UNIVERSITY *of York*

This is a repository copy of *A Jensen-Shannon Divergence Kernel for Directed Graphs*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/105837/

Version: Accepted Version

Proceedings Paper:

A Jensen-Shannon Divergence Kernel for Directed Graphs

Cheng Ye, Richard C. Wilson and Edwin R. Hancock

Department of Computer Science,
University of York,
York, YO10 5GH, UK.
Email: {cy666, richard.wilson, edwin.hancock}@york.ac.uk

Abstract. Recently, kernel methods have been widely employed to solve machine learning problems such as classification and clustering. Although there are many existing graph kernel methods for comparing patterns represented by undirected graphs, the corresponding methods for directed structures are less developed. In this paper, to fill this gap in the literature we exploit the graph kernels and graph complexity measures, and present an information theoretic kernel method for assessing the similarity between a pair of directed graphs. In particular, we show how the Jensen-Shannon divergence, which is a mutual information measure that gauges the difference between probability distributions, together with the recently developed directed graph von Neumann entropy, can be used to compute the graph kernel. In the experiments, we show that our kernel method provides an efficient tool for classifying directed graphs with different structures and analyzing real-world complex data.

Keywords: information theoretic kernel, Jensen-Shannon divergence, von Neumann entropy

1 Introduction

Graph kernels have recently evolved into a rapidly developing branch of pattern recognition. Broadly speaking, there are two main advantages of kernel methods, namely a) they can bypass the need for constructing an explicit high-dimensional feature space when dealing with high-dimensional data and b) they allow standard machine learning techniques to be applied to complex data, which bridges the gap between structural and statistical pattern recognition [1]. However, although there are a great number of kernel methods aimed at quantifying the similarity between structures underpinned by undirected graphs, there is very little work on solving the corresponding problems for directed graphs. This is unfortunate since many real-world complex systems such as the citation networks, communications networks, neural networks and financial networks give rise to structures that are directed [2].

Motivated by the need to fill this gap in literature, in this paper we propose an information theoretic kernel method for measuring the structural similarity

between a pair of directed graphs. Specifically, our goal is to develop a symmetric and positive definite function that maps two directed patterns to a real value, by exploiting the information theoretic kernels that are related to the Jensen-Shannon divergence and a recently developed directed graph structural complexity measure, namely the von Neumann entropy.

1.1   Related Literature

Recently, information theoretic kernels on probability distributions have attracted a great deal of attention and have been extensively employed to the domain of classification of structured data [3]. Martins et al. [4] have generalized the recent research advances, and have proposed a new family of nonextensive information theoretic kernels on probability distributions. Such kernels have proved to have a strong link with the Jensen-Shannon divergence, which measures the mutual information between two probability distributions, and which is quantified by gauging the difference between entropies associated with those probability distributions [1].

Graph kernels, on the other hand, particularly aim to assess the similarity between pairs of graphs by computing an inner product on graphs [5]. For example, the classical random walk kernel [6] compares two labeled graphs by counting the number of paths produced by random walks on those graphs. However, random walk kernel has a serious disadvantage, namely "tottering" [7], caused by the possibility that a random walk can visit the same cycle of vertices repeatedly on a graph. An effective technique to avoid tottering is the backtrackless walk kernel developed by Aziz et al. [8], who have explored a set of new graph characterizations based on backtrackless walks and prime cycles.

Turning attention to the directed graph complexity measures, Ye et al. [2] have developed a novel entropy for quantifying the structural complexity of directed networks. Moreover, Berwanger et al. [9] have proposed a new parameter for the complexity of infinite directed graphs by measuring the extent to which cycles in graphs are intertwined. Recently, Escolano et al. [10] have extended the heat diffusion-thermodynamic depth approach from undirected networks to directed networks and have obtained a means of quantifying the complexity of structural patterns encoded by directed graphs.

