Attributed Graph Similarity from the Quantum Jensen-Shannon Divergence

Luca Rossi[1], Andrea Torsello[1], and Edwin R. Hancock[2]

[1] Department of Environmental Science, Informatics and Statistics,
Ca' Foscari University of Venice, Italy
{lurossi,torsello}@dsi.unive.it
[2] Department of Computer Science, University of York, YO10 5GH, UK
edwin.hancock@york.ac.uk

Abstract. One of the most fundamental problem that we face in the graph domain is that of establishing the similarity, or alternatively the distance, between graphs. In this paper, we address the problem of measuring the similarity between attributed graphs. In particular, we propose a novel way to measure the similarity through the evolution of a continuous-time quantum walk. Given a pair of graphs, we create a derived structure whose degree of symmetry is maximum when the original graphs are isomorphic, and where a subset of the edges is labeled with the similarity between the respective nodes. With this compositional structure to hand, we compute the density operators of the quantum systems representing the evolution of two suitably defined quantum walks. We define the similarity between the two original graphs as the quantum Jensen-Shannon divergence between these two density operators, and then we show how to build a novel kernel on attributed graphs based on the proposed similarity measure. We perform an extensive experimental evaluation both on synthetic and real-world data, which shows the effectiveness the proposed approach.

Keywords: Graph Similarity, Graph Kernel, Continuous-Time Quantum Walk, Quantum Jensen-Shannon Divergence.

1 Introduction

Graph-based representations have become increasingly popular due to their ability to characterize in a natural way a large number of systems which are best described in terms of their structure. Concrete examples include the use of graphs to represent shapes [1], metabolic networks [2], protein structure [3], and road maps [4]. However, the rich expressiveness and versatility of graphs comes at a cost. In fact, our ability to analyse data abstracted in terms of graphs is severely limited by the restrictions posed by standard pattern recognition techniques, which usually require the graphs to be first embedded into a vectorial space, a procedure which is far from being trivial. The reason for this is that there is no canonical ordering for the nodes in a graph and a correspondence order

must be established before analysis can commence. Moreover, even if a correspondence order can be established, graphs do not necessarily map to vectors of fixed length, as the number of nodes and edges can vary.

One of the most fundamental problem that we need to face in the graph domain is that of measuring the similarity, or alternatively the distance, between graphs. Generally, the similarity between two graphs can be defined in terms of the lowest cost sequence of edit operations, for example, the deletion, insertion and substitution of nodes and edges, which are required to transform one graph into the other [5]. Another approach is that of Barrow and Burstall [6], where the similarity of two graphs is characterized using the cardinality of their maximum common subgraphs. Similarly, Bunke and Shearer [7] introduced a metric on unattributed graphs based on the maximum common subgraph, which later Hidović and Pelillo extended to the case of attributed graphs [8]. Unfortunately, both computing the graph edit distance and finding the maximum common subgraphs turn out to be a computationally hard problem.

Closely related to this problem is that of defining a kernel [9] over graphs. Graph kernels are powerful tools that allow the researcher to overcome the restrictions posed by standard pattern recognition techniques by shifting the problem from that of finding an embedding of a graph to that of defining a positive semidefinite kernel, via the well-known kernel trick. In fact, once we define a positive semidefinite kernel $k : X \times X \to \mathbb{R}$ on a set X, then we know that 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 on which we have defined our kernel. However, due to the rich expressiveness of graphs, the problem of defining effective graph kernels has proven to be extremely difficult.

Many different graph kernels have been proposed in the literature [10,11,12]. Graph kernels are generally instances of the family of R-convolution kernels introduced by Haussler [13]. The fundamental idea is that of defining a kernel between two discrete objects by decomposing them and comparing some simpler substructures. For example, Gärtner et al. [10] propose to count the number of common random walks between two graphs, while Borgwardt and Kriegel [11] measure the similarity based on the shortest paths in the graphs. Shervashidze et al. [12], on the other hand, count the number of graphlets, i.e. subgraphs with k nodes. These kernels can be generally defined both on unattributed and attributed graphs, where in the attributed case one simply enumerates the number of substructures which share the same sequence of labels.

In this paper, we introduce a novel similarity measure between attributed graphs which is based on the evolution of a continuous-time quantum walk [14]. In particular, we are taking advantage of the fact that the interference effects introduced by the quantum walk seem to be enhanced by the presence of symmetrical motifs in the graph [15,16]. Thus, given a pair of graphs, we create a derived structure whose degree of symmetry is maximum when the original graphs are isomorphic. To encode the information on the node attributes, in the new structure we will label the edges connecting one graph to the other with the

value of the similarity between the corresponding nodes. With this structure to hand, we will define two continuous-time quantum walks which have orthogonal density operators under the evolution of the walk whenever the two original graphs are isomorphic. Then, to define the similarity measure we make use of the quantum Jensen-Shannon divergence, a measure which has recently been introduced as a means to compute the distance between quantum states [17]. Finally, we use the proposed similarity measure to define a novel kernel for attributed graphs.

