Pattern Recognition Letters
journal homepage: www.elsevier.com

Local-Global Nested Graph Kernels Using Nested Complexity Traces

Lu Bai*[a], Lixin Cui*[a], Luca Rossi[b], Lixiang Xu[c], Xiao Bai[d], Edwin Hancock[e]

[a] *Central University of Finance and Economics, Beijing, China*
[b] *Aston University, Birmingham, UK*
[c] *Hefei University, Anhui, China*
[d] *Beihang University, Beijing, China*
[e] *University of York, York, UK*

ABSTRACT

In this paper, we propose two novel local-global nested graph kernels, namely the nested aligned kernel and the nested reproducing kernel, drawing on depth-based complexity traces. Both of the nested kernels gauge the nested depth complexity trace through a family of K-layer expansion subgraphs rooted at the centroid vertex, i.e., the vertex with minimum shortest path length variance to the remaining vertices. Specifically, for a pair of graphs, we commence by computing the centroid depth-based complexity traces rooted at the centroid vertices. The first nested kernel is defined by measuring the global alignment kernel, which is based on the dynamic time warping framework, between the complexity traces. Since the required global alignment kernel incorporates the whole spectrum of alignment costs between the complexity traces, this nested kernel can provide rich statistic measures. The second nested kernel, on the other hand, is defined by measuring the basic reproducing kernel between the complexity traces. Since the associated reproducing kernel only requires time complexity $O(1)$, this nested kernel has very low computational complexity. We theoretically show that both of the proposed nested kernels can simultaneously reflect the local and global graph characteristics in terms of the nested complexity traces. Experiments on standard graph datasets abstracted from bioinformatics and computer vision databases demonstrate the effectiveness and efficiency of the proposed graph kernels.
Keywords: Graph Kernels, Depth-based Complexity Traces, Nested Kernels

1. Introduction

In pattern recognition, graph kernels are powerful tools for analyzing structured data represented by graphs Riesen and Bunke (2010). This is because graph kernels not only preserve structural information by implicitly mapping graphs to a high dimensional Hilbert space, but also provide a way of directly applying standard kernel methods for vectorial data (e.g., Support Vector Machines, kernel Principle Component Analysis) to graph structures.

1.1. Literature Review

The idea underpinning most existing graph kernels is that of decomposing graphs into substructures and comparing pairs of specific isomorphic substructures, e.g., walks Urry and Sollich (2013), paths Alvarez et al. (2011), and restricted subgraph or subtree substructures Costa and De Grave (2010). Under this scenario, Bach (2008) have developed a family of kernels for comparing point clouds. These kernels are based on a local tree-walk kernel between subtrees, which is defined by a factorization on suitably defined graphical models of the subtrees. Wang and Sahbi (2013), on the other hand, have defined a graph kernel for action recognition. They first describe actions in the videos using directed acyclic graphs (DAGs). The resulting kernel is defined as an extending random walk kernel by counting the number of isomorphic walks of DAGs. Harchaoui and Bach (2007) have proposed a segmentation graph kernel for images by counting the inexact isomorphic subtree patterns between

E-mail: bailu69@hotmail.com; bailucs@cufe.edu.cn
E-mail: cuilixin@cufe.edu.cn (Corresponding Author)
E-mail: l.rossi@aston.ac.uk (Co-corresponding Author)
E-mail: baixiao@huaa.edu.cn
E-mail: erh@cs.york.ac.uk
*These authors are co-first authors

image segmentation graphs. Other state-of-the-art graph kernels based on substructures include the aligned subtree kernel proposed by Bai et al. (2015b), the subgraph matching kernel proposed by Kriege and Mutzel (2012), the fast depth-based subgraph kernel proposed by Bai and Hancock (2016), the optimal assignment kernel proposed by Kriege et al. (2016), and the random walk kernel proposed by Kashima et al. (2003).

Unfortunately, all the aforementioned graph kernels tend to only capture local characteristics of graphs, since they usually use substructures of limited sizes. As a result, these kernels may fail to reflect global graph characteristics. To overcome this shortcoming, a number of graph kernels based on using the adjacency matrix to capture global graph characteristics have been developed by Johansson et al. (2014); Xu et al. (2015); Bai and Hancock (2013). For instance, Johansson et al. (2014) have developed a family of global graph kernels based on the Lovász number and its associated orthonormal representation through the adjacency matrix. Xu et al. (2015) have proposed a local-global mixed reproducing kernel based on the approximate von Neumann entropy through the adjacency matrix. Bai and Hancock (2013) have defined an information theoretic kernel based on the classical Jensen-Shannon divergence between the steady state random walk probability distributions obtained through the adjacency matrix. Recently, there has been increasing interests in continuous-time quantum walks for the analysis of global graph structures Farhi and Gutmann (1998). The continuous-time quantum walk is the quantum analogue of the classical continuous-time random walk. Unlike the classical random walk that is governed by a doubly stochastic matrix, the quantum walk is governed by an unitary matrix and is not dominated by the low frequencies of the Laplacian spectrum. Thus, the continuous-time quantum walk is able to better discriminate different graph structures.

There have been a number of graph kernels developed using the continuous-time quantum walk. For instance, Bai et al. (2015a) have developed a quantum kernel by measuring the similarity between two continuous-time quantum walks evolving on a pair of graphs. Specifically, they associate each graph with a mixed quantum state that represents the evolution of the quantum walk. The resulting kernel is computed by measuring the quantum Jensen-Shannon divergence between the associated density matrices. Rossi et al. (2015) have developed a quantum kernel by exploiting the relation between the continuous-time quantum walk interferences and the symmetries of a pair of graphs, in terms of the quantum Jensen-Shannon divergence. Both of these quantum kernels employ the Laplacian matrix as the required Hamiltonian operator, and thus can naturally reflect global graph characteristics.

1.2. Contributions

The aim of this work is to overcome the gap between local kernels (i.e., kernels based on local substructures of limited sizes) and the global kernels (i.e., global kernels based on either the adjacency matrix or the continuous-time quantum walk). To this end, we propose two novel local-global nested graph kernels, namely the nested aligned kernel and the nested reproducing kernel, drawing on depth-based complexity traces Bai and Hancock (2016). Both of the nested kernels gauge the nested depth complexity trace through a family of K-layer expansion subgraphs rooted at the centroid vertex, that has minimum shortest path length variance to the remaining vertices. Specifically, for a pair of graphs, we commence by computing the centroid depth-based complexity traces rooted at the centroid vertices. The first nested kernel is defined by measuring the global alignment kernel, which is developed through the dynamic time warping framework, between the complexity traces Cuturi (2011). Since the required global alignment kernel incorporates the whole spectrum of alignment costs between the complexity traces, this nested kernel can provide rich statistic measures. The second nested kernel, on the other hand, is defined by measuring the reproducing kernel between the complexity traces Xu et al. (2015, 2017). Since the associated reproducing kernel only requires time complexity $O(1)$, this nested kernel has efficient computational complexity. We theoretically show that both of the proposed nested kernels can simultaneously reflect the local and global graph characteristics in terms of the nested complexity traces. Experiments on standard graph datasets abstracted from bioinformatics and computer vision databases demonstrate the effectiveness and efficiency of the proposed graph kernels.

1.3. Paper Outline

The remainder of this paper is organized as follows. Section 2 reviews the preliminary concepts that will be used in this work. Specifically, we introduce the global alignment kernel through the dynamic time warping framework, the reproducing kernel, the approximate von Neumann entropy, the Shannon entropy associated with steady state random walks, and the centroid depth-based complexity trace. Section 3 defines the proposed local-global nested graph kernels. Section 4 provides the experimental evaluation. Section 5 concludes this work.

