

A Fast Jensen-Shannon Subgraph Kernel

Lu Bai and Edwin R. Hancock*

Department of Computer Science, University of York, York, YO10 5GH, UK

Abstract. In this paper we present a fast subgraph kernel based on Jensen-Shannon divergence and depth-based representations. For graphs with n vertices and m edges, the worst-case time complexity for our kernel is $O(n^3 + mn)$, in contrast to $O(n^6)$ for the classic graph kernel. Key to this efficiency is that we manage to compute the Jensen-Shannon divergence involved in our kernel with $O(n^2)$ operations. This computational strategy enables our subgraph kernel to easily scale up to graphs of reasonably large sizes and thus overcome the size limits arising in state of the art graph kernels. Experiments on standard bioinformatics graph datasets together with graph datasets extracted from images demonstrate the effectiveness and efficiency of our subgraph kernel.

1 Introduction

There has recently been an increased interest in learning graph structures using graph kernels [1, 12]. In the research literature, most graph kernels are formulated as instances of the R-convolution kernel family defined by Haussler [8]. Here R-convolution is a generic way for defining graph kernels based on the similarities of decomposed subgraphs. In the light of this formulation, any new type of graph decompositions or new type of subgraph similarity measures could result in a new graph kernel. Accordingly, existing graph kernels can be generally categorized into three classes [15], i.e. graph kernels based on comparing all pairs of a) walks, b) paths and c) restricted subgraph and subtree structures. One major limitation of these existing graph kernels is that in practical computation they do not easily scale up to substructures of large sizes (e.g. (sub)graphs with hundreds or even thousands vertices). The resulting graph kernels are limited by the size of the substructures used, which only roughly capture the overall topological structure of a graph. Furthermore, even for relatively small subgraphs, most graph kernels require significant computational overheads. In this paper we propose a new graph kernel, which overcomes this problem and builds upon previous work by Bai and Hancock [2] to develop an information theoretic kernel using the Jensen-Shannon divergence and a novel depth-based graph representation of graph substructure. This results in a computationally efficient kernel.

The Jensen-Shannon divergence is a non-extensive mutual information theoretic measure based on non-extensive entropies. An extensive entropy is defined as the sum of the individual entropies of two probability distributions. The definition of non-extensive entropy generalizes the sum operation into composite actions. The classical Jensen-Shannon divergence is a function of probability distributions, and is related to the Shannon entropy [11]. The problem of establishing Jensen-Shannon divergence measures for

* Edwin R. Hancock is supported by a Royal Society Wolfson Research Merit Award.

graphs can thus be posed as that of computing the required entropies for individual and composite graphs. Depth-based representations of undirected graphs have proved powerful for characterizing topological structures in terms of intrinsic complexities [4]. One approach for graph characterization is to gauge information content flow on K -layer subgraphs (e.g. subgraphs around a vertex having a maximum topological distance or minimal path length K) of increasing size and to use the flow as a structural signature. Unfortunately, existing methods for constructing such a depth-based representation of a graph always require a burdensome computational measure of the intrinsic structural complexity. Moreover, straightforward construction of the depth-based representation possessing global characteristics of a graph tends to be elusive, since it is difficult to determine a stable root vertex for graphs with potentially intricate vertex distributions.

The aim of this paper is to develop a novel subgraph kernel for efficient computation, even when a pair of fully sized subgraphs are compared. To this end, we investigate how to kernelize depth-based graph representations by measuring the information content for K -layer subgraphs using the Jensen-Shannon divergence. The contributions of this paper are threefold. First, we describe how to compute the Jensen-Shannon divergence for a pair of (sub)graphs based on the entropy difference between the original and composite (sub)graphs. The required entropies for the individual (sub)graphs are computed using the von Neumann and the Shannon entropy associated with the steady state random walk. Second, we develop a novel depth-based graph representation. We identify the centroid vertex of a graph by selecting the vertex with the minimum variance of shortest path lengths. From the centroid vertex, we derive a family of centroid expansion subgraphs with increasing layer size K . To avoid the burdensome subgraph enumeration of intrinsic complexity, we establish a depth-based representation for a graph by measuring the entropies of its centroid expansion subgraphs. Third, based on the first two contributions, we develop a fast Jensen-Shannon subgraph kernel. We empirically demonstrate the effectiveness and efficiency of our subgraph kernel on several challenging standard datasets furnished by bioinformatics and object recognition.