The remainder of this paper is organized as follows: Section 2 provides an essential introduction to the basic terminology required to understand the proposed quantum mechanical framework. In Section 3 we introduce our similarity measure and we define a novel attributed graph kernel. Section 4 illustrates the experimental results, and the conclusions are presented in Section 5.

2   Quantum Mechanical Background

Quantum walks are the quantum analogue of classical random walks [14]. In this paper we consider only continuous-time quantum 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 (i.e. a matrix whose columns sum to unity), in the quantum case the dynamics of the walker is governed by a complex unitary matrix i.e., a matrix that multiplied by its conjugate transpose yields the identity matrix. Hence, the evolution of the quantum walk is reversible, which implies that quantum walks are non-ergodic and do not possess a limiting distribution. Using Dirac notation, we denote the basis state corresponding to the walk being at vertex $u \in V$ as $|u\rangle$. A general state of the walk is a complex linear combination of the basis states, such that the state of the walk at time t is defined as

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

where the amplitude $\alpha_u(t) \in \mathbb{C}$ and $|\psi_t\rangle \in \mathbb{C}^{|V|}$ are both complex.

At each point in time the probability of the walker being at a particular vertex of the graph is given by the square of the norm of the amplitude of the relative state. More formally, let X^t be a random variable giving the location of the walker at time t. Then the probability of the walker being at the vertex u at time t is given by

$$\Pr(X^t = u) = \alpha_u(t)\alpha_u^*(t) \tag{2}$$

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

Recall that the adjacency matrix of the graph G has elements

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

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

$$\frac{d}{dt} |\psi_t\rangle = -iA |\psi_t\rangle \tag{4}$$

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

$$|\psi_t\rangle = e^{-iAt} |\psi_0\rangle \tag{5}$$

With the graph adjacency matrix to hand, we can compute its spectral decomposition $A = \Phi\Lambda\Phi^\top$, where Φ is the $n \times n$ matrix $\Phi = (\phi_1|\phi_2|...|\phi_n)$ with the ordered eigenvectors as columns and $\Lambda = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$ is the $n \times n$ diagonal matrix with the ordered eigenvalues as elements, such that $0 = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$. Using this spectral decomposition and the fact that $e^{-iAt} = \Phi e^{-i\Lambda t}\Phi^\top$ we can finally re-write Eq. 5 as

$$|\psi_t\rangle = \Phi e^{-i\Lambda t}\Phi^\top |\psi_0\rangle \tag{6}$$

2.1   Quantum Jensen-Shannon Divergence

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

$$\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i| \tag{7}$$

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

$$H_N(\rho) = -Tr(\rho \log \rho) = -\sum_j \lambda_j \log \lambda_j \,, \tag{8}$$

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

$$D_{JS}(\rho, \sigma) = H_N\left(\frac{\rho + \sigma}{2}\right) - \frac{1}{2}H_N(\rho) - \frac{1}{2}H_N(\sigma) \tag{9}$$

This quantity is always well defined, symmetric and negative definite. It can also be shown that $D_{JS}(\rho, \sigma)$ is bounded, i.e.

$$0 \leq D_{JS}(\rho, \sigma) \leq 1 \tag{10}$$

Let $\rho = \sum_i p_i\rho_i$ be a mixture of quantum states ρ_i, with $p_i \in \mathbb{R}^+$ such that $\sum_i p_i = 1$, then we can prove that

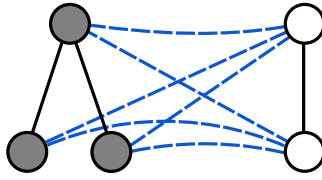$$H_N(\sum_i p_i\rho_i) \leq H_S(p_i) + \sum_i p_i H_N(\rho_i) \tag{11}$$

Fig. 1. Given two graphs $G_1(V_1, E_1, \nu_1)$ and $G_2(V_2, E_2, \nu_2)$ 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 $v \in V_2$.

where the equality is attained if and only if the states ρ_i have support on orthogonal subspaces, where the support of an operator is the subspace spanned by the eigenvectors of the operator with non-zero eigenvalues. By setting $p_1 = p_2 = 0.5$, we see that

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

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

3   A Similarity Measure for Attributed Graphs

Given two graphs $G_1(V_1, E_1, \nu_1)$ and $G_2(V_2, E_2, \nu_2)$, where ν_1 and ν_2 are respectively the functions assigning attributes to the nodes of G_1 and G_2, we build a new graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$ 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 $v \in V_2$ (see Fig. 1 for an example). Moreover, the edges $(u, v) \in E_{12}$ are labeled with a real value $\omega(\nu_1(u), \nu_2(v))$ representing the similarity between $\nu_1(u)$ and $\nu_2(v)$. With this new structure to hand, we define two continuous-time quantum walks $\left|\psi_t^-\right\rangle = \sum_{u \in V} \psi_{0u}^- |u\rangle$ and $\left|\psi_t^+\right\rangle = \sum_{u \in V} \psi_{0u}^+ |u\rangle$ on $\mathcal{G}$ with starting states