1.2   Outline

The remainder of the paper is structured as follows. Section II details the development of an information theoretic kernel for quantifying the structural similarity between a pair of directed graphs. In particular, we show the graph kernel can be computed in terms of the Jensen-Shannon divergence between the von Neumann entropy of the disjoint union of two graphs, and the entropies of individual graphs. In Sec. III, we show the effectiveness of our method by exploring its experimental performance on both artificial and realistic directed network data. Finally, Sec. IV summarizes our contribution present in this paper and points out future research directions.

2 Jensen-Shannon Divergence Kernel for Directed Graphs

In this section, we start from the Jensen-Shannon divergence and explore how this similarity measure can be used to construct a graph kernel method for two directed graphs. In particular, the kernel can be computed in terms of the entropies of two individual graphs and the entropy of their composite graph. We then introduce the recently developed von Neumann entropy which quantifies the structural complexity of directed graphs. This in turn allows us to develop the mathematical expression for the Jensen-Shannon divergence kernel.

2.1 Jensen-Shannon Divergence

Generally speaking, the Jensen-Shannon divergence is a mutual information measure for assessing the similarity between two probability distributions. Mathematically, given a set $M_+^1(X)$ of probability distributions in which X indicates a set provided with some σ-algebra of measurable subsets. Then, the Jensen-Shannon divergence $D_{JS} : M_+^1(X) \times M_+^1(X) \to [0, +\infty)$ between probability distributions P and Q is computed as:

$$D_{JS}(P,Q) = \frac{1}{2}D_{KL}(P\|M) + \frac{1}{2}D_{KL}(M\|Q), \tag{1}$$

where D_{KL} denotes the classical Kullback-Leibler divergence and $M = \frac{P+Q}{2}$. Specifically, let π_p and π_q ($\pi_p + \pi_q = 1$ and $\pi_p, \pi_q \geq 0$) be the weights of the probability distributions P and Q respectively, then the generalization of the Jensen-Shannon divergence can be defined as:

$$D_{JS}(P,Q) = H_S(\pi_p P + \pi_q Q) - \pi_p H_S(P) - \pi_q H_S(Q), \tag{2}$$

where H_S is used to denote the Shannon entropy of the probability distribution.

To proceed, we introduce the Jensen-Shannon kernel method for undirected graphs developed by Bai et al. [1]. Given a pair of undirected graphs $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$ where $\mathcal{V}$ represents the edge set and $\mathcal{E}$ denotes the vertex set of the graph, the Jensen-Shannon divergence between these two graphs is expressed by:

$$D_{JS}(\mathcal{G}_1, \mathcal{G}_2) = H(\mathcal{G}_1 \oplus \mathcal{G}_2) - \frac{H(\mathcal{G}_1) + H(\mathcal{G}_2)}{2}, \tag{3}$$

where $\mathcal{G}_1 \oplus \mathcal{G}_2$ denotes the disjoint union graph of graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ and $H(\cdots)$ is the graph entropy. With this definition to hand and making use of the von Neumann entropy for undirected graphs [11], the Jensen-Shannon diffusion kernel $k_{JS} : \mathcal{G} \times \mathcal{G} \to [0, +\infty)$ is then

$$\begin{aligned} k_{JS}(\mathcal{G}_1, \mathcal{G}_2) &= \exp\{-\lambda D_{JS}(\mathcal{G}_1, \mathcal{G}_2)\} \\ &= \exp\left\{\lambda \frac{H_{VN}(\mathcal{G}_1) + H_{VN}(\mathcal{G}_2)}{2} - \lambda H_{VN}(\mathcal{G}_1 \oplus \mathcal{G}_2)\right\}, \end{aligned} \tag{4}$$

where H_{VN} is the undirected graph von Neumann entropy, and λ is a decay factor $0 < \lambda \leq 1$. Clearly, the diffusion kernel is exclusively dependent on the individual entropies of the two graphs being compared as the union graph entropy can also be expressed in terms of those entropies.

