PHYSICAL REVIEW E **91**, 022815 (2015)

Measuring graph similarity through continuous-time quantum walks and the quantum Jensen-Shannon divergence

Luca Rossi,[1] Andrea Torsello,[2] and Edwin R. Hancock[3]
[1]*School of Computer Science, University of Birmingham, United Kingdom*
[2]*Dipartimento di Scienze Ambientali, Informatica e Statistica, Università Ca' Foscari Venezia, Venezia, Italy*
[3]*Department of Computer Science, University of York, United Kingdom*
(Received 30 July 2014; revised manuscript received 7 December 2014; published 23 February 2015)

In this paper we propose a quantum algorithm to measure the similarity between a pair of unattributed graphs. We design an experiment where the two graphs are merged by establishing a complete set of connections between their nodes and the resulting structure is probed through the evolution of continuous-time quantum walks. In order to analyze the behavior of the walks without causing wave function collapse, we base our analysis on the recently introduced quantum Jensen-Shannon divergence. In particular, we show that the divergence between the evolution of two suitably initialized quantum walks over this structure is maximum when the original pair of graphs is isomorphic. We also prove that under special conditions the divergence is minimum when the sets of eigenvalues of the Hamiltonians associated with the two original graphs have an empty intersection.

DOI: 10.1103/PhysRevE.91.022815 PACS number(s): 89.75.Fb, 89.75.Hc, 89.75.Kd

I. INTRODUCTION

Graph-based representations have long been used as a powerful way to characterize a large number of systems which are best described in terms of their topological or interconnection structure [1–4]. However, the rich expressiveness of graphs usually comes at the cost of an increased difficulty in applying standard pattern recognition techniques and machine learning to them, as these usually require the graphs to be first embedded into a vectorial space, a procedure which is far from being trivial. This is in turn due to the lack of a canonical ordering for the nodes in a graph. In fact, correspondences or alignment to a reference structure must be established before the analysis can commence. Moreover, even if a correspondence order can be established, the dimension of the embedding space may vary, as a result of structural modifications, i.e., changes in the number of nodes and edges.

Kernel methods [5] provide an elegant way to transform the problem at hand from that of finding an embedding of the set of data entities to that of defining a positive semidefinite kernel between them using the well-known kernel trick. The best known example of this approach is probably furnished by support vector machines (SVMs) [6]. The data entities considered can be vectors, graph nodes, or, as in our case, entire graphs. Given a positive semidefinite kernel $k : X \times X \to \mathbb{R}$ on a set X, we know that there exists a map $\phi : X \to H$ into a Hilbert space H, such that $k(x,y) = \langle \phi(x), \phi(y) \rangle$ for all $x, y \in X$, where $\langle \cdot, \cdot \rangle$ denotes the scalar product in H. As a consequence, 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 a kernel is defined. Inspired by the R-convolution kernel introduced by Haussler [7], a number of graph kernels have been proposed in the literature [8–10]. The unifying principle underpinning these kernels is that of defining the similarity between two graphs by decomposing them and then comparing the resulting simpler substructures. This led for example to the introduction of the random walk kernel by Gärtner *et al.* [8], which is based on the enumeration of common random walks between two graphs. Similarly, Borgwardt and Kriegel [9] measure the similarity by comparing the shortest paths in the graphs, while in Shervashidze *et al.* [10] this is related to the presence of small subgraphs. Another interesting approach is that of Bai and Hancock [11], where the authors investigate the possibility of defining a graph kernel based on the Jensen-Shannon kernel. The Jensen-Shannon kernel is a nonextensive information-theoretic kernel, which is defined in terms of the entropy of probability distributions over the structures being compared [12]. Bai and Hancock extend this idea to the graph domain by associating with each graph either its Von Neumann entropy [13], i.e., the Shannon entropy associated with the Laplacian eigenvalues of the graph, or the steady-state distribution of a random walk on the graph. Finally, it is important to note that the problem of defining a *complete* kernel, i.e., a kernel whose implicit map ϕ is injective, is at least as hard as the graph isomorphism problem [8].

Quantum walks have recently emerged as a primitive for designing novel quantum algorithms [14–17] on graph structures. Similarly to a classical random walk, a quantum walk is defined as a dynamical process over the vertices of the graph. However, the two walks possess remarkably different properties. While in the classical case the state vector describing the evolution of the walk is a real-valued probability vector, the quantum walk is characterized by a complex-valued amplitude vector, with no restrictions on the sign and phase. This property in turn allows different paths of the walk to interfere with each other in both constructive and destructive ways, and it is responsible for many of the exotic properties of quantum walks. Moreover, in the classical case the evolution of the walk is governed by a double stochastic matrix, while in the quantum case the evolution is governed by a unitary matrix, which renders the walk reversible. As a consequence, the quantum walk is nonergodic and, most importantly, it does not have a limiting distribution. This lack of convergence makes the behavior of quantum walks on general graphs considerably harder to study than their classical counterparts. For this reason, quantum walks have been extensively studied on a wide number of specific topologies [18,19], such as the infinite line, cycles, regular lattices, star graphs, and complete graphs. Finally, one of the most celebrated properties of

quantum walks is that they can be used to achieve polynomial and sometimes even exponential speedups over classical computation in a number of interesting problems [20,21]. For example, Farhi and Gutmann [21] have shown that if we take two cojoined n-level binary trees that are connected at their leaves, a quantum walk commencing from the root of the first tree can hit the root of the second tree exponentially faster than a similarly defined classical random walk. The major contribution of Farhi and Gutmann's work [21] is to show that one may achieve an exponential speedup without relying on the quantum Fourier transform. The highly symmetrical structure of the cojoined-trees graph described above turns out to be of key importance to the speedup. Recall that given a graph $G = (V,E)$, an automorphism is a permutation τ of the set of vertices V of the graph which preserves the adjacency relations, i.e., if $(u,v) \in E$ then $(\tau(u),\tau(v)) \in E$. The set of symmetries of G can thus be represented by its automorphism group Aut(G). Krovi and Brun [22] have shown that the phenomena of exponential speedup as well as that of infinite hitting times are generally a consequence of the degeneracies in the eigenspace of the evolution operator. These in turn are related to the symmetry group of the graph. Intuitively, as a consequence of symmetry a quantum walker can reach a vertex v along multiple paths with the same phase. In this case, the interference is constructive and the node v has a high probability of being visited. This leads to a faster hitting time for node v. However, depending on the initial state of the walk, the quantum walker can also reach v along paths with phases that correspond to destructive interference. In fact, in some cases the probability of the walker to visiting node v will be zero, i.e., the hitting time can be infinite. Another consequence of the intimate connection between symmetries and quantum walks has been investigated by Emms *et al.* [23]. Specifically, the authors define a quasiquantum analog of the commute time associated with the continuous-time quantum walk and then explore the possibility of using it to embed the nodes of the graph into a low-dimensional vector space. Their work reveals that the symmetries of the graph correspond to degenerate directions in the quantum commute-time embedding space. More recently, Rossi *et al.* [24] proposed a way to detect approximate axial symmetries in networks by measuring the interference patterns of continuous-time quantum walks. However, the analyses of Emms *et al.* [23] and Rossi *et al.* [24] are not based on a principled observable and are hence semi-classical. In order to overcome this limitation, Rossi *et al.* [25] made use of the quantum Jensen-Shannon divergence, effectively rendering the analysis fully based on observable properties.

The quantum Jensen-Shannon divergence (QJSD) has recently been developed as a generalization of the classical Jensen-Shannon divergence to quantum states by Majtey and co-workers [26–28]. Just as the classical Jensen-Shannon divergence [29], the quantum Jensen-Shannon divergence is symmetric, bounded, and always defined. Unlike its classical counterpart, however, it has been proved to be the square of a metric only for pure states [28], whereas for mixed states there is only empirical evidence suggesting that it is [28]. Moreover, it has been shown that for mixed quantum states the quantum Jensen-Shannon divergence has good distinguishability properties. Note that as the QJSD is defined in terms of the Von Neumann entropy it is not directly a quantum-mechanical observable, i.e., there is no operator whose expected value is the QJSD. However, it can be computed from density matrices whose entries are indeed observables.

In this paper, we introduce a kernel on unattributed graphs which evaluates the similarity between two graphs through the evolution of a suitably defined continuous-time quantum walk on their structure. We measure the similarity between two graphs by merging them into a structure whose degree of symmetry will be maximum when the original graphs are isomorphic. With this structure to hand, we define two continuous-time quantum walks such that the density operators of the resulting quantum states are orthogonal whenever the two original graphs are isomorphic. More precisely, we measure the quantum Jensen-Shannon divergence between these states. We stress that, while this analysis is fully based on observable properties, it is not meant to provide an algorithm exhibiting quantum speedup with respect to classical counterparts, but rather to highlight how quantum walks can be used to provide information about the structural similarities between two graphs.

