This is a repository copy of *Entropic Dynamic Time Warping Kernels for Co-evolving Financial Time Series Analysis*.

Entropic Dynamic Time Warping Kernels for Co-evolving Financial Time Series Analysis

Lu Bai, Lixin Cui, Lixiang Xu, Zhihong Zhang, Yue Wang, Edwin R. Hancock, *IEEE Fellow*

Abstract—Network representations are powerful tools to modelling the dynamic time-varying financial complex systems consisting of multiple co-evolving financial time series, e.g., stock prices, etc. In this work, we develop a novel framework to compute the kernel-based similarity measure between dynamic time-varying financial networks. Specifically, we explore whether the proposed kernel can be employed to understand the structural evolution of the financial networks with time associated with standard kernel machines. For a set of time-varying financial networks with each vertex representing the individual time series of a different stock and each edge between a pair of time series representing the absolute value of their Pearson correlation, our start point is to compute the commute time matrix associated with the weighted adjacency matrix of the network structures, where each element of the matrix can be seen as the enhanced correlation value between pairwise stocks. For each network, we show how the commute time matrix allows us to identify a reliable set of dominant correlated time series as well as an associated dominant probability distribution of the stock belonging to this set. Furthermore, we represent each original network as a discrete dominant Shannon entropy time series computed from the dominant probability distribution. With the dominant entropy time series for each pair of financial networks to hand, we develop an Entropic Dynamic Time Warping Kernels through the classical dynamic time warping framework, for analyzing the financial time-varying networks. We show that the proposed kernel bridges the gap between graph kernels and the classical dynamic time warping framework for multiple financial time series analysis. Experiments on time-varying networks extracted through New York Stock Exchange (NYSE) database demonstrate that the proposed method can effectively detect abrupt changes in networks as time series structures and can be used to characterize different stages in time-varying financial network evolutions.

Index Terms—Time-varying Financial Networks, Graph Entropy, Graph Kernels, Time Series.

I. INTRODUCTION

There have been a surge of research interests to employ machine learning algorithms for the analysis of dynamic networks, due to their potential applications in many practical systems, such as social systems, neural networks [1], uncertain systems [2], financial market systems [3], etc. Among these common systems, the financial market can be considered as a complex time-varying system consisting of multiple interacting financial components [4], e.g., the stock trade price and return rate. Due to the evolution of these financial variables with time, multiple co-evolving financial time series can be generated from the original data.

To analyze the time-varying financial market crisis or risk, change point detection has played an important role to identify abrupt changes in the time series properties [5]. Unfortunately, detecting such crucial points remains challenging, since it is difficult to detect the changes that cannot be easily observed for a system consisting of complex interactions between its constituent co-evolving time series [6]. One way to overcome this problem is to represent multiple co-evolving financial time series as a family of dynamic time-varying networks [7]. Specifically, for each dynamic network at time step t, each vertex represents an individual time series of a stock (e.g., stock trading prices over a time window of past W time steps ending at time step t) and each edge between a pair of co-evolving financial time series indicates their correlation (i.e., the absolute value of their Pearson correlation). Note that, this is an effective way to represent multiple co-evolving financial time series as dynamic networks and has been widely employed in most real-world applications [6], [8], [9], [10], [11]. Under this scenario, network-based approaches can be directly employed for the structural analysis of financial risks.

In machine learning, graph kernels are effective approaches to characterize graph structures in a high dimensional Hilbert space for structure analysis [12], [13]. The aim of this paper is to develop a new kernel approach for analyzing multiple co-evolving financial time series that are represented as network structures. Our work is based on representing each financial network as discrete entropy time series as well as the classical dynamic time warping measure between the series. The proposed approach bridges the gap between graph kernels and the dynamic time warping framework for time series analysis.

