Can a Quantum Walk Tell Which Is Which?
A Study of Quantum Walk-Based Graph Similarity

Giorgia Minello [1], Luca Rossi [2],[*] and Andrea Torsello [1]

[1] Dipartimento di Scienze Ambientali, Informatica e Statistica, Universita Ca' Foscari Venezia, via Torino 155, 30170 Venezia Mestre, Italy; giorgia.minello@unive.it (G.M.); andrea.torsello@unive.it (A.T.)

[2] Department of Computer Science and Engineering, Southern University of Science and Technology, Nanshan District, Shenzhen 518055, China

[*] Correspondence: rossil@sustech.edu.cn; Tel.: +86-755-88015162

Abstract: We consider the problem of measuring the similarity between two graphs using continuous-time quantum walks and comparing their time-evolution by means of the quantum Jensen-Shannon divergence. Contrary to previous works that focused solely on undirected graphs, here we consider the case of both directed and undirected graphs. We also consider the use of alternative Hamiltonians as well as the possibility of integrating additional node-level topological information into the proposed framework. We set up a graph classification task and we provide empirical evidence that: (1) our similarity measure can effectively incorporate the edge directionality information, leading to a significant improvement in classification accuracy; (2) the choice of the quantum walk Hamiltonian does not have a significant effect on the classification accuracy; (3) the addition of node-level topological information improves the classification accuracy in some but not all cases. We also theoretically prove that under certain constraints, the proposed similarity measure is positive definite and thus a valid kernel measure. Finally, we describe a fully quantum procedure to compute the kernel.

1. Introduction

In recent years, we have observed rapid advancements in the fields of machine learning and quantum computation. An increasing number of researchers is looking at challenges emerging at the intersection of these two fields, from quantum annealing as an alternative to classical simulated annealing [1,2] to quantum parallelism as a source of algorithmic speedup [3,4]. Quantum walks [5,6], the quantum mechanical analogue of classical random walks, have been shown to provide the potential for exponential speedups over classical computation, in part thanks to an array of exotic properties not exhibited by their classic counterpart, such as interference. The rising popularity of quantum walks can be further understood by looking at the work of Childs [7], who showed that any quantum computation can be efficiently simulated by a quantum walk on a sparse and unweighted graph, thus elevating quantum walks to the status of universal computational primitives.

In machine learning and pattern recognition, graphs are used as convenient representations for systems that are best described in terms of their structure and classical random walks have been repeatedly and successfully used to analyse such structure [8–12]. Indeed, the rich expressiveness of these representations comes with several issues when applying standard machine learning and pattern recognition techniques to them. In fact, these techniques usually require graphs to be mapped to corresponding vectorial representations, which in turn needs a canonical node order to be established first. Graphs with different number of nodes and edges present yet another challenge, as the dimension

of the embedding space depends on this information. In this context, classical random walks provide an effective way to compute graph invariants that can be use to characterise their structure and embed them into a vectorial space [13,14].

Classical random walks have also been used to successfully define graph kernels [9,12,15–18]. Graph kernels offer an elegant way to transform the problem from that of finding an explicit permutation invariant embedding to that of defining a positive semi-definite pairwise kernel measure. One can then choose between an array of different kernel methods, the best known example being support vector machines (SVMs) [19], to solve the pattern analysis task at hand. This is based on the well-known kernel trick. Given a set X and a positive semi-definite kernel $k : X \times X \to \mathbb{R}$, there exists a map $\phi : X \to H$ 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 (e.g., vectors, graphs) on which a kernel is defined. In the case of graphs, examples of kernels include the shortest-path kernel [15], the graphlet kernel [20], and the Weisfeiler-Lehman subtree kernel [16]. The common principle connecting these kernels is that of measuring the similarity between two graphs in terms of the similarity between simpler substructures (e.g., paths, subgraphs, subtrees) contained in the original graphs. When the substructures used to decompose the graphs are classical random walks, we obtain kernels like the random walk kernel [9] or the Jensen-Shannon kernel [12]. The kernel of Gärtner et al. [9] counts the number of matching random walks between two graphs while using a decay factor to downweigh the contribution of long walks to the kernel. The kernel of Bai and Hancock [12], on the other hand, is based on the idea of using classical random walks to associate a probability distribution to each graph and then use the Jensen-Shannon divergence [21,22] between these distributions as a proxy for the similarity between the original graphs.

The quantum analogue of the Jensen-Shannon divergence (QJSD) [23] allows to extend the Jensen-Shannon divergence kernel [12] to the quantum realm [17,18]. While the classical Jensen-Shannon divergence is a pairwise measure on probability distributions, the QJSD is defined on quantum states. As its classical counterparts, the QJSD is symmetric, bounded, always defined, and it has been proved to be a metric for the special case of pure states [23,24]. Unfortunately, there is no theoretical proof yet that the same holds for mixed states. Inspired by this quantum divergence measure, Rossi et al. [18] and Bai et al. [17] have proposed two different graph kernels based on continuous-time quantum walks. While the kernel of Bai et al. was proved to be positive semi-definite [17], it has several drawbacks compared to [18] (for which the positive semi-definiteness has not been proved yet), most notably the need to compute the optimal alignment between the input graphs before the quantum walk-based analysis can commence. Rossi et al. [18], on the other hand, are able to avoid this by exploiting the presence of interference effects. Given a pair of input graphs, their method establishes a complete set of connections between them and defines the initial states of two walks on this structure so as to highlight the presence of structural symmetries [18,25]. However the analysis of Rossi et al. [18] is limited to the case of undirected graphs and, as mentioned, it falls short of proving the positive semi-definiteness of the kernel.

In this paper, we address these issues in the following way:

- We define a novel kernel for directed graphs based on [18] and the work of Chung [26] on directed Laplacians;
- We extend the work of [18] by incorporating additional node-level topological information using two well-known structural signatures, the Heat Kernel Signature [27] and the Wave Kernel Signature [28];
- We give a formal proof of the positive definiteness of the undirected kernel for the case where the Hamiltonian is the graph Laplacian and the starting state satisfies several constraints;
- We propose a simple yet efficient quantum algorithm to compute the kernel;
- We perform an empirical comparison of the performance of the kernel for both directed and undirected graphs and for different choices of the Hamiltonian.

We perform an extensive set of experimental evaluations and we find that:

- Adding the edge directionality information allows to better discriminate between the different classes, even when compared with other commonly used kernels for directed graphs;
- In most cases the incorporation of the node-level topological information results in a significant improvement over the performance of the original kernel;
- The optimal Hamiltonian (in terms of classification accuracy) depends on the dataset, as already suggested in [18];
- The constraints we enforce to ensure the positive definiteness of the kernel disrupt the phase of the initial state, leading to a decrease in classification accuracy.

The remainder of this paper is organised as follows. Section 2 reviews the fundamental concepts of graph theory and quantum mechanics needed to understand the present paper. Section 3 introduces the quantum-walk based similarity measure, analyses it from a theoretical perspectives, and illustrates how it can be adapted to work on both undirected and directed graphs. Finally, Section 4 discusses the results of our experimental evaluation and Section 5 concludes the paper.

2. Graphs and Quantum Walks

In this section we introduce several fundamental graph-theoretical and quantum mechanical notions. In particular, we show how to associate a density matrix to a graph using continuous-time quantum walks. The density matrix representation of the graphs will be then used in Section 3 to define the kernel for both directed and undirected graphs.

2.1. Elementary Graph-Theoretic Concepts

Let $G(V, E)$ denote a directed graph, where V is a set of n vertices and $E \subseteq V \times V$ is a set of edges such that each $e = (u, v) \in E$ has a start-vertex $u \in V$ and an end-vertex $v \in V$. Please note that we sometimes omit (V, E) and simply write G when the vertex and edge sets are implicitely defined. The adjacency matrix A of G has elements

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

If $A_{uv} = A_{vu} = 1$ we say that the edge $e = (u, v)$ is bidirectional. If every edge of a graph G is bidirectional, i.e., the associated adjacency matrix A is symmetric, we say that G is an undirected graph.

For a node u, the degree of u is the number of nodes connected to u by an edge. For undirected graphs this is $d_u = \sum_{v \in V} A(u, v)$, while for directed graphs we distinguish between the in-degree $d_u^{in} = \sum_{v \in V} A(v, u)$ and the out-degree $d_u^{out} = \sum_{v \in V} A(u, v)$, which can be added up to give the total degree $d_u^{tot} = d_u^{in} + d_u^{out}$.