Note that a number of alternative graph kernels based on the classical Jensen-Shannon divergence and its quantum counterpart have been recently introduced in the literature [11,30–32]. In particular, the present paper builds on the work of Rossi *et al.* [30,31], but it differs from it in a number of significant aspects. More specifically, in an attempt to shed light on the general behavior of the kernel, we analyze the relation between the kernel value and the graph spectra, and we show that the divergence is minimum when the sets of eigenvalues of the Hamiltonians associated with the two original graphs have an empty intersection. We also perform an extensive set of experiments to evaluate the impact of the Hamiltonian and the time parameter on the classification accuracy. With respect to the classical Jensen-Shannon kernel of Bai and Hancock [11], here we do not need to construct a product graph from the two input graphs in order to measure the composite entropy. Instead, we naturally handle the computation by comparing quantum states defined over the same state space. Moreover, our work is also significantly different from that of Bai *et al.* [32], where, in order to guarantee the permutational invariance, the authors need to compute the optimal alignment between the input graphs before the analysis can even commence. In this work, on the other hand, we solve the problem by establishing a complete set of connections between the two graphs and carefully crafting the initial states of the walks so as to highlight the presence of structural symmetries. In a nutshell, in the present work the computation of the kernel is naturally handled by means of the interference effects of quantum walks, thus avoiding either using a rotation in Hilbert space [32], or the construction of a product union graph [11] in the classical case.

The remainder of the paper is organized as follows: Sec. II provides a brief introduction to continuous-time quantum walks, while Sec. III reviews the concepts of Von Neumann entropy and quantum Jensen-Shannon divergence. In Sec. IV we propose a method to measure the similarity between two unattributed graphs based on the quantum Jensen-Shannon

divergence. Section V illustrates the experimental results, while the conclusions are presented in Section VI.

II. CONTINUOUS-TIME QUANTUM WALKS

Let $G = (V,E)$ be an undirected graph, where V is a set of n vertices and $E = (V \times V)$ is a set of edges. The adjacency matrix of G is the symmetric matrix with elements

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

and the diagonal matrix D has elements $d_u = \sum_{v=1}^{n} A(u,v)$, where d_u is the degree of the node u. The graph Laplacian is then defined as $L = D - A$, and it can be interpreted as a combinatorial analog of the discrete Laplace-Beltrami operator [33].

A *continuous-time random walk* on the graph G models a Markovian diffusion process over its node set, where the transitions are allowed only along the edges connecting adjacent vertices. Let $\mathbf{p}_t \in \mathbb{R}^n$ be a vector denoting the state of the walk at time t, such that its uth entry gives the probability of the walk being at vertex u at time t. Then the state vector evolves according to the equation

$$\mathbf{p}_t = e^{-Lt}\mathbf{p}_0, \tag{2}$$

where L is the generator matrix of the underlying continuous-time Markov process.

The *continuous-time quantum walk* is the quantum counterpart of the continuous-time random walk, and it is similarly defined as a dynamical process over the vertices of the graph [21]. Here the classical state vector is replaced by a vector of complex amplitudes over V whose squared norm sums to unity, and as such the state of the system is not constrained to lie in a probability space. In fact, the lack of restrictions on the sign and complex phase allows for interference effects to take place. Let us denote, using Dirac notation, 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

$$|\psi_t\rangle = \sum_{u \in V} \alpha_u(t)|u\rangle, \tag{3}$$

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 evolution of the walk is then given by the Schrödinger equation, where we denote the time-independent Hamiltonian as $\mathcal{H}$,

$$\frac{\partial}{\partial t}|\psi_t\rangle = -i\mathcal{H}|\psi_t\rangle. \tag{4}$$

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

$$|\psi_t\rangle = e^{-i\mathcal{H}t}|\psi_0\rangle. \tag{5}$$

By analogy with the case of a particle moving in an empty space with zero potential energy, it is common practice to choose the Laplacian matrix as the system Hamiltonian, i.e., $\mathcal{H} = L$. However, any Hermitian operator encoding the structure of the graph, such as the adjacency matrix of the graph, can be chosen as an alternative.

Note that in the quantum case the evolution of the state vector of the walker is governed by a complex-valued unitary matrix. Hence the evolution of the quantum walk is reversible, implying that quantum walks are nonergodic and do not possess a limiting distribution. This is in stark contrast to the classical case, where the dynamics of the walk is governed by a stochastic matrix. As a result of the unitary evolution and the complex-valued nature of the amplitude vector, the behavior of classical and quantum walks differs significantly, and quantum walks possess a number of interesting properties not exhibited by classical random walks.

Finally, note that we can rewrite Eq. (5) as follows. Let us compute the spectral decomposition of the Hamiltonian $\mathcal{H} = \Phi\Lambda\Phi^\top$, where Φ is the $n \times n$ matrix $\Phi = (\phi_1|\phi_2|\cdots|\phi_j|\cdots|\phi_n)$ with the ordered eigenvectors ϕ_js of $\mathcal{H}$ as columns and

$$\Lambda = \text{diag}(\lambda_1,\lambda_2,\ldots,\lambda_j,\ldots,\lambda_n)$$

is the $n \times n$ diagonal matrix with the ordered eigenvalues λ_j of $\mathcal{H}$ as elements. Using the spectral decomposition of the Hamiltonian and the fact that $\exp[-i\mathcal{H}t] = \Phi\exp[-i\Lambda t]\Phi^\top$ we can then write

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

III. QUANTUM JENSEN-SHANNON DIVERGENCE

The observation process for a quantum system is defined in terms of projections onto orthogonal subspaces associated with operators on the quantum state space called *observables*. Let O be an observable of the system, with spectral decomposition

$$O = \sum_i a_i P_i, \tag{7}$$

where the a_i are the (distinct) eigenvalues of O and the P_i the orthogonal projectors onto the corresponding eigenspaces. The outcome of an observation, or projective measurement, of a quantum state $|\psi\rangle$ is one of the eigenvalues a_i of O, and it has probability

$$P(a_i) = \langle\psi|P_i|\psi\rangle. \tag{8}$$

After the measurement, the state of the quantum system becomes

$$|\bar{\psi}\rangle = \frac{P_i|\psi\rangle}{||P_i|\psi\rangle||}, \tag{9}$$

where $|||\psi\rangle|| = \sqrt{\langle\psi\,|\,\psi\rangle}$ is the norm of the vector $|\psi\rangle$.

The *density operator* (or *density matrix*) is introduced in quantum mechanics to describe a system whose state is an ensemble of pure quantum states $|\psi_i\rangle$, each with probability p_i. The density operator of such a system is defined as

$$\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|. \tag{10}$$

Density operators are positive unit-trace matrices that play an important role in the quantum observation process. In fact, the

expectation value of the measurement can be calculated from the density matrix ρ:

$$\langle O \rangle = \mathrm{tr}(\rho O), \tag{11}$$

where tr is the trace operator. Similarly, the observation probability of a_i can be expressed in terms of the density matrix ρ as

$$P(a_i) = \mathrm{tr}(\rho P_i) \tag{12}$$

where the corresponding density operator after the measurement has taken place will be

$$\rho' = \sum_i P_i \rho P_i. \tag{13}$$

The *von Neumann entropy* [34] H_N of a mixture is defined in terms of the trace and logarithm of the density operator ρ,

$$H_N = -\mathrm{tr}(\rho \ln \rho) = -\sum_i \xi_i \ln \xi_i \tag{14}$$

where $\xi_1, \ldots, \xi_n$ are the eigenvalues of ρ. If $\langle \psi_i | \rho | \psi_i \rangle = 1$, i.e., the quantum system is a pure state $|\psi_i\rangle$ with probability $p_i = 1$, then the Von Neumann entropy $H_N(\rho) = -\mathrm{tr}(\rho \ln \rho)$ is zero. On other hand, for a mixed state described by the density operator σ we have a nonzero Von Neumann entropy associated with it.

With the Von Neumann entropy to hand, the quantum Jensen-Shannon divergence between two density operators ρ and σ is defined as

$$D_{JS}(\rho,\sigma) = H_N\left(\frac{\rho+\sigma}{2}\right) - \frac{1}{2}H_N(\rho) - \frac{1}{2}H_N(\sigma). \tag{15}$$

This quantity is always well defined, symmetric, and positive definite.

It can also be shown that $D_{JS}(\rho,\sigma)$ is bounded, i.e., $0 \leqslant D_{JS}(\rho,\sigma) \leqslant 1$. Let $\rho = \sum_i p_i \rho_i$ be a mixture of quantum states ρ_i, with $p_i \in \mathbb{R}^+$ such that $\sum_i p_i = 1$; then one can prove that

$$H_N\left(\sum_i p_i \rho_i\right) \leqslant H_S(p_i) + \sum_i p_i H_N(\rho_i), \tag{16}$$

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

$$D_{JS}(\rho,\sigma) = H_N\left(\frac{\rho+\sigma}{2}\right) - \frac{1}{2}H_N(\rho) - \frac{1}{2}H_N(\sigma) \leqslant 1. \tag{17}$$

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

Our interest in the quantum Jensen-Shannon divergence lies in the fact that it verifies several interesting properties which are required for a good distinguishability measure between quantum states [27,28]. The distinguishability problem is of central importance in quantum mechanics, and it is related to the concept of distance between states. In the work of Wootters [35], the distance between two states $|\phi\rangle$ and $|\psi\rangle$ of the same physical system is computed by enumerating the distinguishable states between $|\phi\rangle$ and $|\psi\rangle$. It turns out that Wootters' work is fundamentally based on the extension of a distance over the space of probability distributions to the Hilbert space of pure quantum states. Similarly, the relative entropy [36] generalizes to the quantum world the information-theoretic Kullback-Leibler divergence. However, the relative entropy is neither a distance, as it is not symmetric, nor does it satisfy the triangle inequality, and, most importantly, it is unbounded.

