HAQJSK: Hierarchical-Aligned Quantum Jensen-Shannon Kernels for Graph Classification

Lu Bai, Member, IEEE, Lixin Cui, Member, IEEE, Yue Wang, Member, IEEE, Ming Li, Fellow, IEEE, Jing Li, Fellow, IEEE, Philip S. Yu, Fellow, IEEE, and Edwin R. Hancock, Fellow, IEEE

Abstract—In this work, we propose two novel quantum walk kernels, namely the Hierarchical Aligned Quantum Jensen-Shannon Kernels (HAQJSK), between un-attributed graph structures. Different from most classical graph kernels, the proposed HAQJSK kernels can incorporate hierarchical aligned structure information between graphs and transform graphs of random sizes into fixed-size aligned graph structures, i.e., the Hierarchical Transitive Aligned Adjacency Matrix of vertices and the Hierarchical Transitive Aligned Density Matrix of the Continuous-Time Quantum Walks (CTQW). With pairwise graphs to hand, the resulting HAQJSK kernels are defined by computing the Quantum Jensen-Shannon Divergence (QJSD) between their transitive aligned graph structures. We show that the proposed HAQJSK kernels not only reflect richer intrinsic whole graph characteristics in terms of the CTQW, but also address the drawback of neglecting structural correspondence information that arises in most R-convolution graph kernels. Moreover, unlike the previous QJSD based graph kernels associated with the QJSD and the CTQW, the proposed HAQJSK kernels can simultaneously guarantee the properties of permutation invariant and positive definiteness, explaining the theoretical advantages of the HAQJSK kernels. The experiment indicates the effectiveness of the new proposed kernels.

Index Terms—Quantum walks, quantum Jensen-Shannon divergence, graph kernels, graph classification.

I. INTRODUCTION

Graph-based structures are powerful tools in various research domains that focus on modeling pairwise relationships between components, e.g., the analysis of social networks [1], molecule networks [2], 3D shapes [3], traffic networks [4], etc. A central task of graph data analysis is how to compute or learn significant numeric characteristics or features from discrete graph structures for graph classification. To this end, one prevalent way is to adopt machine learning algorithms based on graph kernel methods, which can represent graph characteristics in a meaningful high-dimensional Hilbert space and therefore preserve the structures [5]. This paper aims to propose a family of novel quantum information theoretic graph kernels associated with Continuous-Time Quantum Walks (CTQWs), for the purpose of graph classification. Our kernels are based on measuring the Quantum Jensen-Shannon Divergence (QJSD) between the hierarchical transitive aligned structures of pairwise graphs associated with the CTQW. The proposed kernels not only encapsulate transitive alignment information between graph structures, but also guarantee the positive definiteness.

A. Literature Review

Generally speaking, a graph kernel is essentially a kind of kernel function [6]. However, unlike the traditional kernel functions proposed for vectorial data, graph kernels are developed for graph structure data and measure the positive definite similarity between pairs of graphs [7], [8], [9], bridging the theoretical gap between graph structure data and the large spectrum of machine learning algorithms called kernel methods [10], e.g., the Support Vector Machines (SVMs), the Kernel Regression, or the kernel Principle Component Analysis (kPCA) (see the reference [11] for more details). Perhaps the most successful and widely used manner to define the graph kernel is the principle of R-convolution developed by Haussler [12]. It is a generic principle to propose new graph kernels between a pair of graphs by measuring the isomorphism between decomposed substructures, i.e., counting the pairs of isomorphic substructure between the graphs. Specifically, for a pair of graphs $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$ associated with their vertex sets $V_i$ and edge sets $E_i$ respectively, let $S_p \subseteq G_p$ and $S_q \subseteq G_q$ be a pair of their substructures through a specified graph decomposing method. Then, one R-convolution graph kernel $k_R$ between $G_p$ and $G_q$...
have developed a non-Backtrack Path Kernel through
have proposed a Binary Subtree-based Kernel for
have proposed a family of Global Graph Kernel based

\[ k_R(G_p, G_q) = \sum_{S_p \subseteq G_p} \sum_{S_q \subseteq G_q} s(S_p, S_q). \]

Here, \( s(S_p, S_q) \) is usually set as the Dirac kernel, and \( s(S_p, S_q) \) equals to 1 if \( S_p \) and \( S_q \) are isomorphic to each other (i.e., \( S_p \simeq S_q \)), and 0 otherwise.

With the scenario of R-convolution, one can employ any available graph decomposing method to develop a new R-convolution graph kernel, e.g., the graph kernels based on the decomposed a) path-based substructures [13], b) walk-based substructures [14], c) subtree-based substructures [15], d) subgraph-based substructures [16], etc. For instance, Kashima et al. [17] have defined a Random Walk-based Kernel by decomposing graph structures into random walks of different restricted lengths. The kernel is defined by calculating the number of the random walk pairs with same lengths. Borgwardt et al. [18] have developed a Shortest Path-based Graph Kernel by counting the number of the shortest paths with same lengths. Aziz et al. [13] have developed a non-Backtrack Path Kernel through cycle-based structures that are abstracted from the Ihara zeta function [19]. The kernel is computed by counting the numbers of cycle pairs with same lengths. Costa and Grave [20] have proposed a Neighborhood Subgraph Pairwise Distance Kernel by decomposing graphs into layer-wise expansion neighborhood subgraphs rooted at a pair of vertices with specified distance and measuring the isomorphism between the subgraphs. Gaidon et al. [21] have proposed a Binary Subtree-based Kernel for video classification problems. They commence by representing complicated actions as spatio-temporal parts and transforming them into binary tree patterns. The proposed kernel is defined by comparing the pairs of isomorphic subtrees. Shervashidze et al. [22] have developed a Weisfeiler-Lehman Subtree Kernel by counting the pairs of isomorphic subtree patterns corresponded by the Weisfeiler-Lehman graph invariant.

Other popular R-convolution graph kernels include a) the Segmentation-based Graph Kernel [8], b) the Pyramid Quantized Shortest Path Kernel [23], c) the Pyramid Quantized Weisfeiler-Lehman Subtree Kernel [23], d) the Wasserstein Weisfeiler-Lehman Kernel [24], e) the Isolation Graph Kernel [25], f) the Graph Filtration Kernels [26], h) the Truncated Tree Based Graph Kernels [27], etc.

One shortcoming arising in most of the aforementioned R-convolution graph kernels is that of ignoring the structural correspondence or alignment information between the graph structures. This problem is due to the fact that the R-convolution kernels only augment one unit kernel value if a pair of isomorphic substructures are detected, and the procedures of the isomorphism identification do not consider whether the isomorphic substructures are structurally aligned within the whole graph structures. For an instance of a computer vision problem, there are a pair of graph structures extracted from two images in Fig. 1, where each image include the same house based on different viewpoints. The R-convolution kernels will directly augment one unit kernel value when they identify the isomorphic triangle-based substructures, no matter whether they are structurally aligned to each other through the vision background. Thus, the R-convolution kernels will not reflect the precise similarities between a pair of graphs, and may in turn influence the graph classification performance.

To address the above issues, Bai and Xu et al. [28], [29], [30] have developed a family of Depth-based Alignment Kernels that encapsulate the vertex correspondence information. For a pair of graphs, they first compute the Depth-based Complexity Trace rooted at each vertex as the vectorial vertex representations [31], the resulting alignment kernels are attained by counting the pairs of aligned vertices that are confirmed by evaluating the distance between the vectorial vertex representations in a euclidean space. More specifically, these vertex alignment kernels are theoretically equivalent to aligned subgraph kernels that encapsulate structural correspondence information between substructures, addressing the shortcoming of neglecting structural correspondence information between graphs arising in classical R-convolution kernels.

Unfortunately, both the aforementioned R-convolution kernels and vertex alignment kernels fail to reflect characteristics of whole graph structures. The reasons for this problem are twofold. First, the definition of the R-convolution kernels relies on the graph decomposition, that may cause the notorious computational inefficiency. This is due to the fact that decomposing a large graph (e.g., a graph with with more than hundreds of vertices) into all possible substructures tends to be a NP-hard problem. Thus, the R-convolution kernels usually compromise to use substructures of small sizes, and it is hard to represent the whole graph characteristics with such substructures. Second, the vertex alignment kernels only focus on identifying the correspondence information between local vertices, which only reflect local structural information. Moreover, the above alignment kernels are not positive definite. This is due to the fact that the aligned vertices confirmed by the alignment kernels are not transitive, i.e., if vertices \( a \) and \( b \) are both aligned to vertex \( c \), we cannot guarantee that vertices \( a \) and \( b \) are also aligned to each other. The transitive alignment between vertices is a necessary condition to ensure the positive definiteness for an alignment or matching kernel [32].

To overcome the drawback of neglecting whole graph characteristics that arises in the above R-convolution kernels and vertex alignment kernels, some graph kernels based on the adjacency matrices, that are natural representations of whole graph structures, are developed. For example, Johansson et al. [33] have proposed a family of Global Graph Kernel based on the celebrated Lovász numbers as well as their orthonormal representations computed with the adjacency matrices.
Xu et al. [34] have developed a Hybrid Reproducing Kernel using the global entropy measures of the adjacency matrices.