$$\psi_{0u}^- = \begin{cases} +\frac{d_u}{C} & \text{if } u \in G_1 \\ -\frac{d_u}{C} & \text{if } u \in G_2 \end{cases} \qquad \psi_{0u}^+ = \begin{cases} +\frac{d_u}{C} & \text{if } u \in G_1 \\ +\frac{d_u}{C} & \text{if } u \in G_2 \end{cases} \tag{13}$$

where d_u is the degree of the node u and C is the normalisation constant such that the probabilities sum to one. Note that the walk will spread at a speed proportional to the edge weights, which means that given an edge $(u, v) \in E_{12}$, the more similar $\nu_1(u)$ and $\nu_2(v)$ are, the faster the walker will propagate along the inter graphs connection (u, v). On the other hand, the intra-graph connection weights, which are not dependent on the nodes similarity, will not affect the propagation speed.

Given this setting, we allow the two quantum walks to evolve until a time T, and we define the average density operators ρ_T and σ_T over this time as

$$\rho_T = \frac{1}{T} \int_0^T \left|\psi_t^-\right\rangle\left\langle\psi_t^-\right| \, \mathrm{d}t \qquad \sigma_T = \frac{1}{T} \int_0^T \left|\psi_t^+\right\rangle\left\langle\psi_t^+\right| \, \mathrm{d}t \tag{14}$$

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

In the next section we will prove that, whenever G_1 and G_2 are isomorphic, the quantum Jensen-Shannon divergence between ρ_T and σ_T will be maximum, i.e., it will be equal to 1. Hence, it seems reasonable to use the value of the quantum Jensen-Shannon divergence as a measure of the similarity between the two graphs. In particular, in the next section we use the QJSD to define a novel kernel for attributed graphs.

3.1 A QJSD Kernel for Attributed Graphs

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

$$k_T(G_1, G_2) = D_{JS}(\rho_T, \sigma_T) \tag{15}$$

where ρ_T and σ_T are the density operators defined as in Eq. 14. Note that in this formulation the kernel is parametrised by the time variable T. As it is not clear how we should set this parameter, in this paper we propose to let $T \to \infty$, i.e., we compute $\lim_{T \to +\infty} k_T(G_1, G_2)$. In the following section we will show how to compute analytically this limit.

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

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

Proof. We need to prove that

$$(\rho_T)^\dagger \sigma_T = \frac{1}{T^2} \int_0^T \rho_{t_1} \, dt_1 \int_0^T \sigma_{t_2} \, dt_2 = \mathbf{0} \tag{16}$$

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

If $t_1 = t_2 = t$, then

$$\left\langle \psi_0^- \middle| (U^t)^\dagger U^t \middle| \psi_0^+ \right\rangle = 0 \tag{17}$$

since $(U^t)^\dagger U^t$ is the identity matrix and the initial states are orthogonal by construction. On the other hand, if $t_1 \neq t_2$, we have

$$\left\langle \psi_0^- \middle| U^{\Delta t} \middle| \psi_0^+ \right\rangle = 0 \tag{18}$$

where $\Delta_t = t_2 - t_1$.

To conclude the proof we rewrite the previous equation as

$$\langle \psi_0^- | U^{\Delta t} | \psi_0^+ \rangle = \sum_k \psi_{k0}^+ \sum_l \psi_{l0}^+ U_{lk}^{\Delta t}$$

$$= \sum_{k_1} \psi_{k_10}^+ \sum_l \psi_{l0}^+ U_{lk_1}^{\Delta t} - \sum_{k_2} \psi_{k_20}^+ \sum_l \psi_{l0}^+ U_{lk_2}^{\Delta t}$$

$$= \sum_l \psi_{l0}^+ \left(\sum_{k_1} \psi_{k_10}^+ U_{lk_1}^{\Delta t} - \sum_{k_2} \psi_{k_20}^+ U_{lk_2}^{\Delta t} \right) = 0 \qquad (19)$$

where the indices l, k, run over the nodes of $\mathcal{G}$, and k_1 and k_2 run over the nodes G_1 and G_2 respectively.

To see that Eq. 19 holds, note that U is a symmetric matrix and it is invariant to graph symmetries, i.e., if u and v are symmetric then $U_{uu}^{\Delta t} = U_{vv}^{\Delta t}$, and that if G_1 and G_2 are isomorphic, then $k_1 = k_2$ and $\psi_{1:k_10}^+ = \psi_{k_1+1:k_20}^+$. Recall that $U^t = e^{-iAt}$, where A is the graph adjacency matrix. A symmetry orbit is defined as a group of vertices where v_1 and v_2 belong to the same orbit if there is an automorphism $\sigma \in Aut(G)$ such that $\sigma v_1 = v_2$, where $Aut(G)$ is the set of automorphisms of G. In other words, if u and v belong to a symmetry orbit, there exists an automorphism of the graph with a corresponding permutation matrix P such that

$$A = P^\top A P \qquad (20)$$

and

$$P |e_u\rangle = |e_v\rangle \qquad (21)$$

This in turn implies that the graph adjacency matrix is invariant to symmetries. As we will show, the same holds for the unitary operator of the quantum walk. In fact, given the spectral decomposition of $A = \Phi \Lambda \Phi^\top$, we can see that the following equality holds

$$\Phi \Lambda \Phi^\top = P^\top (\Phi \Lambda \Phi^\top) P \qquad (22)$$

and thus

$$\Phi = P^\top \Phi \qquad (23)$$

Let us now write the unitary operator in terms of the adjacency matrix eigen-decomposition, which yields

$$e^{-iAt} = \Phi e^{-i\Lambda t} \Phi^\top \qquad (24)$$

From Equations 23 and 24 it follows that

$$\Phi e^{-i\Lambda t} \Phi^\top = P^\top \Phi e^{-i\Lambda t} \Phi^\top P \qquad (25)$$

