataset, the best performing kernel is highlighted in bold.

Kernel	PPI	PTC	COIL	Reeb
SA QJSD	75.69 ± 0.85	$\mathbf{60.13 \pm 0.51}$	67.84 ± 0.15	38.50 ± 0.26
QJSD	69.12 ± 1.01	56.06 ± 0.45	$\mathbf{69.90 \pm 0.22}$	35.78 ± 0.42
WL	$\mathbf{79.40 \pm 0.96}$	56.95 ± 0.31	29.00 ± 0.57	50.53 ± 0.41
GR	51.94 ± 0.97	55.22 ± 0.19	66.46 ± 0.44	22.80 ± 0.36
SP	63.31 ± 0.80	56.51 ± 0.36	69.68 ± 0.36	$\mathbf{55.93 \pm 0.36}$
RW	50.37 ± 0.78	55.68 ± 0.14	12.18 ± 0.21	16.47 ± 0.43

Table 2. Runtime comparison on the four graph datasets.

Kernel	PPI	PTC	COIL	Reeb
SA QJSD	4.11”	14.90”	46.49”	20.07”
QJSD	118.16”	39.81”	1911.452”	495.46”
WL	4.55”	4.32”	23.46”	12.29”
GR	2.90”	0.85”	7.99”	2.23”
SP	5.09”	1.03”	22.43”	7.30”
RW	20.19”	246.20”	505.75”	833.16”

Table 1 shows the average classification accuracy ($\pm$ standard error) of the different kernels on the selected datasets. As expected, we see that the state alignment almost invariably yields an increase of the performance with respect to the standard QJSD kernel. Indeed, the localization property of the kernel that results from the quantum state alignment leads to a better discrimination, and thus a higher classification accuracy. Moreover, while the QJSD kernel has not been proven to be positive definite, as the quantum Jensen-Shannon divergence has only been experimentally shown to be negative definite for mixed states, our kernel is indeed positive definite, as proved in the previous Section.

Finally, Table 2 shows the runtimes of the different kernels on the four graph datasets. Note that in terms of runtime the SA QJSD kernel easily outperforms the other spectral methods, i.e., the QJSD kernel and the random walk kernel, and it is still competitive when compared with the remaining kernels.

With respect to the other kernels, the SA QJSD kernel achieves the best accuracy on the PTC dataset, and it remains competitive with the best performing ones on the PPI and COIL dataset. On the Reeb dataset, on the other hand, the shortest-path kernel and the Weisfeiler-Lehman kernel outperform our kernel and the remaining ones. Note also that the Weisfeiler-Lehman mitigates the localization problem by making use of the node labels and thus improving node localization in the evaluation of the kernel. On the other hand, our kernel does not take node attributes into account.

5 Conclusions

In this paper we have generalized a recent structural kernel based on the Jensen-Shannon divergence between quantum walks over the graph by introducing a novel alignment step which, rather than permuting the nodes of the structures, aligns the quantum states of their walks. We proved that the resulting kernel mantains the localization within the structures, but still guarantees positive definiteness. We tested our kernel against a number of alternative graph kernels and we showed its effectiveness in a number of structural classification tasks.

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

References

1. L. Bai, E.R. Hancock, A. Torsello, L. Rossi, A Quantum Jensen-Shannon Graph Kernel Using the Continuous-Time Quantum Walk, in: *Proc. Graph-Based Representations in Pattern Recognition (GbRPR)*, pp. 121–131, 2013.
2. S. Biasotti, S. Marini, M. Mortara, G. Patanè, M. Spagnuolo, B. Falcidieno, 3d shape matching through topological structures, in: *DGCI*, pp. 194–203, 2003.
3. K.M. Borgwardt, H.P. Kriegel, Shortest-path kernels on graphs, in: *Proc. IEEE Int. Conf. Data Mining (ICDM)*, pp. 74–81, 2005.
4. Briët, J., Harremoës, P., Properties of classical and quantum jensen-shannon divergence, *Physical review A* **79**, 052311, 2009.
5. Holger Fröhlich, Jörg K. Wegner, Florian Sieker, Andreas Zell, Optimal assignment kernels for attributed molecular graphs, in: *International Conference on Machine Learning*, pp.: 225-232, 2005.
6. D. Haussler, Convolution kernels on discrete structures, *Technical Report UCS-CRL-99-10*, Santa Cruz, CA, USA, 1999.
7. L. J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, et al., String 8a global view on proteins and their functional interactions in 630 organisms, *Nucleic acids research* **37** (suppl 1), pp. D412–D416, 2009.
8. H. Kashima, K. Tsuda, A. Inokuchi, Marginalized kernels between labeled graphs, in: *Proc. Int. Conf. Machine Learning (ICML)*, pp. 321–328, 2003.
9. J. Kempe, Quantum random walks: an introductory overview. *Contemporary Physics* **44**, pp. 307–327, 2003.
10. R. Konder, J. Lafferty, Diffusion kernels on graphs and other discrete input spaces, in: *Proc. Int. Conf. Machine Learning (ICML)*, pp. 315–322, 2002.
11. A. Mowshowitz, Entropy and the complexity of graphs: I. An index of the relative complexity of a graph, *The bulletin of mathematical biophysics* **30**(1):175–204, 1968.
12. A. Mowshowitz, Graphs, groups and matrices, in: *Proc. 25th Summer Meeting Canad. Math.Congress, Congr. Numer.* **4**, pp. 509–522, 1971.
13. S. A. Nene, S. K. Nayar, H. Murase, Columbia object image library (coil-20), Dept. Comput. Sci., Columbia Univ., New York.[Online] `http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php`.
14. L. Rossi, A. Torsello, E.R. Hancock, R. Wilson, Characterising graph symmetries through quantum Jensen-Shannon divergence, *Physical Review E* **88**(3), 032806, 2013.
15. G. Li, M. Semerci, B. Yener and M.J. Zaki, Effective graph classification based on topological and label attributes, *Statistical Analysis and Data Mining* **5**, pp. 265–283, 2012.
16. N. Shervashidze, P. Schweitzer, E.J. van Leeuwen, K. Mehlhorn, K.M. Borgwardt, Weisfeiler-lehman graph kernels, *Journal of Machine Learning Research* **1**, pp. 1–48, 2010.
17. N. Shervashidze, S. Vishwanathan, T. Petri, K. Mehlhorn, K. Borgwardt, Efficient graphlet kernels for large graph comparison, in: *Proceedings of the International Workshop on Artificial Intelligence and Statistics*, 2009.
18. A. Torsello, On the quantum alignment of the Quantum Jensen Shannon Divergence, Technical report DAIS-2014-1, DAIS, Università Ca' Foscari Venezia. `http://www.dsi.unive.it/~atorsell/DAIS-2014-1.pdf`, 2014.
19. Jean-Philippe Vert, The optimal assignment kernel is not positive definite, *arXiv*:0801.4061, 2008.