A. Literature Review

Network representations have been widely employed for the analysis of time-varying complex systems consisting of multiple co-evolving time series [8], [9], [10], [6], [11], e.g., the stock market with trade price, climate data, and functional magnetic resonance images. These approaches are based on the idea that the structure of so-called time-varying complex networks [14] inferred from the corresponding time series of the system can represent physical interactions between system entities that are richer than the original individual time series. According to this approach, one of the main objectives is to identify the extreme events which may considerably change the network structure. For example, in time-varying financial

Lu Bai, Lixin Cui, and Yue Wang are with Central University of Finance and Economics, Beijing, China. L. Bai and L. Cui (cuilixin@cufe.edu.cn) have equal contributions and are co-first authors.

Lixiang Xu is with Hefei University, Anhui, China.

Zhihong Zhang is with Xiamen University, Fujian, China. zhihong@xmu.edu.cn (Corresponding Author)

Edwin R. Hancock is with University of York, York, UK. Email: edwin.hancock@york.ac.uk
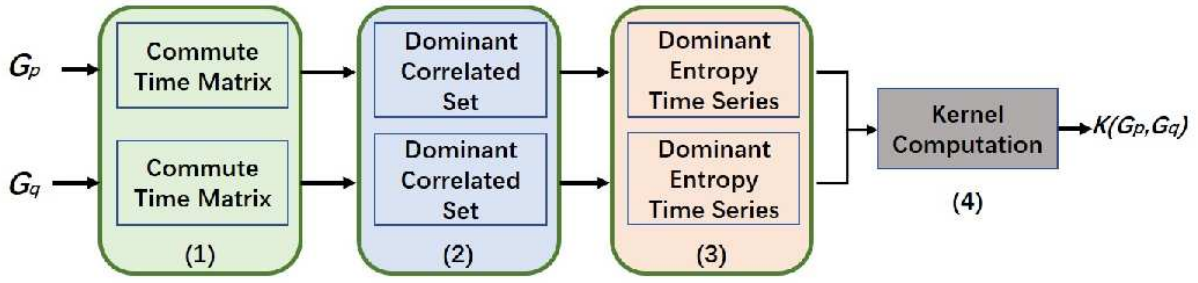
Fig. 1. The architecture to compute the proposed kernel. Given a pair of time-varying financial networks, for each of them we (1) compute the commute time (CT) matrix, (2) identify the set of dominant correlated time series represented by vertices through CT, (3) compute the dominant probability distribution and represent each network as dominant entropy time series, and finally (4) compute the kernel between the dominant entropy time series of two networks.

networks, extreme events corresponding to the financial instability of the stock are of particularly interest [6] and can be inferred by detecting the anomalies in the corresponding networks [11]. The network structure before and after an extreme event should be significantly different.

Broadly speaking, most existing approaches aim to characterize networks based on two principal approaches, namely a) derive network characteristics using connectivity structures, or statistics capturing connectivity structures and b) characterize the networks using statistical physics. Proponents of the former approach focus on capturing network substructures using communities, hubs and clusters [15], [16], [17]. On the other hand, proponents of the latter approach describe the network structures based on the partition function to characterize the network structures, and the corresponding temperature, energy, and entropy measures can be calculated in terms of this function [18], [19], [20], [21], [11]. Unfortunately, both approaches tend to approximate structural relationships of networks in a low dimensional pattern space, hence leading to substantial loss of information. This shortcoming affects the effectiveness of existing network methods for time series analysis. One principal approach to address this drawback is to adopt graph kernels. In machine learning, graph kernels are powerful tools for analyzing graph-based structural data. The main advantage of adopting graph kernels is that they provide an effective way of mapping graph structures into a high dimensional Hilbert space and thus better encapsulate the structural information.