Another alternative way of analyzing whole graph structures is to employ the Continuous-time Quantum Walks (CTQW) [35]. In quantum mechanics and quantum information theory [36], the CTQW represents a quantum analogue of the classical Continuous-time Random Walk (CTRW), it is controlled by the unitary matrix and is not dominated by the low Laplacian spectrum frequency. By contrast, the classical CTRW is controlled by the doubly stochastic matrix. Hence, the CTQW can better distinguish different whole graphs. More specifically, there have been a family of Quantum Jensen-Shannon Kernels (QJSK) defined based on the CTQW. For example, Bai et al. [37] have defined a quantum walk kernel by measuring the Quantum Jensen-Shannon Divergence (QJSD) between the density matrices of the CTQW evolved on graph structures. Rossi et al. [38] have proposed a quantum walk kernel by exploring the relationship between the interferences and symmetries of the CTQW evolved on graphs, in terms of the QJSD. Both the quantum walk kernels adopt the Laplacian matrices of graphs as the required Hamiltonian operators, that can naturally reflect whole characteristics of graph structures. Unfortunately, computing the QJSD between a pair of graph structures requires a composite graph structure that needs the vertex alignment information between the graphs, both the quantum kernels cannot guarantee the transitivity between the aligned vertices or even compute the composite structure by randomly arranging the vertex orders. Thus, these quantum kernels are not positive definite, and can not reflect the precise similarity between graphs. Broadly speaking, developing effective graph kernels is always a theoretical challenge problem.

B. Contributions

The objective of this work is to overcome the drawbacks of the above-mentioned graph kernels by proposing a family of novel Hierarchical-Aligned Quantum Jensen-Shannon Kernels (HAQJSK), that not only encapsulate structural correspondence information of local vertices but also capture structural characteristics of whole graph structures. To this end, we develop our previous Quantum Jensen-Shannon Kernel (QJSK) [37] one step further and theoretically generalize the QJSK kernel as new hierarchical aligned QJSK kernels (i.e., the HAQJSK kernels). One of the key innovations for the proposed HAQJSK kernels is to establish the hierarchical transitive correspondence or alignment information between vertices of graphs. More specifically, the HAQJSK kernels can adopt the correspondence information to transform the original graphs of different sizes into fixed-size hierarchical transitive aligned structures. The resulting HAQJSK kernels are defined by measuring the QJSD between the aligned structures of graphs associated with the CTQW. For the QJSD, since the required composite structure of a pair of graphs is computed based on their aligned graph structures and encapsulates the transitive correspondence information of the graphs, the HAQJSK kernels are theoretically equivalent to a transitive vertex alignment kernel and guarantee the property of positive definiteness. Overall, the contributions of this paper are summarized as follows.

First, we propose a new framework to transform the arbitrary-size graphs into fixed-size hierarchical transitive aligned structures, i.e., the hierarchical transitive aligned density matrix of the CTQW and the hierarchical transitive aligned adjacency matrix of the vertices. This is achieved by further extending the original transitive vertex matching approach [39] into a hierarchical matching method to identify the hierarchical transitive correspondence information between vertices of graphs. The aligned structures are then constructed through the associated hierarchical correspondence matrices. We show that the aligned structures not only remain the topological and quantum interference information residing on original graph structures (see details in Section II-A), but also provide an elegant way to define transitive aligned QJSK kernels.

Second, with the hierarchical transitive aligned structures of graphs in hand, we develop a family of novel HAQJSK kernels by measuring the QJSD associated with the CTQW between the aligned structures. For a pair of graphs, since the required composite structure of the QJSD is computed by summing their aligned structure representations, the composite structure naturally takes into account the transitive vertex correspondences between the pair of graphs, reflecting the locational correspondences between the CTQW evolving on the graphs. Hence, the proposed HAQJSK kernels can be theoretically considered as the transitive aligned version of the previous QJSK kernels, that not only guarantee the positive definiteness but also reflect more precise kernel similarity measures than the existing QJSD based quantum kernels [37], [38]. Furthermore, we show that the proposed HAQJSK kernels can simultaneously reflect the structural information residing on the local vertices as well as the whole graph structures.

Third, we empirically indicate the effectiveness of the HAQJSK kernels associated with the C-SVMs for classification tasks of standard graph datasets. Overall, the HAQJSK kernels have better classification performance than state-of-the-art graph deep learning methods as well as graph kernels.

II. QUANTUM MECHANICAL BACKGROUNDS AND RELATED WORKS

In this section, we briefly review the quantum mechanical background as well as some theoretically related works.

A. Continuous-Time Quantum Walks

Quantum-based methods have been employed to propose novel machine learning and data mining approaches [40], due to the fact that they can capture richer information than their classical counterparts. For example, Melucci [41], [42] has proposed a family of relevance feedback algorithms through the quantum-based probabilistic subspace. Fawaz et al. [43] have proposed a new method to train the binary neural network based on the quantum amplitude amplification. For more details in terms of the quantum computation and quantum algorithms, we refer the readers to the textbook [44]. In this subsection, we mainly review the concepts of the CTQW.
In quantum information theory [36], [45], [46], the CTQW is a quantum analogue of the classical CTRW that simulates the transition process of a Markovian diffusion on the vertices of graphs through the adjacency matrices, and both the CTQW and CTRW are defined on the vertex sets (i.e., the vertex sets are their state space). More formally, given a sample graph structure \( G(V, E) \) with the vertex set \( V \) and the edge set \( E \), the state vector for the CTQW at time \( t \) is a probability distribution over \( V \), i.e., a vector \( \tilde{p}_t \in \mathbb{R}^n \) whose \( u \)-th entry gives the probability of the CTQW visiting vertex \( u \) at time \( t \). Assume \( A \) and \( D \) are the adjacency matrix as well as the diagonal degree matrix of \( G \) respectively, the CTRW on \( G \) evolves based on the following equation

\[
\tilde{p}_t = e^{-Lt} \tilde{p}_0,
\]

where \( L = D - A \) is the graph Laplacian. In this classical case, the CTRW is defined as a dynamical process over the vertex set \( V \) of the sample graph \( G \).

Unlike the classical CTRW where the state vector lies in a probability space, the state of CTQW is defined as a vector of complex amplitudes over \( V \). The squared norm of the amplitudes sums to unity over \( V \), with no restriction on their sign or complex phase. This in turn allows both destructive and constructive interference to take place between the complex amplitudes. Moreover, unlike the classical CTRW whose dynamics is governed by a stochastic matrix, the evolution of the CTQW is governed by a complex valued unitary matrix. Thus, the evolution of the CTQW is reversible, non-ergodic, and not restricted by a limiting distribution, i.e., the behaviours of the classical CTRW and the CTQW are significantly different. In quantum mechanics [47], the basis state of the CTQW is written in Dirac notation, that is also known as bra-ket notation. Specifically, a quantum state can be described as a single ket vector \( |\alpha\rangle \), that is a \( |V| \)-dimensional column vector in a complex-valued Hilbert space. The conjugate transpose of \( |\alpha\rangle \) is a bra (row) vector, denoted as \( \langle \alpha | \). The inner product between \( |\alpha\rangle \) and \( |\beta\rangle \) is written as \( |\alpha\rangle \langle \beta | \) and their outer product is written as \( |\alpha\rangle |\beta\rangle \). Based on Dirac notation, we denote the basis state corresponding to the CTQW at vertex \( u \in V \) as \( |u\rangle \), i.e., \( |u\rangle \) is the vector that is equal to 1 in the position corresponding to the vertex \( u \) and 0 in the position corresponding to other vertices. As a result, the state \( |\psi(t)\rangle \) of the CTQW at time \( t \) is expressed as a complex linear combination of the orthonormal basis states \( |v\rangle \) over all vertices \( v \in V \), i.e.,

\[
|\psi(t)\rangle = \sum_{v \in V} \alpha_v(t) |v\rangle.
\]

Here \( \alpha_v(t) \in \mathbb{C} \) is the complex amplitude of the basis state \( |v\rangle \) and follows the condition \( \sum_{v \in V} |\alpha_v(t)|^2 = 1 \) for any vertex \( v \in V \) and \( t \in \mathbb{R}^+ \), and \( \alpha_v^*(t) \) is the complex conjugate of \( \alpha_v(t) \). Different from the CTRW, we define the evolution of the CTQW using the Schrödinger equation

\[
\frac{\partial |\psi(t)\rangle}{\partial t} = -i\hat{H}|\psi(t)\rangle,
\]

where \( \hat{H} \) is the Hamiltonian of the system and specifies the total system energy. In this paper, we propose to adopt the Laplacian matrix \( L \) as the Hamiltonian. More specifically, assume \( A \) and \( D \) are the vertex adjacency matrix and the vertex degree matrix of \( G \) respectively, \( L \) is defined as \( D - A \). Let \( L = \Phi^T \Lambda \Phi \) be the spectral decomposition of \( L \), where \( \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{|V|-1} , \lambda_{|V|}) \) is a diagonal matrix with the ascending-ordered eigenvalues as the elements (i.e., \( \lambda_1 < \lambda_2 < \cdots < \lambda_{|V|-1} < \lambda_{|V|} \)) and \( \Phi = (\phi_1 | \phi_2 | \cdots | \phi_{|V|-1} | \phi_{|V|} ) \) is the matrix with the corresponding orthonormal eigenvectors as columns. The state \( |\psi(t)\rangle \) of the CTQW at time \( t \) is defined as

\[
|\psi(t)\rangle = \Phi^T e^{-i\Lambda t} \Phi |\psi(0)\rangle,
\]

where \( |\psi(0)\rangle \) denotes an initial state and is computed by taking the square root of the vertex degree distribution of \( G \) [37].