2.2 Von Neumann Entropy for Directed Graphs

Given a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of a vertex set $\mathcal{V}$ together with an edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, its adjacency matrix A is defined as

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

The in-degree and out-degree at vertex u are respectively given as

$$d_u^{in} = \sum_{v \in \mathcal{V}} A_{vu}, \ d_u^{out} = \sum_{v \in \mathcal{V}} A_{uv}.$$

The transition matrix P of a graph is a matrix describing the transitions of a Markov chain on the graph. On a directed graph, P is given as

$$P_{uv} = \begin{cases} \frac{1}{d_u^{out}} & \text{if } (u,v) \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases}$$

As stated in [12], the normalized Laplacian matrix of a directed graph can be defined as

$$\tilde{L} = I - \frac{\Phi^{1/2} P \Phi^{-1/2} + \Phi^{-1/2} P^T \Phi^{1/2}}{2}, \tag{5}$$

where I is the identify matrix and $\Phi = diag(\phi(1), \phi(2), \cdots)$ in which ϕ represents the unique left eigenvector of P. Clearly, the normalized Laplacian matrix is Hermitian, i.e., $\tilde{L} = \tilde{L}^T$ where $\tilde{L}^T$ denotes the conjugated transpose of $\tilde{L}$.

Ye et al. [2] have shown that using the normalized Laplacian matrix to interpret the density matrix, the von Neumann entropy of directed graphs is the Shannon entropy associated with the normalized Laplacian eigenvalues, i.e.,

$$H_{VN}^D = -\sum_{i=1}^{|\mathcal{V}|} \frac{\tilde{\lambda}_i}{|\mathcal{V}|} \ln \frac{\tilde{\lambda}_i}{|\mathcal{V}|}, \tag{6}$$

where $\tilde{\lambda}_i, i = 1, \cdots, |\mathcal{V}|$ are the eigenvalues of the normalized Laplacian matrix $\tilde{L}$. Unfortunately, for large graphs this is not a viable proposition since the time required to solve the eigensystem is cubic in the number of vertices. To overcome this problem we extend the analysis of Han et al. [11] from undirected to directed graphs. To do this we again make use of the quadratic approximation to the Shannon entropy in order to obtain a simplified expression for the von Neumann entropy of a directed graph, which can be computed in a time that is quadratic in the number of vertices. Our starting point is the quadratic approximation to

the von Neumann entropy in terms of the traces of normalized Laplacian and the squared normalized Laplacian

$$H_{VN}^D = \frac{Tr[\tilde{L}]}{|\mathcal{V}|} - \frac{Tr[\tilde{L}^2]}{|\mathcal{V}|^2}. \tag{7}$$

To simplify this expression a step further, we show the computation of the traces for the case of a directed graph. This is not a straightforward task, and requires that we distinguish between the in-degree and out-degree of vertices. We first consider Chung's expression for the normalized Laplacian of directed graphs and write

$$Tr[\tilde{L}] = Tr[I - \frac{\Phi^{1/2}P\Phi^{-1/2} + \Phi^{-1/2}P^T\Phi^{1/2}}{2}]$$
$$= Tr[I] - \frac{1}{2}Tr[\Phi^{1/2}P\Phi^{-1/2}] - \frac{1}{2}Tr[\Phi^{-1/2}P^T\Phi^{1/2}].$$

Since the matrix trace is invariant under cyclic permutations, we have

$$Tr[\tilde{L}] = Tr[I] - \frac{1}{2}Tr[P\Phi^{-1/2}\Phi^{1/2}] - \frac{1}{2}Tr[P^T\Phi^{1/2}\Phi^{-1/2}]$$
$$= Tr[I] - \frac{1}{2}Tr[P] - \frac{1}{2}Tr[P^T].$$

The diagonal elements of the transition matrix P are all zeros, hence we obtain

$$Tr[\tilde{L}] = Tr[I] = |\mathcal{V}|,$$

which is exactly the same as in the case of undirected graphs.