On the other hand, the QJSD between two pure states has been proved to be the square of a metric [28], while for the case of mixed states there is strong numerical evidence that this is also the case. Note that alternative metrics have been proposed in the literature, such as the Bures distance [37], which is defined as

$$B(\rho,\sigma) = \sqrt{2}[1 - \mathrm{tr}((\rho^{1/2}\sigma\rho^{1/2})^{1/2})]^{1/2}. \tag{18}$$

The Bures distance and the QJSD require the same number of observations, as they both need the full density matrices to be computed. However, the QJSD turns out to be faster to compute than the Bures distance. In fact, the latter involves taking the square root of matrices, usually computed through matrix diagonalization which scales as $O(n^3)$, where n is the number of vertices in the graph. On the other hand, to compute the QJSD only the eigenvalues of ρ, σ and $\frac{\rho+\sigma}{2}$ are needed, which can be computed at $O(n^2)$.

IV. THE QJSD KERNEL

Let $G_1(V_1,E_1)$ and $G_2(V_2,E_2)$ be two unattributed graphs, i.e., graphs with no attributes or features attached to their nodes and edges. Given G_1 and G_2, we build 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 $v \in V_2$ (see Fig. 1 for an example). With this structure to hand, we define two independent continuous-time quantum walks with starting states

$$\begin{aligned} |\psi_0^-\rangle &= \frac{\sum_{u\in V_1} d_u|u\rangle - \sum_{v\in V_2} d_v|v\rangle}{C} \\ |\psi_0^+\rangle &= \frac{\sum_{u\in \mathcal{V}} d_u|u\rangle}{C}, \end{aligned} \tag{19}$$

where the basis state corresponding to the walk being at vertex $v \in \mathcal{V}$ is denoted as $|v\rangle$, d_v denotes the degree of vertex v, and C is the normalization constant such that the probabilities sum to 1. Intuitively, we set the initial amplitude on the nodes of G_1 and G_2 to be respectively in antiphase and in phase. That is, we design the initial states of the walks so as to highlight the presence of destructive and constructive interference patterns.

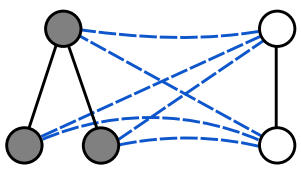


FIG. 1. (Color online) Given two graphs $G_1(V_1,E_1)$ and $G_2(V_2,E_2)$ we construct a new graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$ where $\mathcal{V} = V_1 \cup V_2$, $\mathcal{E} = E_1 \cup E_2$, and we add a new edge (u,v) between each pair of nodes $u \in V_1$ and $v \in V_2$.

We let the two quantum walks evolve under Eq. (5) until a time T and we define the average density operators ρ_T^- and ρ_T^+ as

$$\rho_T^- = \frac{1}{T}\int_0^T |\psi_t^-\rangle\langle\psi_t^-|\,dt, \quad \rho_T^+ = \frac{1}{T}\int_0^T |\psi_t^+\rangle\langle\psi_t^+|\,dt. \tag{20}$$

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

The rationale behind the proposed approach is that, whenever G_1 and G_2 are isomorphic, the distinguishability between the two states ρ_T^- and ρ_T^+, which emphasize respectively destructive and constructive interference, will be maximal. In other words, we design an experiment where the starting states are orthogonal and remain orthogonal during the quantum walk evolution, provided that G_1 and G_2 are isomorphic. Thus, given two unattributed graphs G_1 and G_2, we define the quantum Jensen-Shannon kernel $k_T(G_1,G_2)$ between them as

$$k_T(G_1,G_2) = D_{JS}(\rho_T^-,\rho_T^+), \tag{21}$$

where ρ_T^- and ρ_T^+ are the density operators defined as in Eq. (20). Note that this kernel is parametrized by the time T. As we will show in the next sections, the choice of the time parameter can affect both the computational complexity of the kernel and its classification accuracy. Finally, recall that in Eq. (19) we defined the initial state to be proportional to the node degree in the original graphs. As a consequence, the kernel is not defined on graphs $G = (V,E)$ with $E = \emptyset$, i.e., completely disconnected graphs.

Note that the proposed setting is significantly different from that of the classical Jensen-Shannon (JS) kernel of Bai and Hancock [11]. In fact, in Ref. [11] the authors need to construct a product graph from the two input graphs in order to compute the composite entropy of the system, whereas in our case this is not necessary. In Ref. [32] Bai *et al.* propose a quantum version of the classical JS kernel, however in order to guarantee the permutational invariance the authors need to compute the optimal alignment between the input graphs before the analysis can even commence. In our case, on the other hand, this problem is overcome by allowing the quantum walks to take place over the same merged structure and naturally measuring the similarity between the original graphs using the interference effects of quantum walks.

A. Kernel computation

In this section we evaluate the computational complexity of the kernel. In particular, we show that the solution to Eq. (20) can be computed analytically. Define $P_\lambda = \sum_{k=1}^{\mu(\lambda)} \phi_{\lambda,k}\phi_{\lambda,k}^\top$ to be the projection operator on the subspace spanned by the $\mu(\lambda)$ eigenvectors $\phi_{\lambda,k}$ associated with the eigenvalue $\lambda \in \Lambda$, where Λ is the set of eigenvalues of the Hamiltonian. Given this set of projectors, the unitary operator inducing the quantum walk can be rewritten as

$$U^t = \sum_\lambda e^{-i\lambda t} P_\lambda. \tag{22}$$

Recall that $|\psi_t\rangle = U^t|\psi_0\rangle$. Given Eq. (22) we can express the density matrix at time t in terms of the projectors P_λ, i.e.,

$$\rho_t = U^t\rho_0(U^t)^\dagger = \sum_{\lambda_1\in\Lambda}\sum_{\lambda_2\in\Lambda} e^{-i(\lambda_1-\lambda_2)t} P_{\lambda_1}\rho_0 P_{\lambda_2}^\top. \tag{23}$$

As a consequence, we can reformulate Eq. (20) as

$$\rho_T = \sum_{\lambda_1\in\Lambda}\sum_{\lambda_2\in\Lambda} P_{\lambda_1}\rho_0 P_{\lambda_2}^\top \frac{1}{T}\int_0^T e^{-i(\lambda_1-\lambda_2)t}\,dt. \tag{24}$$

The integral in Eq. (24) can be solved, yielding

$$\rho_T = \sum_{\lambda_1\in\Lambda}\sum_{\lambda_2\in\Lambda} P_{\lambda_1}\rho_0 P_{\lambda_2}^\top \frac{i(1-e^{iT(\lambda_2-\lambda_1)})}{T(\lambda_2-\lambda_1)}. \tag{25}$$

Letting $T\to\infty$, the integral in Eq. (24) reduces to the Dirac delta function $\delta(\lambda_1-\lambda_2)$. Hence, Eq. (24) simplifies to

$$\rho_\infty = \sum_{\lambda\in\tilde\Lambda} P_\lambda\rho_0 P_\lambda^\top, \tag{26}$$

where $\tilde\Lambda$ is the set of distinct eigenvalues of the Hamiltonian, i.e., the eigenvalues λ with multiplicity $\mu(\lambda) = 1$. Finally, along the same lines as Rossi *et al.* [25], one can show that as a consequence of Eq. (26) the infinite-time limit of the average density matrix commutes with the Hamiltonian $\mathcal{H}$, and thus the complexity of computing the Von Neumann entropy of ρ_∞, i.e., the Shannon entropy of its eigenvalues, is $O(\sum_{\lambda\in\tilde\Lambda}\mu(\lambda)^2)$, where $\mu(\lambda)$ is the multiplicity of the eigenvalue λ. As a consequence, we have that the complexity of computing the QJSD kernel with $T\to\infty$ is upper bounded by that of computing the eigendecomposition of $\mathcal{H}$, i.e., $O(|\mathcal{V}|^3)$. In the following sections, unless otherwise stated, we will assume that the kernel is computed for $T\to\infty$.

B. Kernel properties

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

Lemma 1. If G_1 and G_2 are two isomorphic graphs, then ρ_T^- and ρ_T^+ have support on orthogonal subspaces.