This in turn implies that if u and v are symmetrical then $U_{uu}^t = U_{vv}^t$, which concludes the proof.

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

Proof. The first property is trivially proved by noting that, according to Eq. 15, the kernel between G_1 and G_2 is defined as the quantum Jensen-Shannon divergence between two density operators, and then recalling that the value of quantum Jensen-Shannon divergence is bounded to lie between 0 and 1.

The second property follows again from Eq. 15 and Theorem 1. It is sufficient to note that the quantum Jensen-Shannon divergence reaches its maximum value if and only if the density operators have support on orthogonal spaces.

Unfortunately we cannot prove that our kernel is positive semidefinite, but both empirical evidence and the fact that the Jensen-Shannon Divergence is negative semidefinite on pure quantum states [21] while our graph kernel is maximal on orthogonal states suggest that the kernel constraints are never violated in practice.

3.2   Kernel Computation

We conclude this section with a few remarks on the computational complexity of our kernel. In particular, we show that the solutions to Eq. 14 can be computed analytically. Recall that $|\psi_t\rangle = e^{-iAt} |\psi_0\rangle$, then we rewrite Eq. 14 as

$$\rho_T = \frac{1}{T} \int_0^T e^{-iAt} |\psi_0\rangle \langle \psi_0| e^{iAt} \, dt \tag{26}$$

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

$$\rho_T = \frac{1}{T} \int_0^T \Phi e^{-i\Lambda t} \Phi^\top |\psi_0\rangle \langle \psi_0| \Phi e^{i\Lambda t} \Phi^\top \, dt \tag{27}$$

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

$$\rho_T(r, c) = \frac{1}{T} \int_0^T \left(\sum_k \sum_l \phi_{rk} e^{-i\lambda_k t} \phi_{lk} \psi_{0l}^- \right)$$
$$\cdot \left(\sum_m \sum_n \psi_{0m}^\dagger \phi_{mn} e^{i\lambda_n t} \phi_{cn} \right) dt \tag{28}$$

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

$$\rho_T(r, c) = \frac{1}{T} \int_0^T \left(\sum_k \phi_{rk} e^{-i\lambda_k t} \bar{\psi}_k \sum_n \phi_{cn} e^{i\lambda_n t} \bar{\psi}_n \right) dt \tag{29}$$
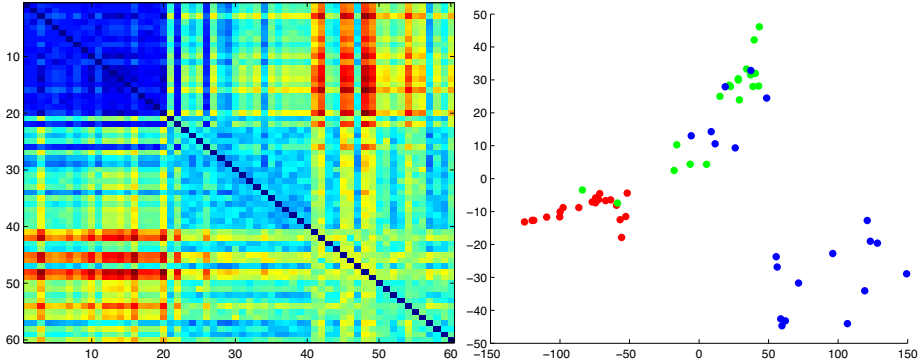
Fig. 2. Edit distance matrix and Multidimensional Scaling of the graph distances for the synthetic dataset

which can be finally rewritten as

$$\rho_T(r,c) = \sum_k \sum_n \phi_{rk} \phi_{cn} \bar{\psi}_k \bar{\psi}_n \frac{1}{T} \int_0^T e^{i(\lambda_n - \lambda_k)t} \, \mathrm{d}t \tag{30}$$

If we let $T \to \infty$, Eq. 30 further simplifies to

$$\rho_T(r,c) = \sum_{\lambda \in \tilde{\Lambda}} \sum_k \sum_n \phi_{rk} \phi_{cn} \bar{\psi}_k \bar{\psi}_n \tag{31}$$

where $\tilde{\Lambda}$ is the set of distinct eigenvalues of A, while k and n are indices which run over the dimensions of the eigenspace associated with $\lambda \in \tilde{\Lambda}$. As a consequence, we see that the complexity of computing the QJSD kernel is upper bounded by that of computing the eigendecomposition of $\mathcal{G}$, i.e. $O(|\mathcal{V}|^3)$.