Next we turn our attention to $Tr[\tilde{L}^2]$:

$$Tr[\tilde{L}^2] = Tr[I^2 - (\Phi^{1/2}P\Phi^{-1/2} + \Phi^{-1/2}P^T\Phi^{1/2}) + \frac{1}{4}(\Phi^{1/2}P\Phi^{-1/2}\Phi^{1/2}P\Phi^{-1/2}$$
$$+ \Phi^{1/2}P\Phi^{-1/2}\Phi^{-1/2}P^T\Phi^{1/2} + \Phi^{-1/2}P^T\Phi^{1/2}\Phi^{1/2}P\Phi^{-1/2}$$
$$+ \Phi^{-1/2}P^T\Phi^{1/2}\Phi^{-1/2}P^T\Phi^{1/2})]$$
$$= Tr[I^2] - Tr[P] - Tr[P^T] + \frac{1}{4}(Tr[P^2] + Tr[P\Phi^{-1}P^T\Phi]$$
$$+ Tr[P^T\Phi P\Phi^{-1}] + Tr[P^{T^2}])$$
$$= |\mathcal{V}| + \frac{1}{2}(Tr[P^2] + Tr[P\Phi^{-1}P^T\Phi]),$$

which is different to the result obtained in the case of undirected graphs.

To continue the development we first partition the edge set $\mathcal{E}$ of the graph $\mathcal{G}$ into two disjoint subsets $\mathcal{E}_1$ and $\mathcal{E}_2$, where $\mathcal{E}_1 = \{(u,v)|(u,v) \in \mathcal{E}$ and $(v,u) \notin \mathcal{E}\}$, $\mathcal{E}_2 = \{(u,v)|(u,v) \in \mathcal{E}$ and $(v,u) \in \mathcal{E}\}$ that satisfy the conditions $\mathcal{E}_1 \bigcup \mathcal{E}_2 = \mathcal{E}$, $\mathcal{E}_1 \bigcap \mathcal{E}_2 = \emptyset$. Then according to the definition of the transition matrix, we find

$$Tr[P^2] = \sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} P_{uv} P_{vu} = \sum_{(u,v) \in \mathcal{E}_2} \frac{1}{d_u^{out} d_v^{out}}.$$

Using the fact that $\Phi = diag(\phi(1),(2),\cdots)$ we have

$$Tr[P\Phi^{-1}P^T\Phi] = \sum_{u\in\mathcal{V}}\sum_{v\in\mathcal{V}} P_{uv}^2 \frac{\phi(u)}{\phi(v)} = \sum_{(u,v)\in\mathcal{E}} \frac{\phi(u)}{\phi(v)d_u^{out2}}.$$

Then, we can approximate the von Neumann entropy of a directed graph in terms of the in-degree and out-degree of the vertices as follows

$$H_{VN}^D = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2}\left\{ \sum_{(u,v)\in\mathcal{E}} \left(\frac{1}{d_u^{out}d_v^{out}} + \frac{d_u^{in}}{d_v^{in}d_u^{out2}} \right) - \sum_{(u,v)\in\mathcal{E}_1} \frac{1}{d_u^{out}d_v^{out}} \right\} \quad (8)$$

or, equivalently,

$$H_{VN}^D = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2}\left\{ \sum_{(u,v)\in\mathcal{E}} \frac{d_u^{in}}{d_v^{in}d_u^{out2}} + \sum_{(u,v)\in\mathcal{E}_2} \frac{1}{d_u^{out}d_v^{out}} \right\}. \quad (9)$$

Clearly, both entropy approximations contain two terms. The first is the graph size while the second one depends on the in-degree and out-degree statistics of each pair of vertices connected by an edge. Moreover, the computational complexity of these expressions is quadratic in the graph size.