Proof. We need to prove that

$$(\rho_T^-)^\dagger\rho_T^+ = \frac{1}{T^2}\int_0^T \rho_{t_1}^-\,dt_1\int_0^T \rho_{t_2}^+\,dt_2 = \mathbf{0}, \tag{27}$$

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

$$\langle\psi_0^-|(U^t)^\dagger U^t|\psi_0^+\rangle = 0 \tag{28}$$

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

$$\langle\psi_0^-|U^{\Delta t}|\psi_0^+\rangle = 0, \tag{29}$$

where $\Delta_t = t_2 - t_1$. Let α_u^+ denote the amplitude of $|\psi_0^+\rangle$ at node u. To conclude the proof we rewrite the previous

equation as

$$\langle\psi_0^-|U^{\Delta t}|\psi_0^+\rangle = \sum_{u\in\mathcal{V}}\alpha_u^- \sum_{v\in\mathcal{V}}\alpha_v^+ U_{vu}^{\Delta t} = \sum_{u_1\in V_1}\alpha_{u_1}^+ \sum_{v_1\in\mathcal{V}}\alpha_{v_1}^+ U_{v_1u_1}^{\Delta t} - \sum_{u_2\in V_2}\alpha_{u_2}^+ \sum_{v_2\in\mathcal{V}}\alpha_{v_2}^+ U_{v_2u_2}^{\Delta t} = 0, \quad (30)$$

where we denote by $U_{uv}^{\Delta t}$ the element of the matrix $U^{\Delta t}$ corresponding to the pair of vertices u and v. Recall that two graphs $G_1(V_1,E_1)$ and $G_2(V_2,E_2)$ are isomorphic if there exists a bijection f between the elements of V_1 and V_2 such that $(u,v)\in E_1$ if and only if $(f(u),f(v))\in E_2$. To see that Eq. (30) holds, we start by noting that $U^{\Delta t}$ is a symmetric matrix and it is invariant to graph symmetries [25]. That is, if $f(u_1)=u_2$ and $f(v_1)=v_2$, then $U_{u_1v_1}^{\Delta t}=U_{u_2v_2}^{\Delta t}$. Moreover, if G_1 and G_2 are isomorphic then $\alpha_{u_1}^+=\alpha_{f(u_1)}^+$. Thus, it follows that in Eq. (30) each term $\alpha_{u_1}^+\alpha_{v_1}^+U_{v_1u_1}^{\Delta t}$ cancels out with a term $\alpha_{f(u_1)}^+\alpha_{f(v_1)}^+U_{f(v_1)f(u_1)}^{\Delta t}$, $\forall\, u_1\in U_1, v_1\in\mathcal{V}$. ■

Corollary 1. Given a pair of graphs G_1 and G_2, the kernel satisfies the following properties: (1) $0\leqslant k(G_1,G_2)\leqslant 1$ 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. (21), 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 the quantum Jensen-Shannon divergence is bounded to lie between 0 and 1.

The second property follows again from Eq. (21) and Lemma 1. In other words, if the density operators have support on orthogonal spaces then the quantum Jensen-Shannon divergence reaches its maximum value. To see this, it is sufficient to note that from $(\rho_T^-)^\dagger\rho_T^+=\mathbf{0}$ it follows that the set of eigenvalues of $\frac{\rho_T^-+\rho_T^+}{2}$ is

$$\Lambda\left(\frac{\rho_T^-+\rho_T^+}{2}\right)=\left\{\frac{\lambda}{2}|\lambda\in\Lambda(\rho_T^+)\cup\Lambda(\rho_T^-)\right\},$$

where $\Lambda(\rho_T^+)$ and $\Lambda(\rho_T^-)$ denote the set of eigenvalues of ρ_T^+ and ρ_T^-, respectively. Let $\lambda^+\in\Lambda(\rho_T^+)$ and $\lambda^-\in\Lambda(\rho_T^-)$. Then, the Von Neumann entropy of $\frac{\rho_T^-+\rho_T^+}{2}$ is

$$\begin{aligned} H_N\left(\frac{\rho_T^-+\rho_T^+}{2}\right) &= -\sum_{\lambda^+\in\Lambda(\rho_T^+)}\frac{\lambda^+}{2}\ln\frac{\lambda^+}{2}-\sum_{\lambda^-\in\Lambda(\rho_T^-)}\frac{\lambda^-}{2}\ln\frac{\lambda^-}{2}\\ &= -\frac{1}{2}\sum_{\lambda^+\in\Lambda(\rho_T^+)}\lambda^+\ln\lambda^+-\frac{1}{2}\sum_{\lambda^-\in\Lambda(\rho_T^-)}\lambda^-\ln\lambda^-+1\\ &= \frac{1}{2}H_N(\rho_T^+)+\frac{1}{2}H_N(\rho_T^-)+1 \end{aligned} \quad (31)$$

and, as a consequence, $D_{JS}(\rho_T^-,\rho_T^+)=1$. ■

Unfortunately, we are currently unable to provide a formal proof of the positive semidefiniteness of our kernel, although we give empirical evidence of this in the experimental section. Some potentially useful results can be found in the work of Briet *et al.* [38], where the authors prove that the classical Jensen-Shannon divergence and its quantum mechanical counterpart can be used as negative semidefinite kernels between probability distributions and pure quantum states, respectively. Note that because of how we set up our experiment the more divergent (i.e., dissimilar) are ρ_T^- and ρ_T^+, the more similar are the input graphs. Recall that whenever G_1 and G_2 are isomorphic, then the states of the walks remain orthogonal during the entire evolution process. Thus, the negative semidefiniteness of the QJSD kernel between quantum states implies positive semidefiniteness of the kernel between graphs.

In order to shed light on the general behavior of the kernel, we now analyze the special case in which we use the Laplacian of the merged graph as the Hamiltonian for the quantum walk. Under these assumptions, we can express the Hamiltonian $\mathcal{H}$ in terms of the Laplacians L_1 and L_2 of graphs G_1 and G_2 respectively:

$$\mathcal{H}=\begin{pmatrix}L_1+nI_m & -\mathbf{1}_m\mathbf{1}_n^T\\ -\mathbf{1}_n\mathbf{1}_m^T & L_2+mI_n\end{pmatrix}, \quad (32)$$

where m and n are the sizes of graphs G_1 and G_2, respectively, I_n denotes an $n\times n$ identity matrix, and $\mathbf{1}_n$ is an n-dimensional vector with all unit entries.

Recall that a graph Laplacian has a single *all 1's* eigenvector with unit constant elements, corresponding to a zero eigenvalue. All the remaining eigenvectors have elements that sum to zero. Let $|\phi\rangle$ be a nonconstant eigenvector of L_1 of eigenvalue λ, and $|\chi\rangle$ a nonconstant eigenvector of L_2 of eigenvalue μ. It is easy to show that the vectors $\bar{\phi}=(\phi^T,\mathbf{0}_n^T)^T$ and $\bar{\chi}=(\mathbf{0}_m^T,\chi^T)^T$ are eigenvectors of $\mathcal{H}$, where $\mathbf{0}_n$ is the vector composed of n zeros, and $\bar{\phi}$ and $\bar{\chi}$ are obtained by adding n trailing zeros to ϕ and m leading zeros to χ, respectively. In fact:

$$\mathcal{H}\bar{\phi}=\begin{pmatrix}L_1+nI_m & -\mathbf{1}_m\mathbf{1}_n^T\\ -\mathbf{1}_n\mathbf{1}_m^T & L_2+mI_n\end{pmatrix}\begin{pmatrix}\phi\\ \mathbf{0}_n\end{pmatrix} = \begin{pmatrix}(\lambda+n)\phi+\mathbf{0}_m\\ \mathbf{0}_n+\mathbf{0}_n\end{pmatrix}=(\lambda+n)\bar{\phi}, \quad (33)$$

$$\mathcal{H}\bar{\chi}=\begin{pmatrix}L_1+nI_m & -\mathbf{1}_m\mathbf{1}_n^T\\ -\mathbf{1}_n\mathbf{1}_m^T & L_2+mI_n\end{pmatrix}\begin{pmatrix}\mathbf{0}_m\\ \chi\end{pmatrix} = \begin{pmatrix}\mathbf{0}_m+\mathbf{0}_m\\ \mathbf{0}_n+(\mu+m)\chi\end{pmatrix}=(\mu+n)\bar{\chi}. \quad (34)$$

This characterizes $m+n-2$ eigenvectors of $\mathcal{H}$. The remaining two eigenvectors are $(\mathbf{1}_m^T,\mathbf{1}_n^T)$ with eigenvalue 0, and $(n\mathbf{1}_m^T,-m\mathbf{1}_n^T)^T$ with eigenvalue $m+n$:

$$\begin{pmatrix}L_1+nI_m & -\mathbf{1}_m\mathbf{1}_n^T\\ -\mathbf{1}_n\mathbf{1}_m^T & L_2+mI_n\end{pmatrix}\begin{pmatrix}\mathbf{1}_m\\ \mathbf{1}_n\end{pmatrix} = \begin{pmatrix}n\mathbf{1}_m-n\mathbf{1}_m\\ -m\mathbf{1}_n+m\mathbf{1}_n\end{pmatrix}=\mathbf{0}_{n+m}, \quad (35)$$

$$\begin{pmatrix}L_1+nI_m & -\mathbf{1}_m\mathbf{1}_n^T\\ -\mathbf{1}_n\mathbf{1}_m^T & L_2+mI_n\end{pmatrix}\begin{pmatrix}n\mathbf{1}_m\\ -m\mathbf{1}_n\end{pmatrix} = \begin{pmatrix}n^2\mathbf{1}_m+nm\mathbf{1}_m\\ -nm\mathbf{1}_n-m^2\mathbf{1}_n\end{pmatrix}=(m+n)\begin{pmatrix}n\mathbf{1}_m\\ -m\mathbf{1}_n\end{pmatrix}. \quad (36)$$

Note that $m+n$ is an upper bound for an eigenvalue of the Laplacian of a graph of size $m+n$, and it is reached only if G_1 and G_2 have nodes that are connected to every other node in the graph, while the eigenvalue 0 is isolated as the merged graph is connected. In this analysis we will ignore the possibility of multiple $m+n$ eigenvectors and will assume that there is no node connected to every other node. However, even with these restrictions, the analysis will be sufficient to draw some conclusions that apply in the general case as well.