Note that, the state \( |\psi(t)\rangle \) at time \( t \) is a pure state. But we usually use the mixed state, that is a statistical ensemble of pure states \( |\psi(t)\rangle \) each with a probability \( p_t \) at time \( t \). The density matrix (i.e., the density operator) of such a system is defined as \( \rho = \sum_{t} p_t |\psi(t)\rangle \langle \psi(t)| \), where \( |\psi(t)\rangle \) is the conjugate transpose of \( |\psi(t)\rangle \) and \( |\psi(t)\rangle \) is the outer product between them. Assume the CTQW evolves from time 0 to time \( T \). The mixed density matrix of \( G \) can be defined as a time-averaged density matrix

\[
\rho_G^T = \frac{1}{T} \int_0^T \Phi^T e^{-i\Lambda t} \Phi |\psi(0)\rangle \langle \psi(0)| \Phi^T e^{i\Lambda t} \Phi \, dt.
\]

Assume \( \phi_{xa} \) and \( \phi_{yb} \) are the \( (xa) \)-th and \( (yb) \)-th elements of the matrix of eigenvectors \( \Phi \) of \( L \). When \( T \) approaches \( \infty \), we can employ the closed form solution to calculate the \( (x,y) \)-th element of \( \rho_G^T \) as

\[
\rho^T_{G(x,y)} = \sum_{\lambda \in \Lambda} \sum_{x \in B_x} \sum_{y \in B_y} \phi_{xa} \phi_{yb} \tilde{\alpha}_x \tilde{\alpha}_y,
\]

where \( \tilde{\alpha}_x \) and \( \tilde{\alpha}_y \) are the distinct eigenvalues of \( L \), and \( B_x \) refers to the basis of the eigenspace with \( \lambda \).

**Remarks:** The CTQW has some attractive properties that are not available for the classical CTRW. First, as we have introduced previously, the CTQW permits the interference to take place during the evolution. Thus, when the CTQW backtracks on an edge it does so with opposite phase, reducing the tottering problem on edges. Moreover, the evolution of the CTQW is not governed by the low frequency components of the Laplacian spectrum. Hence, the CTQW can better discriminate various graph structures than the classical CTRW. Overall, the CTQW provides an elegant way to define novel algorithms for graph-based structure data analysis.

### B. The Quantum Jensen-Shannon Divergence

In quantum information theory [48], the QJSD is defined as a quantum generalization of its classical version, i.e., the classical Jensen-Shannon Divergence (JSD). As in the classical case, the QJSD is always well defined, symmetric, negative definite and bounded [48]. However, unlike the classical JSD defined to probability distributions, the QJSD is defined on the density matrices for quantum states. Specifically, for a pair of density matrices \( \rho \) and \( \sigma \), the QJSD is

\[
D_{QJS}(\rho, \sigma) = H_N \left( \frac{\rho + \sigma}{2} \right) - \frac{1}{2} H_N(\rho) - \frac{1}{2} H_N(\sigma),
\]

where \( H_N \) is the Shannon entropy for probabilities.
where \( H_N(\rho) \) is the von Neumann entropy of the quantum state (i.e., a density matrix) \([36]\) and is defined as
\[
H_N = -\text{tr}(\rho \log \rho) = -\sum_i \xi_i \ln \xi_i,
\]
and \( \xi_1, \ldots, \xi_n \) are the eigenvalues of \( \rho \). Generally, the entropy \( H_N \) of a pure state is zero, but it has a non-zero value associated with the mixed state. Thus, in this work we compute \( H_N \) of each graph \( G \) associated with the density matrix \( \rho_G^\infty \) of the CTQW defined in \((6)\). Specifically, the von Neumann entropy \( H_N \) is defined as
\[
H_N(\rho_G) = -\text{tr}(\rho_G \log \rho_G) = -\sum_j^{[V]} \lambda_j^G \log \lambda_j^G,
\]
where \( \lambda_j^G \) are the eigenvalues of \( \rho_G^\infty \).

C. The Quantum Jensen-Shannon Kernels for Graphs

This subsection reviews the concepts for the QJS kernel proposed in the previous work \([37]\), that is also based on the QJSD associated with the CTQW. Assume we evolve the CTQW on a pair of graphs \( G_p(V_p, E_p) \) and \( G_q(V_q, E_q) \), and \( \rho_p \) and \( \rho_q \) are their associated mixed state density matrices defined by \((6)\). The QJS kernel between \( G_p \) and \( G_q \) is defined as
\[
k_{\text{QJSU}}(G_p, G_q) = \exp \left[ -\mu D_{\text{QJS}}(\rho_p, \rho_q) \right],
\]
where
\[
D_{\text{QJS}}(\rho_p, \rho_q) = H_N \left( \frac{\rho_p + \rho_q}{2} \right) - \frac{1}{2} H_N(\rho_p) - \frac{1}{2} H_N(\rho_q).
\]
is the QJSD between the density matrices \( \rho_p \) and \( \rho_q \), \( H_N(\cdot) \) represents the von Neumann entropy given by \((9)\), and \( \mu \) is the decay factor and is usually set as 1. Note that, since the sizes of the graphs \( G_p \) and \( G_q \) are usually different, straightforwardly computing their composite density matrices \((\frac{\rho_p + \rho_q}{2})\) (i.e., a kind of composite structure of \( G_p \) and \( G_q \)) tends to be elusive. One way to solve this problem is to expand the density matrix of the smaller graph with zero elements, so that the composite density matrix can be seen as a macro characteristic of a graph structure. As a result, comparing to the state-of-the-art graph kernels reviewed in Section I-A, the QJS kernels not only reflect more whole structural information than most existing R-convolution kernels that compromise to employ small-size substructures, but also better discriminate different graph structures than the graph kernels focusing on capturing whole characteristics through the adjacency matrix. Unfortunately, both the QJS kernels \( k_{\text{QJSU}} \) and \( k_{\text{QJSA}} \) still suffer from a number of theoretical drawbacks. First, the unaligned QJS kernel \( k_{\text{QJSU}} \) does not encapsulate any correspondence information between vertices into the computation of the required QJSD. Thus, \( k_{\text{QJSU}} \) cannot guarantee the positive definiteness and is not permutation invariant to the vertex orders. Second, although the aligned QJS kernel \( k_{\text{QJSU}} \) employs the Umeyama matching method to identify the vertex correspondence information between pairs of graphs, and thus overcomes the permutation invariant problem. The aligned vertices are not guaranteed to be transitive. As a result, \( k_{\text{QJSU}} \) is also not a positive definite kernel. This indicates that both the unaligned and aligned QJS kernels \( k_{\text{QJSU}} \) and \( k_{\text{QJSA}} \) cannot reflect a precise similarity measure between graphs. As an evidence, the experimental evaluations in \([37]\) demonstrate that the classification performance of the C-SVMs associated with the aligned QJS kernel \( k_{\text{QJSU}} \) is not significantly better than that associated with the unaligned QJS kernel \( k_{\text{QJSU}} \).

In this paper, we will propose a new variant of the HAQJSK kernel to overcome the above drawbacks.

D. The Transitive Vertex Matching Method for Graphs

In this subsection, we review the transitive vertex matching approach proposed in the previous work \([39]\). The key idea of this matching method is to align the vertices of each graph to a family of prototype representations, that are identified by computing the centroids over the vectorial vertex representations of all sample graphs through the classical k-means clustering method. Thus, the prototype representations reflect the main and representative structure characteristics over all graphs.

Assume the \( k \)-dimensional vectorial representations of the \( N \) vertices \( V \) over all available graphs in \( G \) are denoted as
\[
R^k(V) = \{R^k(v_1), R^k(v_2), \ldots, R^k(v_3), \ldots, R^k(v_N)\},
\]
where \( Q \in \{0, 1\}^{[V_p] \times [V_q]} ([V_p] \geq [V_q]) \) is the vertex correspondence matrix between \( G_p \) and \( G_q \). Here, the correspondence matrix \( Q \) is computed based on the Umeyama Spectral Matching Method \([49]\), that directly utilizes the eigendecomposition of the density matrices. More specifically, it is clear that vertices \( u \in V_p \) and \( v \in V_q \) are aligned to each other, if the \((u, v)\)-th element of \( Q \) is 1.