4   Experimental Results

In this section, we evaluate the performance of the proposed kernel and we compare it with a number of well-known alternative graph kernels, namely the classic random walk kernel [10], the shortest-path kernel [11] and the 3-nodes graphlet kernel [12], both in their unattributed and attributed versions. Note that since the attributed versions of these kernels are defined only on graphs with categorically labeled nodes, in our experiments we will need to bin the node attributes before computing the kernels.

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 sample. The process is averaged over 100 random partitions of the data, and the results are reported in terms of average accuracy ± standard error.
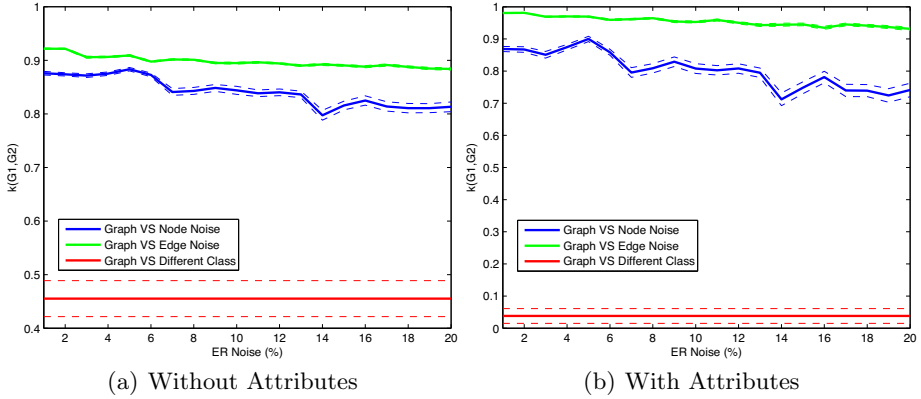
(a) Without Attributes   (b) With Attributes

Fig. 3. The effects of Erdös-Rényi structural noise applied to the nodes and edges of the graph on the kernel value. Using the proposed similarity measure, the noisy versions of the graph belonging to the first class are clearly distinguishable from the instances of the second class. As expected, taking the attributes into account (right) makes the distinction even clearer (note the difference in the scale).

4.1 Synthetic Data

We start by evaluating the proposed kernel on a set of synthetically generated graphs. To this end, we have randomly generated 3 different weighted graph prototypes with size 16, 18 and 20 respectively. For each prototype we started with an empty graph and then we iteratively added the required number of nodes each labeled with a random mean and variance. Then we added the edges and their associated observation probabilities up to a given edge density. Given the prototypes, we sampled 20 observations from each class being careful to discard graphs that were disconnected. Details about the generative model used to sample the graphs can be found in [19]. Figure 2 shows the edit distance matrix of the dataset and the Multidimensional Scaling [20] of the graph distances.

With the synthetic graphs to hand, we initially investigate how the value of the kernel between two graphs varies as we apply Erdös-Rényi noise to the graph structure. In this case the similarity between two nodes u and v is defined as $\omega(u,v) = e^{-\lambda(\nu_1(u)-\nu_2(v))^2}$, where $\nu_1(u)$ and $\nu_2(v)$ are the real-valued attributes associated with u and v respectively. Figure 3 shows the result of this experiment. Here we randomly pick a graph G belonging to class 1, and we compute a number of increasingly noisy versions of it. The noise is applied either to the edges only, i.e. adding or deleting edges, or to the nodes as well, i.e. adding or deleting nodes and edges. We then compute the average value of the kernel between G and its corrupted versions, and we plot it against the average similarity between G and the graphs of class 2. Figure 3 shows that, even at considerably high levels of noise, G is clearly distinguishable from the instances of the second class. As expected, taking the attributes into account renders the distinction even clearer (note the change in the y-scale). However, when augmented with the attributes information, our simi-
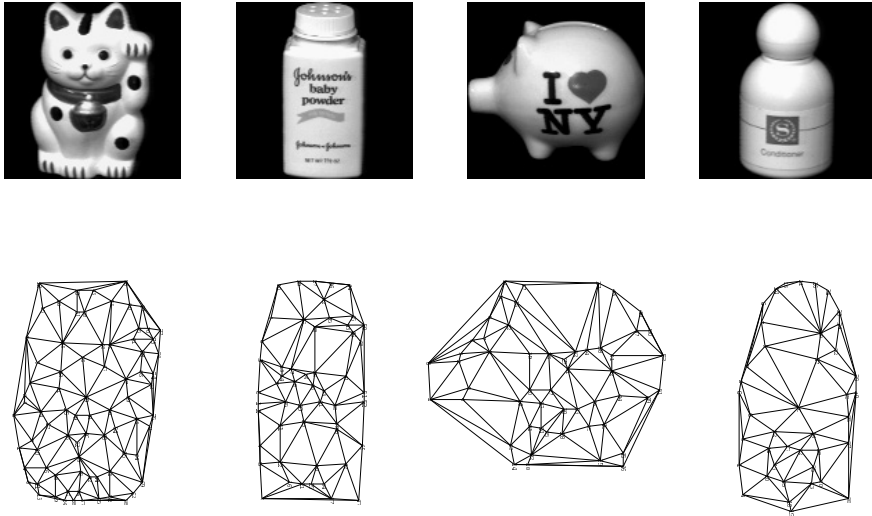
Fig. 4. The four selected objects from the COIL [22] dataset and a sample of their associated Delaunay graphs. Each node of the graphs is labeled with the (x, y) coordinates of the corresponding feature point.

larity measure seems to be slightly more sensitive to structural noise, in particular when the noise is affecting the nodes of the graph.