For undirected graphs, we can define the degree matrix D as the matrix with the d_us as the diagonal elements and zero elsewhere. The graph Laplacian of G is then defined as $L = D - A$, and it can be interpreted as a combinatorial analogue of the discrete Laplace-Beltrami operator [29]. Similarly, we can define the normalised Laplacian $\mathcal{L} = D^{-1/2} L D^{1/2}$. For directed graphs, things are less straightforward. In Section 3.1 we illustrate how to define both Laplacian and normalised Laplacian on directed graphs based on the work of Chung [26].

2.2. Quantum Walks on Graphs

Using the Dirac notation to denote as $|u\rangle$ the basis state corresponding to the walk being at vertex $u \in V$, the state of a continuous-time quantum walk on G at time is

$$|\psi_t\rangle = \sum_{u \in V} \alpha_u(t) |u\rangle, \tag{2}$$

where the amplitude $\alpha_u(t) \in \mathbb{C}$ and $|\psi_t\rangle \in \mathbb{C}^{|V|}$ are both complex. Moreover, we have that $\alpha_u(t)\alpha_u^*(t)$ gives the probability that at time t the walker is at the vertex u, and thus $\sum_{u \in V} \alpha_u(t)\alpha_u^*(t) = 1$ and $\alpha_u(t)\alpha_u^*(t) \in [0,1]$, for all $u \in V$, $t \in \mathbb{R}^+$.

The state of the walk evolves through time according to the Schrödinger equation

$$\frac{\partial}{\partial t} |\psi_t\rangle = -i\mathcal{H} |\psi_t\rangle \tag{3}$$

where $\mathcal{H}$ denotes the time-independent Hamiltonian. Given an initial state $|\psi_0\rangle$, the solution of the previous equation for time t is then

$$|\psi_t\rangle = e^{-i\mathcal{H}t} |\psi_0\rangle . \tag{4}$$

In the case of undirected graphs, the Laplacian matrix is usually chosen as the system Hamiltonian, i.e., $\mathcal{H} = L$. However in theory any Hermitian operator encoding the graph structure can be used instead, e.g., the adjacency matrix or the normalised Laplacian matrix. Directed graphs pose an additional challenge, as the asymmetricity of the adjacency matrix implies that it cannot be used as the system Hamiltonian. However it is possible to define symmetric Laplacian and normalised Laplacian matrix representations for directed graphs, as explained in Section 3.1.

Equation (4) can be also rewritten in terms of the spectral decomposition of the Hamiltonian $\mathcal{H} = \Phi\Lambda\Phi^\top$, where Φ is the $|V| \times |V|$ matrix with the ordered eigenvectors of $\mathcal{H}$ as columns and Λ is the $|V| \times |V|$ diagonal matrix with the ordered eigenvalues λ_j of $\mathcal{H}$ as elements. Since $\exp[-i\mathcal{H}t] = \Phi\exp[-i\Lambda t]\Phi^\top$, we can write the solution of the Schrödinger equation at time t as

$$|\psi_t\rangle = \Phi e^{-i\Lambda t}\Phi^\top |\psi_0\rangle . \tag{5}$$

2.3. From Quantum Walks to Graph Density Matrices

The density matrix representation is introduced in quantum mechanics as a way describe a system that is a statistical ensemble of pure states $|\psi_i\rangle$, each with probability p_i, i.e.,

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i| . \tag{6}$$

Please note that density matrices are positive unit trace matrices. Given a density matrix ρ with eigenvalues $\lambda_1, \ldots, \lambda_n$, we can compute its von Neumann entropy [30] $S(\rho)$ as

$$S(\rho) = -\text{tr}(\rho \log \rho) = -\sum_i \lambda_i \ln \lambda_i . \tag{7}$$

Moreover, given two density matrices ρ and σ we can compute their quantum Jensen-Shannon divergence (QJSD) [23,24] as

$$QJSD(\rho,\sigma) = S\left(\frac{\rho+\sigma}{2}\right) - \frac{1}{2}S(\rho) - \frac{1}{2}S(\sigma) . \tag{8}$$

The QJSD is always well defined, symmetric and positive definite [23,24]. It can also be shown that $QJSD(\rho,\sigma)$ is bounded, i.e., $0 \leq QJSD(\rho,\sigma) \leq 1$, with the latter equality attained only if ρ and σ have support on orthogonal subspaces [18].

The QJSD has been shown to have several interesting properties required for a good distinguishability measure between quantum states [23,24]. In pattern recognition, it has been used to define several graph kernels [17,18,31]. This in turn is achieved by associating quantum states to graphs and using the QJSD between the states as a proxy of the similarity between the graphs.

More specifically, given a graph G and a continuous-time quantum walk with starting state $|\psi_0\rangle$, the idea is to associate a quantum state to G as follows. First, we let the quantum walk evolve

according to Equation (4) until a time T is reached. Then, we define the quantum state with density matrix representation

$$\rho_T = \frac{1}{T} \int_0^T |\psi_t\rangle \langle \psi_t| \, \mathrm{d}t \,. \tag{9}$$

This state represents a uniform ensemble of pure states $|\psi_t\rangle$ each corresponding to a different stage of the quantum walk evolution. Different versions of QJSD-based graph kernels are based on different definitions of this density matrix [17,18]. In this paper we focus on the kernel of Rossi et al. [18], which was shown to have several advantages compared to its alternatives. However, while the original work of Rossi et al. [18] focused only on undirected graph, in the next sections we show how it can be extended to directed graphs. We also study how the choice of the Hamiltonian influences the kernel performance in a graph classification task and we investigate if the addition of structural information at node level has any benefit. While Rossi et al. [18] were unable to prove the positive definiteness of the kernel, we are able to provide a proof under the assumption that the Hamiltonian is the graph Laplacian and the starting state satisfies certain conditions. Finally, we propose a fully quantum procedure to compute the kernel.

3. Graph Similarity from Quantum Walks

Given two undirected graphs $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$, we merge them into a larger structure by establishing a complete set of connections between the two node sets. More specifically, we construct a 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 $v \in V_2$. Figure 1 shows an example pair of graphs and their union.
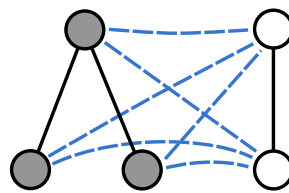
Figure 1. Given two graphs $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ we build a new graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = V_1 \cup V_2$, $\mathcal{E} = E_1 \cup E_2$ and we add a new edge (u,v) between each pair of nodes $u \in V_1$ and $v \in V_2$.

The intuition of Rossi et al. [18] was to define two mixed quantum states on this new graph describing the time-evolution of two quantum walks specifically designed to amplify constructive and destructive interference effects. To this end, we define two independent quantum walks with starting states

$$|\psi_0^-\rangle = \frac{\sum_{u \in V_1} d_u |u\rangle - \sum_{v \in V_2} d_v |v\rangle}{C} \qquad |\psi_0^+\rangle = \frac{\sum_{u \in \mathcal{V}} d_u |u\rangle}{C} \,. \tag{10}$$

We can then associate the following quantum states to the graph,

$$\rho_T^- = \frac{1}{T} \int_0^T |\psi_t^-\rangle \langle \psi_t^-| \, \mathrm{d}t \qquad \rho_T^+ = \frac{1}{T} \int_0^T |\psi_t^+\rangle \langle \psi_t^+| \, \mathrm{d}t \,. \tag{11}$$

Given these two states, we finally make use of the QJSD to compute the similarity between the input graphs in terms of the dissimilarity between the ρ_T^- and ρ_T^+, i.e.,

$$k_T(G_1, G_2) = QJSD(\rho_T^-, \rho_T^+) = S\left(\frac{\rho_T^- + \rho_T^+}{2}\right) - \frac{1}{2}\left(S(\rho_T^-) + S(\rho_T^+)\right). \tag{12}$$

Since the quantum states ρ_T^- and ρ_T^+ represent the evolution of two quantum walks that emphasize destructive and constructive interference, respectively, we expect that the more similar the input graphs are, the more dissimilar ρ_T^- and ρ_T^+ are. In the extreme case where G_1 and G_2 and isomorphic, it can be shown that $k_T(G_1, G_2) = 1$ [18].