With the decomposition of the Hamiltonian to hand, we can compute the infinite-time average density matrix analytically. In fact, let P^1_λ be the projection onto the eigenspace corresponding to eigenvalue λ of graph G_1 and, equivalently, P^2_μ be the projector onto the μ eigenspace of graph G_2. Further, let $\bar{P}^1_\lambda$ and $\bar{P}^2_\mu$ be the extension of the projector on the merged graph, and

$$\bar{P}_0 = \frac{\left|\left(\mathbf{1}_m^T,\mathbf{1}_n^T\right)^R\right\rangle\!\left\langle\left(\mathbf{1}_m^T,\mathbf{1}_n^T\right)^T\right|}{m+n},$$

$$\bar{P}_{m+n} = \frac{\left|\left(n\mathbf{1}_m^T,-m\mathbf{1}_n^T\right)^T\right\rangle\!\left\langle\left(n\mathbf{1}_m^T,-m\mathbf{1}_n^T\right)^T\right|}{mn(m+n)}$$

be the projectors onto the 0 and $m+n$ eigenspaces of the merged graph. Let us now define the following sets of eigenvalues of the merged graph:

$$\Lambda_{1\setminus 2} = \{\lambda \in \tilde{\Lambda}(G_1)\setminus\{0\} \,|\, \forall\, \mu \in \tilde{\Lambda}(G_2)\setminus 0, \lambda + n \neq \mu + m\}, \tag{37}$$

$$\Lambda_{2\setminus 1} = \{\mu \in \tilde{\Lambda}(G_2)\setminus\{0\} \,|\, \forall\, \lambda \in \tilde{\Lambda}(G_1)\setminus 0, \lambda + n \neq \mu + m\}, \tag{38}$$

$$\Lambda_{1\cap 2} = \{\lambda \in \tilde{\Lambda}(G_1)\setminus\{0\} \,|\, \exists\, \mu \in \tilde{\Lambda}(G_2)\setminus 0, \lambda + n = \mu + m\}. \tag{39}$$

Further, let ϕ_0 and χ_0 be the restrictions of the initial pure state on G_1 and G_2, respectively; then we have $\rho_0^+ = |(\phi_0^T,\chi_0^T)^T\rangle\langle(\phi_0^T,\chi_0^T)^T|$ and $\rho_0^- = |(\phi_0^T,-\chi_0^T)^T\rangle\langle(\phi_0^T,-\chi_0^T)^T|$.

From Eq. (26), we have

$$\begin{aligned}
\rho_\infty^+ &= \sum_{\lambda\in\Lambda_{1\setminus 2}} \bar{P}^1_\lambda\rho_0^+\bar{P}^1_\lambda + \sum_{\mu\in\Lambda_{1\setminus 2}} \bar{P}^2_\mu\rho_0^+\bar{P}^2_\mu + \sum_{\lambda\in\Lambda_{1\cap 2}} \left(\bar{P}^1_\lambda + \bar{P}^2_\mu\right)\rho_0^+\left(\bar{P}^1_\lambda + \bar{P}^2_\mu\right) + \bar{P}_0\rho_0^+\bar{P}_0 + \bar{P}_{m+n}\rho_0^+\bar{P}_{m+n} \\
&= \sum_{\lambda\in\tilde{\Lambda}(G_1)\setminus\{0\}} \left|\left(\left(P^1_\lambda\phi_0\right)^T,\mathbf{0}_n^T\right)^T\right\rangle\!\left\langle\left(\left(P^1_\lambda\phi_0\right)^T,\mathbf{0}_n^T\right)^T\right| + \sum_{\mu\in\tilde{\Lambda}(G_2)\setminus\{0\}} \left|\left(\mathbf{0}_m^T,\left(P^2_\mu\chi_0\right)^T\right)^T\right\rangle\!\left\langle\left(\mathbf{0}_m^T,\left(P^2_\mu\chi_0\right)^T\right)^T\right| \\
&\quad + \sum_{\lambda\in\Lambda_{1\cap 2}} \left[\left|\left(\left(P^1_\lambda\phi_0\right)^T,\mathbf{0}_n^T\right)^T\right\rangle\!\left\langle\left(\mathbf{0}_m^T,\left(P^2_\mu\chi_0\right)^T\right)^T\right| + \left|\left(\mathbf{0}_m^T,\left(P^2_\mu\chi_0\right)^T\right)^T\right\rangle\!\left\langle\left(\left(P^1_\lambda\phi_0\right)^T,\mathbf{0}_n^T\right)^T\right|\right] \\
&\quad + \left(\frac{\langle\mathbf{1}_m\,|\,\phi_0\rangle^2}{m+n} + \frac{\langle\mathbf{1}_n\,|\,\chi_0\rangle^2}{m+n} + 2\frac{\langle\mathbf{1}_m\,|\,\phi_0\rangle\langle\mathbf{1}_n\,|\,\chi_0\rangle}{m+n}\right)\bar{P}_0 + \left(\frac{n}{m}\frac{\langle\mathbf{1}_m\,|\,\phi_0\rangle^2}{(m+n)} + \frac{m}{n}\frac{\langle\mathbf{1}_n\,|\,\chi_0\rangle^2}{(m+n)} - 2\frac{\langle\mathbf{1}_m\,|\,\phi_0\rangle\langle\mathbf{1}_n\,|\,\chi_0\rangle}{(m+n)}\right)\bar{P}_{m+n}
\end{aligned} \tag{40}$$

and, symmetrically,

$$\begin{aligned}
\rho_\infty^- &= \sum_{\lambda\in\tilde{\Lambda}(G_1)\setminus\{0\}} \left|\left(\left(P^1_\lambda\phi_0\right)^T,\mathbf{0}_n^T\right)^T\right\rangle\!\left\langle\left(\left(P^1_\lambda\phi_0\right)^T,\mathbf{0}_n^T\right)^T\right| + \sum_{\mu\in\tilde{\Lambda}(G_2)\setminus\{0\}} \left|\left(\mathbf{0}_m^T,\left(P^2_\mu\chi_0\right)^T\right)^T\right\rangle\!\left\langle\left(\mathbf{0}_m^t,\left(P^2_\mu\chi_0\right)^T\right)^T\right| \\
&\quad - \sum_{\lambda\in\Lambda_{1\cap 2}} \left[\left|\left(\left(P^1_\lambda\phi_0\right)^T,\mathbf{0}_n^T\right)^T\right\rangle\!\left\langle\left(\mathbf{0}_m^T,\left(P^2_\mu\chi_0\right)^T\right)^T\right| + \left|\left(\mathbf{0}_m^T,\left(P^2_\mu\chi_0\right)^T\right)^T\right\rangle\!\left\langle\left(\left(P^1_\lambda\phi_0\right)^T,\mathbf{0}_n^T\right)^T\right|\right] \\
&\quad + \left(\frac{\langle\mathbf{1}_m\,|\,\phi_0\rangle^2}{m+n} + \frac{\langle\mathbf{1}_n\,|\,\chi_0\rangle^2}{m+n} - 2\frac{\langle\mathbf{1}_m\,|\,\phi_0\rangle\langle\mathbf{1}_n\,|\,\chi_0\rangle}{m+n}\right)\bar{P}_0 + \left(\frac{n}{m}\frac{\langle\mathbf{1}_m\,|\,\phi_0\rangle^2}{(m+n)} + \frac{m}{n}\frac{\langle\mathbf{1}_n\,|\,\chi_0\rangle^2}{(m+n)} + 2\frac{\langle\mathbf{1}_m\,|\,\phi_0\rangle\langle\mathbf{1}_n\,|\,\chi_0\rangle}{(m+n)}\right)\bar{P}_{m+n}.
\end{aligned} \tag{41}$$

Hence, for the mixed density matrix, we have

$$\begin{aligned}
\frac{\rho_\infty^+ + \rho_\infty^-}{2} &= \sum_{\lambda\in\tilde{\Lambda}(G_1)\setminus\{0\}} \left|\left(\left(P^1_\lambda\phi_0\right)^T,\mathbf{0}_n^T\right)^T\right\rangle\!\left\langle\left(\left(P^1_\lambda\phi_0\right)^T,\mathbf{0}_n^T\right)^T\right| + \sum_{\mu\in\tilde{\Lambda}(G_2)\setminus\{0\}} \left|\left(\mathbf{0}_m^T,\left(P^2_\mu\chi_0\right)^T\right)^T\right\rangle\!\left\langle\left(\mathbf{0}_m^T,\left(P^2_\mu\chi_0\right)^T\right)^T\right| \\
&\quad + \left(\frac{\langle\mathbf{1}_m\,|\,\phi_0\rangle^2}{m+n} + \frac{\langle\mathbf{1}_n\,|\,\chi_0\rangle^2}{m+n}\right)\bar{P}_0 + \left(\frac{n}{m}\frac{\langle\mathbf{1}_m\,|\,\phi_0\rangle^2}{(m+n)} + \frac{m}{n}\frac{\langle\mathbf{1}_n\,|\,\chi_0\rangle^2}{(m+n)}\right)\bar{P}_{m+n}.
\end{aligned} \tag{42}$$