Most existing state-of-the-art graph kernels fall into the scenario of R-convolution kernels, that are originally proposed by Haussler in 1999 [22]. The main idea underpinning R-convolution kernels is based on decomposing graphs into substructures and measuring the similarity between each pair of input graphs in terms of their isomorphic substructures, e.g., graph kernels based on comparing pairs of isomorphic a) walks, b) subgraphs, and c) subtrees. Representative R-convolution graph kernels based on substructures include the Weisfeiler-Lehman subtree kernel [23], the tree-based continuous attributed kernel [13], the aligned subtree kernel [24], the Jensen-Tsallis q-difference graph kernel [25], the optima assignment Weisfeiler-Lehman kernel [26], the core variants-based shortest path kernel [27], the random walk graph kernel [28], etc. Unfortunately, directly employing these graph

kernels to analyze time-varying network structures inferred from the original time series tends to be difficult. Because such network structures in most real-world applications are by nature complete weighted graphs, i.e., each vertex is adjacent to all remaining vertices, whereas the edge weights between the vertices may be rather different. It is difficult to decompose such a graph into substructures. This in turn influences the effectiveness of most existing graph kernels.

One way to address the problem is to discard the less interacted information between a pair of vertices and adopt the sparser versions of the original time-varying networks, i.e., the sparser networks only preserve the original edges indicating pairs of more interacted vertices. Under this scenario, Ye et al. [11], Silva et al. [6] and Wang et al. [29] have taken the widely adopted threshold-based methods and preserved the edges whose weights fall into the larger 10% of correlation-based weights. Although this strategy provides a way of directly employing existing graph kernels to accommodate time-varying networks for multiple co-evolving time series analysis, these sparse structures rely on the selection of the threshold. Thus, it is not clear how to preliminarily select a suitable threshold. Moreover, these sparser structures also lead to significant information loss, because many weighted edges are discarded. To further reveal this weakness, assume a sample financial network exhibited by Figure 2, where each vertex represents a stock and each edge corresponds to the correlation value between a pair of stocks. Clearly, one can only preserve 2 edges (i.e., 10% of the edges) through the above threshold-based strategy. In summary, analyzing time-varying networks associated with state-of-the-art graph kernels remains challenges.

B. Contributions

The objective of this paper is to address the aforementioned problems and develop a new kernel-based approach for analyzing multiple co-evolving financial time series. Specifically, we propose an Entropic Dynamic Time Warping Kernel (EDTWK) for time-varying financial networks, with each vertex representing the individual time series of a different stock (e.g., stock trading price) and each edge between a pair of co-evolving financial time series representing the absolute value of their Pearson correlation. One key innovation of

the proposed EDTWK kernel is the automatic identification of the dominant correlated vertex subset for each of the financial networks, i.e., the proposed kernel incorporates the process of identifying the most mutually correlated stocks specified by the vertex subset. In contrast, the aforementioned methods through the threshold-based strategy cannot guarantee that the preserved vertices correspond to a more mutually correlated vertex subset. This is because these methods tend to individually select each edge with a higher correlation weight and many edges between the preserved vertices may not exist. Based on financial risk theory [30], the financial crises are usually caused by a set of the most mutually correlated stocks while having less uncertainty. As a result, the proposed EDTWK kernel not only overcomes the shortcoming of heuristically selecting the threshold value that arises in the threshold-based approach for time-varying network analysis [11], [6], but also captures more reliable information concerning the evolution of the financial system to hand. The computational framework of the proposed EDTWK kernel is shown in Fig.1. Specifically, the main contributions of this work are threefold.

First, for a family of time-varying financial networks, our start point is to compute the commute time matrices associated with their original weighted adjacency matrices, i.e., the absolute Pearson correlation based matrices. The reason of using the commute time matrix as the representation of each network structure is that each element of this matrix represents the average path length between a pair of vertices over all possible paths residing on the original weighted edges [31]. Thus, the commute time can be seen as the enhanced absolute Pearson correlation value between the time series of pairwise stocks, i.e., it integrates the effectiveness of all possible correlation-based paths between a pair of vertices of the original network. Moreover, the commute time is robust under the perturbation of the network structure (e.g., the changes of edges or paths [31]). As a result, the commute time matrix can provide a more stable representation for the financial network structure that may accumulate a lot of noises over time. In summary, the commute time matrix offers an elegant way to probe the original structure of the time-varying financial networks (see details in Section II-A). More specifically, the proposed approach associated with the commute time matrix will be more effective than that associated with the original absolute Pearson correlation matrix (see details in Section IV-B and Section IV-C).