**Remarks:** Both the QJS kernels \( k_{\text{QJSU}} \) and \( k_{\text{QJSA}} \) can pad the density matrix of the small graph to the same size of the large graph. This process is equivalent to extending the adjacency matrix of the small graph to the same size of the large graph. Since the density matrix of the CTQW can encapsulate more complicated structure information than the original adjacency matrix, and the required von Neumann entropy of the density matrix can be seen as a macro characteristic of a graph structure. As a result, comparing to the state-of-the-art graph kernels reviewed in Section I-A, the QJS kernels not only reflect more whole structural information than most existing R-convolution kernels that compromise to employ small-size substructures, but also better discriminate different graph structures than the graph kernels focusing on capturing whole characteristics through the adjacency matrix. Unfortunately, both the QJS kernels \( k_{\text{QJSU}} \) and \( k_{\text{QJSA}} \) still suffer from a number of theoretical drawbacks. First, the unaligned QJS kernel \( k_{\text{QJSU}} \) does not encapsulate any correspondence information between vertices into the computation of the required QJSD. Thus, \( k_{\text{QJSU}} \) cannot guarantee the positive definiteness and is not permutation invariant to the vertex orders. Second, although the aligned QJS kernel \( k_{\text{QJSU}} \) employs the Umeyama matching method to identify the vertex correspondence information between pairs of graphs, and thus overcomes the permutation invariant problem. The aligned vertices are not guaranteed to be transitive. As a result, \( k_{\text{QJSU}} \) is also not a positive definite kernel. This indicates that both the unaligned and aligned QJS kernels \( k_{\text{QJSU}} \) and \( k_{\text{QJSA}} \) cannot reflect a precise similarity measure between graphs. As an evidence, the experimental evaluations in \([37]\) demonstrate that the classification performance of the C-SVMs associated with the aligned QJS kernel \( k_{\text{QJSU}} \) is not significantly better than that associated with the unaligned QJS kernel \( k_{\text{QJSU}} \).

In this paper, we will propose a new variant of the HAQJSK kernel to overcome the above drawbacks.
where $|V| = N$ and $v_i \in V$. To construct the transitive vertex correspondence information, Cui et al. [39] have aligned the vertices of each individual graph in $G$ to a common set of $k$-dimensional prototype representations $P^k(G)$ over all graphs in $G$. This is done by employing the classical $k$-means clustering method on $R^k(V)$ to locate a set of $k$-dimensional mean vectors, through minimizing the following objective function

$$
\arg \min_{\Omega} \sum_{j=1}^{N} \sum_{R^k(v_i) \in c_j} \|R^k(v_i) - \mu_j\|_2^2,
$$

(14)

where $\Omega = (c_1, c_2, \ldots, c_N)$ corresponds to $N$ clusters over $R^k(V)$, and $\mu_j$ is the mean of all $R^k(v_i)$ belonging to the $j$-th cluster $c_j$. As a result, the set of the $k$-dimensional prototype representations $P^k(G)$ are defined as

$$
P^k(G) = \{\mu^k_1, \mu^k_2, \ldots, \mu^k_j, \ldots, \mu^k_N\}.
$$

(15)

For a graph $G_p(V_p, E_p) \in G$, we align its $k$-dimensional vectorial vertex representations $R^k(V_p)$ to $P^k(G)$ and compute the correspondence matrix $C^k_p(i, j) = \{1, 0\}^{V_p \times P^k(G)}$ as

$$
C^k_p(i, j) = \begin{cases} 
1 & \text{if } R^k(v_i) \in P^k(V_p) \text{ of } v_i \in V_p \text{ belongs to the } j \text{-th cluster } c_j, \text{ i.e., vertex } v_i \in V_p \text{ is aligned to } \mu^k_j \in P^k(G) \\
0 & \text{otherwise.}
\end{cases}
$$

(16)

If $C^k_p(i, j)$ is equal to 1, there exists one-to-one correspondence information between the $i$-th vertex $v_i \in V_p$ and the $j$-th prototype representation $\mu^k_j \in P^k(G)$. Because $\mu^k_j$ is the nearest prototype representation to $v_i$ based on (14). More specifically, assume a pair of graphs $G_p$ and $G_q$ from $G$, if their vertices $v_i \in V_p$ and $v_j \in V_q$ are aligned to the same prototype representation $\mu^k_j \in P^k(G)$, their vertices $v_p$ and $v_q$ are transitively aligned to each other.

### III. THE HIERARCHICAL ALIGNED QUANTUM JENSEN-SHANNON GRAPH KERNELS

In this section, we define a family of novel Hierarchical Aligned Quantum Jensen-Shannon Kernels (HAQJSK) for unattributed graphs. We commence by developing a new framework to transform random sized graphs into fixed-size hierarchical transitive aligned structures. Moreover, for a pair of graphs, we define the HAQJSK kernels by measuring the QJSD between their aligned structures. Finally, we indicate the theoretical advantages of the proposed HAQJSK kernels.

#### A. The Hierarchical Transitive Aligned Graph Structures

In this subsection, we develop a new framework to convert a set of graphs $G$ with arbitrary sizes into fixed-size hierarchical transitive aligned structures, that can reflect rich multi-scale structural characteristics of the graphs. To this end, we propose to hierarchically employ the transitive vertex matching method reviewed in Section II-D, and identify the hierarchical transitive correspondence information between graphs.

More specifically, let the previously defined $k$-dimensional vectorial representations $R^k(V)$ of the vertices over all graphs in $G$ be the set of 0-level prototype representations $P^{0,k}$. By hierarchically performing the $k$-means clustering on the $h - 1$-level prototype representations $P^{h-1,k}$, we compute a family of different $h$-level prototype representations over the vertices of all graphs in $G$ as

$$
HP^{h,k}(G) = \{P^{1,k}(G), P^{h,k}(G), \ldots, P^{H,k}(G)\},
$$

(17)

where $1 \leq h \leq H$, the set of 1-level prototype representations $P^{1,k}(G)$ are essentially the set of original $k$-dimensional prototype representations $P^{k}(G)$ defined by (15), and each $h$-level prototype representations $P^{h,k}$ in $HP^{h,k}(G)$ are computed as the centroid vectors of the clusters identified by performing the $k$-means method on the $h - 1$-level prototype representations $P^{h-1,k}$ in $HP^{h=k}(G)$. As a result, each $h$-level prototype representations $P^{h,k}$ capture the main characteristics over the $h - 1$-level prototype representations $P^{h-1,k}$, naturally forming a family of hierarchical prototype representations. Fig. 2 exhibits a detailed example of computing a family of different $h$-level prototype representations $HP^{h,k}(G)$ defined by (17).

Similar to the definition of (16), we align the $k$-dimensional vectorial vertex representations $R^k(V_p)$ of each graph $G_p(V_p, E_p) \in G$ to each set of $h$-level prototype representations $P^{h,k}$ in $HP^{h,k}(G)$, and compute a family $h$-level hierarchical correspondence matrices as

$$
C^{h,k}(G_p) = \{C^{1,k}_p, C^{2,k}_p, \ldots, C^{h,k}_p, \ldots, C^{H,k}_p\},
$$

(18)

where the 1-level correspondence matrix $C^{1,k}_p \in \{0, 1\}^{V_p \times P^{1,k}(G)}$ is essentially the correspondence matrix $C^{h,k}_p \in \{0, 1\}^{V_p \times P^{h,k}(G)}$ defined by (16).

**Definition 3.1 (The Hierarchical Transitive Aligned Graph Structures):** With the family of $h$-level hierarchical correspondence matrices $C^{h,k}$ in hand, we transform each graph $G_p \in G$ into fixed-size transitive aligned structures using the correspondence information. Specifically, for each graph $G_p$, assume $A_p$ is its vertex adjacency matrix, and $\rho_p$ is its density matrix associated with the CTQW defined in (6). We transform $G_p$ into a family of fixed-size $h$-level transitive aligned adjacency matrices as

$$
A = \{A^{1,k}_p, A^{2,k}_p, \ldots, A^{h,k}_p, \ldots, A^{H,k}_p\},
$$

(19)

where each $h$-level transitive aligned adjacency matrix $A^{h,k}_p \in \mathbb{R}^{P^{h,k} \times P^{h,k}}$ is

$$
A^{h,k}_p = C^{h,k}_p^{T} A_p C^{h,k}_p.
$$

(20)

Similarly, we transform $G_p$ into a family of fixed-size $h$-level transitive aligned density matrices as

$$
P = \{\rho^{1,k}_p, \rho^{2,k}_p, \ldots, \rho^{h,k}_p, \ldots, \rho^{H,k}_p\},
$$

(21)

where each $h$-level transitive aligned density matrix $\rho^{h,k}_p \in \mathbb{R}^{P^{h,k} \times P^{h,k}}$ is defined as

$$
\rho^{h,k}_p = C^{h,k}_p^{T} \rho_p C^{h,k}_p.
$$

(22)