2 Jensen-Shannon Divergence on Graphs

In this section, we exploit the Jensen-Shannon divergence to develop a fast valid kernel measure for (sub)graphs. To commence, consider the graph $G(V, E)$ with vertex set V and edge set $E \subseteq V \times V$, the adjacency matrix A for $G(V, E)$ has elements

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

The vertex degree matrix of $G(V, E)$ is a diagonal matrix D with diagonal elements given by $D(i, i) = d(i) = \sum_{j \in V} A(i, j)$. The Laplacian matrix L is defined as $L = D - A$. The normalized Laplacian matrix is given by $\hat{L} = D^{-1/2} L D^{-1/2}$. The spectral decomposition of the normalized Laplacian matrix is $\hat{L} = \hat{\Phi} \hat{\Lambda} \hat{\Phi}^T$ where $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_{|V|})$ is a diagonal matrix with the ordered eigenvalues as elements ($0 = \hat{\lambda}_1 < \hat{\lambda}_2 < \dots < \hat{\lambda}_{|V|}$) and $\hat{\Phi} = (\hat{\phi}_1 | \hat{\phi}_2 | \dots | \hat{\phi}_{|V|})$ is a matrix with the corresponding ordered orthonormal eigenvectors as columns. The normalized Laplacian matrix is positive semi-definite and so has all eigenvalues non-negative. The number of zero eigenvalues is the number of connected components in $G(V, E)$.

2.1 A Jensen-Shannon Divergence on Graphs

The classical Jensen-Shannon divergence is a non-extensive mutual information similarity measure defined on probability distributions. Assume $M_+^1(\chi)$ is a set of probability distributions where χ is a set provided with some σ -algebra of measurable subsets, the Jensen-Shannon divergence $D_{JS} : M_+^1(\chi) \times M_+^1(\chi) \rightarrow R$ between the (discrete) probability distributions $P = (p_1, p_2, \dots, p_M)$ and $Q = (q_1, q_2, \dots, q_M)$, is negative definite (**nd**) with the following function

$$D_{JS}(P, Q) = H_S\left(\frac{P+Q}{2}\right) - \frac{H_S(P) + H_S(Q)}{2}. \quad (2)$$

where $H_S(P) = \sum_{m=1}^M p_m \log p_m$ is the entropy of the probability distribution P . In our work, we develop a Jensen-Shannon divergence for a pair of graphs. Given a pair of graphs $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$, the Jensen-Shannon divergence is

$$D_{JS}(G_p, G_q) = H(G_p \oplus G_q) - \frac{H(G_p) + H(G_q)}{2}. \quad (3)$$

where $G_p \oplus G_q$ is a composite structure obtained from the two graphs $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$. Here we use the disjoint union defined in Sec.2.3 as the composite structure, and explore the use of both the von Neumann entropy and Shannon entropy over the graphs. The Jensen-Shannon divergence D_{JS} for graphs defined in Eq.(3) is symmetric. With the Jensen-Shannon divergence for graphs defined in Eq.(2) to hand, we define a Jensen-Shannon diffusion graph kernel $k_{JS}: G_p \times G_q \rightarrow R$ with the kernel value

$$k_{JS}(G_p, G_q) = \exp(-\lambda D_{JS}(G_p, G_q)). \quad (4)$$

where λ is a decay factor and satisfies $0 < \lambda < 1$. The resulting diffusion kernel is positive definite.

2.2 Graph Entropies

To compute the Jensen-Shannon diffusion kernel, we require a means of evaluating the entropy of a graph. There are several alternative routes, but here we consider two computationally efficient strategies.