2.3   Jensen-Shannon Divergence Kernel for Directed Graphs

It is interesting to note that the von Neumann entropy is related to the sum of the entropy contribution from each directed edge, and this allows us to compute a local entropy measure for edge $(u,v) \in \mathcal{E}$,

$$I_{uv} = \frac{d_u^{in}}{2|\mathcal{E}||\mathcal{V}|d_v^{in}d_u^{out2}}.$$

If this edge is bidirectional, i.e., $(u,v) \in \mathcal{E}_2$, then we add an additional quantity to I_{uv},

$$I'_{uv} = \frac{1}{2|\mathcal{E}_2||\mathcal{V}|d_u^{out}d_v^{out}}.$$

Clearly, the local entropy measure represents the entropy associated with each directed edge and more importantly, it avoids the bias caused by graph size, which means that it is the edge entropy contribution determined by the in and out-degree statistics, and neither the vertex number nor edge number of the graph that distinguishes a directed edge.

In our analysis, since we are measuring the similarity between the structures of two graphs, the vertex and edge information plays a significant role in the comparison. As a result, we use the following measure for quantifying the complexity of a directed graph, which is affected by the vertex and edge information,

$$H_E(\mathcal{G}) = \sum_{(u,v)\in\mathcal{E}} \frac{d_u^{in}}{d_v^{in}d_u^{out2}} + \sum_{(u,v)\in\mathcal{E}_2} \frac{1}{d_u^{out}d_v^{out}}. \quad (10)$$

We let $\mathcal{G}_1 \oplus \mathcal{G}_2 = \{\mathcal{V}_1 \bigcup \mathcal{V}_2, \mathcal{E}_1 \bigcup \mathcal{E}_2\}$, i.e., the disjoint union graph (or composite graph) of $\mathcal{G}_1$ and $\mathcal{G}_2$. Then, we compute the Jensen-Shannon divergence between $\mathcal{G}_1$ and $\mathcal{G}_2$ as

$$D_{JS}(\mathcal{G}_1, \mathcal{G}_2) = H_E(\mathcal{G}_1 \oplus \mathcal{G}_2) - \frac{H_E(\mathcal{G}_1) + H_E(\mathcal{G}_2)}{2}.$$

As a consequence, we finally have the Jensen-Shannon diffusion kernel

$$k_{JS}(\mathcal{G}_1, \mathcal{G}_2) = \exp\left\{ \frac{H_E(\mathcal{G}_1) + H_E(\mathcal{G}_2)}{2} - H_E(\mathcal{G}_1 \oplus \mathcal{G}_2) \right\}. \qquad (11)$$

The decay factor λ is set to be 1 for simplifying matters. Clearly, since Jensen-Shannon divergence is a dissimilarity measure and is symmetric, the diffusion kernel associated with the divergence is positive definite. Moreover, for a pair of graphs with N vertices, the computational complexity of the kernel is $O(N^2)$.

3 Experiments

In this section, we evaluate the experimental performance of the proposed directed graph Jensen-Shannon divergence kernel. Specifically, we first explore the graph classification performance of our method on a set of random graphs generated from different models. Then we apply our method to the real-world data, namely the NYSE stock market networks, in order to explore whether the kernel method can be used to analyze the complex data effectively.

3.1 Datasets

We commence by giving a brief overview of the datasets used for experiments in this paper. We use two different datasets, the first one is synthetically generated artificial graphs, while the other one is extracted from real-world financial system.

Random Directed Graph Dataset. Contains a large number of directed graphs which are randomly generated according to two different directed random graph models, namely a) the classical Erdős-Rényi model and b) the Barabási-Albert model [13]. The different directed graphs in the database are created using a variety of model parameters, e.g., the graph size and the connection probability in the Erdős-Rényi model and the number of added connections at each time step in the Barabási-Albert model.