In this work, we use the $k$-dimensional Depth-based Complexity Traces (DCT) of vertices as the vectorial representations $R^k(V)$ of vertices, following the statement in [39]. It has been shown...
that the DCT can preserve meaningful entropic information flows rooted from each local vertex to the whole graph structure through a family of $k$-layer expansion subgraphs rooted at the local vertex [31], [50]. To encapsulate more nested structural information, we propose to vary $k$ from 1 to $K$ (e.g., the largest layer of the expansion subgraphs, that corresponds to the greatest shortest path length over all graphs in $G$). More specifically, we compute each set of $h$-level prototype representations associated with the $K$-dimensional DCT as $P^{h,k}(G)$. Moreover, we employ the first $k$ ($1 \leq k \leq K$) dimensions of the prototype representations belonging to $P^{h,k}(G)$ as the set of $h$-level prototype representations $P^{h,k}(G) \in H^{p,k}(G)$ defined by (17), and thus use $P^{h,k}(G)$ to compute the $h$-level hierarchical correspondence matrix $C^{h,k} \in C^{h,k}(G_p)$ defined by (18). As a result, based on (20) we compute the family of Hierarchical Transitive Aligned Adjacency Matrices for the graph $G_p \in G$ as

$$A_p = \{ \tilde{A}^1_p, \tilde{A}^2_p, \ldots, \tilde{A}^h_p, \ldots, \tilde{A}^H_p \},$$

(23)

where

$$\tilde{A}^h_p = \sum_{k=1}^{K} \frac{A^{h,k}_p}{K}. $$

(24)

Similarly, based on (22) we also compute the family of Hierarchical Transitive Aligned Density Matrices for the graph $G_p \in G$ as

$$\hat{P}_p = \{ \hat{\rho}^1_p, \hat{\rho}^2_p, \ldots, \hat{\rho}^h_p, \ldots, \hat{\rho}^H_p \},$$

(25)

where

$$\hat{\rho}^h_p = \sum_{k=1}^{K} \frac{\rho^{h,k}_p}{K}. $$

(26)

**Remarks:** Clearly, both the Hierarchical Transitive Aligned Adjacency Matrix $\tilde{A}^h_p \in \tilde{A}$ and the Hierarchical Transitive Aligned Density Matrix $\hat{\rho}^h_p \in \hat{P}$ encapsulate the transitive aligned vertex correspondence information between all graphs in $G$, i.e., for any pair of graphs in $G$, their aligned adjacency matrices or aligned density matrices are transitively aligned to each other. This is because $\tilde{A}^h_p \in \tilde{A}$ and $\hat{\rho}^h_p \in \hat{P}$ are both computed by transforming the adjacency matrix and the density matrix of original graphs through the hierarchical correspondence matrices defined by (18), and their rows and columns correspond to the same hierarchical prototype representations defined by (17). As a result, the aligned structures $\tilde{A}$ and $\hat{P}$ provide a natural way to define new transitive aligned quantum Jensen-Shannon graphs kernels by measuring the QISD between the aligned structures.

### B. The Proposed HAQJSK Kernels

In this subsection, we define a family of HAQJSK kernels between graphs in $G$ through the hierarchically aligned graph structures defined in Section III-A.

**Definition 3.2 (The HAQJSK Kernel based on Hierarchical Transitive Aligned Adjacency Matrices):** For the pair of graphs $G_p \in G$ and $G_q \in G$ defined previously, let

$$\tilde{A}_p = \{ \tilde{A}^1_p, \tilde{A}^2_p, \ldots, \tilde{A}^h_p, \ldots, \tilde{A}^H_p \}$$

and

$$\tilde{A}_q = \{ \tilde{A}^1_q, \tilde{A}^2_q, \ldots, \tilde{A}^h_q, \ldots, \tilde{A}^H_q \}$$

denote their associated families of fixed-sized Hierarchical Transitive Aligned Adjacency Matrices defined by (23). We commence by employing (6) to compute the density matrices of the CTQW evolving on the aligned adjacency matrices in $\tilde{A}_p$ and $\tilde{A}_q$ as

$$\hat{Q}_p = \{ \hat{\rho}^1_p, \hat{\rho}^2_p, \ldots, \hat{\rho}^h_p, \ldots, \hat{\rho}^H_p \}$$

and

$$\hat{Q}_q = \{ \hat{\rho}^1_q, \hat{\rho}^2_q, \ldots, \hat{\rho}^h_q, \ldots, \hat{\rho}^H_q \}$$
and
\[
\bar{Q}_q = \{ \bar{q}_1^h, \bar{q}_2^h, \ldots, \bar{q}_q^h \}.
\]

The HAQJSK kernel \(K_{HAQJSK}^A\) evolving on the adjacency matrix \(A_p\) and \(A_q\) associated with \(A_p^\alpha\) and \(A_q^\alpha\) is defined as
\[
K_{HAQJSK}^A (G_p, G_q) = K_{HAQJSK}^A (\bar{Q}_p, \bar{Q}_q)
= \sum_{h=1}^{H} \exp \left( -D_{QJS} (\bar{q}_q^h, \bar{q}_q^h) \right),
\]
(27)
where
\[
D_{QJS} (\bar{q}_q^h, \bar{q}_q^h) = H_N \left( \frac{\bar{q}_q^h + \bar{q}_q^h}{2} - \frac{1}{2} H_N (\bar{q}_q^h) - \frac{1}{2} H_N (\bar{q}_q^h) \right),
\]
(28)

Let \(\delta (A) : A \rightarrow \rho\) represent the mapping function (i.e., (6)) that computes the density matrix \(\rho\) of the CTQW evolving on the adjacency matrix \(A\) of a graph \(G\). Based on (20) and (24), (28) is further written as
\[
D_{QJS} (\bar{q}_q^h, \bar{q}_q^h) = D_{QJS} (\delta (A_p^\alpha), \delta (A_p^\alpha))
= D_{QJS} \left[ \delta \left( \frac{\sum_{k=1}^{K} C_{p}^{h,k} T A_{p} C_{p}^{h,k}}{K} \right), \delta \left( \frac{\sum_{k=1}^{K} C_{q}^{h,k} T A_{q} C_{q}^{h,k}}{K} \right) \right]
= H_N \left[ \delta \left( \frac{\sum_{k=1}^{K} C_{p}^{h,k} T A_{p} C_{p}^{h,k}}{K} \right) \right] - \frac{1}{2} H_N \left( \frac{\sum_{k=1}^{K} C_{p}^{h,k} T A_{p} C_{p}^{h,k}}{K} \right)
- \frac{1}{2} H_N \left( \frac{\sum_{k=1}^{K} C_{q}^{h,k} T A_{q} C_{q}^{h,k}}{K} \right),
\]
(29)
where \(A_p\) and \(A_q\) are the original adjacency matrices of \(G_a\) and \(G_p\), and \(C_{p}^{h,k}\) and \(C_{q}^{h,k}\) are the \(h\)-level correspondence matrices of \(G_a\) and \(G_p\).

**Definition 3.3 (The HAQJSK Kernel based on Hierarchical Transitive Aligned Density Matrices):** For the pair of graphs \(G_p \in \mathcal{G} \) and \(G_q \in \mathcal{G}\) defined previously, let
\[
\bar{p}_p = \{ p_1^h, p_2^h, \ldots, p_p^h \}
\]
and
\[
\bar{p}_q = \{ q_1^h, q_2^h, \ldots, q_q^h \}
\]
denote their associated families of fixed-size Hierarchical Transitive Aligned Density Matrices defined by (25). The HAQJSK kernel \(K_{HAQJSK}^D\) between \(G_p\) and \(G_q\) associated with \(\bar{A}_p\) and \(\bar{A}_q\) is defined as
\[
K_{HAQJSK}^D (G_p, G_q) = K_{HAQJSK}^D (\bar{p}_p, \bar{p}_q)
= \sum_{h=1}^{H} \exp \left( -D_{QJS} (\bar{q}_q^h, \bar{q}_q^h) \right),
\]
(30)
where
\[
D_{QJS} (\bar{q}_q^h, \bar{q}_q^h) = H_N \left( \frac{\bar{q}_q^h + \bar{q}_q^h}{2} - \frac{1}{2} H_N (\bar{q}_q^h) - \frac{1}{2} H_N (\bar{q}_q^h) \right).
\]
(31)

Based on (20) and (26), (31) is further written as
\[
D_{QJS} (\bar{q}_q^h, \bar{q}_q^h) = D_{QJS} \left[ \frac{\sum_{k=1}^{K} C_{p}^{h,k} T p_{p} C_{p}^{h,k}}{K}, \frac{\sum_{k=1}^{K} C_{q}^{h,k} T p_{q} C_{q}^{h,k}}{K} \right]
= H_N \left( \frac{\sum_{k=1}^{K} C_{p}^{h,k} T p_{p} C_{p}^{h,k}}{K} \right) + \frac{1}{2} H_N \left( \frac{\sum_{k=1}^{K} C_{q}^{h,k} T p_{q} C_{q}^{h,k}}{K} \right),
\]
(32)
where \(p_{p}\) and \(p_{q}\) are the original density matrices of the CTQW evolving on \(G_p\) and \(G_q\), and \(C_{p}^{h,k}\) and \(C_{q}^{h,k}\) are the \(h\)-level correspondence matrices of \(G_a\) and \(G_p\).