In [18], the authors suggest to consider the limit of ρ_T^+ for $T \to \infty$, which in turn allows to rewrite Equation (11) as

$$\rho_\infty^+ = \sum_{\lambda \in \tilde{\Lambda}(\mathcal{H})} P_\lambda \rho_0^+ P_\lambda^\top , \qquad (13)$$

where $\tilde{\Lambda}(\mathcal{H})$ is the set of distinct eigenvalues of the Hamiltonian (H), i.e., the eigenvalues λ with multiplicity $\mu(\lambda) = 1$, and $P_\lambda = \sum_{k=1}^{\mu(\lambda)} \phi_{\lambda,k} \phi_{\lambda,k}^\top$ is the projection operator on the subspace spanned by the $\mu(\lambda)$ eigenvectors $\phi_{\lambda,k}$ associated with λ. Similarly, one can work out the limit of ρ_T^- for $T \to \infty$. This in turn has the effect of making the computation of the kernel easier in addition to allowing us to drop the time parameter T, yielding

$$k_\infty(G_1, G_2) = QJSD(\rho_\infty^-, \rho_\infty^+) = S\left(\frac{\rho_\infty^- + \rho_\infty^+}{2}\right) - \frac{1}{2}\left(S(\rho_\infty^-) + S(\rho_\infty^+)\right). \qquad (14)$$

3.1. Extension to Directed Graphs

The kernel described in the previous subsection can only be computed for undirected graphs. In order to cope with directed graphs, we propose using the directed graph Laplacian introduced by Chung [26]. We first merge the two directed input graphs $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ as done in the undirected case, i.e., we create a complete set of undirected edges connecting the nodes of V_1 to the nodes of V_2. Note, however, that the resulting merge graph is directed, as a consequence of the edges in E_1 and E_2 being directed (see Figure 2). Hence we define the starting states $|\psi_0^-\rangle$ and $|\psi_0^+\rangle$ similarly to Equation (10), but using the out degree instead:

$$|\psi_0^-\rangle = \frac{\sum_{u \in V_1} d_u^{out} |u\rangle - \sum_{v \in V_2} d_v^{out} |v\rangle}{C} \qquad |\psi_0^+\rangle = \frac{\sum_{u \in V} d_u^{out} |u\rangle}{C}. \qquad (15)$$

To compute the Hamiltonian and thus ρ_T^- and ρ_T^+, we proceed as follows. Let us define the classical random walk transition matrix M for the directed graph G as the matrix with elements $M_{uv} = A_{uv}/d_u^{out}$, where we use the convention that when the out-degree of u is zero, i.e., $d_u^{out} = 0$, we let $\mathrm{M}_{uv} = 0$.
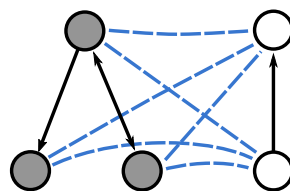

Figure 2. The graph obtained by merging two directed graphs.

From the Perron-Frobenius theorem we know that for a strongly connected directed graph, i.e., a directed graph where there exists a path between every pair of vertices and there are no sinks, the transition matrix M has a unique non-negative left eigenvector π [32], i.e., $\pi M = \lambda \pi$, where λ is the eigenvalue associated with π. According to the Perron-Frobenius theorem, if M is aperiodic we also have that $\lambda = 1$, and thus $\pi M = \pi$, implying that the random walk on the directed graph G with transition matrix M will converge to a stationary probability distribution.

Let Π be the diagonal matrix with the elements of π on the diagonal. Then Chung [26] defines the Laplacian matrix of the directed graph G as

$$L = \Pi - \frac{1}{2}\left(\Pi M + M^\top \Pi\right). \qquad (16)$$

Similarly, the normalised Laplacian of G is defined as

$$\tilde{L} = I - \frac{1}{2}\left(\Pi^{\frac{1}{2}} M \Pi^{-\frac{1}{2}} + \Pi^{-\frac{1}{2}} M^\top \Pi^{\frac{1}{2}}\right), \tag{17}$$

where I denotes the identity matrix.

Please note that while the adjacency matrix of a directed graph is clearly not Hermitian, both the normalised Laplacian and the Laplacian defined in Equations (16) and (17) are symmetric and thus can used as the Hamiltonian governing the quantum walks evolution. An alternative to this would have been to symmetrize the graph edges effectively making the graph undirected, however as we will show in the experimental part this causes the loss of important structural information.

We would like to stress that the method of Chung [26] is not the only way to associate a (normalised) Laplacian matrix to a directed graph. We are aware of at least one different method proposed by Bauer [33]. However we decide to focus on the definition proposed by Chung [26] as this has been successfully applied to the analysis of directed graphs in pattern recognition and network science [34–36].

3.2. Integrating Local Topological Information

In [31] it was shown that for the case of undirected attributed graphs $G_1(V_1, E_1, f_1)$ and $G_2(V_2, E_2, f_2)$, where f_i is a function assigning attributes to the nodes of the graph G_i, it is possible to effectively incorporate the information on the pairwise similarity between the attributes of the two graphs by allowing the adjacency matrix of the merged graph to be weighted. More specifically, for each $u \in V_1$ and $v \in V_2$, the authors proposed to label the edge (u, v) with a real value $\omega(f_1(u), f_2(v))$ representing the similarity between $f_1(u)$ and $f_2(v)$.

While in this paper we only deal with unattributed graphs, we propose using a similar method to incorporate additional node-level structural information into the kernel. More specifically, given an undirected graph $G(V, E)$ and a node $v \in V$, we capture the structure of the graph from the perspective of v using two well known spectral signatures based on the graph Laplacian, the Heat Kernel Signature (HKS) [27] and the Wave Kernel Signature (WKS) [28]. The signatures simulate a heat diffusion and wave propagation process on the graph, respectively. The result is in both cases a series of vectors characterising the graph topology centered at each node. In graph matching, particularly in computer vision applications [27,28], this is used to match pairs of nodes with similar (in an Euclidean sense) vectorial representations under the assumption that these describe structurally similar nodes in the original graph.

Given two nodes $u, v \in \mathcal{V}$, we compute their spectral signatures $f_1(u)$ and $f_1(v)$ using either HKS or WKS and we label the edge (u, v) with the real value

$$\omega(f_1(u), f_2(v)) = ||f_1(u) - f_2(v)||_2. \tag{18}$$

Please note that the signatures are computed on the original graphs, not on the merged one. Then the adjacency matrix of the graph obtained by merging G_1 and G_2 becomes

$$A_{uv} = \begin{cases} \omega(f_1(u), f_2(v)) & \text{if } u, v \in \mathcal{V} \text{ and } u \neq v \\ 0 & \text{otherwise}. \end{cases} \tag{19}$$

Figure 3 shows an example of two graphs with 2-dimensional signatures and their merged graph. For ease of presentation, the weight of each edge is shown by varying its thickness, i.e., the thicker the edge the higher the weight and the more similar the signatures of the connected nodes. Here we decided to compute the similarity between each pair of nodes in $\mathcal{V}$, effectively turning G_1 and G_2 into two weighted cliques, i.e., complete weighted graphs, where the weights on the edges encode the pairwise nodes similarity. While at first sight it may seem like we are discarding the original structure of G_1 and G_2, note that this is actually encoded in the node signatures and thus the edge weights.

In Section 4 we run an extensive set of experiments to determine if integrating node-level structural information improves the ability of the kernel to distinguish between structurally similar graphs.
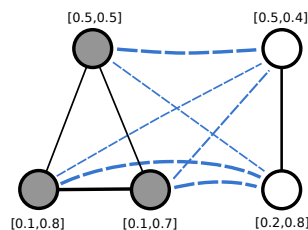
Figure 3. The graph obtained by merging two undirected graphs with 2-dimensional node signatures. The tickness of the edges is proportional to the similarity between the signatures of the nodes being connected.

3.3. Kernel Properties

Consider a pair of graphs G_1 and G_2, with n and m nodes respectively. Here we consider the case where we take the graph Laplacian to be the Hamiltonian of the system and we assume that the input graphs are undirected. Let L_1 and L_2 be the Laplacian of G_1 and G_2 respectively. Then the Laplacian of the merged graph is

$$L = \left[\begin{array}{c|c} L_1 + mI_n & -\mathbf{1}_n\mathbf{1}_m{}^\top \\ \hline -\mathbf{1}_m\mathbf{1}_n{}^\top & L_2 + nI_m \end{array} \right] . \tag{20}$$

We can provide a full characterization of the eigensystem of L. Let $|v\rangle$ and $|w\rangle$ be eigenvectors of L_1 and L_2, respectively, with corresponding eigenvalues v and w. Further, assume that $|v\rangle$ and $|w\rangle$ are not Fiedler vectors, i.e., $\langle \mathbf{1}_n|v\rangle = 0$ and $\langle \mathbf{1}_m|w\rangle = 0$. Then $|v, \mathbf{0}_m\rangle$ is an eigenvector of L with eigenvalue $v + m$ and $|\mathbf{0}_m, w\rangle$ is an eigenvector of L with eigenvalue $w + n$. The remaining eigenvectors of L are $|\mathbf{1}_{m+n}\rangle$ of eigenvalue 0 and $|m\mathbf{1}_n, -n\mathbf{1}_m\rangle$ of eigenvalue $m + n$. In the following, we will denote with P_λ^1 the orthogonal projector on the eigenspace of L_1 of eigenvalue λ, if λ is an eigenvalue of L_1, 0 otherwise. For the special case of the 0 eigenspace, we eliminate the constant eigenvector $|\mathbf{1}_n\rangle$. P_λ^2 is similarly defined for L_2.