Second, with the commute time matrix of each time-varying financial network to hand, we employ this matrix to automatically identify a set of dominant correlated vertices in the network structure (i.e., a set of the most mutually correlated time series represented by the set of vertices), by maximizing a quadratic programming problem associated with the commute time matrix. Specifically, we compute a dominant probability distribution of these time series belonging to the most mutually correlated set. We show that this strategy not only overcomes the shortcoming of existing threshold-based approaches [11], [6] that roughly select pairs of relatively more correlated time series, but also encapsulates reliable information in terms of the evolution of the financial system to hand. Furthermore, we transform each original time-varying financial network

into a discrete dominant entropy time series associated with the dominant probability distribution, i.e., we characterize the uncertainty of each network structure within the financial system to hand in terms of the classical Shannon entropy associated with the probability distribution. With each pair of entropy time series to hand, we compute the EDTWK kernel through the classical dynamic time warping framework. We show that the proposed kernel not only accommodates the complete weighted graphs through the commute time matrix, but also bridges the gap between graph kernels and the classical dynamic time warping framework for time series analysis (see details in Section III-B).

Third, we perform the proposed kernel on time-varying financial networks extracted from New York Stock Exchange (NYSE) data. Experimental results demonstrate that the proposed method can preserve the ordinal arrangement of the time-varying financial networks, and thus well understand the structural evolution of the networks with time, i.e., the proposed kernel can effectively detect abrupt changes in networks as time series structures and can be used to characterize different stages in time-varying financial network evolutions.

C. Paper Outline

This paper is organized as follows. Section II reviews the preliminary concepts. Section III defines the EDTWK kernel for time series analysis. Section IV provides the empirical evaluation results. Section V provides the conclusion and future work of this paper.

II. PRELIMINARY CONCEPTS

In this section, we briefly review preliminary concepts which will be utilized in this paper. We first review the concept of the commute time. Furthermore, we review the concept of a dynamic time warping framework inspired kernel.

A. Commute Time on Graphs

As we have stated, one main objective of this work is to automatically identify a set of the most mutually correlated stocks in terms of their time series. To this end, we require a correlation matrix as the weighted vertex adjacency matrix of the corresponding time-varying financial network to represent the physical interactions between multiple co-evolving time series. Broadly speaking, most state-of-the-art approaches adopt the aforementioned absolute Pearson correlation based matrix as the network representation [11], [29], [6]. In order to capture a reliable and robust mutually correlated stock set, in this work we propose to utilize the commute time matrix associated with the original correlation matrix as the network representation.

The main reasons of employing the commute time matrix are threefold. First, the commute time averages the time taken for a random walk to travel between a pair of vertices over all connecting paths residing on the original correlation based weighted adjacency matrix. Thus, the commute time can be considered as the enhanced correlation matrix. Second, since the commute time amplifies the correlation based affinity

between a pair of vertices, it is robust under the perturbation of the graph structure, e.g., the changes of edges or paths. Thus, the commute time based enhanced correlation matrix is robust and provides a stable correlation representation for the time-varying financial network that may accumulate a lot of noises over time. Third, the commute time is calculated through the Laplacian matrix of the original correlation based weighted adjacency matrix. In Section III, we will show how the commute time matrix can be employed to identify a set of the most mutually correlation stocks specified by a set of dominant vertices, associated with a quadratic problem.