2. Preliminary Concepts

In this section, we review some preliminary concepts that will be used in this work. We commence by reviewing the dynamic time warping framework. Specifically, we introduce the global alignment kernel based on this framework Cuturi (2011). Moreover, we review a reproducing kernel that is an extension of the H^1-reproducing kernel to the graph kernel realm. Finally, we review the concept of the depth-based complexity trace that naturally forms a nested sequence of a graph in terms of the entropy measure.

2.1. Global Alignment Kernels from the Dynamic Time Warping Framework

In this subsection, we review the global alignment kernel based on the dynamic time warping framework proposed by Cuturi (2011). Let $\mathbf{T}$ be a set of discrete time series that take values in a space $\mathcal{X}$. For a pair of discrete time series $\mathbf{P} = (p_1, \ldots, p_m) \in \mathbf{T}$ and $\mathbf{Q} = (q_1, \ldots, q_n) \in \mathbf{T}$ with lengths m and n respectively, the alignment π between $\mathbf{P}$ and $\mathbf{Q}$ is

defined as a pair of increasing integral vectors (π_p, π_q) of length $l \leq m + n - 1$, where

$$1 = \pi_p(1) \leq \cdots \leq \pi_p(l) = m$$

and

$$1 = \pi_q(1) \leq \cdots \leq \pi_q(l) = n$$

such that (π_p, π_q) is defined to have unitary increments and no simultaneous repetitions. For any index $1 \leq i \leq l - 1$, the increment vector of $\pi = (\pi_p, \pi_q)$ satisfies

$$\begin{pmatrix} \pi_p(i+1) - \pi_p(i) \\ \pi_q(i+1) - \pi_q(i) \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}. \quad (1)$$

In the dynamic time warping framework, the coordinates π_p and π_q of the alignment π define the warping function Cuturi (2011). Let $\mathcal{A}(m, n)$ be the set of all possible alignments between $\mathbf{P}$ and $\mathbf{Q}$. The dynamic time warping distance between $\mathbf{P}$ and $\mathbf{Q}$ is defined as

$$\mathrm{DTW}(\mathbf{P}, \mathbf{Q}) = \min_{\pi \in \mathcal{A}(m,n)} D_{\mathbf{P},\mathbf{Q}}(\pi), \quad (2)$$

where the cost

$$D_{\mathbf{P},\mathbf{Q}}(\pi) = \sum_{i=1}^{|\pi|} \varphi(p_{\pi_p(i)}, q_{\pi_q(i)}), \quad (3)$$

is defined by a local divergence φ that measures the discrepancy between any pair of elements $p_i \in \mathbf{P}$ and $q_i \in \mathbf{Q}$. Generally, φ can be defined as the squared Euclidean distance, i.e., $\varphi(p, q) = \|p - q\|^2$.

Based on the dynamic time warping distance defined in Eq.(2), Haasdonk and Bahlmann (2004) have defined a dynamic time warping kernel k_{DTW} between $\mathbf{P}$ and $\mathbf{Q}$ as

$$k_{\mathrm{DTW}}(\mathbf{P}, \mathbf{Q}) = e^{-\mathrm{DTW}(\mathbf{P},\mathbf{Q})}. \quad (4)$$

Unfortunately, this kernel is not positive definite. This is because the optimal alignment required by the dynamic time warping cannot guarantee transitivity. To overcome the shortcoming, Cuturi (2011) considers all possible alignments in $\mathcal{A}(m, n)$ and proposes another dynamic time warping inspired kernel, i.e., the global alignment kernel, as

$$k_{\mathrm{GA}}(\mathbf{P}, \mathbf{Q}) = \sum_{\pi \in \mathcal{A}(m,n)} e^{-D_{\mathbf{P},\mathbf{Q}}(\pi)}, \quad (5)$$

where k_{GA} is positive definite, since it quantifies the quality of both the optimal alignment and all other alignments $\pi \in \mathcal{A}(m, n)$. The kernel k_{GA} elaborates on the dynamic time warping distance by considering the same set of elementary operations Cuturi et al. (2007). However k_{GA} not only generalizes the dynamic time warping kernel k_{DTW}, but also provides richer statistic measures by incorporating the whole spectrum of alignment costs $\{D_{\mathbf{P},\mathbf{Q}}(\pi), \pi \in \mathcal{A}(m, n)\}$.

Intuitively, the global alignment kernel k_{GA} allows one to define a new graph kernel, by measuring the warping alignment π between any types of graph characteristic sequences that have certain element orders with increasing structural variables, e.g, the graph embedding vectors proposed by Conte et al. (2013), the depth-based complexity traces from expansion subgraphs of increasing sizes proposed by Bai and Hancock (2016) , or cycle characteristics with increasing lengths identified from the Ihara zeta function proposed by Ren et al. (2011).

2.2. The Reproducing Kernel

In mathematics, a Hilbert Space is an inner product space that is complete and separable with respect to the norm defined by the inner product. If the Hilbert space contains complex-valued functions associated with a reproducing kernel, we call it as a reproducing kernel Hilbert space (RKHS) or a proper Hilbert space. Generally speaking, an RKHS has nice properties if a function $f(x)$ in the RKHS is close to a function $g(x)$ in the sense of the distance derived from the inner product.

Definition 1. (The reproducing kernel) A function $K : E \times E \to C$, $(s, t) \mapsto K(s, t)$ is a reproducing kernel of the Hilbert space H if and only if
(i) $\forall t \in E$, $K(., t) \in H$;
(ii) $\forall t \in E$, $\forall \phi \in H$ $\langle \phi, K(., t) \rangle = \phi(t)$.
The last condition (ii) is called the reproducing property, i.e., the value of the function ϕ at the point t is reproduced by the inner product of ϕ with $K(., t)$. □

In this subsection, we review how to compute a basic reproducing kernel for graphs based on the work of Xu et al. (2018). We start with the concept of the H^1-reproducing kernel in $H^1(R)$ space, which can be seen as an extension of the H^1-reproducing kernel to the graph kernel realm. Specifically, in the following Lemma 1, we obtain the basic solution of the generalized differential operator using the Delta function based on the work of Xu et al. (2015, 2017). The Delta function $\sigma(x)$ physically represents the density of an idealized point mass or a point charge. In practice, the Delta function plays an important role in partial differential equations, mathematical physics, Fourier analysis, and theory of probability Aronszajn (1950). Assume the real number set and the integer set are denoted by R and Z, respectively. Let $H^n(R) = \{u(x) | u(x), u'(x), u''(x), \ldots, u^{(n-1)}\}$ are absolutely continuous functions in $\{R, u'(x), u''(x), \ldots, u^{(n)} \in L^2(R)\}$, where $n \in Z^+$. The inner product in $H^n(R)$ space is defined as

$$\langle u, v \rangle_{H^n(R)} = \int_R (\sum_{i=1}^{n} c_i u^{(i)} v^{(i)}) \mathrm{d}x, \forall u, v \in H^n(R), \quad (6)$$

where $C_i (i = 0, 1, 2, \ldots, n)$ is the coefficient of

$$(a + b)^n = \sum_{i=0}^{n} c_i a^i b^{n-i}. \quad (7)$$

Lemma 1. *Let $K_1(x)$ be the basic solution of the operator $L = 1 - \frac{d^2}{dx^2}$, then the basic reproducing kernel of $H^1(R)$ is $K_1(x - y)$.*

By Xu et al. (2015), we know the function

$$K_1(x, y) = K_1(x - y) = \frac{1}{2} e^{-|x-y|}, \quad (8)$$

which obviously satisfies condition (i) and (ii) of Definition 1. So $K_1(x, y) = K_1(x - y)$ is a H^1-reproducing kernel in $H^1(R)$ space. □

Intuitively, the basic reproducing kernel K_1 allows one to define a new basic reproducing graph kernel associated with any type of graph characteristics values, e.g., the graph entropy measures suggested Bai and Hancock (2014). Moreover, since the computation of the basic reproducing kernel K_1 only requires time complexity $O(1)$ and the computation of some entropy measures are only quadratic in the number of vertices.