Under this eigendecomposition we can write ρ_∞^+ and ρ_∞^- as the following sums of rank-1 matrices:

$$\begin{aligned} \rho_\infty^+ &= \sum_{\lambda \in (\tilde{\Lambda}(L_1)+m) \cup (\tilde{\Lambda}(L_2)+n)} \left| P_{\lambda-m}^1 |\xi_0\rangle, P_{\lambda-n}^2 |\chi_0\rangle \right\rangle \left\langle P_{\lambda-m}^1 |\xi_0\rangle, P_{\lambda-n}^2 |\chi_0\rangle \right| + \\ &\quad \frac{(\langle \mathbf{1}_n|\xi_0\rangle + \langle \mathbf{1}_m|\chi_0\rangle)^2}{(m+n)^2} |\mathbf{1}_{m+n}\rangle \langle \mathbf{1}_{m+n}| + \frac{(\frac{1}{n}\langle \mathbf{1}_n|\xi_0\rangle - \frac{1}{m}\langle \mathbf{1}_m|\chi_0\rangle)^2}{(m+n)^2} |m\mathbf{1}_n, -n\mathbf{1}_m\rangle \langle m\mathbf{1}_n, -n\mathbf{1}_m| \\ &= \sum_{\mu \in (\tilde{\Lambda}(L_1)-n) \cup (\tilde{\Lambda}(L_2)-m)} \left| P_{\mu+n}^1 |\xi_0\rangle, P_{\mu+m}^2 |\chi_0\rangle \right\rangle \left\langle P_{\mu+n}^1 |\xi_0\rangle, P_{\mu+m}^2 |\chi_0\rangle \right| + \\ &\quad \frac{(\langle \mathbf{1}_n|\xi_0\rangle + \langle \mathbf{1}_m|\chi_0\rangle)^2}{(m+n)^2} |\mathbf{1}_{m+n}\rangle \langle \mathbf{1}_{m+n}| + \frac{(\frac{1}{n}\langle \mathbf{1}_n|\xi_0\rangle - \frac{1}{m}\langle \mathbf{1}_m|\chi_0\rangle)^2}{(m+n)^2} |m\mathbf{1}_n, -n\mathbf{1}_m\rangle \langle m\mathbf{1}_n, -n\mathbf{1}_m| , \end{aligned} \tag{21}$$

with the eigenvalues of L being $\lambda = \mu + m + n$. Similarly,

$$\begin{aligned} \rho_\infty^- &= \sum_{\mu \in (\tilde{\Lambda}(L_1)-n) \cup (\tilde{\Lambda}(L_2)-m)} \left| P_{\mu+n}^1 |\xi_0\rangle, -P_{\mu+m}^2 |\chi_0\rangle \right\rangle \left\langle P_{\mu+n}^1 |\xi_0\rangle, -P_{\mu+m}^2 |\chi_0\rangle \right| + \\ &\quad \frac{(\langle \mathbf{1}_n|\xi_0\rangle - \langle \mathbf{1}_m|\chi_0\rangle)^2}{(m+n)^2} |\mathbf{1}_{m+n}\rangle \langle \mathbf{1}_{m+n}| + \frac{(\frac{1}{n}\langle \mathbf{1}_n|\xi_0\rangle + \frac{1}{m}\langle \mathbf{1}_m|\chi_0\rangle)^2}{(m+n)^2} |m\mathbf{1}_n, -n\mathbf{1}_m\rangle \langle m\mathbf{1}_n, -n\mathbf{1}_m| . \end{aligned} \tag{22}$$

Please note that each matrix in the summation is a rank-1 matrix and the matrices are orthogonal to each other as a consequence of the projectors being orthogonal to each others. This in turn implies that the size of the spectra of ρ_∞^+ and ρ_∞^- is the same and it is equal to the number of distinct eigenvalues in $(\tilde{\Lambda}(L_1) - n) \cup (\tilde{\Lambda}(L_2) - m)$. More precisely, for each $\mu \in (\tilde{\Lambda}(L_1) - n) \cup (\tilde{\Lambda}(L_2) - m)$, excluding those derived from the two Fiedler vectors of L_1 and L_2, there exists a non-zero eigenvalue of both ρ_∞^+ and ρ_∞^- of the form

$$\langle \xi_0| P_{\mu+n}^1 |\xi_0\rangle + \langle \chi_0| P_{\mu+m}^2 |\chi_0\rangle .$$

In addition, we have two eigenvalues of the form

$$\frac{(\langle \mathbf{1}_n | \xi_0 \rangle + \langle \mathbf{1}_m | \chi_0 \rangle)^2}{m+n} \qquad \text{and} \qquad mn\frac{(\frac{1}{n}\langle \mathbf{1}_n | \xi_0 \rangle - \frac{1}{m}\langle \mathbf{1}_m | \chi_0 \rangle)^2}{m+n},$$

for L_1, and

$$\frac{(\langle \mathbf{1}_n | \xi_0 \rangle - \langle \mathbf{1}_m | \chi_0 \rangle)^2}{m+n} \qquad \text{and} \qquad mn\frac{(\frac{1}{n}\langle \mathbf{1}_n | \xi_0 \rangle + \frac{1}{m}\langle \mathbf{1}_m | \chi_0 \rangle)^2}{m+n},$$

for L_2.

In other words the two density matrices, despite having different eigenspaces, have identical spectra induced from the eigenspaces orthogonal to the Fiedler vector and differ only in the values due to the extremal eigenvalues of L.

Turning our attention to the arithmetic mean of the two density matrices in Equation (14), we have

$$
\begin{aligned}
\frac{\rho_\infty^+ + \rho_\infty^-}{2} &= \sum_{\mu \in (\tilde{\Lambda}(L_1) - n)} \left| P_{\mu+n}^1 |\xi_0\rangle, \mathbf{0}_m \right\rangle \left\langle P_{\mu+n}^1 |\xi_0\rangle, \mathbf{0}_m \right| + \\
&\quad \sum_{\mu \in (\tilde{\Lambda}(L_2) - m)} \left| \mathbf{0}_n, -P_{\mu+m}^2 |\chi_0\rangle \right\rangle \left\langle \mathbf{0}_n, -P_{\mu+m}^2 |\chi_0\rangle \right| + \\
&\quad \frac{\langle \mathbf{1}_n | \xi_0 \rangle^2 + \langle \mathbf{1}_m | \chi_0 \rangle^2}{(m+n)^2} |\mathbf{1}_{m+n}\rangle \langle \mathbf{1}_{m+n}| + \frac{\left(\frac{\langle \mathbf{1}_n | \xi_0 \rangle}{n}\right)^2 + \left(\frac{\langle \mathbf{1}_m | \chi_0 \rangle}{m}\right)^2}{(m+n)^2} |m\mathbf{1}_n, -n\mathbf{1}_m\rangle \langle m\mathbf{1}_n, -n\mathbf{1}_m|.
\end{aligned}
\tag{23}
$$

Again, this is a summation of rank-1 matrices orthogonal to one-another, resulting in eigenvalues of the form

$$\langle \xi_0 | P_{\mu+n}^1 |\xi_0\rangle \qquad \text{and} \qquad \langle \chi_0 | P_{\mu+m}^2 |\chi_0\rangle,$$

corresponding to the (distinct) eigenvalues of L_1 and L_2, respectively, excluding the eigenspace component induced by the Fiedler vectors, as well as the two eigenvector from the extremal eigenvalues of L

$$\frac{\langle \mathbf{1}_n | \xi_0 \rangle^2 + \langle \mathbf{1}_m | \chi_0 \rangle^2}{m+n} \qquad \text{and} \qquad \frac{\frac{m}{n}\langle \mathbf{1}_n | \xi_0 \rangle^2 + \frac{n}{m}\langle \mathbf{1}_m | \chi_0 \rangle^2}{m+n}.$$