In this subsection, we briefly introduce the concept of the commute time. Assume $G(V, E, A)$ is a complete weighted graph, where E is the edge set, V is the vertex set, and each vertex of V is connected by all the remaining vertices. Let A be the associated weighted adjacency matrix of G. If $A(u, v) = A(v, u) > 0$, we say that the vertices $v \in V$ and $u \in V$ are adjacent. Let D denote the degree matrix of G. D is a diagonal matrix and each of its diagonal element $D(u, u)$ corresponds to the sum of the corresponding row or column of A, i.e., $D(u, u) = \sum_v A(u, v)$. Then, the graph Laplacian matrix L is computed by $L = D - A$. The spectral decomposition of L is defined as $L = \Phi \Lambda \Phi^T$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$ is a $|V| \times |V|$ diagonal matrix with ascending eigenvalues as elements, i.e., $0 = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{|V|}$, and Φ is a $|V| \times |V|$ matrix $\Phi = (\phi_1|\phi_2|...|\phi_{|V|})$ with the corresponding ordered eigenvectors as columns. For G, the hitting time $H(u, v)$ between each pair of vertices v and u is computed as the expected number of steps taken by a classical random walk commencing from u and ending at v. Likewise, the commute time $C(u, v)$ is defined as the expected number of steps of the random walk commencing from u and ending at v, and then coming back to u again, i.e., $C(u, v) = H(u, v) + H(v, u)$. Thus, the commute time $C(u, v)$ can be calculated through the unnormalized Laplacian eigenvalues and eigenvectors [31] as

$$C(u, v) = \sum_{u \in V} D(u, u) \sum_{j=2}^{|V|} \frac{1}{\lambda_j} (\phi_j(u) - \phi_j(v))^2. \quad (1)$$

Remarks: The commute time has been proven to be a powerful tool to extract rich characteristics from complete weighted graphs. In previous works, Bai et al. [32] have employed the commute time matrix to develop a new quantum-inspired kernel for dynamic financial network analysis. Specifically, for the original complete weighted adjacency matrix of each financial network, they commence by abstracting the minimum or maximum spanning tree associated with the commute time matrix. For a pair of complete weighted graphs to be compared, the resulting quantum kernel is defined by measuring the similarity between their associated commute time spanning tree structures in terms of a new developed evolving model of discrete-time quantum walks. This approach significantly reduces the problem of information loss that arises in previously mentioned threshold-based methods for financial network analysis [11], [29], [6]. This is because the weights of the preserved edges on spanning tree structures correspond to the commute time values between corresponding pairs of vertices, and the commute time values integrates the

effectiveness over all possible paths residing on the original weighted edges. However, similar to these threshold-based approaches [11], [29], [6], the quantum kernel [32] cannot guarantee that the preserved vertices correspond to a more mutually correlated vertex subset, since the spanning tree is a very sparse structure (only $n - 1$ edges preserved for the network with n vertices) and many edges between the vertices do not exist. In other words, this kernel approach cannot reflect the most mutually correlated time series specified by the vertices, and will influence the effectiveness. To overcome this problem, in Section III, we will develop a new kernel-based approach for financial network analysis that can integrate the process of adaptively identifying the most mutually correlated financial time series of stocks associated with the commute time matrix.

B. The Dynamic Time Warping Framework

We review the global alignment kernel that is defined through the classical dynamic time warping framework [33]. Assume $\mathbf{T}$ is a set of discrete time series that take values in a space $\mathcal{X}$. For each 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 computed 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 assumed to possess unitary increments and no simultaneous repetitions. For $\mathbf{P}$ and $\mathbf{Q}$, each of their elements can be **an observation vector with fixed dimensions** at a corresponding time step. For any index i that is between 1 and $l - 1$ (i.e., $1 \leq i \leq l - 1$), the following condition holds for the increment vector of $\pi = (\pi_p, \pi_q)$, i.e.,

$$\begin{pmatrix} \pi_p(i+1) - \pi_p(i) \\ \pi_q(i+1) - \pi_q(i) \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}. \quad (2)$$

Within the framework of the classical dynamic time warping [33], the coordinates π_p and π_q of the alignment π define the warping function. Assume $\mathcal{A}(m, n)$ corresponds to a set of all possible alignments between $\mathbf{P}$ and $\mathbf{Q}$, Cuturi [33] has proposed a dynamic time warping inspired kernel, namely the Global Alignment Kernel, by considering all the possible alignments in $\mathcal{A}(m, n)$. The kernel is defined as

$$k_{\text{GA}}(\mathbf{P}, \mathbf{Q}) = \sum_{\pi \in \mathcal{A}(m,n)} e^{-D_{\mathbf{P}, \mathbf{Q}}(\pi)}, \quad (3)$$