The basic reproducing kernel K_1 provides us a way of defining new fast graph kernel associated with graph entropy measures. For instance, Xu et al. (2018) have proposed a hybrid reproducing kernel by measuring the basic reproducing kernel K_1 between the entropies of global graphs. Since the associated entropy measures only require time complexity $O(n^2)$ where n is the vertex number of the graph, their hybrid reproducing kernel only requires time complexity $O(n^2+1)$. Unfortunately, the hybrid reproducing kernel between global graph entropies cannot reflect local characteristics from the global graph structure.

2.3. The Entropy Measure of A Graph

We review the concepts of two state-of-the-art graph entropy measures, namely the approximate von Neumann entropy and the Shannon entropy associated with the steady state random walk proposed by Han et al. (2012) and Bai and Hancock (2014) respectively. Assume we have a sample undirected graph denoted as $G(V,E)$ where V is the vertex set and $E \subseteq V \times V$ is the undirected edge set. The adjacency matrix A of the graph $G(V,E)$ is a $|V| \times |V|$ symmetric matrix and each element satisfies

$$A(i,j) = \begin{cases} 1 & \text{if}(v_i, v_j) \in E; \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The vertex degree matrix D of G is a diagonal matrix whose elements are defined by

$$D(v_i, v_i) = d(i) = \sum_{v_j \in V} A(i,j). \tag{10}$$

Definition 2. (The Approximate Von Neumann Entropy) Based on the definition in the work of Han et al. (2012), we can compute a fast approximate von Neumann entropy for the graph $G(V,E)$ in terms of its degree matrix D as

$$H_{VN}(G) = 1 - \frac{1}{|V|} - \sum_{(v_i,v_j)\in E} \frac{1}{|V|^2 d(i) d(j)}, \tag{11}$$

where each edge $(v_i, v_j) \in E$ is indicated by the adjacency matrix A defined by Eq.(9). □

Definition 3. (The Shannon Entropy) For each vertex $v_i \in V$ of the graph $G(V,E)$, the probability of a steady state random walk on $G(V,E)$ visiting v_i is

$$P(i) = d(i) / \sum_{v_j \in V} d(j). \tag{12}$$

From this probability distribution P, we can straightforwardly compute the Shannon entropy as

$$H_S(G) = -\sum_{i=1}^{|V|} P(i) \log P(i). \tag{13}$$

□

Both the approximate von Neumann entropy and the Shannon entropy require computational complexity $O(n^2)$, where n is the vertex number. This is because both the entropy measures rely on the vertex degree statistics computed from the pairs of vertices connected by edges, and the number of such edges is at most $\frac{n(n-1)}{2}$. This observation indicates that both the entropy measures can be efficiently computed. Finally, note that, based on the observations by Bai and Hancock (2013), the approximate von Neumann entropy and the Shannon entropy also have different properties. Eq.(11) indicates that the approximate von Neumann entropy is computed through the reciprocals of connected vertex degrees and is thus sensitive to edges connected by vertices with low degrees. Because such edges usually form bridges between vertex clusters, the von Neumann entropy is sensitive to the interconnections between vertex clusters within a graph. On the other hand, Eq.(13) indicates that the Shannon entropy is dominated by vertices with large degrees. As a result, this entropy measure responds most to graph structures consisting of groups with highly intra-connected vertices.

2.4. Centroid Depth-based Complexity Traces

We review the concept of the centroid depth-based complexity trace of a graph rooted at the centroid vertex proposed by Bai and Hancock (2014). Let $G(V,E)$ be an undirected graph with vertex set V and edge set E. Based on Dijkstra's algorithm, we commence by computing the shortest path matrix S_G, where each element $S_G(v,u)$ of S_G represents the length of the shortest path between vertices $v \in V$ and $u \in V$. For each vertex $v \in V$, let $S(v)$ be the average length of the shortest paths from v to the remaining vertices, i.e.,

$$S(v) = \frac{1}{|V|} \sum_{u \in V} S_G(v,u). \tag{14}$$

As discussed by Bai and Hancock (2014), the centroid vertex $\hat{v}_C$ of $G(V,E)$ can be identified by selecting the vertex that has the minimum variance of shortest path lengths to the remaining vertices, i.e., the index of $\hat{v}_C$ is

$$\hat{v}_C = \arg\min_{v} \sum_{u \in V} [S_G(v,u) - S_V(v)]^2. \tag{15}$$

Let $N^K_{\hat{v}_C}$ be a vertex subset of $G(V,E)$ satisfying

$$N^K_{\hat{v}_C} = \{u \in V \mid S_G(\hat{v}_C, u) \leq K\}. \tag{16}$$

For $G(V,E)$ and its centroid vertex $\hat{v}_C$, we construct a family of K-layer centroid expansion subgraphs $\mathcal{G}_K(\mathcal{V}_K; \mathcal{E}_K)$ as

$$\begin{cases} \mathcal{V}_K &= \{u \in N^K_{\hat{v}_C}\}; \\ \mathcal{E}_K &= \{(u,v) \subset N^K_{\hat{v}_C} \times N^K_{\hat{v}_C} \mid (u,v) \in E\}. \end{cases} \tag{17}$$

Note that the number of expansion subgraphs is equal to the greatest length L of the shortest paths from the centroid vertex to the remaining vertices of $G(V,E)$. Moreover, the L-layer expansion subgraph is $G(V,E)$ itself. An example of constructing a K-layer expansion subgraph is shown in Fig.1.

Definition 4. (Centroid Depth-based complexity traces) For a graph $G(V,E)$, let $\{\mathcal{G}_1, \cdots, \mathcal{G}_K, \cdots, \mathcal{G}_L\}$ be the family of K-layer centroid expansion subgraphs rooted at the centroid vertex of $G(V,E)$. Based on Bai and Hancock (2016), the centroid depth-based complexity trace $DB(G)$ of $G(V,E)$ is computed by measuring the entropies of the subgraphs, i.e.,

$$\text{DB}(G) = \{H(\mathcal{G}_1), \cdots, H(\mathcal{G}_K), \cdots, H(\mathcal{G}_L)\}, \tag{18}$$

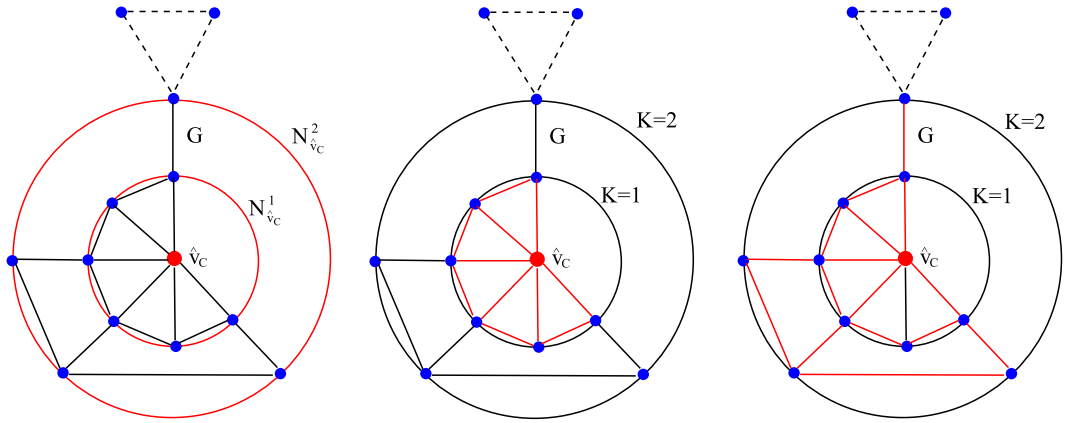


Fig. 1. The left-most figure shows the determination of K-layer centroid expansion subgraphs for a graph $G(V,E)$ which hold $|N^1_{\hat{v}_C}| = 6$ and $|N^2_{\hat{v}_C}| = 10$ vertices. While the middle and the right-most figure show the corresponding 1-layer and 2-layer subgraphs regarding the centroid vertex $\hat{v}_C$, and are depicted by red-colored edges. In this example, the vertices of different K-layer subgraphs regarding the centroid vertex $\hat{v}_C$ are calculated by Eq.(15), and pairwise vertices possess the same connection information in the original graph $G(V,E)$.