Von Neumann Entropy. The classical von Neumann entropy of $G(V, E)$ associated with the normalized Laplacian eigenspectrum is defined as $H_N = -\sum_{i=1}^{|V|} \frac{\hat{\lambda}_i}{|V|} \log \frac{\hat{\lambda}_i}{|V|}$. The computation of the von Neumann entropy requires cubic number of vertices operations. Han et al. [7] have shown how the computation can be rendered quadratic in the number of the vertices. By approximating the von Neumann entropy by its quadratic counterpart, the simplified von Neumann entropy is

$$H_N(G) = 1 - \frac{1}{|V|} - \frac{1}{|V|^2} \sum_{(v_i, v_j) \in E} \frac{1}{d(i)d(j)}. \quad (5)$$

Shannon Entropy. An alternative entropy for $G(V, E)$ is to exploit steady state random walks on $G(V, E)$. The probability of the steady state random walks on $G(V, E)$ visiting vertex v_i is $P_G(i) = d(i) / \sum_{v_j \in V} d(j)$. And the Shannon entropy for $G(V, E)$ is

$$H_S(G) = - \sum_{i=1}^{|V|} P_G(i) \log P_G(i). \quad (6)$$

For the graph $G(V, E)$ with $n = |V|$ vertices, both the von Neumann entropy $H_N(G)$ and the Shannon entropy $H_S(G)$ require time complexity $O(n^2)$. This comes from the fact that the degree matrix D of $G(V, E)$ can be computed by visiting every entries $A(i, j)$ in the adjacency matrix A . Thus both entropies $H_N(G)$ and $H_S(G)$ can be directly computed by visiting all the n^2 pairs of vertices. \square

2.3 A Composite Entropy of a Pair of Graphs

We now turn our attention to computing the composite entropy $H(G_p \oplus G_q)$ for the pair of graphs $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$. Since we aim to develop an efficient kernel, we seek a computationally inexpensive route and use the disjoint union $G_{DU} = G_p \cup G_q = \{V_p \cup V_q, E_p \cup E_q\}$ for constructing our composition of graphs $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$. Let graphs G_p and G_q be the connected components of the disjoint union graph G_{DU} , then the composite entropy is $H(G_{DU}) = \rho_p H(G_p) + \rho_q H(G_q)$, where $\rho_p = |V(G_p)| / |V(G_{DU})|$ and $\rho_q = |V(G_q)| / |V(G_{DU})|$. Here the entropy function $H(\cdot)$ could be either the von Neumann entropy $H_N(\cdot)$ defined in Eq.(5) or the Shannon entropy $H_S(\cdot)$ defined in Eq.(6).

3 Fast Jensen-Shannon Subgraph Kernel

The idea of using Jensen-Shannon divergence to develop graph kernels has previously been explored by Lu and Hancock [2]. However, the kernel is slow since it requires a global computation of the divergence. In this section we overcome this problem and aim to develop a novel and fast subgraph kernel based on the Jensen-Shannon divergence (JSD). Our idea is to decompose a graph into substructures (i.e. subgraphs) spanned from a root vertex to the remaining vertices with respect to the minimal path length K , and measure the entropies on these subgraphs as a depth-based signature of the graph. To obtain a family of subgraphs capturing fine structures of a graph, we identify a centroid vertex as the root vertex. In the literature of graph theory, the centroid of a graph has been defined as a structure composed of vertices closest to all others [17]. Here we present a novel method to identify the centroid vertex of a graph by evaluating the shortest path length distribution around a vertex. We select the vertex with the minimum variance of shortest path lengths to the remaining vertices as the centroid vertex. The vertices surrounding the centroid vertex in a graph lie along the different shortest paths from the vertex, and the centroid vertex has a global view of the vertex path length distribution surrounding it. Finally, we establish an information theoretic subgraph kernel by applying the Jensen-Shannon diffusion kernel to subgraphs derived from the centroid vertices.