where $D_{\mathbf{P}, \mathbf{Q}}(\pi)$ is the alignment cost given by

$$D_{\mathbf{P}, \mathbf{Q}}(\pi) = \sum_{i=1}^{|\pi|} \varphi(p_{\pi_p(i)}, q_{\pi_q(i)}), \quad (4)$$

and is defined through a local divergence φ that quantifies the discrepancy between each pair of elements $p_i \in \mathbf{P}$ and $q_i \in \mathbf{Q}$. In general, φ is defined as the squared Euclidean distance [33]. Note that, the kernel k_{GA} measures the quality

of both the optimal alignment and all other alignments $\pi \in \mathcal{A}(m, n)$, thus it is positive definite. Moreover, k_{GA} provides richer statistical measures of similarity by encapsulating the overall spectrum of the alignment costs $\{D_{\mathbf{P}, \mathbf{Q}}(\pi), \pi \in \mathcal{A}(m, n)\}$.

Remarks: The dynamic time warping based global alignment kernel k_{GA} has been proven to be a powerful tool of analyzing vectorial time series [33]. To extend the effective kernel k_{GA} for time series analysis domain into the graph kernel domain, Bai et al. [34] have developed a family of nested graph kernels through k_{GA}. Specifically, they commenced by decomposing each graph structure into a family of K-layer expansion subgraphs rooted at the centroid vertex. The nested depth-based complexity trace of each graph is computed by measuring the entropy on the family of K-layer expansion subgraphs. Since the parameter K varies from 1 to K, this complexity trace naturally forms a one-dimensional sequence-based characterization vector, that is similar to the one-dimensional time series vector. As a result, for a pair of graphs the resulting dynamic time warping based kernel can be directly computed by measuring the global alignment kernel k_{GA} between their complexity traces. Although, they demonstrated that the nested graph kernels outperform state-of-the-art graph kernels [28], [35], [36] on graph classification tasks. Unfortunately, as we have stated, the financial networks are by nature complete weighted graphs and it is difficult to decompose such network structures into the required expansion subgraphs rooted at the centroid vertex. As a result, directly preforming the dynamic time warping inspired kernel k_{GA} for time-varying financial networks tends to be elusive and remains challenges.

III. THE KERNEL FOR TIME-VARYING NETWORKS

In this section, we propose a kernel-based similarity measure for time-varying networks representing multiple co-evolving financial time series. Specifically, we commence by identifying a set of the most mutually correlated time series through maximizing a quadratic programming method on the commute time matrix. We exhibit how this allows us to compute a probability distribution for the time series belonging to the dominant set. Finally, we characterize each time-varying network as a discrete dominant entropy time series through the Shannon entropy associated with the probability distribution, and in turn develop a new kernel-based approach in terms of the classical dynamic time warping framework [33].

A. Identifying Dominant Correlated Time Series

We identify a set of the most mutually correlated time series for each time-varying financial network. Let $\mathbf{G} = \{G_1, \dots, G_t, \dots, G_T\}$ be a family of time-varying financial networks extracted from a complex system $\mathbf{S}$, and $G_t(V_t, E_t, A_t)$ be the sample network extracted from the system at time step t based on the absolute Pearson correlation measure stated earlier (i.e., the corresponding statement in Section I). Note that, in this work we assume that the time-varying network structures have fixed numbers of vertices, i.e., these networks have the same vertex set V_t, whereas the
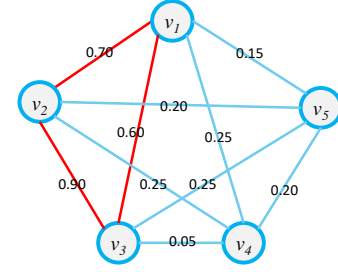

Fig. 2. The subset of financial time series specified by vertices $\{v_1, v_2, v_3\}$ is dominant.

edge sets E_t are quite different with time t. In real-world application, this a very common situation and usually appears where the time-varying networks are extracted from complex systems with a specified set of co-evolving time series, i.e., the system $\mathbf{S}$ has a fixing number of components co-evolving with time.