NYSE Stock Market Network Dataset. Is extracted from a database consisting of the daily prices of 3799 stocks traded on the New York Stock Exchange (NYSE). In our analysis we employ the correlation-based network to represent the structure of the stock market since many meaningful economic insights can be extracted from the stock correlation matrices [14]. To construct the dynamic network, 347 stocks that have historical data from January 1986 to February

2011 are selected. Then, we use a time window of 20 days (i.e., 4 weeks) and move this window along time to obtain a sequence (from day 20 to day 6004) in which each temporal window contains a time-series of the daily return stock values over a 20-day period. We represent trades between different stocks as a network. For each time window, we compute the cross-correlation coefficients between the time-series for each pair of stocks, and create connections between them if the maximum absolute value of the correlation coefficient is among the highest 5% of the total cross correlation coefficients. This yields a time-varying stock market network with a fixed number of 347 vertices and varying edge structure for each of 5976 trading days.

3.2   Graph Classification

To investigate the classification performance of our proposed kernel method, we first apply it to the *Random Directed Graph Dataset*. Given a dataset consisting of N graphs, we construct a $N \times N$ kernel matrix

$$K = \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}$$

in which k_{ij}, $i, j = 1, 2, \cdots, N$, denotes the graph kernel value between graphs $\mathcal{G}_i$ and $\mathcal{G}_j$. In our case, we have $k_{ij} = k_{JS}(\mathcal{G}_i, \mathcal{G}_j)$. With the kernel matrix to hand, we then perform the kernel principal component analysis (kernel PCA) [15] in order to extract the most important information contained in the matrix and embed the data to a low-dimensional principal component space.

Figure 1 gives two kernel PCA plots of kernel matrix computed from a) Jensen-Shannon divergence kernel and b) shortest-path kernel [7]. In particular, the kernels are computed from 600 graphs that belong to six groups (100 graphs in each group): a) Erdős-Rényi (ER) graphs with $n = 30$ vertices and connection probability $p = 0.3$; b) Barabási-Albert (BA) graphs with $n = 30$ vertices and average degree $\bar{k} = 9$; c) ER graphs with $n = 100$ and $p = 0.3$; d) BA graphs with $n = 100$ and $\bar{k} = 30$; e) ER graphs with $n = 30$ and $p = 0.6$; and f) BA graphs with $n = 30$ and $\bar{k} = 18$. From the left-hand panel, the six groups are clearly separated very well, implying that the Jensen-Shannon divergence kernel is effective in comparing not only the structural properties of directed graphs that belong to different classes, but also the structure difference between graphs generated from the same model. However, the panel on the right-hand side suggests that the shortest-path kernel cannot efficiently distinguish the difference between graphs of the six groups since the graphs cannot be clearly separated.

To better evaluate the properties of the proposed kernel method. We proceed to compare the classification result of the Jensen-Shannon divergence kernels developed for directed and undirected graphs. To this end, we employ the undirected Jensen-Shannon divergence kernel and repeat the above analysis on the undirected version of graphs in *Random Directed Graph Dataset* and report
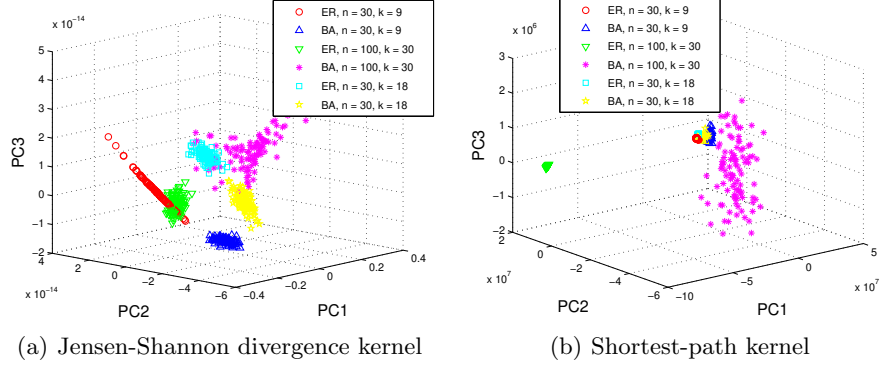
(a) Jensen-Shannon divergence kernel   (b) Shortest-path kernel