3.1 Subgraphs from the Centroid Vertex

We commence by identifying the centroid vertex of a graph. For an undirected graph $G(V, E)$, the shortest path $S_G(i, j)$ for a pair of vertices v_i and v_j can be computed by using Johnson’s algorithm [9]. The matrix S_G whose element $S_G(i, j)$ represents the shortest path length between vertices v_i and v_j is referred to as the shortest path matrix for graph $G(V, E)$. The average-shortest-path vector S_V for $G(V, E)$ is a vector with the same vertex order as S_G and with element $S_V(i) = \sum_{j=1}^{|V|} S_G(i, j)/|V|$ representing the average shortest path length from vertex v_i to the remaining vertices. We then locate the centroid vertex \hat{v}_i for $G(V, E)$ as follows

$$\hat{v}_i = \arg \min_i \sum_{j=1}^{|V|} [S_G(i, j) - S_V(i)]^2. \tag{7}$$

The centroid vertex \hat{v}_i of $G(V, E)$ is identified by selecting the vertex with the minimum variance of shortest path lengths over all vertices in $G(V, E)$. Therefore, the shortest paths starting from the centroid vertex \hat{v}_i form a *steady* path set that exhibits the least length variability compared with those path sets originating from the other vertices. Thus the centroid vertex has a global view of the path length distribution. Let $N_{\hat{v}_C}^K$ be a subset of V satisfying $N_{\hat{v}_C}^K = \{u \in V \mid S_G(\hat{v}_C, u) \leq K\}$. For $G(V, E)$ with the centroid vertex \hat{v}_C , the K -layer centroid expansion subgraph $\mathcal{G}_K(\mathcal{V}_K; \mathcal{E}_K)$ is

$$\begin{cases} \mathcal{V}_K = \{u \in N_{\hat{v}_C}^K\}; \\ \mathcal{E}_K = \{(u, v) \subset N_{\hat{v}_C}^K \mid (u, v) \in E\}. \end{cases} \tag{8}$$

The number of centroid expansion subgraphs is equal to the greatest length L of the shortest paths from the centroid vertex to the remaining vertices of the graph $G(V, E)$. The L -layer expansion subgraph is the graph $G(V, E)$ itself.

For graph $G(V, E)$ ($n = |V|$ and $m = |E|$), computing entropies for its centroid expansion subgraphs requires time complexity $O(n^2L + mn)$.

3.2 Depth-Based Representation

In this subsection, we develop a fast Jensen-Shannon subgraph kernel ($k_{\mathcal{J}\mathcal{S}}$) as an information theoretic decomposition kernel. The proposed kernel $k_{\mathcal{J}\mathcal{S}}$ is defined by kernelizing depth-based graph representations in terms of measuring information content similarities for K -layer subgraphs using the Jensen-Shannon divergence. For a sample graph $G(V, E)$, we commence by identifying the centroid vertex \hat{v}_C using Eq.(7). Based on \hat{v}_C we construct the K -layer centroid expansion subgraph \mathcal{G}_K of $G(V, E)$ using Eq.(8). As we increase K from 1 to the greatest shortest path length L with respect to the centroid vertex \hat{v}_C , we obtain a family of centroid expansion subgraphs $\{\mathcal{G}_1, \dots, \mathcal{G}_K, \dots, \mathcal{G}_L\}$. We then measure the entropies of the subgraphs and establish the depth-based representation $D(G)$ of $G(V, E)$ as $D(G) = \{H(\mathcal{G}_1), \dots, H(\mathcal{G}_K), \dots, H(\mathcal{G}_L)\}$, where $H(\mathcal{G}_K)$ is the entropy for the K -layer subgraph \mathcal{G}_K of $G(V, E)$. For a pair of

graphs $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$, we develop a similarity measure between their depth-based representations $D(G_p)$ and $D(G_q)$ as follows

$$s(D(G_p), D(G_q)) = \sum_{K=1}^L s_H(H(\mathcal{G}_{p;K}), H(\mathcal{G}_{q;K})). \quad (9)$$