For each network G_t, we first compute its commute time matrix as C_t associated with its original absolute Pearson correlation based adjacency matrix. As we have stated previously, the commute time not only reflects the integrated effectiveness of all possible weighted paths between a pair of vertices of the original network structure, but is also robust with the perturbation of the network structure (i.e., the changes of edge weight on the original weighted adjacency matrix). As a result, the commute time matrix C_t can be seen as a reliable enhanced absolute Pearson correlation matrix for G_t. In other words, the commute time matrix provides a stable representation to further characterize the dynamic network G_t associated with time-varying correlations between vertices.

With the commute time matrix C_t of each network G_t to hand, we automatically identify a set of dominant correlated time series through the dominant set problem proposed by Pavan et al. [37]. The definition of the dominant set simultaneously emphasizes internal homogeneity and together with external inhomogeneity, and can be employed as a general definition of a cluster. An instance is exhibited in Fig.2. Here, assume a time-varying financial network consisting of 5 vertices denoted as v_1, v_2, v_3, v_4 and v_5. Each weight of this network represents the correlation between pairwise vertices. For this instance, the subset DS = $\{v_1, v_2, v_3\}$ forms the dominant set, i.e., the internal set. This is because the sum of the edge weights between the internal set $\{v_1, v_2, v_3\}$ is larger than the sum of those between the internal and external sets. Here, the edges between the internal set $\{v_1, v_2, v_3\}$ are marked as red color. As a result, the time series specified by DS can seen as the set of the most mutually correlated time series. To automatically identify the most mutually correlated time series from G_t, we can solve the corresponding dominant set problem by maximizing a quadratic program problem [37]. More specifically, associated with C_t, we compute the solution $\mathbf{a}$ of the following quadratic program problem [37]

$$\mathrm{argmax} \frac{1}{2} \mathbf{a}^T \mathbf{C}_t \mathbf{a} \qquad (5)$$

subject to $\mathbf{a} \in \mathbb{R}^{|V_t|}$, $a_i \geq 0$ and $\sum_{i=1}^{|V_t|} a_i = 1$. The solution vector $\mathbf{a}$ of Eq.(5) is an $|V_t|$-dimensional vector. When $a_i > 0$,

the i-th time series represented by the vertex $v_i \in V_t$ belongs to the most correlated time series subset of G_t. Thus, the number of the selected time series n is specified by counting the number of all positive components of $\mathbf{a}$. Based on the definition of Pavan and Pelillo [37], we can solve the local maximum of $f(\mathbf{a})$ by

$$a_i(k+1) = a_i(k) \frac{(\mathbf{C}_t \mathbf{a}(k))_i}{\mathbf{a}(k)^T \mathbf{C}_t \mathbf{a}(k)} \ . \tag{6}$$

where $a_i(k)$ corresponds to the i-th time series represented by $v_i \in V_t$ at iteration k. Based on the element value of $\mathbf{a}$, all time series represented by the vertices $v_1, \ldots, v_{|V_t|}$ fall into two disjoint subsets, i.e.,

$$\mathbf{S}_1(a) = \{v_i \in V_t \mid a_i > 0\}$$

and

$$\mathbf{S}_2(a) = \{v_j \in V_t \mid a_j = 0\}.$$

Clearly, the set $\mathbf{S}_1$ with nonzero values indicates the set of dominant correlated time series, i.e., the set of the most mutually correlated time series. Finally, note that, the solution vector $\mathbf{a}$ also corresponds to a probability distribution of the time series belonging to the dominant set $\mathbf{S}_1$, i.e., each element a_i corresponds to the probability of the i-th time series belonging to $\mathbf{S}_1$.