From these equations we can see that there are two sources for the differences between ρ_∞^+ and ρ_∞^-. The first is in the interference in the eigenspaces associated with the 0 and $n+m$ eigenvalues of the Hamiltonian resulting in a movement of a component of magnitude $\frac{2}{m+n}\langle\mathbf{1}_m\,|\,\phi_0\rangle\langle\mathbf{1}_n\,|\,\chi_0\rangle$ from one space to the other, and is thus linked with the average degree of the two graphs. The second is associated with a form of size-adjusted cospectrality, i.e., the situation in which $\lambda + n = \mu + m$ holds. In this situation, graphs with disjoint size-adjusted spectrum will have minimal values of the kernel.

In the situation of a general Hamiltonian, we cannot provide such a detailed result. However, we note that the situation that allows for differences between ρ_∞^+ and ρ_∞^- arises when the supports of the eigenspaces of the Hamiltonian span both

graphs, and in that sense, the kernel measures the extent to which the eigenspaces of the Hamiltonian simultaneously span both graphs.

V. EXPERIMENTAL RESULTS

In this section we measure the performance of the QJSD kernel in a classification task. More specifically, we make use of the following five different standard graph data sets:

MUTAG. This is a data set of 188 chemical compounds labeled according to whether or not they affect the frequency of genetic mutations in the bacterium *Salmonella typhimurium* [39].

PPIs. The PPIs data set consists of protein-protein interaction networks (PPIs) [40]. The nodes of the graphs correspond to proteins, and two nodes are connected if they have a direct (physical) or indirect (functional) association. The original data set consists of 219 PPIs collected from five different kinds of bacteria. More specifically, there are eight PPIs from *Aquifex aelicus* and *Thermotoga maritima*, 52 PPIs from *Gram-positive Staphylococcus aureus* and *Cyanobacteria*, 73 PPIs from *Anabaena variabilis* and *Proteobacteria*, and 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. Here we consider the task of discriminating among the 40 PPIs from *Acidovorax avenae* and the 46 PPIs from *Acidobacteria*.

PTC. The predictive toxicology challenge data set (PTC) records the carcinogenicity of several hundred chemical compounds for male rats (MRs), female rats, male mice, and female mice [41]. For our experiments we select the graphs of male rats. There are a total of 344 graphs in the MR class, with the smallest one having only two nodes and the largest one having 109 nodes.

COIL5. The COIL database consists of images of 100 objects [42]. In our experiments, we use the images for the first five objects. For each of these objects we consider 72 images from different viewpoints. Each image is converted into a graph by computing the Delaunay triangulation of the set of corner points extracted using the Harris detector.

Shock. The shock data set consists of graphs from a database of two-dimensional (2D) shapes [43]. 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. Some statistics concerning the datasets are given in Table I.

TABLE I. Information on the graph data sets.

Data sets	MUTAG	PPI	PTC	COIL	Shock
Max no. of vertices	28	232	109	241	33
Min no. of vertices	10	3	2	72	4
Average no. of vertices	17.93	109.60	25.56	144.97	13.16
No. of graphs	188	86	344	360	150
No. of classes	2	2	2	5	10

To these real-world data sets we add a set of 30 synthetically generated graphs belonging to three different classes, where the graphs belonging to each class were sampled from a generative model with 12, 14, and 16 nodes, respectively. Here a generative model consists of a graph where each node and edge is labeled with the probability of observing that node and edge, respectively. Details concerning the generative model can be found in Ref. [44].

We use a binary C-support vector machine (C-SVM) to test the efficacy of the QJSD kernel [45]. More specifically, we perform tenfold 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.

Furthermore, we compare the performance of the kernel with that of a number of well-known alternative graph kernels, namely, the shortest-path kernel [9], the classic random walk kernel [8], the graphlet kernel [10], and the the Weisfeiler-Lehman subtree kernel [46], where here a *graphlet* refers to a subgraph with $k \in \{3,4,5\}$ nodes.

With the exception of the Weisfeiler-Lehman kernel, none of the kernels examined in these experiments take graph attributes into account. In the case of the Weisfeiler-Lehman, however, each node is labeled with its degree. As for the maximum subtree height h, in our experiments we let $h = \{1–3\}$ and we choose the optimal value by cross-validation [46]. Finally,

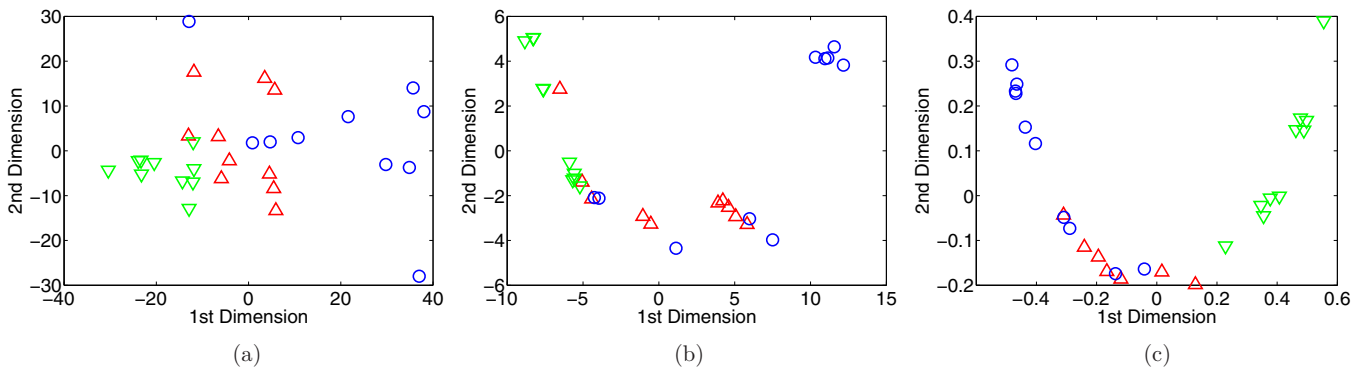
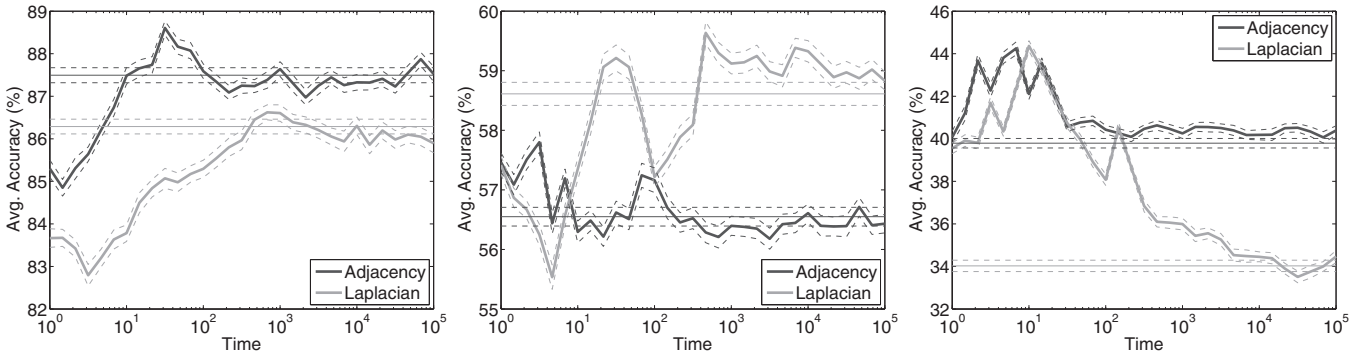


FIG. 2. (Color online) Two-dimensional MDS embeddings of the synthetic data, where the elements of the three classes are indicated by blue circles, red up triangles, and green down triangles, respectively. Here the axes correspond to the embedding dimensions given by the two largest eigenvalues of the similarity matrix $M = -\frac{1}{2}HDH$, where the distance matrix D denotes (a) the edit distance, (b) the distance between the graph spectra, and (c) the QJSD kernel, respectively.



(a) MUTAG (b) PTC (c) Shock

FIG. 3. The average classification accuracy (± standard error, dashed line) as a function of the stopping time T. Here the dark gray and light gray lines refer to the cases in which the Hamiltonians are the adjacency and Laplacian matrix, respectively, and the horizontal line shows the classification accuracy when $T \to \infty$.

note that the implementation of the graphlet kernel used in this paper counts the instances of graphlets of size 3.