Let us define the function

$$\text{xlog}(x) = -x \log(x). \tag{24}$$

We can now compute the QJSD kernel between the graphs G_1 and G_2 as

$$
\begin{aligned}
k_\infty(G_1, G_2) &= QJSD(\rho_\infty^-, \rho_\infty^+) = S\left(\frac{\rho_\infty^- + \rho_\infty^+}{2}\right) - \frac{1}{2}\left(S(\rho_\infty^-) + S(\rho_\infty^+)\right) \\
&= \sum_{\mu \in (\tilde{\Lambda}(L_1) - n)} \text{xlog}\left(\langle \xi_0 | P_{\mu+n}^1 |\xi_0\rangle\right) + \sum_{\mu \in (\tilde{\Lambda}(L_2) - m)} \text{xlog}\left(\langle \chi_0 | P_{\mu+m}^2 |\chi_0\rangle\right) \\
&\quad - \sum_{\mu \in (\tilde{\Lambda}(L_1) - n) \cup (\tilde{\Lambda}(L_2) - m)} \text{xlog}\left(\langle \xi_0 | P_{\mu+n}^1 |\xi_0\rangle + \langle \chi_0 | P_{\mu+m}^2 |\chi_0\rangle\right) \\
&\quad + \text{xlog}\left(\frac{\langle \mathbf{1}_n | \xi_0 \rangle^2 + \langle \mathbf{1}_m | \chi_0 \rangle^2}{m+n}\right) + \text{xlog}\left(\frac{\frac{m}{n}\langle \mathbf{1}_n | \xi_0 \rangle^2 + \frac{n}{m}\langle \mathbf{1}_m | \chi_0 \rangle^2}{m+n}\right) \\
&\quad - \frac{1}{2}\text{xlog}\left(\frac{(\langle \mathbf{1}_n | \xi_0 \rangle + \langle \mathbf{1}_m | \chi_0 \rangle)^2}{m+n}\right) - \frac{1}{2}\text{xlog}\left(mn\frac{(\frac{1}{n}\langle \mathbf{1}_n | \xi_0 \rangle - \frac{1}{m}\langle \mathbf{1}_m | \chi_0 \rangle)^2}{m+n}\right) \\
&\quad - \frac{1}{2}\text{xlog}\left(\frac{(\langle \mathbf{1}_n | \xi_0 \rangle - \langle \mathbf{1}_m | \chi_0 \rangle)^2}{m+n}\right) - \frac{1}{2}\text{xlog}\left(mn\frac{(\frac{1}{n}\langle \mathbf{1}_n | \xi_0 \rangle + \frac{1}{m}\langle \mathbf{1}_m | \chi_0 \rangle)^2}{m+n}\right).
\end{aligned}
\tag{25}
$$

Assuming that the initial states are normalised in such a way that $\langle \mathbf{1}_n | \xi_0 \rangle = \langle \mathbf{1}_m | \chi_0 \rangle = 0$ and $\langle \xi_0 | \xi_0 \rangle = \langle \chi_0 | \chi_0 \rangle = \frac{1}{2}$, the last terms disappear and we have

$$
\begin{aligned}
k_\infty(G_1, G_2) &= \sum_{\mu \in (\tilde{\Lambda}(L_1) - n)} \mathrm{xlog}\left(\langle \xi_0 | P^1_{\mu+n} | \xi_0 \rangle \right) + \sum_{\mu \in (\tilde{\Lambda}(L_2) - m)} \mathrm{xlog}\left(\langle \chi_0 | P^2_{\mu+m} | \chi_0 \rangle \right) \\
&\quad - \sum_{\mu \in (\tilde{\Lambda}(L_1) - n) \cup (\tilde{\Lambda}(L_2) - m)} \mathrm{xlog}\left(\langle \xi_0 | P^1_{\mu+n} | \xi_0 \rangle + \langle \chi_0 | P^2_{\mu+m} | \chi_0 \rangle \right) \\
&= \tfrac{1}{2} \sum_{\mu \in (\tilde{\Lambda}(L_1) - n)} \mathrm{xlog}\left(2 \langle \xi_0 | P^1_{\mu+n} | \xi_0 \rangle \right) + \tfrac{1}{2} \log(2) + \tfrac{1}{2} \sum_{\mu \in (\tilde{\Lambda}(L_2) - m)} \mathrm{xlog}\left(2 \langle \chi_0 | P^2_{\mu+m} | \chi_0 \rangle \right) \\
&\quad + \tfrac{1}{2}\log(2) - \sum_{\mu \in (\tilde{\Lambda}(L_1) - n) \cup (\tilde{\Lambda}(L_2) - m)} \mathrm{xlog}\left(\frac{2 \langle \xi_0 | P^1_{\mu+n} | \xi_0 \rangle + 2 \langle \chi_0 | P^2_{\mu+m} | \chi_0 \rangle}{2} \right) \\
&= \log(2) - \left[\sum_{\mu \in (\tilde{\Lambda}(L_1) - n) \cup (\tilde{\Lambda}(L_2) - m)} \mathrm{xlog}\left(\frac{2 \langle \xi_0 | P^1_{\mu+n} | \xi_0 \rangle + 2 \langle \chi_0 | P^2_{\mu+m} | \chi_0 \rangle}{2} \right) \right. \\
&\quad \left. - \tfrac{1}{2} \sum_{\mu \in (\tilde{\Lambda}(L_1) - n)} \mathrm{xlog}\left(2 \langle \xi_0 | P^1_{\mu+n} | \xi_0 \rangle \right) - \tfrac{1}{2} \sum_{\mu \in (\tilde{\Lambda}(L_2) - m)} \mathrm{xlog}\left(2 \langle \chi_0 | P^2_{\mu+m} | \chi_0 \rangle \right) \right].
\end{aligned}
\tag{26}
$$

This leads us to the following final observation. Let f_1 and f_2 be two discrete distributions with values in $\tilde{\Lambda}(L_1) - n$ and $\tilde{\Lambda}(L_2) - m$, respectively, defined as

$$
f_1(x) = \sum_{\mu \in (\tilde{\Lambda}(L_1) - n)} \left(2 \langle \xi_0 | P^1_{\mu+n} | \xi_0 \rangle \right) \delta(x - \mu)
\tag{27}
$$

$$
f_2(x) = \sum_{\mu \in (\tilde{\Lambda}(L_2) - m)} \left(2 \langle \chi_0 | P^2_{\mu+m} | \chi_0 \rangle \right) \delta(x - \mu),
\tag{28}
$$

where $\delta(x)$ is the Dirac delta function. Then we have

$$
k_\infty(G_1, G_2) = QJSD(\rho_\infty^-, \rho_\infty^+) = 1 - CJSD(f_1, f_2),
\tag{29}
$$

where $CJSD$ denotes the classical Jensen-Shannon divergence. Now we can prove the following:

Theorem 1. *Using the Laplacian as the Hamiltonian and normalizing $|\xi_0\rangle$ and $|\chi_0\rangle$ such that $\langle \xi_0 | \xi_0 \rangle = \langle \xi_0 | \xi_0 \rangle = \frac{1}{2}$ and $\langle \mathbf{1}_n | \xi_0 \rangle = \langle \mathbf{1}_m | \xi_0 \rangle = 0$, then the quantum Jensen-Shannon kernel $k_\infty(G_1, G_2) = QJSD(\rho_\infty^-, \rho_\infty^+)$ is positive definite.*

Proof. The proof descends directly from the derivations above and from the fact that the classical Jensen-Shannon divergence is conditionally negative-definite [37]. $\square$

3.4. Kernel Computation

The issue of efficiently computing the QJSD kernel is that of computing the entropies of ρ_∞^-, ρ_∞^+, and of their mean. Here we will discuss a general quantum approach to estimate the entropy of an infinite-time average mixing matrix starting from a pure state. Following [18], we write the infinite-time average mixing matrix as

$$
\rho_\infty = \sum_{\lambda \in \tilde{\Lambda}(\mathcal{H})} P_\lambda \rho_0 P_\lambda,
\tag{30}
$$

where $\mathcal{H}$ is the Hamiltonian and P_λ is the projector onto its λ eigenspace. If $\rho_0 = |\phi\rangle \langle \phi|$ is a pure state, then the eigenvectors corresponding to non-zero eigenvalues of ρ_∞ are of the form

$$
P_\lambda | \phi \rangle,
\tag{31}
$$

for $\lambda \in \tilde{\Lambda}(\mathcal{H})$, and the corresponding eigenvalues are

$$
\langle \phi | P_\lambda | \phi \rangle.
\tag{32}
$$

We already used this property in the previous section for the special case of the graph Laplacian as the Hamiltonian, but it holds in general.

It is important to note that the lambdas here are energy levels of the system, i.e., possible observed values using the Hamiltonian as an observable, and the probability of observing an energy state λ from the state $\rho_0 = |\phi\rangle \langle\phi|$ is $\langle\phi| P_\lambda |\phi\rangle$. This means that the von Neumann entropy of ρ_∞ is equivalent to the classic Shannon entropy of the energy levels of the (pure) initial state, i.e., the entropy of the possible observations from the pure initial state using the Hamiltonian as the observable.

The quantum algorithm to estimate the von Neumann entropy of the infinite-time average mixing matrix $\rho_\infty = |\phi\rangle \langle\phi|$ is as follows. First, we create several systems in the initial pure state $|\phi\rangle$ and we observe their energy using $\mathcal{H}$ as the observable. Then we use any entropy estimator for finite (discrete) distributions, using the energies as samples. Algorithms such as [38] or [39] are particularly interesting because they work well in the under-sampled case.