**Lemma:** The HAQJSK kernels \(K_{HAQJSK}^A\) and \(K_{HAQJSK}^D\) are positive definite (pd).

**Proof:** The reasons of the positive definiteness for the proposed HAQJSK kernels \(K_{HAQJSK}^A\) and \(K_{HAQJSK}^D\) are twofold. First, the associated QJSD measure for the HAQJSK kernels is a symmetric dissimilarity measure [51]. Thus, the HAQJSK kernels defined as the negative exponential of the QJSD measure will be positive definite, if they are transitive aligned kernels [32], [37], following the theoretical statement in [37], [52]. Second, (29) and (32) indicate that the \(h\)-level correspondence matrices \(C_{p}^{h,k}\) and \(C_{q}^{h,k}\) can transform the original adjacency matrices or density matrices of any pair of graphs \(G_p\) and \(G_q\) into fixed-size aligned structures. Since \(C_{p}^{h,k}\) and \(C_{q}^{h,k}\) are computed by aligning each graph to a common \(h\)-level prototype representations, the correspondence information identified by \(C_{p}^{h,k}\) and \(C_{q}^{h,k}\) are transitive. Thus, the transformed aligned structures of different pairs of graphs are also transitively aligned to each other. Moreover, for (29) and (32), since the required composite structures
\[
\left[ \delta \left( \frac{\sum_{k=1}^{K} C_{p}^{h,k} T A_{p} C_{p}^{h,k}}{K} \right) \right] + \frac{1}{2} \left[ \delta \left( \frac{\sum_{k=1}^{K} C_{q}^{h,k} T A_{q} C_{q}^{h,k}}{K} \right) \right]
\]
and
\[
\left( \frac{\sum_{k=1}^{K} C_{p}^{h,k} T p_{p} C_{p}^{h,k}}{K} \right) + \frac{1}{2} \left( \frac{\sum_{k=1}^{K} C_{q}^{h,k} T p_{q} C_{q}^{h,k}}{K} \right)
\]
are computed by summing the aligned structures, these composite structures are also transitively aligned to the aligned structures of different graphs. Thus, the proposed HAQJSK kernels can be seen as transitive aligned kernels, that compute the similarity between transitive aligned structures. In summary, the proposed HAQJSK kernels are positive definite.

**Remarks:** Because the proposed HAQJSK kernel \(K_{HAQJSK}^A\) is computed based on the density matrices of the CTQW evolving on the Hierarchical Transitive Aligned Adjacency Matrices, that
are transformed from the original graph structures (i.e., the original adjacency matrices). $K_{HAQJSK}^\alpha$ focuses more on reflecting hierarchical topological information of the original graphs. On the other hand, the proposed HAQJSK kernel $K_{HAQJSK}^\alpha$ is computed based on the Hierarchical Aligned Density Matrices, that are transformed from the original density matrices of the CTQW evolving on the original graph structures. $K_{HAQJSK}^\alpha$ focuses more on reflecting hierarchical quantum walk information of the original graphs.

C. Discussions of the Proposed HAQJSK Kernels

Unlike some existing classical graph kernels, the proposed HAQJSK kernels have a number of important theoretical properties, that explain their effectiveness of the proposed kernels. These properties are shown in Table I and briefly discussed as follows.

First, unlike the existing Depth-based Alignment Kernels (DBAK) [28], [29], [30] that can not guarantee the transitivity between aligned vertices, the proposed HAQJSK kernels can encapsulate transitive vertex correspondence information into the kernel computation process. This is because the vertex correspondence information is identified by aligning each graph to a common set of prototype representations, i.e., only the vertices aligning to the same prototype representation will be considered as aligned vertices. Thus, the proposed HAQJSK kernels not only guarantee the positive definiteness that is not available for these DBAK kernels, but also reflect more precise kernel-based similarity measures. Moreover, unlike the DBMK kernels that only identify correspondence information between original graph structures, the proposed HAQJSK kernels can reflect hierarchical correspondence information by aligning the graphs to the hierarchical prototype representations. Hence, only the proposed HAQJSK kernels can reflect multi-scale structure information of graphs.

Second, unlike the family of R-convolution Kernels (RK) [12] that compromise to use substructures of limited sizes and only capture local structural information of graphs, the proposed HAQJSK kernels can capture whole structural information by measuring the QSD between the von Neumann entropies of whole graph structures. Thus, the proposed HAQJSK kernels can reflect richer whole structural information than the R-conK kernels. Furthermore, since the von Neumann entropy of a graph is associated with the CTQW, which can reflect intrinsic structural information of graphs, the proposed kernels can capture more complicated topological information of graphs compared to the R-conK kernels.

Third, unlike the existing Quantum Jensen-Shannon Kernels (QJSK) [37], [38], [53] that are also defined based on the QJS measure associated with the CTQW, only the proposed HAQJSK kernel can encapsulate the transitive vertex alignment information to compute the required composite structure of the QJS. Hence, only the proposed HAQJSK kernels can be considered as a kind of transitive alignment kernels, simultaneously guaranteeing the permutation invariant and the positive definiteness. Moreover, the proposed HAQJSK kernels are defined based on the hierarchical transitive aligned structures, that are transformed from the original graphs structures through the hierarchical correspondence information. By contrast, the QJSK kernels are defined based on the original graph structures. Thus, only the proposed HAQJSK kernels can reflect hierarchical structure information of graphs.

Fourth, unlike the Global Graph Kernels [33], [34] that focus on capturing whole structure information of graphs, the proposed HAQJSK kernels not only reflect whole structure information through the CTQW, but also encapsulate local structure information through the hierarchical correspondence information of local vertices. In other words, the proposed HAQJSK kernels have better trade-off between the whole and local structure information.

Finally, note that, it has been shown that some existing R-convolution graph kernels based on ordinary random walks [17] and Weisfeiler-Lehman subtree patterns [22] suffer from the notorious tottering problem [37]. This drawback is due to the fact that the substructures employed by these R-convolution kernels may contain multiple paths or edges connected by the same vertex pairs, causing redundant information and influencing the effectiveness of these R-convolution kernels. By contrast, the CTQW can theoretically reduce the problem of tottering through the quantum interference between vertices. Thus, the proposed HAQJSK kernels can reduce the problem of tottering through the CTQW. On the other hand, although some existing R-convolution kernels can also overcome the tottering problem associated with the backtrackless paths. For example, the non-Backtrack Path Kernel [13] based on the cycles, the Shortest Path Kernel [18] absed on the shortest paths. Unfortunately, the backtrackless paths of these R-convolution kernels only reflect local structure information. By contrast, the proposed HAQJSK kernels not only reduce the tottering problem, but also simultaneously capture whole and local structure information.

D. The Computational Complexity

Assume the set of graphs $G$ has $N$ graphs, each graph in $G$ has $n$ vertices and $m$ edges, computing the proposed HAQJSK kernels over $G$ needs three computational steps, i.e., a) compute the DB representations rooted at each vertex, b) construct the hierarchical prototype representations, c) compute the hierarchical correspondence matrix, and d) compute the QJS between hierarchical transitive aligned structures. The first computational step relies on evaluating the shortest path between vertices of each graph, and thus needs time complexity $O(N^2 n \log n + Nnm)$. The second computational step relies on the $\kappa$-means clustering method and requires time complexity $O(H^2 Mn)$, where $H$ is the greatest hierarchical level of the parameter $h$. 

| Kernel Properties       | HAQJSK | QJSK | DBAK | RK | GOK |
|------------------------|--------|------|------|----|-----|
| Positive Definite      | Yes    | No   | No   | Yes| Yes |
| Reduce Toffing         | Yes    | Yes  | No   | No | No  |
| Structural Alignment   | Yes    | Yes  | Yes  | No | No  |
| Transitive Alignment   | Yes    | No   | No   | Yes| No  |
| Local Information      | Yes    | Yes  | Yes  | Yes| No  |
| Whole Information      | Yes    | Yes  | No   | Yes| No  |
| Hierarchical Alignment | Yes    | No   | No   | --|--   |

*-- indicate that these kernels do not refer to this problem.*
defined previously, \( M \) is the number of the 1-level prototype representations (i.e., \( M = |P^{1,k}_h| \) defined previously), and \( I \) is the iteration number of \( \kappa \)-means. The third computational step relies on computing the \( h \)-level hierarchical correspondence matrix between each graph to the \( h \)-level hierarchical prototype representations, and thus needs time complexity \( O(HMn) \). The fourth computational step depends on on the spectral decomposition of the CTQW, and thus needs time complexity \( O(N^2n^3) \). Thus, the whole time complexity is \( O(Nn \log n + NMn + HIMn + HNMn + N^2n^3) \). Because \( N \gg M, N \gg I, N \gg H, \) and \( n^2 \gg m \), the resulting time complexity of the HAQJSK kernels are \( O(N^2n^3) \), indicating that the proposed kernels have a polynomial time.

IV. EXPERIMENTS

We evaluate the classification performance of the proposed kernels on twelve benchmark graph datasets extracted from computer vision (CV) [3], [59], bioinformatics (Bio) [60], and social networks (SN) [60], respectively. Table II exhibits the detailed statistical information of these datasets.