where $H(\mathcal{G}_K)$ is the entropy measure of the K-layer expansion subgraph $\mathcal{G}_K$, and it can be either the approximate von Neumann entropy defined by Eq.(11) or the Shannon entropy defined by Eq.(13). □

Based on Bai and Hancock (2016), the centroid depth-based complexity trace has a number of interesting properties. First, it encapsulates the entropy-based information content flow through the family of K-layer centroid expansion subgraphs rooted at the centroid vertex, and thus reflects rich intrinsic depth topology information of a graph. Second, it can be efficiently computed on large graphs. This is because it is computed on a small set of expansion subgraphs rooted at the centroid vertex, and the computational complexity is polynomial.

Furthermore, we observe that the centroid depth-based complexity trace $\mathrm{DB}(G) = \{H(\mathcal{G}_1), \cdots, H(\mathcal{G}_K), \cdots, H(\mathcal{G}_L)\}$ of each graph G also preserves ***nest property***, i.e., the entropy-based complexity information of each K-layer expansion subgraph encapsulates the information of the 1-layer to $K-1$-layer expansion subgraphs. This follows the fact that the family of K-layer expansion subgraphs rooted at the centroid vertex $\hat{v}_C$ of G constructs a nested sequence. Specifically, based on Eq.(17), we observe that the family of expansion subgraphs satisfies

$$\hat{v}_C \in \mathcal{G}_1 \cdots \subseteq \mathcal{G}_K \subseteq \cdots \subseteq \mathcal{G}_L \subseteq G.$$

In other words, it represents a sequence of subgraphs that gradually expand from the centroid vertex to the global graph structure, and each K-layer expansion subgraph completely encapsulates the structure information from the lower layer expansion graphs, i.e., the 1-layer to $K-1$-layer expansion subgraphs. As a result of its nested nature, the centroid depth-based complexity trace can be seen as a nested complexity traces that naturally reflects both the local and global structure information of a graph.

In a summary, the centroid depth-based complexity trace provides an elegant way of developing novel fast graph kernels that simultaneously consider local and global graph structures.

3. The Local-Global Nested Graph Kernel

In this section, we introduce two novel local-global nested graph kernels, namely the nested aligned kernel and the nested reproducing kernel, that can reflect both the local and global graph characteristics through the centroid depth-based representations. Specifically, the first nested graph kernel is based on the dynamic time warping measure between the centroid depth-based complexity traces. On the other hand, the second nested graph kernel is based on the basic reproducing kernel between the centroid depth-based complexity traces. We show that both of the kernels can be computed in polynomial time.

3.1. The Nested Graph Kernels

Let $G_P(V_P, E_P)$ and $G_Q(V_Q, E_Q)$ be a pair of graphs, from a graph set $\mathbf{G}$. We commence by computing the centroid depth-based complexity traces of G_P and G_Q rooted at their centroid vertices as

$$\mathrm{DB}(G_P) = \{H(\mathcal{G}_{P;1}), \cdots, H(\mathcal{G}_{P;K}), \cdots, H(\mathcal{G}_{P;L_{\max}})\}$$

and

$$\mathrm{DB}(G_Q) = \{H(\mathcal{G}_{Q;1}), \cdots, H(\mathcal{G}_{Q;K}), \cdots, H(\mathcal{G}_{Q;L_{\max}})\},$$

respectively, where $\mathcal{G}_{P;K}$ and $\mathcal{G}_{Q;K}$ are the K-layer expansion subgraphs rooted at the centroid vertices of G_P and G_Q, $L_{\max}$ is the greatest length of the shortest paths rooted at the centroid vertices over all graphs in $\mathbf{G}$, and the entropy measure $H(.)$ of a K-layer expansion subgraph can be either the approximate von Neumann entropy defined in Eq.(11) or the Shannon entropy defined in Eq.(13). Note that, for G_P and G_Q, if their greatest lengths M and N of the shortest paths rooted at their centroid vertices satisfy $K \geq M$ and $K \geq N$, their K-layer expansion subgraphs are their global structures. We consider two alternative ways to define nested graph kernels based on the centroid depth-based representations.

Definition 5. (The Nested Aligned Kernel) Based on the global alignment kernel defined in Section 2.1, we develop a new

nested aligned graph kernel k_{NAK} between G_P and G_Q as

$$k_{\mathrm{NAK}}(G_P, G_Q) = k_{\mathrm{GA}}\{\mathrm{DB}(G_P), \mathrm{DB}(G_Q)\} = \sum_{\pi \in \mathcal{A}(L_{\max}, L_{\max})} e^{-D_{\mathbf{P},\mathbf{Q}}(\pi)}, \quad (19)$$

where π denotes the warping alignment between $\mathrm{DB}(G_P)$ and $\mathrm{DB}(G_Q)$, $\mathcal{A}(L_{\max}, L_{\max})$ denotes all possible alignments, and $D_{\mathbf{P},\mathbf{Q}}(\pi)$ is the alignment cost defined in Eq.(3). Note that, although our kernel is based on the global alignment kernel k_{GA} that is a positive definite kernel, the time series compared by k_{NAK} are not defined over the same underlying space but on two different graphs. Thus, we cannot prove that the the proposed kernel k_{NAK} is positive definite. In our future work, we will further explore the possibility of creating a positive definite kernel by computing the depth-based complexity traces over a common structure obtained by combining the input graphs, e.g., a union graph that preserves the structural information of all graphs.

Assume we have a pair of graphs each having n vertices and m edges, computing the nested aligned kernel k_{NAK} requires time complexity $O(m \log n + L_{\max}^2)$. This is because identifying the centroid vertex and computing the centroid depth-based complexity trace of a graph rely on the computation of the shortest path matrix and both of the two processes require time complexity $O(m \log n)$. Furthermore, computing all possible alignments between the depth-based complexity traces has time complexity $O(L_{\max}^2)$, where $L_{\max}$ is the greatest length of the shortest paths rooted at the centroid vertices of all graphs (note that $L_{\max}$ is usually much lower than the vertex number n and the edge number m). As a result, the proposed kernel k_{NAK} has polynomial time complexity $O(m \log n + L_{\max}^2)$.