where $s_H(H(\mathcal{G}_{p;K}), H(\mathcal{G}_{q;K}))$ is an entropy-based similarity measure for the K -layer subgraphs $\mathcal{G}_{p;K}$ and $\mathcal{G}_{q;K}$ of $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$. By using the Jensen-Shannon diffusion kernel $k_{JS}(\cdot, \cdot)$ in Eq.(4) as the entropy-based similarity measure $s_H(\cdot, \cdot)$ in Eq.(9), the similarity between the depth-based representations $D(G_p)$ and $D(G_q)$ is formulated as the sum of the diffusion kernel measures for all the pairs of K -layer subgraphs of $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$.

For the graphs $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$ the Jensen-Shannon subgraph kernel $k_{JS}(G_p, G_q)$ is defined as

$$k_{JS}(G_p, G_q) = s(D(G_p), D(G_q)) = \sum_{K=1}^L k_{JS}(\mathcal{G}_{p;K}, \mathcal{G}_{q;K}). \quad (10)$$

where $\mathcal{G}_{p;K}(\mathcal{V}_{p;K}, \mathcal{E}_{p;K})$ and $\mathcal{G}_{q;K}(\mathcal{V}_{q;K}, \mathcal{E}_{q;K})$ are the K -layer centroid expansion subgraphs of $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$ rooted from their centroid vertices $\hat{v}_{p;C}$ and $\hat{v}_{q;C}$, respectively, and $k_{JS}(\mathcal{G}_{p;K}, \mathcal{G}_{q;K})$ is the Jensen-Shannon diffusion kernel between $\mathcal{G}_{p;K}(\mathcal{V}_{p;K}, \mathcal{E}_{p;K})$ and $\mathcal{G}_{q;K}(\mathcal{V}_{q;K}, \mathcal{E}_{q;K})$. According to Eq.(2.3), Eq.(3) and Eq.(4), $k_{JS}(\mathcal{G}_{p;K}, \mathcal{G}_{q;K})$ is

$$k_{JS}(\mathcal{G}_{p;K}, \mathcal{G}_{q;K}) = \exp \left[-\lambda \left\{ \frac{2|\mathcal{V}_{p;K}| - |\mathcal{V}_{q;K}|}{2|\mathcal{V}_{p;K}| + 2|\mathcal{V}_{q;K}|} H(\mathcal{G}_{p;K}) + \frac{2|\mathcal{V}_{q;K}| - |\mathcal{V}_{p;K}|}{2|\mathcal{V}_{p;K}| + 2|\mathcal{V}_{q;K}|} H(\mathcal{G}_{q;K}) \right\} \right]. \quad (11)$$

The resulting Jensen-Shannon subgraph kernel k_{JS} is positive semidefinite. The computation of the proposed subgraph kernel between a pair of graphs, each of which has n vertices and m edges, requires time complexity $O(n^2L + mn)$.

One extreme case for the time complexity $O(n^2L + mn)$ is the chain structure, where the centroid vertex is the intermediate vertex along the chain and L is $\lfloor n/2 \rfloor$. Hence, the worst-case time complexity of our subgraph kernel is $O(n^3 + mn)$. But for most real-world graphs, the vertices surrounding the centroid vertex tend to be well distributed, and then our subgraph kernel usually only requires time complexity $O(n^2 + mn)$ for the practical kernel computation.

For a pair of graphs $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$ with different sizes, the greatest sizes of the expansion subgraphs for the two graphs could be different. Suppose that $\hat{v}_{C;p}$ and $\hat{v}_{C;q}$ are the centroid vertices of $G_p(V_p, E_p)$ and $G_q(V_q, E_q)$, and the lengths of the greatest shortest paths from the centroid vertices $\hat{v}_{C;p}$ and $\hat{v}_{C;q}$ are L_p and L_q , respectively, where $L_p > L_q$. In practical computations, to balance the size difference between largest centroid expansion subgraphs of the two graphs, we use the graph $G_q(V_q, E_q)$ as the $(L_q + 1)$ -layer to L_p -layer expansion subgraphs of $G_q(V_q, E_q)$.