Fig. 1. Kernel PCA plots of two graph kernel methods on a number of directed graphs generated from Erdős-Rényi model and Barabási-Albert model.

the result in Fig. 2. The undirected kernel method does not separate different groups of graphs well as there are a large number of data points overlapping in the kernel principal component space. Moreover, the BA graphs represented by yellow stars and the ER graphs symbolled by cyan squares ($n = 30$ and $\bar{k} = 18$) are completely mixed. This comparison clearly indicates that for directed and undirected random graphs that are generated using the same model and parameters, the directed graph kernel method is more effective in classifying them than its undirected analogue.
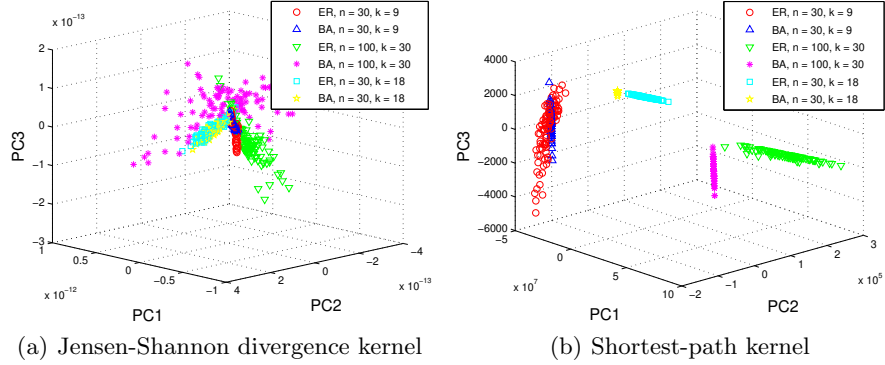

(a) Jensen-Shannon divergence kernel   (b) Shortest-path kernel

Fig. 2. Kernel PCA plots of two graph kernel methods on a number of undirected graphs generated from Erdős-Rényi model and Barabási-Albert model.

3.3   Financial Data Analysis

We turn our attention to the financial data, and explore whether the proposed kernel method can be used as an efficient tool for visualizing and studying the time evolution of the financial networks. To commence, we select two financial crisis time series from the data, namely a) Black Monday (from day 80 to day 120) and b) September 11 attacks (from day 3600 to day 3650) and construct their corresponding kernel matrices respectively. Then, in Fig. 3 we show the path of the financial network in the PCA space during the selected financial crises respectively. The number beside each data point represents the day number in the time-series. From the left-hand panel we observe that before Black Monday, the network structure remains relatively stable. However, during Black Monday (day 117), the network experiences a considerable change in structure since the graph Jensen-Shannon divergence kernel changes significantly. After the crisis, the network structure returns to its normal state. A similar observation can also be made concerning the September 11 attacks which is shown in the right-hand panel. The stock network again undergoes a significant crash in which the network structure undergoes a significant change. The crash is also followed by a quick recovery.
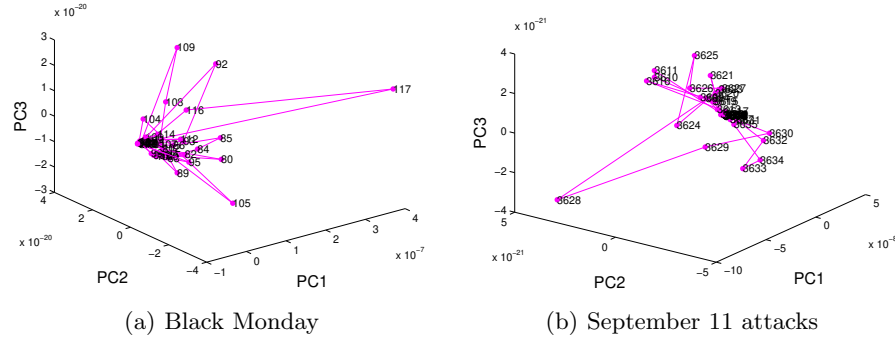

(a) Black Monday   (b) September 11 attacks

Fig. 3. Path of the time-evolving financial network in the PCA space during different financial crises.