Figure 2 shows multidimensional scaling (MDS) [47] embeddings of the synthetic graphs. MDS is a standard approach for linear dimensionality reduction which aims at locating the low-dimensional embedding into a vector space that best preserves the interpoint distances given by a distance matrix D. To this end, the matrix $M = -\frac{1}{2}HDH$ is first computed, where $H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ is the centering matrix, $\mathbf{1}$ denotes the m-dimensional vector of all 1's, and n is the number of points to embed. Then, for each embedded point the jth coordinate is determined by the jth eigenvector and the jth eigenvalue of M, where the eigenvalues are sorted in decreasing order. Let K denote the QJSD kernel over the set of synthetic graphs, where the ijth element of K is denoted as k_{ij}, i.e., k_{ij} is the kernel value between graph i and graph j. We then compute the distance matrix $D = (d_{ij})$ with $d_{ij} = \sqrt{k_{ii} + k_{jj} - 2k_{ij}} = 2(1 - kij)$, to which we apply MDS to identify the principal axes of structural variation represented by the similarity data. Note that here we set the Hamiltonian to be the adjacency matrix of the graph. We then compare the resulting embedding with those obtained starting from the edit distance matrix [48] and the graph spectra distance matrix. Given a set of permitted edit operations, i.e., addition or deletion of a vertex or an edge of a graph, the edit distance between two graphs is defined as the least-cost edit operations sequence needed that transforms a graph into another one [49]. To compute the distance between the graph spectra, on the other hand, we adopt the following procedure. For each graph G with adjacency matrix A, we compute the column vector s_G of the ordered eigenvalues of A. As the graphs are of different sizes and thus their spectra are of different lengths, the vectors are all made to be the same length by padding zeros to the end of the shorter vector. The (i, j)th element of the distance matrix is then $d_{ij} = ||s_i - s_i||$. Figure 2 from left to right shows the MDS embeddings associated with (a) the edit distance, (b) the distance between the graph spectra, and (c) the QJSD kernel, respectively. Compared to the distance between the graph spectra, the QJSD kernel provides a clearer separation between the different classes. This in turn suggests that the Hilbert space induced by the QJSD kernel efficiently captures the class separation of the data set. Finally, note that while the edit distance also yields a good class separation, it requires computing a mapping between the node sets of the graphs. This in turn is known to be a particularly hard task, due to the combinatorial nature of the problem.

We now evaluate how the accuracy of the QJSD kernel varies as we let the quantum walks evolve for longer times T. Recall that the value of the QJSD kernel depends on the choice of the stopping time of the quantum walks. In the previous section, we showed that in the large-time limit the density matrix of the quantum walk commutes with the Hamiltonian, and thus the value of the QJSD kernel can be readily computed given the eigendecomposition of $\mathcal{H}$. This reduced computational complexity in turn motivates the choice of $T \to \infty$. However, as seen in Fig. 3, this does not necessarily coincide with the optimal time, i.e., the time that leads to the highest classification performance. Recall that in a classification task we can measure the performance of a classifier in terms of *classification accuracy*. This is defined as the fraction of the data instances that are assigned the correct class label, where in our case a data instance corresponds to a graph. Here we plot the average classification accuracy (± standard error) for the QJSD kernel as a function of the time parameter T on the MUTAG, PTC, and shock data sets, where the Hamiltonian is chosen to be either the adjacency matrix or the Laplacian. Note that these three data sets are the only ones that allow us to perform an extensive analysis of the parameters. The plots show that almost invariably the accuracy is reaching a maximum before stabilizing around a limit.

Interestingly, for the MUTAG and shock data sets, when we take the adjacency matrix as the Hamiltonian the maximum accuracy is achieved for a time T which is of the same order of magnitude as the average graph size in the data set. However, the same does not hold for the PTC data set, where we note that the graph size shows a larger variability. Future work will try to identify a possible connection between this observed behavior and the average mixing time of the continuous-time quantum walks on the graphs. Note also that for the PTC data set the best classification accuracy is achieved when the Laplacian is the Hamiltonian of the system. In the shock data set, on the other hand, in the large-time limit the adjacency-based QJSD kernel outperforms the Laplacian-based one, while the maximum accuracy of the Laplacian-based QJSD kernel equals that of

TABLE II. Classification accuracy (± standard error) on unattributed graph data sets. $\text{QJSD}_{\mathcal{H}}$ is the QJSD kernel, where $\mathcal{H}$ denotes the Hamiltonian, i.e., adjacency matrix (A) or Laplacian (L), SP is the shortest-path kernel [9], RW is the random walk kernel [8], GR_3 denotes the graphlet kernel [10] computed using all graphlets of size 3, while WL is the Weisfeiler-Lehman subtree kernel [46]. We also denote by $\text{QJSD}^*_{\mathcal{H}}$ the optimal QJSD kernel (with respect to the time) with Hamiltonian $\mathcal{H}$.

Kernel	MUTAG	PPI	PTC	Shock	COIL5
QJSD_A	87.41 ± 0.18	75.90 ± 0.39	56.38 ± 0.15	40.70 ± 0.21	69.17 ± 0.11
QJSD_L	86.16 ± 0.19	71.26 ± 0.46	$\mathbf{58.76 \pm 0.17}$	33.96 ± 0.24	$\mathbf{69.80 \pm 0.10}$
QJSD^*_A	$\mathbf{88.81 \pm 0.17}$			$\mathbf{44.21 \pm 0.22}$	
QJSD^*_L	86.56 ± 0.20			44.12 ± 0.20	
SP	83.02 ± 0.21	63.81 ± 0.39	56.07 ± 0.13	38.50 ± 0.16	69.44 ± 0.16
RW	66.41 ± 0.07	51.21 ± 0.37	55.79 ± 0.03	0.41 ± 0.05	12.11 ± 0.10
GR_3	81.31 ± 0.20	49.96 ± 0.43	55.41 ± 0.08	26.05 ± 0.32	66.62 ± 0.15
WL	84.54 ± 0.26	$\mathbf{82.81 \pm 0.41}$	55.66 ± 0.19	37.68 ± 0.24	31.59 ± 0.21

the adjacency-based one. Finally, note that with the exception of the PTC data set, the difference between the asymptote and the peak is relatively small, albeit statistically significant. This in turn suggests that the choice of $T \to \infty$ produces a classification accuracy close to the optimum.

Finally, we report the average classification accuracies (± standard error) of the different kernels in Table II. We evaluate the accuracy of the QJSD kernel for different choices of the time T and of the Hamiltonian. More specifically, we let the Hamiltonian be either the adjacency or the Laplacian matrix, and the time be either the optimal one or $T \to \infty$. We select the value of T through an exhaustive search, where the optimal value is chosen by cross-validation on the training set. Here we let T vary logarithmically between 1 and 10^6, which is indeed computationally very expensive. However, at the present moment we do not have any exact or even heuristic criterion to reduce the search space, and future work will be aimed at this. As shown in Table II, in fact, we were able to complete the simulation only on the MUTAG and Shock data sets. Interestingly, in both cases we achieved the best performance by fine tuning the value of T. With respect to the Hamiltonian, we observe once again that the adjacency matrix is not always the best choice. Generally, however, the QJSD kernel obtains a classification accuracy which is better than or sometimes comparable to that of the other kernels, regardless of how we set T and $\mathcal{H}$.

In terms of computational complexity, recall that, for each pair of graphs $G_1(V_1,E_1)$ and $G_2(V_2,E_2)$, we need to compute the complete eigendecomposition of the merged graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$. For the sake of clarity, let us assume that we are given a data set of N graphs each with $|V|$ nodes. When the Hamiltonian is the adjacency matrix, we need to explicitly compute the eigendecomposition of the merged graph for each pair of input graphs, and thus the overall complexity is $O(N^2|V|^3)$. However, when the Hamiltonian is the graph Laplacian, it is possible to precompute the eigendecomposition of the individual graphs, thus lowering the overall complexity to $O(N|V|^3)$. Similarly, the most efficient implementations of the shortest-path kernel and the graphlet kernel [10,46] with graphlets of size 3 scale as $O(N|V|^3)$. The random walk kernel [50], on the other hand, scales as $O(N^2|V|^3))$, while the Weisfeiler-Lehman subtree kernel [46] scales as $O(h|V|^2N^2)$, with h denoting the maximum height of the subtrees.

VI. CONCLUSIONS

Graph-based representations are widely used as a powerful tool for modeling and analyzing real-world complex systems. However, the rich expressiveness of graphs poses a number of problems when the application of pattern recognition and machine learning techniques is considered. Although kernel methods provide a way to shift this representational problem, the design of novel and efficient graph kernels remains an open challenge. In this paper, we proposed a quantum-inspired kernel for unattributed graphs where we gauged the similarity between the input structures through continuous-time quantum walks. More specifically, we computed the divergence between two suitably defined quantum states, and we showed that this measure has a number of interesting properties related to the spectrum of the system Hamiltonian. In particular, we have shown that the kernel value can be interpreted as a measure of the extent to which the eigenspaces of the Hamiltonian simultaneously span both graphs.