Table 1. Information of the Graph-based Datasets

Datasets	MUTAG	D&D	ENZYMES	PPIs
Max # vertices	28	5748	126	232
Min # vertices	10	30	2	3
Mean # vertices	17.9	284.3	32.6	109.6
# graphs	188	1178	600	86
# disjoint graphs	0	21	31	0
# classes	2	2	6	2

4 Experimental Results

4.1 Standard Graph Datasets

We commence by demonstrating the performance of our Jensen-Shannon subgraph kernel and compare it to several state of the art graph kernels on four standard graph based datasets abstracted from bioinformatics databases. These datasets include: MUTAG, D&D, ENZYMES and PPIs. The MUTAG dataset contains graphs representing 188 chemical compounds to predict mutagenicity. The D&D dataset contains 1178 protein structures. Each protein is represented by a graph. The ENZYMES dataset contains graphs representing protein tertiary structures consisting of 600 enzymes from the BRENDA enzyme database. The PPIs dataset consists of protein-protein interaction networks (PPIs). Here we select Proteobacteria40 PPIs and Acidobacteria46 PPIs as the testing graphs. Details about the datasets are shown in Table.1.

We evaluate the performance of our proposed Jensen-Shannon subgraph kernel using the von Neumann entropy (JSSV) or Shannon entropy (JSSS) separately on the graph datasets abstracted from the bioinformatics databases, and then compare it with several alternative state of the art graph kernels. These graph kernels for comparison include 1) the Weisfeiler-Lehman subtree kernel (WL) [15], 2) the shortest path graph kernel (SPGK) [3], 3) the graphlet count graph kernel (GCGK) [16], 4) the random walk graph kernel (RWGK) [6], 5) the p-random walk graph kernel (PWGK) [10], and 6) the Ramon & Ganter graph kernel (RGGK) [5]. We compute the kernel matrix on each dataset by using our proposed method. We also compute the kernel matrices of the Weisfeiler-Lehman subtree kernel and shortest path graph kernel on the PPIs dataset. For each kernel matrix we perform Principle Component Analysis (PCA) to embed the graphs into feature space as vectors. We then perform 10-fold cross-validation using the Support Vector Machine (SVM) Classification associated with the Sequential Minimal Optimization (SMO) and the PUK kernel[13] to evaluate the performance of our method and those of the alternative methods. We use nine samples for training and one for testing. All the SMO-SVMs were performed along with their parameters optimized on each dataset. We report the average classification accuracies and runtime of establishing the kernel matrices of each method in Table.2, with the runtime measured under Matlab R2011a running on a 2.2GHz Intel 2-Core processor. Shervashidze et al. [15] have reported the accuracies of state of the alternative art graph kernels on each dataset excluding the PPIs dataset, by performing 10-fold cross-validation associated with the C-Support Vector Machine Classification (C-SVM). The runtime for establishing the kernel matrices using these methods was measured under Matlab R2008a running on

Table 2. Accuracy and CPU Runtime Comparisons on Graph Datasets

Datasets	MUTAG	D&D	ENZYMES	PPIs
JSSV	84.04	76.40	33.16	74.41
JSSS	85.10	78.50	38.00	80.01
WL	82.05	79.78	46.42	75.90
SPGK	87.28	78.45	41.68	70.93
GCGK	75.61	78.59	32.70	–
RWGK	80.72	71.70	21.68	–
PWGK	79.19	66.64	27.67	–
RGGK	85.72	57.27	13.35	–

Datasets	MUTAG	D&D	ENZYMES	PPIs
JSSV	1"	2'50"	4"	1"
JSSS	1"	2'49"	4"	1"
WL	1"	11'	20"	27"
SPGK	2"	23h17'2"	5"	47"
GCGK	3"	30'21"	25"	–
RWGK	12"	48days	12'19"	–
PWGK	4'42"	4days	10'	–
RGGK	40'60"	103days	38days	–

a 3.0GHz Intel 8-Core processor with 16GB RAM. We report these accuracies and runtime in Table.2.