4   Conclusion

In this paper, we present a novel kernel method for comparing the structures of a pair of directed graphs. In essence, this kernel is based on the Jensen-Shannon divergence between two probability distributions associated with graphs, which are represented by the recently developed von Neumann entropy for directed graphs. Specifically, the entropy approximation formula allows the entropy to be

expressed in terms of a sum over all edge-based entropies of a directed graph, which in turn allows the kernel to be computed from the edge entropy sum of two individual graphs and their disjoint union graph. In the experiments, we have evaluated the effectiveness of our kernel method in terms of classifying structures that belong to various classes and analyzing time-varying realistic networks. In the future, it would be interesting to see what features the Jensen-Shannon divergence kernel reveal in additional domains, such as human functional magnetic resonance imaging data. Another interesting line of investigation would be to explore if the kernel method can be extended to the domains of edge-weighted graphs, labeled graphs and hypergraphs.

References

1. Bai, L., Hancock, E.R.: Graph kernels from the Jensen-Shannon divergence. Journal of Mathematical Imaging and Vision **47**(1-2) (2013) 60–69
2. Ye, C., Wilson, R.C., Comin, C.H., Costa, L.d.F., Hancock, E.R.: Approximate von Neumann entropy for directed graphs. Physical Review E **89**(052804) (2014)
3. Cuturi, M., Fukumizu, K., Vert, J.P.: Semigroup kernels on measures. Journal of Machine Learning Research **6** (2005) 1169–1198
4. Martins, A.F.T., Smith, N.A., Xing, E.P., Aguiar, P.M.Q., Figueiredo, M.A.T.: Nonextensive information theoretic kernels on measures. Journal of Machine Learning Research **10** (2009) 935–975
5. Vishwanathan, S.V.N., Schraudolph, N.N., Kondor, R., Borgwardt, K.M.: Graph kernels. Journal of Machine Learning Research **11** (2010) 1201–1242
6. Kashima, H., Tsuda, K., Inokuchi, A.: Marginalized kernels between labeled graphs. Proceedings in 20th International Conference on Machine Learning (2003) 321–328
7. Borgwardt, K.M., Kriegel, H.P.: Shortest-path kernels on graphs. Proceedings in the 5th IEEE International Conference on Data Mining (2005) 74–81
8. Aziz, F., Wilson, R.C., Hancock, E.R.: Backtrackless walks on a graph. IEEE Transactions on Neural Networks and Learning Systems **24**(6) (2013) 977–989
9. Berwanger, D., Gradel, E., Kaiser, L., Rabinovich, R.: Entanglement and the complexity of directed graphs. Theoretical Computer Science **463** (2012) 2–25
10. Escolano, F., Hancock, E.R., Lozano, M.A.: Heat diffusion: Thermodynamic depth complexity of networks. Physical Review E **85**(036206) (2012)
11. Han, L., Escolano, F., Hancock, E.R., Wilson, R.C.: Graph characterizations from von Neumann entropy. Pattern Recognition Letters **33** (2012) 1958–1967
12. Chung, F.: Laplacians and the Cheeger inequailty for directed graphs. Annals of Combinatorics **9** (2005) 1–19
13. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science **286** (1999) 509–512
14. Battiston, S., Caldarelli, G.: Systemic risk in financial networks. Journal of Financial Managements Markets and Institutions **1** (2013) 129–154
15. Schölkopf, B., Smola, A., Müller, K.B.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation **10** (1998) 1299–1319