Definition 6. (The Nested Reproducing Kernel) Based on the basic reproducing kernel K_1 defined in Section 2.2, we develop a new nested reproducing graph kernel k_{NRK} between G_P and G_Q as

$$k_{\mathrm{NRK}}(G_P, G_Q) = k_{\mathrm{NRK}}\{\mathrm{DB}(G_P), \mathrm{DB}(G_Q)\} = \sum_{K=1}^{L_{\max}} K_1\{H(\mathcal{G}_{P;K}), H(\mathcal{G}_{Q;K})\} = \frac{1}{2}\sum_{K=1}^{L_{\max}} e^{-|H(\mathcal{G}_{P;K}) - H(\mathcal{G}_{Q;K})|}. \quad (20)$$

The entropy measure $H(.)$ can be either the approximate von Neumann entropy or the Shannon entropy. Moreover, unlike k_{NAK}, we can guarantee that the proposed nested reproducing kernel k_{NRK} is positive definite (***pd***). This is because the associated basic reproducing kernel $K_1\{H_S(\mathcal{G}_{P;K}), H_S(\mathcal{G}_{Q;K})\}$ between each pair of K-layer expansion subgraphs is ***pd***, and the resulting nested reproducing kernel $k_{\mathrm{NRK}}(G_P, G_Q)$ can be seen as the sum of the ***pd*** kernel K_1 between $L_{\max}$ pairs of K-layer expansion subgraphs.

Assume we have a pair of graphs each having n vertices and m edges. Computing the nested reproducing kernel k_{NRK} requires time complexity $O(m \log n + L_{\max})$. This is because, as we have stated, computing the centroid depth-based complexity trace of each graph requires time complexity $O(m \log n)$. Moreover, computing the reproducing kernel K_1 between the entropies of the L pairs of K-layer expansion subgraphs requires time complexity $O(L_{\max})$. As a result, the proposed kernel k_{NRK} has polynomial time complexity $O(m \log n + L_{\max})$. Comparing to the proposed nested aligned kernel k_{NAK} that has time complexity $O(m \log n + L_{\max}^2)$, the proposed nested reproducing kernel k_{NRK} has more efficient computational complexity.

3.2. Discussions and Related Works

As we have stated in Section 2.4, the required centroid depth-based complexity trace for the proposed nested kernels gauges the entropy-based complexities on the K-layer expansion subgraphs rooted at the centroid vertex. Since the family of these expansion subgraphs gradually lead the centroid vertex to the global graph structure, the centroid depth-based complexity trace can be seen as a nested complexity trace that naturally reflects both the local and global structure information of a graph. As a result, either of the proposed graph kernels can be seen as a local-global nested graph kernel that simultaneously captures local and global graph characteristics. Our proposed kernels overcome the gap between the local kernels (i.e., kernels based on local substructures of limited sizes proposed by Harchaoui and Bach (2007); Kriege and Mutzel (2012); Kriege et al. (2016)) and the global kernels (i.e., global kernels based on either the adjacency matrix or the continuous-time quantum walk proposed by Johansson et al. (2014); Xu et al. (2015); Bai and Hancock (2013)). Furthermore, our proposed local-global nested graph kernels are based on a small number of dominant K-layer expansion subgraphs rooted at the centroid vertex, and thus have more efficient computational complexity than state-of-the-art graph kernels based on a large number of substructures.

The proposed local-global nested graph kernels are related to some state-of-the-art methods. Specifically, the proposed nested aligned kernel is related to the global alignment kernel developed by Cuturi (2011), since they are both based on the dynamic time wrapping framework. On the other hand, the nested reproducing kernel is related to the hybrid reproducing kernel developed by Xu et al. (2018), since both of the kernels are based on the basic reproducing kernel proposed by Xu et al. (2018). However, the proposed nested kernels are still theoretically different from these existing methods. First, the original global alignment kernel is designed for vectorial time series and thus is not available for graphs. By contrast, the proposed nested aligned kernel transforms each graph structure into a nested centroid depth-based complexity trace, and measures the alignment score between the traces of pairwise graphs. ***The nested aligned kernel thus bridges the gap between graph kernels and the dynamic time wrapping framework***. Second, the hybrid reproducing kernel is based on both the Shannon and von Neumann entropies of global graph structures, and is only restricted to reflecting global graph characteristics. By contrast, the proposed nested reproducing kernel is based on the nested centroid depth-based complexity trace that gradually leads the entropy measures from the local centroid vertex to the global graph structure, and thus overcomes the restriction of the original hybrid reproducing kernel. Furthermore, as we have stated,

the largest layer expansion subgraph of a graph rooted at the centroid vertex is just the global structure of the graph, thus the hybrid reproducing kernel can be seen as the basic reproducing kernel between both the von Neumann and Shannon entropies of the largest layer expansion subgraphs. As a result, ***the original hybrid reproducing kernel is just a special case of the proposed nested reproducing kernel and only reflects a part of information of the proposed nested reproducing kernel***.

Finally, note that, both the proposed local-global nested graph kernels are also related to the fast Jensen-Shannon subgraph kernel developed by Bai and Hancock (2016), since they are all based on the similarity measure between the nested centroid depth-based complexity traces. However, the fast Jensen-Shannon subgraph kernel is based on the Jensen-Shannon divergence measure between each pair of K-layer expansion subgraphs, and this divergence measure does not encapsulate the alignment information between the probability distributions of the subgraphs. As a result, unlike the proposed nested graph kernels, this subgraph kernel cannot reflect precise kernel based similarity measures between the centroid depth-based complexity traces of graphs. These above observations indicate the theoretical advantages of the proposed local-global nested graph kernels.

4. Experimental Evaluations

In this section, we evaluate the performance of the proposed local-global nested graph kernels. We commence the by exhibiting the nest property of the centroid depth-based complexity traces. Finally, we perform the proposed kernels on graph classification tasks.

4.1. Graph Datasets

We evaluate our kernels on standard graph datasets. These datasets include: MUTAG, PTC, COIL5, Shock, CATH2, Reeb and D&D. Details of these datasets are shown in Table 1.

MUTAG: The MUTAG dataset consists of graphs representing 188 chemical compounds labeled according to whether or not they affect the frequency of genetic mutations in the bacterium *Salmonella typhimuriums* and aims to predict whether each compound is associated with mutagenicity.

PTC: The PTC (The Predictive Toxicology Challenge) dataset records the carcinogenicity of several hundred chemical compounds for male rats (MR), female rats (FR), male mice (MM) and female mice (FM). These graphs are very small, i.e., $20-30$ vertices, and sparsem, i.e., $25-40$ edges. We select the graphs of male rats (MR) for evaluation. There are 344 test graphs in the MR class.

COIL5: The COIL5 dataset is abstracted from the COIL image database. The COIL database consists of images of 100 3D objects. In our experiments, we use the images for the first five objects. For each of these objects we employ 72 images captured from different viewpoints. For each image we first extract corner points using the Harris detector, and then establish Delaunay graphs based on the corner points as vertices. Each vertex is used as the seed of a Voronoi region, which expands radially with a constant speed. The linear collision fronts of the regions delineate the image plane into polygons, and the Delaunay graph is the region adjacency graph for the Voronoi polygons.

Shock: The Shock dataset consists of graphs from the Shock 2D shape database. Each graph is a skeletal-based representation of the differential structure of the boundary of a 2D shape. There are 150 graphs divided into 10 classes.

CATH2: The CATH2 dataset is harder to classify, since the proteins in the same topology class are structurally similar. The protein graphs are 10 times larger in size than chemical compounds, with 200 . 300 vertices. There is 190 testing graphs in the dataset.

Reeb: The SHREC 3D Shape database consists of 15 classes and 20 individuals per class, that is 300 shapes Biasotti et al. (2003). This is a standard benchmark in 3D shape recognition. From the SHREC 3D Shape database, we establish a Reeb graph datasets through a mapping functions. This functions is ERG barycenter that computes the distance from the center of mass/barycenter. Details of the three mapping function can be found in Biasotti et al. (2003). The number of maximum, minimum and average vertices for the three datasets are 220, 41 and 95.42 respectively.