There are a number of conclusions that can be drawn from Table 2. First, on the MUTAG, D&D, PPIs datasets, our subgraph kernel outperforms or is competitive with the alternatives. Secondly, on the ENZYMES dataset, the performance of our subgraph kernel is lower than that of the Weisfeiler-Lehman subtree kernel, is competitive with the shortest path graph kernel, and outperform the remaining methods. Thirdly, the runtime of our subgraph kernel is better than that of each of the alternatives on all datasets studied. This includes the challenging D&D dataset which contains graphs with thousands of vertices. Finally, the average greatest layer sizes of the subgraphs for the MUTAG, D&D, ENZYMES and PPIs datasets are 8.53, 24.14, 6.45 and 8.79, respectively.

The runtimes of our subgraph kernel using the von Neumann and Shannon entropies are similar. However the classification accuracies are quite dissimilar. The reason for this is that according to Eq.(5) the von Neumann entropy is more sensitive to variable graph size than the Shannon entropy. This implies that the Shannon entropy is better suited to graph datasets containing graphs of variable size. On the ENZYMES data-set, all methods perform relatively poorly. The reason for this is that there are six classes of graph in the ENZYMES dataset, and only two classes in the remaining datasets.

4.2 View-Based Object Recognition

In this subsection, we focus on graphs extracted from images and assess the effectiveness of the proposed subgraph kernel in detecting object clusters. The first dataset (i.e. ALOI) consists of graphs extracted from images of three similar boxes in the ALOI

database. The second dataset (i.e. MOVI-CMU) consists of graphs extracted from images of three similar toy houses in the MOVI and CMU databases. The third dataset (i.e. COIL) consists of graphs extracted from images of three similar cups in the COIL database. For each object there are 18 images captured from different viewpoints. The graphs are the Delaunay triangulations of feature points extracted from the different images. The minimum and maximum vertices of the three datasets are 295 (min) and 1288 (max) for ALOI, 249 (min) and 734 (max) for MOVI-CMU, and 13 (min) and 80 (max) for COIL.

For our subgraph kernel with the von Neumann entropy (JSSV) or the Shannon entropy (JSSS), we compute the kernel matrix on each dataset. We embed the testing graphs into feature space as vectors by performing the Principle Component Analysis (PCA) on each kernel matrix. Furthermore, we also compare our methods against a number of alternative graph characterisations, including a) the Ihara zeta function of graphs (CIZF) [14], and b) the pattern vectors from algebraic graph theory (PVAG) [18]. For these alternative methods we compute the feature vectors of graphs on each dataset. To evaluate the performance of all the methods, we perform the K -means clustering method to compute the classification accuracies for the three datasets. The experimental results are shown in Table 3

For the graphs in the COIL dataset of which the average number of vertices is less than 60, all the methods give good performance and can achieve 100% recognition accuracy. For the graphs in the ALOI and MOVI-CMU datasets of which the average number of vertices is between 200 to 700, the proposed methods outperform all the alternatives, only CIZF is competitive to our methods on the MOVI-CMU dataset. The CIZF and PVAG can not finish the computation on the ALOI dataset, since they generate infinite feature values for a graph of large size. The evaluation reveals that our subgraph kernel can easily scale to large graphs.

Table 3. Experimental Comparisons on Image Based Graph Dataset

Datasets	JSSV	JSSS	CIZF	PVAG
ALOI	87.03	90.73	—	—
MOVI-CMU	96.29	100	96.29	81.49
COIL	100	100	100	100

5 Conclusion

In this paper, we have shown how to construct a fast Jensen-Shannon subgraph kernel using depth-based representations of graphs. The proposed method is based on a fast Jensen-Shannon diffusion kernel measure defined in terms of the Jensen-Shannon divergence on (sub)graphs and an elegant graph decomposition through the depth-based representation. The proposed subgraph kernel overcomes the subgraph size restrictions arising in state of the art graph kernels, and also renders an efficient computation. The experimental results have demonstrated the effectiveness and efficiency of the proposed subgraph kernel.