In the special case considered in Theorem 1, we can make the estimation even more efficient noting that we can observe independently from $L_1 - nI_n$ and $L_2 - mI_m$, and in particular we can perform a single set of observation for each graph and reuse them in every pairwise computation for the kernel. In this case, the samples from the first graph are used to estimate the entropy $S(f_1)$, the samples from the second graph are used to estimate $S(f_2)$, and the two sample-sets are merged to estimate $S\left(\frac{f_1+f_2}{2}\right)$.

4. Experiments

In this section, we evaluate the accuracy of the proposed kernel in several classification tasks. Our aim is to investigate (1) the importance of the choice of the Hamiltonian; (2) the integration of node-level structural signatures; and (3) the ability to incorporate edge directionality information using the Laplacians proposed by Chung [26]. To this end, we make use of the following datasets of undirected graphs (see Table 1):

MUTAG [40] is a dataset consisting originally of 230 chemical compounds tested for mutagenicity in Salmonella typhimurium [41]. Among the 230 compounds, however, only 188 (125 positive, 63 negative) are considered to be learnable and thus are used in our simulations. The 188 chemical compounds are represented by graphs. The aim is predicting whether each compound possesses mutagenicity.

PPIs (Protein-Protein Interaction) is a dataset collecting protein-protein interaction networks related to histidine kinase [42] (40 PPIs from Acidovorax avenae and 46 PPIs from Acidobacteria) [43]. The graphs describe the interaction relationships between histidine kinase in different species of bacteria. Histidine kinase is a key protein in the development of signal transduction. If two proteins have direct (physical) or indirect (functional) association, they are connected by an edge. The original dataset comprises 219 PPIs from 5 different kinds of bacteria with the following evolution order (from older to more recent): Aquifex 4 and Thermotoga 4 PPIs from Aquifex aelicus and Thermotoga maritima, Gram-Positive 52 PPIs from Staphylococcus aureus, Cyanobacteria 73 PPIs from Anabaena variabilis and Proteobacteria 40 PPIs from Acidovorax avenae. There is an additional class (Acidobacteria 46 PPIs) which is more controversial in terms of the bacterial evolution since they were discovered.

Table 1. Information on the graph datasets.

Datasets	MUTAG	PPI	PTC	COIL	NCI1	SHOCK	ALZ
Max # vertices	28	232	109	241	106	33	96
Min # vertices	10	3	2	72	3	4	96
Avg # vertices	17.93	109.60	25.56	144.97	29.27	13.16	96
# graphs	188	86	344	360	3530	150	149
# classes	2	2	2	5	2	10	4

PTC (Predictive Toxicology Challenge) dataset records the carcinogenicity of several hundred chemical compounds for Male Rats (MR), Female Rats (FR), Male Mice (MM) and Female Mice (FM) [44]. These graphs are very small and sparse. We select the graphs of Male Rats (MR) for evaluation. There are 344 test graphs in the MR class.

COIL Columbia Object Image Library consists of 3D objects images of 100 objects [45]. There are 72 images per object taken in order to obtain 72 views from equally spaced viewing directions. For each view a graph was built by triangulating the extracted Harris corner points. In our experiments, we use the gray-scale images of five objects.

NCI1 The anti-cancer activity prediction dataset consists of undirected graphs representing chemical compounds screened for activity against non-small cell lung cancer lines [46]. Here we use only the connected graphs in the dataset (3530 out of 4110).

We also make use of the following directed graphs datasets:

Shock The Shock dataset consists of graphs from a database of 2D shapes [47]. Each graph is a medial axis-based representation of the differential structure of the boundary of a 2D shape. There are 150 graphs divided into 10 classes, each containing 15 graphs. The original version contains directed trees each with a root node, the undirected version has been created by removing the directionality.

Alzheimer The dataset is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [48] and concerns interregional connectivity structure for functional magnetic resonance imaging (fMRI) activation networks for normal and Alzheimer subjects. Each image volume is acquired every two seconds with blood oxygenation level dependent signals (BOLD). The fMRI voxels here have been aggregated into larger regions of interest (ROIs). The different ROIs correspond to different anatomical regions of the brain and are assigned anatomical labels to distinguish them. There are 96 anatomical regions in each fMRI image. The correlation between the average time series in different ROIs represents the degree of functional connectivity between regions which are driven by neural activities [49]. Subjects fall into four categories according to their degree of disease severity: AD—full Alzheimer's (30 subjects), LMCI—Late Mild Cognitive Impairment (34 subjects), EMCI—Early Mild Cognitive Impairment (47 subjects), HC—Normal Healthy Controls (38 subjects). The LMCI subjects are more severely affected and close to full Alzheimerś, while the EMCI subjects are closer to the healthy control group (Normal). A directed graph with 96 nodes is constructed for each patient based on the magnitude of the correlation and the sign of the time-lag between the time-series for different anatomical regions. To model causal interaction among ROIs, the directed graph uses the time lagged cross-correlation coefficients for the average time series for pairs of ROIs. We detect directed edges by finding the time-lag that results in the maximum value of the cross-correlation coefficient. The direction of the edge depends on whether the time lag is positive or negative. We then apply a threshold to the maximum values to retain directed edges with the top 40% of correlation coefficients. This yields a binary directed adjacency matrix for each subject, where the diagonal elements are set to zero. Those ROIs which have missing time series data are discarded. In order to fairly evaluate the influence caused by edges directionality, an undirected copy has been created as well. In particular, let A_d be the adjacency matrix of a directed graph. Then its projection over the symmetric matrices space will be given by $(A_d + A_d^\top)/2$.

4.1. Experimental Framework

With these datasets to hand, we aim to solve a graph classification task using as binary C-Support Vector Machine (C-SVM) [19]. 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, and the results are reported in terms of average accuracy $\pm$ standard error. Recall that in classification tasks, the accuracy is intended as the fraction of the data occurrences that are assigned the correct class label. In our case, an occurrence is a graph. Moreover, we contrast the performance of the kernel with that of

other well-established alternative graph kernels, namely the shortest-path kernel [15] (*SP*), the classic random walk kernel [9] (*RW*) and the the Weisfeiler-Lehman subtree kernel [16] (*WL*).

We divide our experiments into two parts. We first evaluate the influence of the Hamiltonian on the classification accuracy, where the Hamiltonian is chosen to be one between (1) the adjacency matrix ($QJSD_A$); (2) the Laplacian ($QJSD_L$); or (3) the normalised Laplacian of the graph ($QJSD_{NL}$). Here we also consider the positive definite version of the kernel discussed in Section 3.3 ($QJSD_*$). In the same experiment, we test if the addition of node signatures benefits the kernel performance, where the signature is either the (1) HKS or the (2) WKS, denoted as ^{hk} and ^{wk} respectively. In the second experiment we test the performance of our kernel for the classification of directed graphs. Please note that all the alternative kernels we consider are also able to cope with directed graphs. Here we test both our competitiveness against these kernels as well as the increase or decrease in performance compared to the case where we remove the edge directionality information.

4.2. Undirected Graphs

Table 2 shows the results of the classification experiments on undirected graphs. The best and second best performing kernels are highlighted in bold and italic, respectively. While the kernels based on the quantum Jensen-Shannon divergence yield the highest accuracy in 4 out of 5 datasets, it is clear that no specific combination of Hamiltonian and structural signatures is optimal in all cases. However, we observe that with the exception of the PPI dataset, both the choice of the Hamiltonian and the use or not of structural signatures has little or no impact on the performance of the kernel, which remains highly competitive when compared to the shortest-path, random walks, graphlet, and Weisfeiler-Lehman kernels. The performance gain is particularly evident when compared to the kernel based on classical random walks, which is consistently outperformed by the kernels based on quantum walks. This in turn is a confirmation of the increased ability of capturing subtle structural similarities of quantum walks with respect to their classical counterpart, as well as a clear statement of the robustness of the proposed kernel to the choice of its parameters. We do observe a tendency of the Laplacian + WKS combination to perform better on datasets with average graph size >100 (PPI and COIL), however this would need more extensive experiments to be validated.

Table 2. Classification accuracy (±standard error) on undirected graph datasets. The best and second best performing kernels are highlighted in bold and italic, respectively.