D&D: The D&D dataset contains 1178 protein structures. Each protein is represented by a graph, in which the vertices are amino acids and two vertices are connected by an edge if they are less than 6 Angstroms apart. The prediction task is to classify the protein structures into enzymes and non-enzymes. The maximum, minimum and average number of vertices are 5748, 30 and 284.32 respectively.

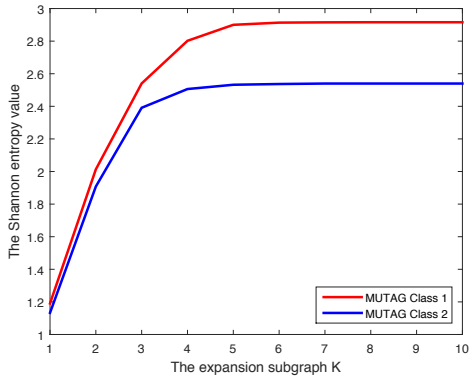
4.2. Evaluations of the Nested Centroid Depth-based Complexity Traces

In this subsection, we investigate the nest property of the centroid depth-based complexity trace. In the experiment, we utilize the testing graphs in the MUTAG and COIL5 datasets. Based on Table 1, the MUTAG and COIL5 datasets represent the testing graphs with low and high average degrees (i.e., 1.10 versus 2.89) respectively. For each testing graph, we commence by identifying the centroid vertex and establish a family of K-layer expansion subgraphs rooted at the vertex. Moreover, we compute the approximate von Neumann entropy or the Shannon entropy associated with the steady state random walk on each of the expansion subgraphs, as the centroid depth-based complexity trace of the graph. For each dataset, we compute the mean centroid depth-based complexity trace of the graphs from the same class. We draw the mean complexity trace and the experimental results are shown in Fig. 2.

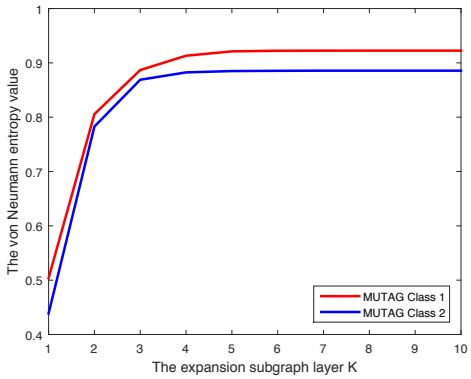
The subfigures of Fig. 2 exhibit the mean centroid depth-based complexity trace, and each colorized line represents the mean complexity trace of the graphs belonging to the same class in a dataset. Here, the subfigures (a) and (b) are for the MUTAG dataset using the Shannon and von Neumann entropies respectively. The subfigures (c) and (d) are for the COIL5 dataset using the Shannon and von Neumann entropies respectively. For each subfigure, the x-axis shows the order of the K-layer centroid expansion subgraph for each individual graph, while the y-axis shows the mean entropy value as a function of the expansion order.

Table 1. Information on the selected graph based bioninformatics datasets

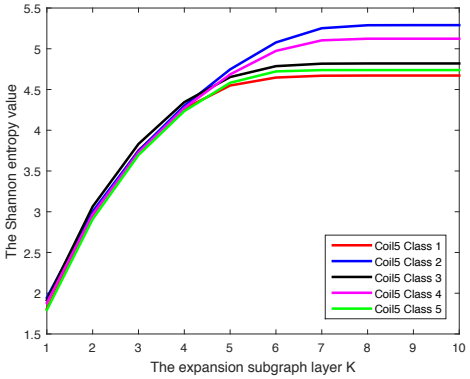
Datasets	MUTAG	PTC	COIL	Shock	CATH2	Reeb	D&D
Max # vertices	28	109	241	33	568	220	5748
Min # vertices	10	2	72	4	143	41	30
Mean # vertices	17.93	25.60	144.90	109.63	308.03	95.42	284.3
Max # edges	33	108	702	32	2220	219	14267
Min # edges	10	1	206	3	556	40	63
Mean # edges	19.79	25.96	419	12.16	1254.80	94.59	715.65
# graphs	188	344	360	150	190	300	1178
# classes	2	2	5	5	2	15	2
Mean#edges/Mean#vertices	1.10	1.00	2.89	0.92	4.07	0.99	2.52



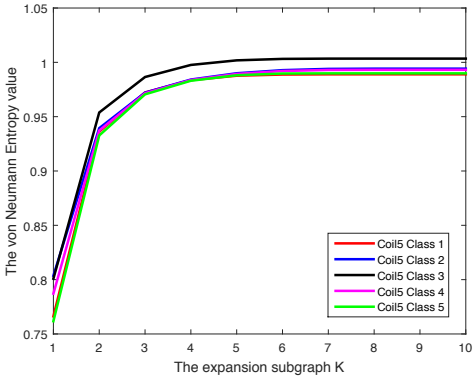
(a) For MUTAG using the Shannon Entropy



(b) For MUTAG using the von Neumann Entropy



(c) For COIL5 using the Shannon Entropy



(d) For COIL5 using the von Neumann Entropy

Fig. 2. Evaluation of the nest property of the centroid depth-based complexity trace).

From Fig. 2, we observe that the mean entropy values tend to be larger with the increasing layer size K of the expansion subgraphs. This is because the family of the K-layer expansion subgraphs rooted at the centroid vertex tend to gradually lead the local centroid vertex to the global graph structure, and each K-layer expansion subgraph encapsulates the structure information of the 1-layer to $K-1$-layer expansion subgraphs. As a result, Fig. 2 demonstrates the nest property of the centroid depth-based complexity traces, i.e., the entropy-based complexity information of each K-layer expansion subgraph encapsulates the information of the 1-layer to $K-1$-layer expansion subgraphs. Furthermore, it is clear that the mean centroid depth-based complexity traces of graphs from different classes are dissimilar. This indicates that the centroid depth-based complexity traces have good ability to distinguish graphs from different classes. Finally, we observe that, for the COIL5 dataset the complexity traces using the Shannon entropy can better distinguish the graphs from different classes than that using the von Neumann entropy. On the other hand, for the MUTAG dataset the complexity traces using either the Shannon entropy or the von Neumann entropy can both well distinguish the graphs from different classes. This is because the average vertex degree of the COIL5 dataset is obviously larger than that of the MUTAG dataset. As we have stated in Section 2.3, the von Neumann entropy is sensitive to graphs having low vertex degrees, while the Shannon entropy is suited to characterizing graphs with high vertex degrees. As a result, the centroid depth-based complexity trace using the Shannon entropy is more suitable to the dataset having graphs with high vertex degrees than that using the von Neumann entropy.

4.3. Experiments on Graph Classification

In this subsection, we investigate the performance of the proposed local-global nested graph kernels. Specifically, we evaluate the classification accuracies of the nested aligned kernel associated with the von Neumann entropy (NAKV) and the Shannon entropy (NAKS), and those of the nested reproducing kernel associated with the von Neumann entropy (NRKV) and the Shannon entropy (NRKS) on a number of graph classification tasks. Furthermore, we also compare our proposed nested graph kernels with seven state-of-the-art graph kernels, including 1) the Jensen-Shannon graph kernel (JSGK) Bai and Hancock (2013), 2) the random walk graph kernel (RWGK) Kashima et al. (2003), 3) the unaligned quantum Jensen-Shannon graph kernel (QJSK) Bai et al. (2015a), 4) the Lovász graph kernel (LGK) Johansson et al. (2014), 5) the fast Jensen-Shannon subgraph kernel Bai and Hancock (2016), 6) the Weisfeiler-Lehman subtree kernel (WLSK) Shervashidze et al. (2010), and 7) the reproducing graph kernel (RGK) Xu et al. (2018). For the WLSK kernel, we set the largest iteration of the required vertex label strengthening methods as 10, and we use the vertex degree as the original vertex label.