A. Experimental Setups With Graph Kernels

Experimental Setup: We compare the classification performance of the proposed HAQJSK kernels with some classical graph kernels. These kernels are 1) the Quantum Jensen-Shannon Kernel (QJSK) associated with the CTQW [37], 2) the Aligned-Subtree Kernel (ASK) [28] with the highest subtree layer 50, 3) the Jensen-Tsallis \( q \)-difference Kernel (JTQK) [54] with \( q = 2 \) and the subtrees of height 10, 4) the Graphlet Count Graph Kernel (GCGK) [55] with graphlet of size 4, 5) the classical Weisfeiler-Lehman Subtree Kernel (WLSK) [56] with the subtrees of height 10, 6) the WLSK kernel associated with Core-Variants (CORE WL) [57], 7) the Shortest Path Graph Kernel (SPGK) [18], 8) the SPGK kernel associated with Core-Variants (CORE SP) [57], 9) the Pyramid Match Graph Kernel (PMGK) [58], and 10) the depth-based Rényi Entropy Matching Kernel (REMK) [30]. More detailed properties of the kernels for comparisons are shown in Table III.

For the proposed HAQJSK kernels, we set the number of the 1-level prototype representations \( |P^{1,k}_h| \) (\( h = 1 \)), and the decay factor \( \lambda \) between \( P^{h-1,k} \) and \( P^{h,k} \) as \( 0.5 \), i.e., \( |P^{h-1,k}|| = |P^{h,k}|^{0.5} \). Moreover, we set the hierarchical parameter \( H \) as 5 (i.e., the parameter \( h \) varies from 1 to 5), and thus compute 5 kernel matrices for each datasets. The reasons of setting the above parameters are twofolds. First, Table II indicates that the number 256 is larger than the averaged vertex numbers of most datasets. Thus, for the proposed kernel on each dataset, the sizes of the required Transitive Aligned Adjacency Matrices and Density Matrices are 256, 128, 64, 32 and 16 respectively, covering the

| Kernel Methods | Kernel Frameworks | Alignment | Transitive | Structure Patterns | Computing Model | Computational Complexity |
|----------------|------------------|-----------|------------|-------------------|----------------|-------------------------|
| HAQJSK(A)      | Information Theory | Yes       | Yes        | Whole Structures  | Quantum Walks  | \( O(N^2n^3) \)         |
| QJSK [37]      | Information Theory | No        | No         | Local (Vertices)  | Quantum Walks  | \( O(N^2n^3) \)         |
| ASK [28]       | Information Theory | Yes       | No         | Local (Vertices)  | Quantum Walks  | \( O(N^2n^3) \)         |
| JTQK [54]      | Information Theory | No        | No         | Local (Subtrees) | Quantum Walks  | \( O(N^2n^3) \)         |
| GCGK [55]      | R-convolution     | No        | No         | Local (Subgraphs) | Classical      | \( O(N^2n^3) \)         |
| WLSK [56]      | R-convolution     | No        | No         | Local (Subtrees) | Classical      | \( O(N^2n^2 + N^3n) \)  |
| CORE WL [57]   | R-convolution     | No        | No         | Local (Subtrees) | Classical      | \( O(N^2n^3) \)         |
| CORE SP [57]   | R-convolution     | No        | No         | Local (Paths)    | Classical      | \( O(N^2n^3) \)         |
| PMGK [58]      | R-convolution     | Yes       | No         | Local (Vertices) | Classical      | \( O(N^2n^3) \)         |
| REMK [30]      | Information Theory | Yes       | No         | Local (Vertices) | Classical      | \( O(N^2n^2 + N^3n) \)  |

Notes: The symbols \( n \) and \( m \) corresponds to the vertex number and edge number of each graph, and the symbol \( N \) corresponds to the graph number.
TABLE IV
CLASSIFICATION COMPARISONS USING GRAPH KERNELS

| Datasets     | MUTAG | PPIs  | CATH2 | PTC(MR) | GatorBat | BAR31 |
|--------------|-------|-------|-------|---------|----------|-------|
| HAQJSK(A)    | 85.83 ± 0.72 | 89.71 ± 0.54 | 83.47 ± 0.88 | 62.35 ± 0.51 | 20.00 ± 0.84 | 68.00 ± 0.60 |
| HAQJSK(D)    | 86.33 ± 0.81 | 87.89 ± 0.35 | 59.05 ± 0.62 | 23.80 ± 0.89 | 71.70 ± 0.61 |
| QSK          | 82.72 ± 0.44 | 65.61 ± 0.77 | 71.11 ± 0.88 | 56.70 ± 0.49 | 9.00 ± 0.89 | 30.80 ± 0.61 |
| ASK          | 87.50 ± 0.65 | 80.14 ± 0.73 | 78.52 ± 0.67 | 56.22 | 7.50 ± 0.74 | 73.10 ± 0.67 |
| JTQK         | 85.50 ± 0.55 | 88.47 ± 0.47 | 68.70 ± 0.69 | 58.50 ± 0.39 | 11.40 ± 0.52 | 60.56 ± 0.35 |
| GCGK         | 83.66 ± 2.11 | 46.61 ± 0.47 | 73.68 ± 1.09 | 53.76 ± 1.41 | 8.40 ± 0.83 | 73.96 ± 0.65 |
| WLSK         | 82.88 ± 0.57 | 88.09 ± 0.41 | 67.36 ± 0.63 | 56.26 ± 0.47 | 10.19 ± 0.61 | 58.53 ± 0.53 |
| CORE WL      | 87.47 ± 1.08 | 80.94 ± 0.41 | 81.89 ± 0.64 | 55.52 ± 0.46 | 9.00 ± 0.75 | 55.73 ± 0.44 |
| SPGK         | 83.38 ± 0.81 | 90.04 ± 0.44 | 81.19 ± 0.63 | 55.96 ± 0.93 |
| CORE SP      | 88.29 ± 1.55 | - | - | - |
| PMGK         | 86.67 ± 0.60 | - | - | - |
| REMK         | 86.35 | 84.13 | 83.58 | 56.79 | 14.40 | 70.08 |

Datasets: RSPHERET1, BEOD31, IMDB-B, IMDB-M, RED-B, COLLAB

- HAQJSK(A) and HAQJSK(D) denote the proposed kernels with and without the hierarchical structure information, respectively.
- QSK, ASK, JTQK, GCGK, WLSK, CORE WL, SPGK, CORE SP and PMGK are the instances of R-convolution graph kernels, and are defined based on substructures of small sizes.
- The number 256 not only guarantees that the proposed kernel can capture dominant local structure information of graphs, but also ensures the computational efficiency of the proposed kernel, having a good trade-off between the performance and the efficiency.
- We perform the 10-fold cross-validation strategy to compute the classification accuracy through the C-Support Vector Machine (C-SVM) [61] associated with the graph kernels. For each kernel, we employ the optimal C-SVMs parameters and repeat the experiment for 10 times on each dataset. We show the average classification accuracy (± standard error) in Table IV. For some alternative kernels, the experimental results are directly from the original references or the comprehensive review paper [62], following the same experimental setups with our HAQJSK kernels.

Experimental Results and Analysis: In terms of the classification performance, it is clear that the proposed HAQJSK kernels can significantly outperform the alternative graph kernels on nine of the twelve datasets. Although, the classification accuracies of the proposed HAQJSK kernels are lower than those of some alternative kernels on the MUTAG, BAR31, and IMDB-M datasets, the HAQJSK kernels are still competitive to these kernels and better than other alternative kernels.

The reasons of the effectiveness for the proposed HAQJSK kernels are due to the advantages that have been well explained in Section III-C. First, the proposed HAQJSK kernels can either capture the complicated characteristics of whole graph structures from the CTQW or reflect local structural information through the correspondence information between vertices. By contrast, the alternative graph kernels ASK, JTQK, GCGK, WLSK, CORE WL, SPGK, CORE SP and PMGK are the instances of R-convolution graph kernels, and are defined based on substructures of small sizes. As a result, these alternative kernels can only reflect restricted local structural information. Second, as a quantum R-convolution kernel, although the JTQK kernel can also reflect comprehensive whole structural information through the CTQW. Similar to the other alternative R-convolution kernels, the JTQK kernel mainly focuses on measuring the isomorphism between substructures and ignores the structural correspondence information between the substructures. Thus, all the alternative R-convolution kernels, including the JTQK kernel, can not reflect precise kernel-based similarity measure between graphs. Third, although the QSK kernel is also defined based on the QJSD measure between whole graphs associated with the CTQW, it cannot integrate the transitive correspondence information between graphs. As a result, the QSK kernel is not a positive definite kernel, and can not reflect precise similarity measure between graphs. Moreover, the QSK kernel focuses on the whole structure information through the CTQW evolved on the graphs, but it lacks local structural information. Fourth, although the ASK and the REMK kernels can also identify the correspondence information between graphs, these two kernels can not guarantee the transitivity between the aligned vertices. Finally, note that, the proposed HAQJSK kernels are not attributed kernels and cannot accommodate the vertex attributed information. By contrast, the ASK, JTQK, WLSK and CORE WL kernels are attributed kernels. But the proposed HAQJSK kernels can still outperform these attributed kernels. In summary, the proposed HAQJSK kernels are more effective than the alternative kernels, and the strategy of the hierarchical structures associated with the CTQW can really improve the performance of graph kernels.
TABLE V