Kernel	MUTAG	PPI	PTC	NCI1	COIL
$QJSD_A$	86.72 ± 0.14	78.71 ± 0.30	56.09 ± 0.15	66.90 ± 0.03	69.90 ± 0.08
$QJSD_L$	84.92 ± 0.18	73.79 ± 0.42	$\mathbf{59.70 \pm 0.16}$	$\mathit{69.48 \pm 0.03}$	$\mathit{70.72 \pm 0.07}$
$QJSD_{NL}$	87.10 ± 0.14	74.63 ± 0.37	55.16 ± 0.18	66.36 ± 0.03	69.61 ± 0.10
$QJSD_A^{hk}$	$\mathbf{88.51 \pm 0.13}$	$\mathit{81.56 \pm 0.34}$	58.76 ± 0.14	63.88 ± 0.03	69.68 ± 0.06
$QJSD_L^{hk}$	86.36 ± 0.16	77.08 ± 0.28	57.63 ± 0.13	64.96 ± 0.02	70.24 ± 0.06
$QJSD_{NL}^{hk}$	$\mathit{87.79 \pm 0.12}$	74.38 ± 0.37	58.45 ± 0.16	64.01 ± 0.04	70.55 ± 0.09
$QJSD_A^{wk}$	85.97 ± 0.14	74.91 ± 0.32	$\mathit{58.91 \pm 0.13}$	63.52 ± 0.05	70.48 ± 0.06
$QJSD_L^{wk}$	85.81 ± 0.16	$\mathbf{84.66 \pm 0.26}$	58.01 ± 0.13	64.45 ± 0.03	$\mathbf{71.34 \pm 0.05}$
$QJSD_{NL}^{wk}$	87.61 ± 0.16	74.74 ± 0.30	57.46 ± 0.16	63.34 ± 0.04	70.31 ± 0.06
$QJSD_*$	75.47 ± 0.18	56.50 ± 0.41	58.15 ± 0.12	62.89 ± 0.02	13.51 ± 0.16
$QJSD_*^{hk}$	70.99 ± 0.12	70.53 ± 0.31	55.29 ± 0.13	67.19 ± 0.01	34.24 ± 0.22
$QJSD_*^{wk}$	79.44 ± 0.16	70.21 ± 0.40	57.65 ± 0.14	64.20 ± 0.02	58.63 ± 0.12
SP	84.98 ± 0.16	66.40 ± 0.31	56.89 ± 0.71	65.44 ± 0.04	70.50 ± 0.13
RW	78.02 ± 0.20	69.94 ± 0.27	55.59 ± 0.01	58.80 ± 0.04	21.03 ± 0.22
GR	81.93 ± 0.17	52.34 ± 0.42	56.20 ± 0.10	62.28 ± 0.02	67.22 ± 0.11
WL	84.62 ± 0.23	79.93 ± 0.35	55.64 ± 0.20	$\mathbf{78.55 \pm 0.04}$	31.33 ± 0.21

We would like to stress that there are indeed alternative strategies to the one proposed in Section 3.2 to integrate the node signatures information. For example, instead of creating a pair of weighted cliques, one could retain the original graphs structure and compute the similarity only

between nodes connected by edges in E_1, E_2, and E_{12}. Another approach could have been to restrict the weighting only to the edges in E_{12}, similarly to what done by Rossi et al. in [31]. However we found little or no difference in terms of classification accuracy between these alternative solutions and thus we decided to omit these results for ease of presentation.

Finally, the performance of the positive definite kernel $QJSD_*$ deserves a closer inspection. In contrast to the other quantum-walk based kernels, in this case the performance is often significantly lower and the dependency on the addition of the node-level structural signatures higher. Unlike the other kernels, in this case we had to enforce several constraints on the initial state in order to guarantee the positive definiteness of the resulting kernel (see Section 3.3). However one of this constraints, i.e., the requirement that the two parts of the initial amplitude vector corresponding to G_1 and G_2 both sum to zero, causes the phase of the initial state to differ from that of the other kernels. While the initial states defined in Equation (10) are such that $\left|\psi_0^+\right\rangle$ has equal phase on both G_1 and G_2, while $\left|\psi_0^-\right\rangle$ has opposite phase on G_1 and G_2, enforcing the constraint of Section 3.3 leads to a change of sign (and thus of phase) for the smallest (in absolute value) components of the initial state. This in turn invalidates the assumption that ρ_∞^+ and ρ_∞^- are designed to stress constructive and destructive interference, respectively, and thus affects the ability of the kernel to distinguish between similar/dissimilar graphs.

4.3. Directed Graphs

Table 3 shows the results of the classification experiments on directed graphs. We compare the kernels on the two directed graphs dataset, as well as their undirected versions, and we highlight again the best performing kernel for each dataset. Please note that when the graphs are directed we cannot choose the (potentially asymmetric) adjacency matrix as the Hamiltonian. We first observe that preserving the edge directionality information leads to a higher or similar classification accuracy for most kernels, with the exception of the shortest-path kernel on the Shock dataset. Please note that the graphs in this dataset are trees, and as such in the directed version there may be pairs of nodes for which a shortest-path does not exist, leading to the observed decrease in classification performance. In general, we observe that both in the directed and undirected case our kernel is the best performing, with the normalised Laplacian being the Hamiltonian that leads to the highest classification accuracy in 3 out of 4 datasets.

Table 3. Classification accuracy ($\pm$standard error) on directed graph datasets, where $_d$ and $_u$ denote the directed and undirected versions of the datasets, respectively. The best and second best performing kernels are highlighted in bold and italic, respectively.

Kernel	ALZ_d	ALZ_u	$SHOCK_d$	$SHOCK_u$
$QJSD_A$	-	**65.87 $\pm$ 0.25**	-	41.48 $\pm$ 0.15
$QJSD_L$	79.26 $\pm$ 0.24	60.42 $\pm$ 0.23	45.89 $\pm$ 0.23	35.77 $\pm$ 0.21
$QJSD_{NL}$	**82.07 $\pm$ 0.17**	61.45 $\pm$ 0.22	**46.05 $\pm$ 0.20**	**44.38 $\pm$ 0.21**
SP	59.86 $\pm$ 0.25	58.00 $\pm$ 0.29	22.09 $\pm$ 0.29	40.16 $\pm$ 0.24
RW	79.06 $\pm$ 0.21	60.75 $\pm$ 0.25	30.45 $\pm$ 0.26	24.34 $\pm$ 0.28
GR	79.00 $\pm$ 0.20	64.34 $\pm$ 0.27	28.91 $\pm$ 0.30	29.39 $\pm$ 0.28
WL	70.87 $\pm$ 0.27	59.46 $\pm$ 0.35	38.74 $\pm$ 0.27	35.78 $\pm$ 0.26

5. Conclusions

We studied the use of continuous-time quantum walks and the quantum Jensen-Shannon divergence for the purpose of measuring the similarity between two graphs. We performed a thorough analysis of the similarity measure originally introduced in [18], showing that: (1) the kernel can be extended to deal with directed graphs, allowing to better discriminate between the different classes, even when compared with other commonly used kernels for directed graphs; (2) while it is possible to incorporate additional node-level topological information in the kernel, this results in a significant improvement over the performance of the original kernel; (3) the optimal Hamiltonian (in terms of classification accuracy) depends on the dataset; (4) when appropriate conditions are met, the kernel

is proved to be positive-definite, however (5) its performance suffers from a loss of structure in the design of the initial state phase; (6) it is possible to efficiently compute the kernel using a fully quantum procedure.

Despite having made important progress toward a better understanding of the quantum Jensen-Shannon divergence kernel, several questions remain unanswered. As such, future work should aim at understanding the role played by the initial state in the performance of the kernel, as well as considering alternative divergence measures and different ways of encoding the edge directionality information. Another idea could be to investigate if we can link the similarity between two graphs to the quantumness [50] of a quantum walk on their merged graph, based on the observation that isomorphic graphs result in a highly symmetric merged graph where strong interference effects can lead the quantum walk to behave significantly differently from its classical counterpart.