We compute the kernel matrix associated with each kernel on each dataset. We perform 10-fold cross-validation using a C-Support Vector Machine (C-SVM) to compute the classification accuracies, using LIBSVM software library Chang and Lin (2011). We use nine samples for training and one for testing. The parameters of the C-SVMs are optimized on each training set using cross-validation. We report the average classification accuracy and the runtime for each kernel in Table 2 and Table 3. The runtime is measured under Matlab R2015a running on a 2.5GHz Intel 2-Core processor (i.e., i5-3210m).

In terms of classification accuracies on different graph datasets, Table 2 indicates that the proposed local-global nested graph kernels outperform or are competitive to most alternative graph kernels. Especially, the proposed NRKV kernel significantly outperforms all alternative graph kernels on most datasets. Only the classification accuracy of the WLSK kernel on the Reeb dataset and that of the JSSK kernel on the D&D dataset are a little higher than the proposed NRKV kernel, but our NRKV kernel is still competitive. The reason of the effectiveness is that, as we have stated, our proposed nested graph kernels can simultaneously capture the local and global graph characteristics. By contrast, the local graph kernels RWGK and WLSK can only reflect local characteristics of graphs based on substructures, while the global graph kernels JSGK, QJSK, LGK and RGK can only capture global characteristics based on global matrix representations of graphs (e.g., the Laplacian or the adjacency matrix). On the other hand, similar to the proposed nested graph kernel the JSSK can also reflect both the local and global graph characteristics, relying on the depth-based representation. Unfortunately, as we have stated in Section 3.2, the required Jensen-Shannon divergence measure of the JSSK kernel does not establish the alignment information between the probability distributions of pairwise graphs. As a result, the JSSK kernel cannot reflect the precise kernel based similarity information between graphs.

Overall, the classification accuracies of the nested reproducing kernels (i.e., the NRKS and NRKV kernels) are significantly better than those of the nested aligned kernels (i.e., the NAKS and NAKV kernels), especially on the PTC, Shock and Reeb datasets. Through Table 1, we observe that the average vertex degrees of the PTC, Shock and Reeb datasets are smaller than 1. On the other hand, the average vertex degrees of the COIL5, CATH2 and D&D datasets are significantly higher, and the classification accuracies of the nested reproducing kernels and the nested aligned kernels on these datasets are competitive. This observation indicates that the nested aligned kernel is not suitable for graph datasets have low average vertex degrees, and the nested reproducing kernels has better applicability on any dataset.

Furthermore, we observe that either of the proposed nested kernels associated with the von Neumann entropy significantly outperforms that associated with the von Shannon entropy on the MUTAG, PTC, Shock and Reeb datasets. This is because the graphs from these datasets have lower average vertex degrees than the remaining datasets. As we have stated in Section 2.3, the von Neumann entropy is suitable to graphs having low vertex degrees. This suggests that for graph datasets having low vertex degrees the proposed nested kernels associated with the von Neumann entropy are more preferable.

In terms of runtime, the proposed nested kernels are not the fastest kernel. However, we can observe that the proposed nested aligned kernels NAKS and NAKV kernels can always com-

Table 2. Classification Accuracy (In % ± Standard Error) Runtime in Second.

Datasets	MUTAG	PTC	COIL5	Shock	CATH2	Reeb	D&D
NAKS	84.22 ± .50	58.00 ± .64	69.75 ± .65	37.60 ± .62	74.00 ± .83	45.20 ± .33	75.52 ± .31
NAKV	83.83 ± .61	57.67 ± .72	52.50 ± .61	37.65 ± .60	70.52 ± .90	48.43 ± .78	75.30 ± .17
NRKS	82.25 ± .63	55.05 ± .77	**71.41** ± .49	43.60 ± .65	74.21 ± .81	52.67 ± .36	75.44 ± .33
NRKV	**84.37** ± .78	**59.26** ± .75	69.44 ± .40	**44.26** ± .99	**75.57** ± .62	55.96 ± .34	75.37 ± .33
JSGK	83.11 ± .80	57.29 ± .41	69.13 ± .79	21.73 ± .76	72.26 ± .76	21.73 ± .76	72.26 ± .76
RWGK	80.77 ± .75	53.97 ± .31	14.21 ± .65	0.33 ± .37	–	43.23 ± .30	–
QJSK	82.72 ± .44	56.70 ± .49	70.11 ± .61	40.60 ± .92	71.11 ± .88	30.80 ± .61	–
LGK	80.83 ± .43	56.29 ± .47	–	31.80 ± .89	–	–	–
JSSK	82.77 ± .74	56.94 ± .43	67.75 ± .67	37.66 ± .80	75.42 ± .76	52.76 ± .47	**76.32** ± .46
WLSK	82.05 ± .57	56.85 ± .52	33.16 ± 1.01	36.40 ± .99	67.36 ± .63	**58.53** ± .53	73.52 ± .20
RGK	84.35 ± .51	58.23 ± .55	70.66 ± .49	37.93 ± .70	71.15 ± .68	27.40 ± .35	75.36 ± .54

Table 3. Runtime for Various Kernels.

Datasets	MUTAG	PTC	COIL5	Shock	CATH2	Reeb	D&D
NAKS	$8.6 \cdot 10^2$	$2.3 \cdot 10^3$	$3.3 \cdot 10^3$	$3.8 \cdot 10^2$	$9.4 \cdot 10^2$	$1.3 \cdot 10^1$	$4.6 \cdot 10^1$
NAKV	$8.6 \cdot 10^2$	$2.3 \cdot 10^3$	$3.3 \cdot 10^3$	$3.8 \cdot 10^2$	$9.4 \cdot 10^2$	$1.3 \cdot 10^1$	$4.6 \cdot 10^1$
NRKS	$4.0 \cdot 10^0$	$1.3 \cdot 10^1$	$1.6 \cdot 10^1$	$2.0 \cdot 10^0$	$8.0 \cdot 10^0$	$1.1 \cdot 10^1$	$2.1 \cdot 10^2$
NRKV	$4.0 \cdot 10^0$	$1.3 \cdot 10^1$	$1.6 \cdot 10^1$	$2.0 \cdot 10^0$	$8.0 \cdot 10^0$	$1.1 \cdot 10^1$	$2.1 \cdot 10^2$
JSGK	$1.0 \cdot 10^0$	$1.0 \cdot 10^0$	$1.0 \cdot 10^0$	$1.0 \cdot 10^0$	$1.0 \cdot 10^0$	$1.0 \cdot 10^0$	$1.0 \cdot 10^0$
RWGK	$4.6 \cdot 10^1$	$6.7 \cdot 10^1$	$1.1 \cdot 10^3$	$2.3 \cdot 10^1$	–	$1.2 \cdot 10^3$	–
QJSK	$2.0 \cdot 10^1$	$1.0 \cdot 10^2$	$1.0 \cdot 10^3$	$1.4 \cdot 10^1$	$4.4 \cdot 10^3$	$6.3 \cdot 10^3$	–
LGK	$1.0 \cdot 10^3$	$7.4 \cdot 10^3$	–	$1.0 \cdot 10^3$	–	–	–
JSSK	$1.0 \cdot 10^0$	$4.0 \cdot 10^0$	$3.0 \cdot 10^0$	$1.0 \cdot 10^0$	$4.0 \cdot 10^0$	$1.0 \cdot 10^0$	$4.5 \cdot 10^1$
WLSK	$3.0 \cdot 10^0$	$9.0 \cdot 10^1$	$6.5 \cdot 10^1$	$3.0 \cdot 10^0$	$5.3 \cdot 10^1$	$3.0 \cdot 10^1$	$4.6 \cdot 10^2$
RGK	$3.0 \cdot 10^0$	$1.3 \cdot 10^1$	$1.5 \cdot 10^1$	$2.0 \cdot 10^0$	$4.0 \cdot 10^0$	$9.0 \cdot 10^0$	$1.5 \cdot 10^2$

plete the computation of the kernel matrices. By contrast, some alternative graph kernels (e.g., the LGK and RWGK kernels) fail to complete the computation in a reasonable time. On the other hand, the proposed nested reproducing kernels have competitive computational efficiency to the fast alternative JSGK and JSSK kernels.