As a second experiment, we test the accuracy of our kernel in a classification task. The results are shown in Table 1. As we can see, our kernel outperforms or is competitive with the alternatives, and yields a close to 100% average accuracy. Note also that, as expected, taking the similarity between the node attributes into account results in a marked increase in the kernel performance. Quite surprisingly, however, we found that the random walk kernel on the categorically labeled graphs yields a lower performance than its unattributed version.

4.2   Delaunay Graphs

We then tested the efficacy of the proposed kernel on the COIL [22] dataset, which consists of images of different objects, with 72 views of each object obtained from equally spaced viewing directions over 360°. For each image, a graph is obtained by computing the Delaunay triangulation of the corner points extracted by the Harris corner detection algorithm. Moreover, each node is labeled with the (x, y) coordinates of the corresponding feature point. The similarity between two nodes is $\omega(u, v) = e^{-\lambda ||\nu_1(u) - \nu_2(v)||_2^2}$, where $||\nu_1(u) - \nu_(v)||_2$ is the Euclidean distance between the two feature points u and v. Here we choose 4 different objects, each with 21 different 5° rotated views. Figure 4 shows the four selected objects together with their associated graphs, while Figure 5 shows the edit distance matrix and the MDS of the graph distances.
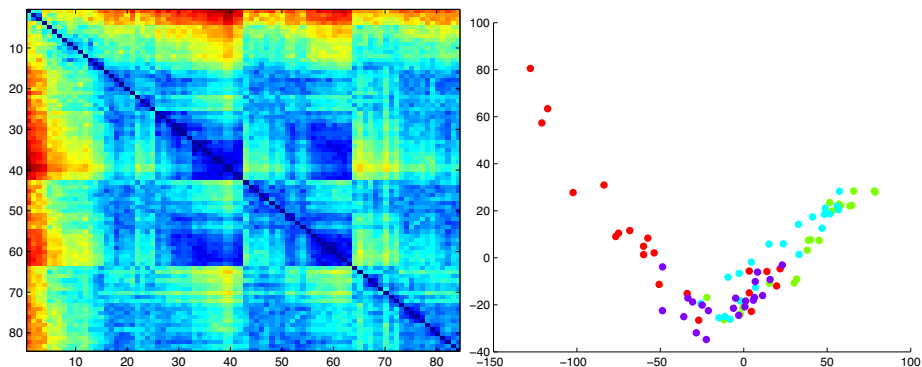
Fig. 5. Edit distance matrix and Multidimensional Scaling of the graph distances for the COIL dataset

We first investigate how integrating the information on the nodes attributes influences the expressive power of our kernel. Figure 6 shows the MDS embedding on the graph distances computed from the unattributed kernel (left) and the attributed one (right). Although the embedding shows that a considerable overlap remains between the different classes, taking the node attributes similarities into account adds a further dimension which can help to discriminate better among the 4 selected objects.

This is indeed reflected in the results of the classification task shown in Table 1. In the attributed case, in fact, the average accuracy of the QJSD kernel is increased by more than 10%, and it outperforms that of all the remaining kernels. Note, however, that if the node labels are dropped, the performance of the QJSD kernel is among the lowest, which once again underlines the importance of incorporating the attributes similarities in the compositional structure.
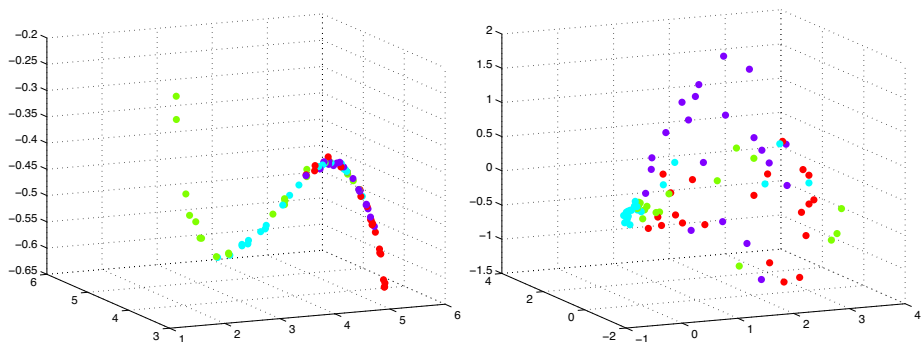