References

1. Kadowaki, T.; Nishimori, H. Quantum annealing in the transverse Ising model. *Phys. Rev. E* **1998**, *58*, 5355. [CrossRef]
2. Boixo, S.; Rønnow, T.F.; Isakov, S.V.; Wang, Z.; Wecker, D.; Lidar, D.A.; Martinis, J.M.; Troyer, M. Evidence for quantum annealing with more than one hundred qubits. *Nat. Phys.* **2014**, *10*, 218–224. [CrossRef]
3. Grover, L.K. A fast quantum mechanical algorithm for database search. In Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, Philadelphia, PA, USA, 22–24 May 1996; ACM: New York, NY, USA, 1996; pp. 212–219.
4. Dong, D.; Chen, C.; Li, H.; Tarn, T. Quantum reinforcement learning. *IEEE Trans. Syst. Man Cybern. Part B* **2008**, *1220*, 38–1207. [CrossRef]
5. Kempe, J. Quantum random walks: An introductory overview. *Contemp. Phys.* **2003**, *44*, 327–307. [CrossRef]
6. Ambainis, A. Quantum walks and their algorithmic applications. *Int. J. Quantum Inf.* **2003**, *1*, 507–518. [CrossRef]
7. Childs, A.M. Universal computation by quantum walk. *Phys. Rev. Lett.* **2009**, *102*, 180501. [CrossRef] [PubMed]
8. Page, L.; Brin, S.; Motwani, R.; Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*; Technical report; Stanford InfoLab: Stanford, CA, USA, 1999.
9. Gärtner, T.; Flach, P.; Wrobel, S. On graph kernels: Hardness results and efficient alternatives. *Learn. Theory Kernel Mach.* **2003**, *277*, 129–143.
10. Qiu, H.; Hancock, E.R. Clustering and embedding using commute times. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1873–1890. [CrossRef] [PubMed]
11. Estrada, E.; Hatano, N. Communicability in complex networks. *Phys. Rev. E* **2008**, *77*, 036111. [CrossRef]
12. Lu, B.; Hancock, E. Graph kernels from the Jensen-Shannon divergence. *J. Math. Imaging Vis.* **2013**, *47*, 60–69.
13. Xiao, B.; Hancock, E.R.; Wilson, R.C. Graph characteristics from the heat kernel trace. *Pattern Recognit.* **2009**, *42*, 2589–2606. [CrossRef]
14. Aziz, F.; Wilson, R.C.; Hancock, E.R. Backtrackless walks on a graph. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *24*, 977–989. [CrossRef] [PubMed]
15. Borgwardt, K.M.; Kriegel, H. Shortest-path kernels on graphs. In Proceedings of the Fifth IEEE International Conference on Data Mining, IEEE Computer Society, Washington, DC, USA, 27–30 November 2005; pp. 74–81.
16. Shervashidze, N.; Schweitzer, P.; jan van Leeuwen, E.; Mehlhorn, K.; Borgwardt, K.M. Weisfeiler-lehman graph kernels. *J. Mach. Learn. Res.* **2011**, *12*, 2539–2561.
17. Bai, L.; Rossi, L.; Torsello, A.; Hancock, E.R. A quantum jensen—Shannon graph kernel for unattributed graphs. *Pattern Recognit.* **2015**, *48*, 344–355. [CrossRef]

18. Rossi, L.; Torsello, A.; Hancock, E.R. Measuring graph similarity through continuous-time quantum walks and the quantum Jensen-Shannon divergence. *Phys. Rev. E* **2015**, *91*, 022815. [CrossRef]

19. Chang, C.; Chih-Jen, L. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27. [CrossRef]

20. Shervashidze, N.; Vishwanathan, S.V.N.; Petri, T.; Mehlhorn, K.; Borgwardt, K.M. Efficient graphlet kernels for large graph comparison. *Artif. Intell. Stat.* **2009**, *5*, 488–495.

21. Martins, A.F.T.; Smith, N.A.; Xing, E.P.; Aguiar, P.M.Q.; Mário, A.T. Nonextensive information theoretic kernels on measures. *J. Mach. Learn. Res.* **2009**, *10*, 935–975.

22. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151. [CrossRef]

23. Majtey, A.P.; Lamberti, P.W.; Prato, D.P. Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states. *Phys. Rev. A* **2005**, *72*, 052310. [CrossRef]

24. Lamberti, P.W.; Majtey, A.P.; Borras, A.; Casas, M.; Plastino, A. Metric character of the quantum Jensen-Shannon divergence. *Phys. Rev. A* **2008**, *77*, 052311. [CrossRef]

25. Rossi, L.; Torsello, A.; Hancock, E.R.; Wilson, R.C. Characterizing graph symmetries through quantum jensen-shannon divergence. *Phys. Rev. E* **2013**, *88*, 032806. [CrossRef] [PubMed]

26. Chung, F. Laplacians and the Cheeger inequality for directed graphs. *Ann. Comb.* **2005**, *9*, 1–19. [CrossRef]

27. Sun, J.; Ovsjanikov, M.; Guibas, L. A concise and provably informative multi-scale signature based on heat diffusion. *Comput. Graph. Forum* **2009**, *28*, 1383–1392. [CrossRef]

28. Aubry, M.; Schlickewei, U.; Cremers, D. The wave kernel signature: A quantum mechanical approach to shape analysis. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Barcelona, Spain, 6–13 November 2011; pp. 1626–1633.

29. Jürgen, J. *Riemannian Geometry and Geometric Analysis*; Springer: Belin, Germany, 2008.

30. Nielsen, M.A.; Chuang, I.L. *Quantum Computation and Quantum Information*; Cambridge University Press: Cambridge, UK, 2010.

31. Rossi, L.; Torsello, A.; Hancock, E.R. Attributed graph similarity from the quantum Jensen-Shannon divergence. In Proceedings of the International Workshop on Similarity-Based Pattern Recognition, York, UK, 3–5 July 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 204–218.

32. Horn, R.A.; Johnson, C.R. *Matrix Analysis*; Cambridge University Press: Cambridge, UK, 1990.

33. Bauer, F. Normalized graph Laplacians for directed graphs. *Linear Algebra Its Appl.* **2012**, *436*, 4193–4222. [CrossRef]

34. Ye, C.; Wilson, R.C.; Comin, C.H.; Costa, L.D.; Hancock, E.R. Approximate von Neumann entropy for directed graphs. *Phys. Rev. E* **2014**, *89*, 052804. [CrossRef] [PubMed]

35. Ye, C.; Wilson, R.; Hancock, E.R. Network analysis using entropy component analysis. *J. Complex Netw.* **2017**, *6*, 404–429. [CrossRef]

36. Ye, C.; Wilson, R.; Rossi, L.; Torsello, A.; Hancock, E. Thermodynamic Analysis of Time Evolving Networks. *Entropy* **2018**, *20*, 759. [CrossRef]

37. Briët, J.; Harremoës, P. Properties of classical and quantum Jensen-Shannon divergence. *Phys. Rev. A* **2009**, *79*, 052311. [CrossRef]

38. Hausser, J.; Strimmer, K. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *J. Mach. Learn. Res.* **2009**, *10*, 1469–1484.

39. Nemenman, I.; Shafee, F.; Bialek, W. Entropy and inference, revisited. *Adv. Neural Inf. Process. Syst.* **2002**, *14*, 471–478.

40. Debnath, A.K.; de Compadre, R.L.L.; Debnath, G.; Shusterman, A.J.; Hansch, C. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. Correlation with molecular orbital energies and hydrophobicity. *J. Med. Chem.* **1991**, *34*, 797–786. [CrossRef]

41. Ralaivola, L.; Swamidass, S.J.; Saigo, H.; Baldi, P. Graph kernels for chemical informatics. *Neural Netw.* **2005**, *8*, 786–797. [CrossRef] [PubMed]

42. Jensen, L.J.; Kuhn, M.; Stark, M.; Chaffron, S.; Creevey, C.; Muller, J.; Doerks, T.; Julien, P.; Roth, A.; Simonovic, M.; et al. STRING 8—A global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **2008**, *37* (Suppl. 1), D412–D416. [CrossRef] [PubMed]

43. Escolano, F.; Hancock, E.R.; Lozano, M.A. Heat diffusion: Thermodynamic depth complexity of networks. *Phys. Rev. E* **2012**, *85*, 036206. [CrossRef]

44. Li, G.; Semerci, M.; Yener, B.; Zaki, M.J. Effective Graph Classification Based on Topological and Label Attributes. *Stat. Anal. Data Min.* **2012**, *5*, 265–283. [CrossRef]

45. Nayar, S.K.; Nene, S.A.; Murase, H. *Columbia Object Image Library (Coil 100)*; Tech. Rep. CUCS-006-96; Department of Comp. Science, Columbia University: New York, NY, USA, 1996.

46. Wale, N.; Karypis, G. *Acyclic Subgraph Based Descriptor Spaces for Chemical Compound Retrieval and Classification*; Minnesota Univ Minneapolis Dept of Computer Science: Minneapolis, MN, USA, 2006.

47. Torsello, A.; Hancock, E. Correcting Curvature-Density Effects in the Hamilton-Jacobi Skeleton. *IEEE Trans. Image Process.* **2006**, *15*, 877–891. [CrossRef]

48. Alzheimer's Disease Neuroimaging Initiative. Available online: http://adni.loni.usc.edu/ (accessed on 26 September 2018).

49. Wang, J.; Wilson, R.C.; Hancock, E.R. fMRI activation network analysis using bose-einstein entropy. In Proceedings of the Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR), Mérida, Mexico, 29 November– 2 December 2016; pp. 218–228.

50. Shahbeigi, F.; Akhtarshenas, S.J.; Rezakhani, A.T. How Quantum is a "Quantum Walk"? *arXiv* **2018**, arXiv:1802.07027.