5. Conclusion

In this paper, we have proposed two novel local-global nested graph kernels, namely the nested aligned kernel and the nested reproducing kernel. Both of the nested kernels are based on the centroid depth-based complexity traces, that gauge the nested depth complexity trace through a family of K-layer expansion subgraphs rooted at the centroid vertex. Unlike most existing state-of-the-art graph kernels that only probe local or global graph characteristics, the proposed nested kernels simultaneously consider the local and global graph characteristics and thus reflect the presence of richer structural patterns. The experiments have demonstrated the effectiveness and efficiency of the proposed nested kernels.

Our future work is to extend the proposed kernel to attributed graphs that encapsulate vertex and edge labels. Moreover, we would also like to further develop novel graph kernels through the dynamic time warping framework associated with other types of (hyper)graph characteristic sequences, e.g., the cycle numbers identified by the Ihara zeta function, the time-varying entropies computed from the continuous-time or discrete-time quantum walk Bai et al. (2015a, 2017). Finally, we are also interested in developing novel graph kernels for time-varying financial market networks Ye et al. (2015), using the dynamic time warping framework.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant no. 61503422, 61602535, 61773415, 61370123 and 61772057), the Open Projects Program of National Laboratory of Pattern Recognition, the program for innovation research in Central University of Finance and Economics, and the Beijing Natural Science Foundation project (no. 4162037).

References

Alvarez, M.A., Qi, X., Yan, C., 2011. A shortest-path graph kernel for estimating gene product semantic similarity. J. Biomedical Semantics 2, 3.

Aronszajn, N., 1950. Theory of reproducing kernels. Transactions of the American mathematical society .

Bach, F.R., 2008. Graph kernels between point clouds, in: Proceedings of ICML, pp. 25–32.

Bai, L., Hancock, E.R., 2013. Graph kernels from the jensen-shannon divergence. Journal of mathematical imaging and vision 47, 60–69.

Bai, L., Hancock, E.R., 2014. Depth-based complexity traces of graphs. Pattern Recognition 47, 1172–1186.

Bai, L., Hancock, E.R., 2016. Fast depth-based subgraph kernels for unattributed graphs. Pattern Recognition 50, 233–245.

Bai, L., Rossi, L., Cui, L., Zhang, Z., Ren, P., Bai, X., Hancock, E.R., 2017. Quantum kernels for unattributed graphs using discrete-time quantum walks. Pattern Recognition Letters 87, 96–103.

Bai, L., Rossi, L., Torsello, A., Hancock, E.R., 2015a. A quantum jensen-shannon graph kernel for unattributed graphs. Pattern Recognition 48, 344–355.

Bai, L., Rossi, L., Zhang, Z., Hancock, E.R., 2015b. An aligned subtree kernel for weighted graphs, in: Proceedings of ICML, pp. 30–39.

Biasotti, S., Marini, S., Mortara, M., Patanè, G., Spagnuolo, M., Falcidieno, B., 2003. 3d shape matching through topological structures, in: Proceedings of DGCI, pp. 194–203.

Chang, C.C., Lin, C.J., 2011. Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27.

Conte, D., Ramel, J., Sidere, N., Luqman, M.M., Gaüzère, B., Gibert, J., Brun, L., Vento, M., 2013. A comparison of explicit and implicit graph embedding methods for pattern recognition, in: Proceedings of GbRPR, pp. 81–90.

Costa, F., De Grave, K., 2010. Fast neighborhood subgraph pairwise distance kernel, in: Proceedings ICML, pp. 255–262.

Cuturi, M., 2011. Fast global alignment kernels, in: Proceedings of ICML, pp. 929–936.

Cuturi, M., Vert, J., Birkenes, Ø., Matsui, T., 2007. A kernel for time series based on global alignments, in: Proceedings of ICASSP, pp. 413–416.

Farhi, E., Gutmann, S., 1998. Quantum computation and decision trees. Physical Review A 58, 915.

Haasdonk, B., Bahlmann, C., 2004. Learning with distance substitution kernels, in: Proceedings of DAGM, pp. 220–227.

Han, L., Escolano, F., Hancock, E.R., Wilson, R.C., 2012. Graph characterizations from von neumann entropy. Pattern Recognition Letters 33, 1958–1967.

Harchaoui, Z., Bach, F., 2007. Image classification with segmentation graph kernels, in: Proceedings of CVPR, pp. 1–8.

Johansson, F.D., Jethava, V., Dubhashi, D.P., Bhattacharyya, C., 2014. Global graph kernels using geometric embeddings, in: Proceedings of ICML, pp. 694–702.

Kashima, H., Tsuda, K., Inokuchi, A., 2003. Marginalized kernels between labeled graphs, in: Proceedings of ICML, pp. 321–328.

Kriege, N., Mutzel, P., 2012. Subgraph matching kernels for attributed graphs, in: Proceedings of ICML.

Kriege, N.M., Giscard, P., Wilson, R.C., 2016. On valid optimal assignment kernels and applications to graph classification, in: Proceedings of NIPS, pp. 1615–1623.

Ren, P., Aleksić, T., Wilson, R.C., Hancock, E.R., 2011. A polynomial characterization of hypergraphs using the ihara zeta function. Pattern Recognition 44, 1941–1957.

Riesen, K., Bunke, H., 2010. Graph classification and clustering based on vector space embedding. World Scientific Publishing Co., Inc.

Rossi, L., Torsello, A., Hancock, E.R., 2015. Measuring graph similarity through continuous-time quantum walks and the quantum jensen-shannon divergence. Physical Review E 91, 022815.

Shervashidze, N., Schweitzer, P., van Leeuwen, E.J., Mehlhorn, K., Borgwardt, K.M., 2010. Weisfeiler-lehman graph kernels. Journal of Machine Learning Research 1, 1–48.

Urry, M., Sollich, P., 2013. Random walk kernels and learning curves for gaussian process regression on random graphs. Journal of Machine Learning Research 14, 1801–1835.

Wang, L., Sahbi, H., 2013. Directed acyclic graph kernels for action recognition, in: Proceedings of ICCV, pp. 3168–3175.

Xu, L., Jiang, X., Bai, L., Xiao, J., Luo, B., 2018. A hybrid reproducing graph kernel based on information entropy. Pattern Recognition 73, 89–98.

Xu, L., Niu, X., Xie, J., Abel, A., Luo, B., 2015. A local–global mixed kernel with reproducing property. Neurocomputing 168, 190–199.

Xu, L., Niu, X.C.X., Zhang, C., Luo, B., 2017. A multiple attributes convolution kernel with reproducing property. Pattern Analysis and Applications 20, 485C494.

Ye, C., Comin, C.H., Peron, T.K.D., Silva, F.N., Rodrigues, F.A., da F. Costa, L., Tosello, A., Hancock, E.R., 2015. Thermodynamic characterization of networks using graph polynomials. Physical Review E 92, 032810.