Fig. 6. Multidimensional Scaling of the graph distances computed from the kernel matrix of the COIL dataset. Left, completely structural approach; right, including the information on the nodes attributes.
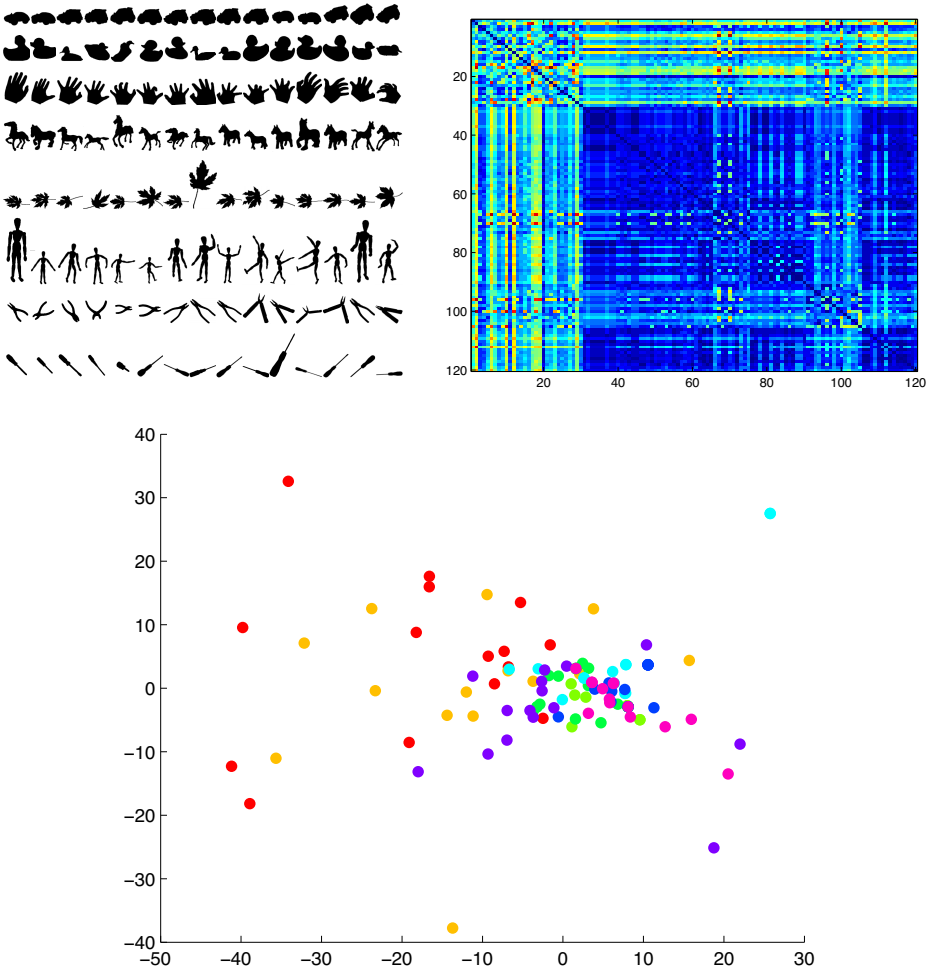
Fig. 7. Top row: Left, shape database; right, edit distance matrix. Bottom row: Multi-dimensional Scaling of the edit distances. As we can see, the class structure is not very clear and there is a considerable overlap between different classes.

4.3   Shock Graphs

Finally, we experimented using shock graphs, a skeletal-based representation of the differential structure of the boundary of a 2D shape. We extracted graphs from a database composed of 120 shapes divided into 8 classes of 15 shapes each. Each graph has a node attribute that reflects the size of the boundary feature generating the corresponding skeletal segment. Figure 7 shows the shape database, the edit distances matrix between the shock graphs and the corresponding MDS. As we can see, the class structure is not very clear, and there is a considerable overlap between different classes. This is reflected in the average accuracy of the kernels, which is the

Table 1. Classification accuracy ($\pm$ standard error) on attributed graph datasets. QJSD is the proposed kernel, SP is the shortest-path kernel of Borgwardt and Kriegel [11], RW is the random walk kernel of Gartner et al. [10], while G_3 denotes the graphlet kernel computed using all graphlets of size 3 described in Shervashidze et al. [12]. The subscript w identifies the kernels which make use of the attributes information. The best performing kernel for each dataset is highlighted in bold.

Kernel	Synth	Shock	COIL
QJSD_w	95.87 ± 0.14	$\mathbf{66.65 \pm 0.22}$	$\mathbf{95.56 \pm 0.20}$
QJSD	84.57 ± 0.25	53.97 ± 0.19	84.05 ± 0.22
SP_w	$\mathbf{96.36 \pm 0.12}$	65.05 ± 0.25	94.40 ± 0.14
SP	91.13 ± 0.15	52.62 ± 0.32	85.25 ± 0.21
RW_w	92.97 ± 0.18	53.26 ± 0.29	90.78 ± 0.26
RW	80.23 ± 0.30	26.11 ± 0.32	78.60 ± 0.25
G3_w	88.75 ± 0.25	41.18 ± 0.27	89.25 ± 0.21
G3	85.60 ± 0.25	38.85 ± 0.32	84.20 ± 0.22

lowest among the three datasets, as Table 1 shows. However, the proposed kernel still outperforms or is competitive with the others.