| Datasets | MUTAG | PTAM(B) | MDDB-B | MDDB-M | RDDB-B | COLLAB |
|----------|-------|---------|--------|--------|--------|--------|
| HAQJSK(A) | 86.83 ± 0.72 | 62.35 ± 0.51 | 74.50 ± 0.45 | 60.08 ± 0.20 | 78.93 ± 0.13 | 79.20 ± 0.17 |
| HAQJSK(D) | 86.34 ± 0.81 | 59.05 ± 0.62 | 72.51 ± 0.31 | 49.30 ± 0.49 | 89.50 ± 0.17 | 78.82 ± 0.14 |
| DGCNN | 85.83 ± 1.66 | 58.09 ± 2.47 | 70.03 ± 0.86 | 47.83 ± 0.85 | 70.02 ± 1.73 | 73.76 ± 0.49 |
| PSGCNN | 85.95 ± 4.37 | 62.29 | 71.00 ± 2.29 | 49.23 ± 2.84 | 86.30 ± 1.58 | 72.60 ± 2.15 |
| DCNN | 66.38 | 58.09 ± 0.55 | 49.66 ± 1.37 | 33.49 ± 1.42 | 52.11 ± 0.71 |
| DGRK | 82.66 ± 1.45 | 57.32 ± 1.13 | 66.46 ± 0.56 | 44.55 ± 0.52 | 78.30 ± 0.30 | 74.09 ± 0.25 |
| AWE | 87.87 ± 4.76 | - | 73.13 ± 3.28 | 51.55 ± 6.66 | 82.97 ± 2.86 | 70.99 ± 1.49 |
| 1-RWNN | 89.2 ± 4.3 | - | 70.4 ± 4.8 | 47.8 ± 3.8 | 90.4 ± 1.9 | 71.7 ± 2.1 |
| 2-RWNN | 86.1 ± 4.8 | - | 70.6 ± 4.4 | 48.8 ± 2.9 | 90.3 ± 1.8 | 71.3 ± 2.1 |
| 3-RWNN | 86.6 ± 4.1 | - | 70.7 ± 4.5 | 47.8 ± 3.5 | 89.7 ± 1.2 | 71.9 ± 2.5 |
| GIN | 71.43 ± 3.9 | 48.53 ± 3.3 | 89.34 ± 1.9 | 75.61 ± 2.1 |

--- : indicate that the method was not performed on the dataset by the original author.

B. Experimental Setups With Graph Deep Learning

Experimental Setup: We compare the classification performance of the proposed HAQJSK kernels with some graph deep learning methods. These methods are 1) the Deep Graph Kernel (DGK) [63], 2) the Diffusion Convolutional Neural Network (DCNN) [64], 3) the PATCHY-SAN based Convolutional Neural Network for graphs (PSGCNN) [65], 4) the Deep Graph Convolutional Neural Network (DGCNN) [66], 5) the Anonymous Walk Embeddings based on feature driven (AWE) [67], 6) the Random Walk Graph Neural Networks (p-RWNN) [68] associated with three different random walk length p (p = 1, 2, 3), and 7) the Graph Isomorphism Network (GIN) [69]. Since the alternative deep learning methods follow the same experimental setup with our proposed kernels, we directly report the classification accuracies from the original corresponding references in Table V. Note that, the PSGCNN model can adopt the edge attributes. However, since most datasets and other methods do not leverage the edge attributes, we only report the results of the PSGCNN model with vertex features for the purpose of fair comparisons. Finally, Errica et al. [70] have pointed out that the experiments of many existing graph deep learning methods often lack rigorousness and are hardly reproducible. Thus, they provide some fair experimental evaluations for some popular methods based on the same experimental setting with ours, and Nikolentzos et al. [68] also compare their p-RWNN model with other methods associated with the results reported in the reference [68]. As a result, for fair comparisons, we also directly cite the results from the reference [68] for the GIN model.

Experimental Results and Analysis: In terms of the classification performance, it is clear that the proposed HAQJSK kernels can significantly outperform the alternative graph deep learning methods on most of the datasets. In fact, the C-SVMs associated with the graph kernel can be seen as a kind of methods based on the shallow learning strategy, that may have lower performance than the deep learning methods. But the proposed HAQJSK kernels still have better classification performance than these graph deep learning methods. This may due to the fact that some of these graph deep learning methods (i.e., the graph neural network models DGCNN, PSGCNN, GIN, p-RWNN and DCN) are theoretically related to the Weisfeiler-Lehman (WL) isomorphism test of the WLSK kernel [66] or the random walks, since they all rely on the information propagation between adjacent vertices. As a result, similar to the WLSK kernel, these graph convolution network models may also suffer from the totering problem. By contrast, the CTQW of the proposed HAQJSK kernels can reduce the problem of totering. Moreover, most of the alternative DGCNN, PSGCNN, GIN, p-RWNN and DCN models need to construct the fix-sized grid structures of graphs for classification task associated with the SoftMax layer. This is usually achieved by reordering the vertices for each graph structure, and then preserve a specified number of top ranked vertices. As a result, similar to the R-convolution kernels mentioned in Section IV-A, these graph neural network models cannot guarantee that the vertices at the same spatial position are also aligned to each other in terms of the structures. At the meanwhile, this may also lead to significant information loss. By contrast, the proposed HAQJSK kernels not only identify the transitive structural correspondence information between graphs, but also preserve the original structure information. Overall, this evaluation again indicates that the strategy of the hierarchical structures associated with the CTQW can improve the performance of graph kernels.

V. CONCLUSIONS AND FUTURE WORKS

In this work, we propose a family of HAQJSK kernels based on measuring the QJS between hierarchical transitive aligned structured graphs associated with the CTQW. The proposed HAQJSK kernels not only overcome the shortcoming of ignoring structural correspondence information arising in classical R-convolution kernels, but also simultaneously reflect richer local and whole structural information than the R-convolution kernels that only reflect local structural information on substructures. Moreover, unlike most existing alignment kernels, the proposed HAQJSK kernels can guarantee the transitivity between the correspondence information. Thus, the proposed HAQJSK kernels not only positive definite kernels, but also reflect more precise similarity measure between graphs. The experimental evaluations indicate the effectiveness.

The proposed kernels have demonstrated the theoretical advantages of the CTQW. Our future work is to further explore the theoretical relationship between the CTQW and the WL method, or between the CTQW and the random walks. This may help us to further improve the WL or random walk inspired graph neural networks through the CTQW.
...
Philip S. Yu (Fellow, IEEE) received the BS degree in electrical engineering from National Taiwan University, the MS and PhD degrees in EE from Stanford University, and the MBA degree from New York University. He is currently a distinguished professor of computer science with the University of Illinois at Chicago (UIC), and holds the Wexler Chair in information technology. He has published more than 970 papers in refereed journals and conferences. He holds or has applied for more than 300 US patents. He was a member of the Steering Committee of the IEEE Data Engineering and the IEEE Conference on Data Mining. He is a fellow of the ACM. He is on the Steering Committee of the ACM Conference on Information and Knowledge Management. He received the ACM SIGKDD 2016 Innovation Award for his influential research and scientific contributions on mining, fusion, and anonymization of Big Data, the IEEE Computer Society’s 2013 Technical Achievement Award for “pioneering and fundamentally innovative contributions to the scalable indexing, querying, searching, mining, and anonymization of Big Data”, and the Research Contributions Award from ICDM 2003, for his pioneering contributions to the field of data mining. He also received the ICDM 2013 10-year Highest-Impact Paper Award, and the EDBT Test of Time Award (2014). He has received several IBM honors, including two IBM Outstanding Innovation Awards, an Outstanding Technical Achievement Award, two Research Division Awards, and the 94th plateau of Invention Achievement Awards. He was the editor-in-chief of IEEE Transactions on Knowledge and Data Engineering (2001–2004).

Edwin R. Hancock (Fellow, IEEE) received the BSc degree in physics, the PhD degree in high-energy physics, and the DSc degree from Durham University, Durham, U.K., in 1977, 1981, and 2008, respectively, and the Doctorate Honoris Causa from the University of Alicante, Alicante, Spain, in 2015. He is currently an emeritus professor with the Department of Computer Science, University of York, York, U.K., an adjunct professor with Beihang University, Beijing, China, a distinguished visiting professor with Xiamen University, Xiamen, China and a Shanghai Science and Technology Commission Overseas visiting fellow with Shanghai University. His main research interests include pattern recognition, machine learning, and computer vision, in which he has made sustained contributions to the use of graph-based methods and physics-based vision. He has published more than 200 journal articles and 650 conference papers. He was the 2016 distinguished fellow of the British Machine Vision Association (BMVA). He was elected as a fellow of the Royal Academy of Engineering (the U.K.’s National Academy of Engineering) in 2021. He is also a fellow of the International Association for Pattern Recognition (IAPR). He has also been a member of the editorial boards of the following journals: IEEE Transactions on Pattern Analysis and Machine Intelligence, Pattern Recognition, Computer Vision and Image Understanding, Image and Vision Computing, and International Journal of Complex Networks. He was a recipient of the Pattern Recognition Medal in 1992, the IAPR Piero Zamperoni Award in 2006, the Royal Society Wolfson Research Merit Award in 2008, and the IAPR Pierre Devijver Award in 2018. He was the founding editor-in-chief of the IET Computer Vision from 2006 to 2012 and was the editor-in-chief of the Journal Pattern Recognition (2016–2024). He was the vice president of the IAPR from 2016 to 2018. He is an IEEE Computer Society distinguished visitor for the period 2021–2023.