Our experimental validation has shown that the QJSD kernel can outperform state-of-the-art kernels in a graph classification task. Although the overall computational complexity is bounded by that of computing the eigendecomposition of the Hamiltonian, we have shown that we can compute the eigenvalues of ρ_∞ in $O(\sum_{\lambda\in\tilde{\Lambda}} \mu(\lambda)^2)$, where $\tilde{\Lambda}$ denotes the set of distinct eigenvalues λ of the Hamiltonian, each with multiplicity $\mu(\lambda)$. Despite leading to a faster computation, we observed that the choice of letting $T \to \infty$ does not necessarily coincide with the optimal one in terms of classification accuracy. We noted that, when we let the adjacency matrix be the Hamiltonian of the system, the maximum accuracy is actually achieved for a time T which is of the same order of magnitude as the average graph size in the data set. Future work should investigate the possibility of a relation between the average mixing time of the continuous-time quantum walks on the merged graph and the optimal time for the kernel. Finally, the positive semidefiniteness of the kernel, which at the moment we are unable to formally prove, will be the subject of further studies.

ACKNOWLEDGMENT

E.H. was supported by a Wolfson Research Merit Award from the UK's Royal Society (ref: WRMA09R2/HLL).

[1] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker, Shock graphs and shape matching, Int. J. Comput. Vision **35**, 13 (1999).

[2] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási, The large-scale organization of metabolic networks, Nature (London) **407**, 651 (2000).

[3] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, A comprehensive two-hybrid analysis to explore the yeast protein interactome, Proc. Natl. Acad. Sci. USA **98**, 4569 (2001).

[4] V. Kalapala, V. Sanwalani, A. Clauset, and C. Moore, Scale invariance in road networks, Phys. Rev. E **73**, 026130 (2006).

[5] B. Schölkopf and A. J. Smola, *Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, MA, 2001).

[6] V. N. Vapnik, *Statistical Learning Theory* (Wiley-Interscience, 1998).

[7] D. Haussler, Convolution kernels on discrete structures, Technical report UCS-CRL-99-10, UC Santa Cruz, 1999 (unpublished).

[8] T. Gärtner, P. Flach, and S. Wrobel, On graph kernels: Hardness results and efficient alternatives, in *Learning Theory and Kernel Machines*, edited by B. Schölkopf and M. K. Warmuth (Springer, Berlin, 2003), pp. 129–143.

[9] K. M. Borgwardt and H. P. Kriegel, Shortest-path kernels on graphs, in *Fifth IEEE International Conference on Data Mining*, edited by J. Han and B. Wah (IEEE, New York, 2005) p. 8.

[10] N. Shervashidze, S. Vishwanathan, T. Petri, K. Mehlhorn, and K. Borgwardt, Efficient graphlet kernels for large graph comparison, in *Proceedings of the International Workshop on Artificial Intelligence and Statistics*, edited by D. van Dyk and M. Welling (Society for Artificial Intelligence and Statistics, Clearwater Beach, Florida, 2009).

[11] L. Bai and E. R. Hancock, Graph kernels from the Jensen-Shannon divergence, J. Math. Imaging Vision **47**, 60 (2013).

[12] A. F. T. Martins, N. A. Smith, E. P. Xing, P. M. Q. Aguiar, and M. A. T. Figueiredo, Nonextensive information theoretic kernels on measures, J. Mach. Learn. Res. **10**, 935 (2009).

[13] F. Passerini and S. Severini. The Von Neumann entropy of networks, arXiv:0812.2597v2.

[14] J. Kempe, Quantum random walks: An introductory overview, Contemp. Phys. **44**, 307 (2003).

[15] A. Ambainis, Quantum walks and their algorithmic applications, Int. J. Quantum. Inf. **01**, 507 (2003).

[16] A. M. Childs, Universal computation by quantum walk, Phys. Rev. Lett. **102**, 180501 (2009).

[17] M. Santha, Quantum walk based search algorithms, in *Theory and Applications of Models of Computation*, edited by M. Agrawal, D.-Z. Du, Z. Duan, and A. Li (Springer, Berlin, 2008), pp. 31–46.

[18] O. Mülken and A. Blumen, Continuous-time quantum walks: Models for coherent transport on complex networks, Phys. Rep. **502**, 37 (2011).

[19] V. Kendon, Decoherence in quantum walks—-a review, Math. Structures Comput. Sci. **17**, 1169 (2007).

[20] N. Shenvi, J. Kempe, and K. Birgitta Whaley, Quantum random-walk search algorithm, Phys. Rev. A **67**, 052307 (2003).

[21] E. Farhi and S. Gutmann, Quantum computation and decision trees, Phys. Rev. A **58**, 915 (1998).

[22] H. Krovi and T. A. Brun, Quantum walks with infinite hitting times, Phys. Rev. A **74**, 042334 (2006).

[23] D. Emms, R. C. Wilson, and E. R. Hancock, Graph embedding using a quasi-quantum analog of the hitting times of continuous time quantum walks, Quantum Inf. Comput. **9**, 231 (2009).

[24] L. Rossi, A. Torsello, and E. R. Hancock, Approximate axial symmetries from continuous time quantum walks, in *Structural, Syntactic, and Statistical Pattern Recognition*, edited by G. L. Gimel'farb, E. R. Hancock, A. Imiya, A. Kuijper, M. Kudo, S. Omachi, T. Windeatt, and K. Yamada (Springer, Berlin Heidelberg, 2012), pp. 144–152.

[25] L. Rossi, A. Torsello, E. R. Hancock, and R. C. Wilson, Characterizing graph symmetries through quantum Jensen-Shannon divergence, Phys. Rev. E **88**, 032806 (2013).

[26] A. Majtey, P. W. Lamberti, M. T. Martin, and A. Plastino, Wootters' distance revisited: A new distinguishability criterium, Eur. Phys. J. D **32**, 413 (2005).

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

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

[29] J. Lin, Divergence measures based on the Shannon entropy, IEEE Trans. Inf. Theory **37**, 145 (1991).

[30] L. Rossi, A. Torsello, and E. R. Hancock, A continuous-time quantum walk kernel for unattributed graphs, in *Graph-Based Representations in Pattern Recognition*, edited by W. G. Kropatsch, N. M. Artner, Y. Haxhimusa, and X. Jiang (Springer, Berlin, 2013), pp. 101–110.

[31] L. Rossi, A. Torsello, and E. R. Hancock, Attributed graph similarity from the quantum Jensen-Shannon divergence, in *Similarity-Based Pattern Recognition*, edited by E. R. Hancock and M. Pelillo (Springer, Berlin, 2013), pp. 204–218.

[32] L. Bai, L. Rossi, A. Torsello, and E. R. Hancock, A quantum Jensen-Shannon graph kernel for unattributed graphs, Pattern Recognition **48**, 344 (2015).

[33] J. Jost, *Riemannian Geometry and Geometric Analysis* (Springer, Berlin, 2011).

[34] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[35] W. K. Wootters, Statistical distance and Hilbert space, Phys. Rev. D **23**, 357 (1981).

[36] G. Lindblad, Entropy, information and quantum measurements, Commun. Math. Phys. **33**, 305 (1973).

[37] D. Bures, An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w^* algebras, Trans. AMS **135**, 199 (1969).

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

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

[40] F. Escolano, E. R. Hancock, and M. A. Lozano, Heat diffusion: Thermodynamic depth complexity of networks, Phys. Rev. E **85**, 036206 (2012).

[41] G. Li, M. Semerci, B. Yener, and M. J. Zaki, Effective graph classification based on topological and label attributes, Stat. Anal. Data Mining **5**, 265 (2012).

[42] S. K. Nayar, S. A. Nene, and H. Murase, Columbia object image library (coil 100), Technical Report No. CUCS-006-96, Department of Computer Science, Columbia University, 1996 (unpublished).

[43] A. Torsello and E. R. Hancock, Correcting curvature-density effects in the Hamilton-Jacobi skeleton, IEEE Trans. Image Process. **15**, 877 (2006).

[44] A. Torsello and L. Rossi, Supervised learning of graph structure, in *Similarity-Based Pattern Recognition*, edited by M. Pelillo and E. R. Hancock (Springer, Berlin, 2011), pp. 117–132.

[45] C. C. Chang and C. J. Lin, Libsvm: A library for support vector machines, ACM Trans. Intelligent Syst. Technol. (TIST) **2**, 27 (2011).

[46] N. Shervashidze, P. Schweitzer, E. J. Van Leeuwen, K. Mehlhorn, and K. M. Borgwardt, Weisfeiler-Lehman graph kernels, J. Machine Learning Res. **12**, 2539 (2011).

[47] Trevor F. Cox and Michael AA Cox, *Multidimensional Scaling* (CRC Press, Boca Raton, FL, 2010).

[48] X. Gao, B. Xiao, D. Tao, and X. Li, A survey of graph edit distance, Pattern Anal. Applic. **13**, 113 (2010).

[49] H. Bunke, On a relation between graph edit distance and maximum common subgraph, Pattern Recognition Lett. **18**, 689 (1997).

[50] S. Vichy N. Vishwanathan, Nicol N. Schraudolph, Risi Kondor, and Karsten M. Borgwardt, Graph kernels, J. Machine Learning Res. **11**, 1201 (2010).