5   Conclusions and Future Work

In this paper, we have introduced a novel similarity measure for attributed graphs based on the time evolution of a continuous-time quantum walk. More precisely, given a pair of graphs we computed the quantum Jensen-Shannon divergence between the evolution of two quantum walks on a suitably defined union of the original graphs. With the quantum Jensen-Shannon divergence to hand, we then established our similarity measure. Finally, we introduced a novel kernel on attributed graphs based on the proposed similarity measure. We performed an extensive experimental evaluation both on synthetic and real-world datasets, and we demonstrated the effectiveness of the proposed approach.

However, in this paper we limited our definition of the kernel to the case where the time parameter T is taken to the limit, i.e., $T \to \infty$. Future work will focus on studying the role of the time parameter more in depth, and it will try to develop a heuristic to establish the optimal time T in terms of classification accuracy.

References

1. Siddiqi, K., Shokoufandeh, A., Dickinson, S., Zucker, S.: Shock graphs and shape matching. International Journal of Computer Vision 35, 13–32 (1999)
2. Jeong, H., Tombor, B., Albert, R., Oltvai, Z., Barabási, A.: The large-scale organization of metabolic networks. Nature 407, 651–654 (2000)

3. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y.: A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences 98, 4569 (2001)
4. Kalapala, V., Sanwalani, V., Moore, C.: The structure of the united states road network. University of New Mexico (2003) (preprint)
5. Shapiro, L., Haralick, R.: Structural descriptions and inexact matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, 504–519 (1981)
6. Barrow, H.G., Burstall, R.M.: Subgraph isomorphism, matching relational structures and maximal cliques. Inf. Process. Lett. 4, 83–84 (1976)
7. Bunke, H., Shearer, K.: A graph distance metric based on the maximal common subgraph. Pattern Recognition Letters 19, 255–259 (1998)
8. Hidović, D., Pelillo, M.: Metrics for attributed graphs based on the maximal similarity common subgraph. International Journal of Pattern Recognition and Artificial Intelligence 18(3), 299–313 (2004)
9. Smola, A., Schölkopf, B.: Learning with kernels. Citeseer (1998)
10. Gärtner, T., Flach, P., Wrobel, S.: On graph kernels: Hardness results and efficient alternatives. In: Schölkopf, B., Warmuth, M.K. (eds.) COLT/Kernel 2003. LNCS (LNAI), vol. 2777, pp. 129–143. Springer, Heidelberg (2003)
11. Borgwardt, K., Kriegel, H.: Shortest-path kernels on graphs. In: Fifth IEEE International Conference on Data Mining, p. 8. IEEE (2005)
12. Shervashidze, N., Vishwanathan, S., Petri, T., Mehlhorn, K., Borgwardt, K.: Efficient graphlet kernels for large graph comparison. In: Proceedings of the International Workshop on Artificial Intelligence and Statistics. Society for Artificial Intelligence and Statistics (2009)
13. Haussler, D.: Convolution kernels on discrete structures. Technical report, UC Santa Cruz (1999)
14. Kempe, J.: Quantum random walks: an introductory overview. Contemporary Physics 44, 307–327 (2003)
15. Emms, D., Wilson, R.C., Hancock, E.R.: Graph embedding using quantum commute times. In: Escolano, F., Vento, M. (eds.) GbRPR. LNCS, vol. 4538, pp. 371–382. Springer, Heidelberg (2007)
16. Rossi, L., Torsello, A., Hancock, E.R.: Approximate axial symmetries from continuous time quantum walks. In: Gimel'farb, G., Hancock, E.R., Imiya, A., Kuijper, A., Kudo, M., Omachi, S., Windeatt, T., Yamada, K. (eds.) SSPR&SPR 2012. LNCS, vol. 7626, pp. 144–152. Springer, Heidelberg (2012)
17. Lamberti, P., Majtey, A., Borras, A., Casas, M., Plastino, A.: Metric character of the quantum Jensen-Shannon divergence. Physical Review A 77, 052311 (2008)
18. Nielsen, M., Chuang, I.: Quantum computation and quantum information. Cambridge university press (2010)
19. Torsello, A., Rossi, L.: Supervised learning of graph structure. In: Pelillo, M., Hancock, E.R. (eds.) SIMBAD 2011. LNCS, vol. 7005, pp. 117–132. Springer, Heidelberg (2011)
20. Wish, M., Carroll, J.D.: 14 Multidimensional scaling and its applications. Handbook of Statistics 2, 317–345 (1982)
21. Briët, J., Harremoës, P.: Properties of classical and quantum jensen-shannon divergence. Physical Review A 79, 052311 (2009)
22. Nayar, S., Nene, S., Murase, H.: Columbia object image library (coil 100). Technical report, Tech. Report No. CUCS-006-96. Department of Comp. Science, Columbia University (1996)