Acknowledgments. We thank Prof. Francisco Escolano, Dr. Peng Ren and Dr. Boyan Bonev for the constructive discussion and suggestions.

References

1. Amizadeh, S., Wang, S., Hauskrecht, M.: An efficient framework for constructing generalized locally-induced text metrics. In: Proceedings of International Joint Conference on Artificial Intelligence, pp. 1159–1164 (2011)
2. Bai, L., Hancock, E.: Graph kernels from the Jensen-Shannon divergence. *Journal of Mathematical Imaging and Vision* (to appear 2013)
3. Borgwardt, K.M., Kriegel, H.-P.: Shortest-path kernels on graphs. In: Proceedings of the IEEE International Conference on Data Mining, pp. 74–81 (2005)
4. Crutchfield, J.P., Shalizi, C.R.: Thermodynamic depth of causal states: Objective complexity via minimal representations. *Physical Review E* 59, 275–283 (1999)
5. Gärtner, T., Driessens, K., Ramon, J.: Graph kernels and gaussian processes for relational reinforcement learning. In: Proceedings of International Conference on Inductive Logic Programming, pp. 146–163 (2003)
6. Gärtner, T., Flach, P.A., Wrobel, S.: On graph kernels: Hardness results and efficient alternatives. In: Schölkopf, B., Warmuth, M.K. (eds.) COLT/Kernel 2003. LNCS (LNAI), vol. 2777, pp. 129–143. Springer, Heidelberg (2003)
7. Han, L., Escolano, F., Hancock, E.R., Wilson, R.C.: Graph characterizations from von neumann entropy. *Pattern Recognition Letters* 33, 1958–1967 (2012)
8. Haussler, D.: Convolution kernels on discrete structures. In: Technical Report UCS-CRL-99-10 (1999)
9. Johnson, D.B.: Efficient algorithms for shortest paths in sparse networks. *Journal of the ACM* 24, 1–13 (1977)
10. Kashima, H., Tsuda, K., Inokuchi, A.: Marginalized kernels between labeled graphs. In: Proceedings of International Conference on Machine Learning, pp. 321–328 (2003)
11. Martins, A.F., Smith, N.A., Xing, E.P., Aguiar, P.M., Figueiredo, M.A.: Nonextensive information theoretic kernels on measures. *Journal of Machine Learning Research* 10, 935–975 (2009)
12. Nori, N., Bollegala, D., Ishizuka, M.: Interest prediction on multinomial, time-evolving social graph. In: Proceedings of International Joint Conference on Artificial Intelligence, pp. 2507–2512 (2011)
13. Platt, J.C.: Fast training of support vector machines using sequential minimal optimization. In: Schölkopf, B., Burges, C.J.C., Smola, A.J. (eds.) *Advances in Kernel Methods*, pp. 185–208 (1999)
14. Ren, P., Aleksic, T., Wilson, R.C., Hancock, E.R.: A polynomial characterization of hypergraphs using the Ihara zeta function. *Pattern Recognition* 44, 1941–1957 (2011)
15. Shervashidze, N., Schweitzer, P., van Leeuwen, E.J., Mehlhorn, K., Borgwardt, K.M.: Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research* 1, 1–48 (2010)
16. Shervashidze, N., Vishwanathan, S.V.N., Petri, T., Mehlhorn, K., Borgwardt, K.M.: Efficient graphlet kernels for large graph comparison. *Journal of Machine Learning Research* 5, 488–495 (2009)
17. Slater, P.J.: Centers to centroids in graphs. *Journal of Graph Theory* 2, 209–222 (1978)
18. Wilson, R.C., Hancock, E.R., Luo, B.: Pattern vectors from algebraic graph theory. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1112–1124 (2005)