B. The Entropic Dynamic Time Warping Kernel

In this subsection, we develop a new kernel method for analyzing time-varying financial networks based on the classical dynamic time warping framework. To this end, we commence by representing the complex networks as discrete dominant entropy time series using the Shannon entropy through the most mutually correlated time series set introduced in Section III-A. The reason of characterizing the network using the entropy measure is that the Shannon entropy is an effective way of measuring the uncertainty in the corresponding financial system, associated with the probability distribution of the stocks belonging to the correlated set. Specifically, for each sample network $G_t(V_t, E_t)$ from $\mathbf{G}$ at time step t, we first compute the associated commute time matrix $\mathbf{C}_t$. Moreover, by solving the quadratic program problem [37] on the commute time matrix $\mathbf{C}_t$, we identify the set of dominant correlated time series $\mathbf{S}_1$ and compute the associated probability distribution $\mathbf{a}$ of the time series belonging to $\mathbf{S}_1$. Based on Section III-A, the remaining non-dominant correlated time series are included in the set $\mathbf{S}_2$. With the probability distribution $\mathbf{a}$ to hand, based on [38] we compute a dominant Shannon entropy for G_t as

$$H_S(\mathbf{a}) = - \sum_{i=1}^{|V_t|} a_i \log a_i, \tag{7}$$

where a_i is the probability of the i-th time series represented by vertex $v_i \in V_t$. Eq.(7) indicates that the dominant Shannon entropy is computed by the sum of elements $-a_i \log a_i$, thus each element $-a_i \log a_i$ can be seen as a **dominant sub-entropy** $\bar{H}_S(v_i)$ of the i-th time series represented by v_i, i.e.,

$$\bar{H}_S(v_i) = -a_i \log a_i. \tag{8}$$

a) A Sample Dynamic Network G_t at Time Step t (The Commute Time Matrix)

b) The Probability Distribution of the Vertices Belonging to the Dominant Set

v_1	v_2	v_3	v_4	v_5
0.24	0.40	0.36	0	0

c) The Dominant Entropy Characteristics Vector E_t for G_t

$E_t =$

v_1	v_2	v_3	v_4	v_5
0.34	0.37	0.36	0	0

d) The Dominant Entropy Time Series S_t for G_t

$S_t = \begin{cases} \end{cases}$

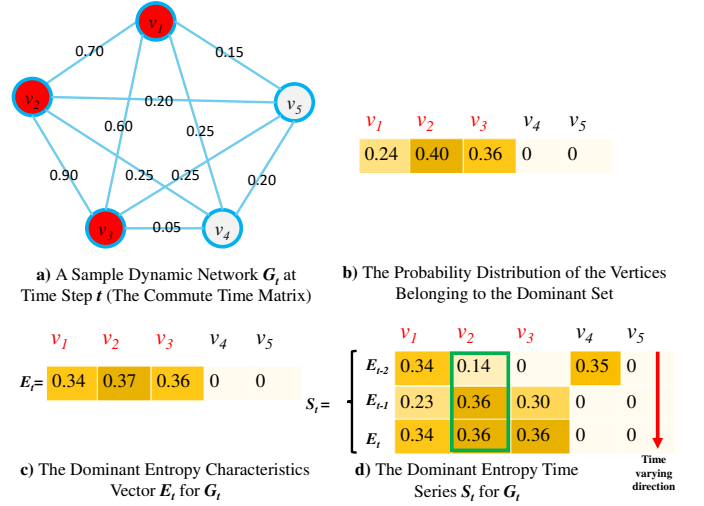|	v_1	v_2	v_3	v_4	v_5
E_{t-2}	0.34	0.14	0	0.35	0
E_{t-1}	0.23	0.36	0.30	0	0
E_t	0.34	0.36	0.36	0	0

Time varying direction

Fig. 3. Instance of Characterizing the Time-varying Networks as the Dominant Entropy Time Series. The procedure of characterizing each time-varying network $G_t \in \mathbf{G}$ at time t consists of three steps. 1) The first step computes probability distribution of the dominant correlated time series over the commute time matrix representation based on Eq.(6). These dominant correlated time series are marked by the red color. 2) The second step computes the dominant entropy characteristics vector E_t of G_t based on Eq.(9). Here, each element of the vector indicates the dominant sub-entropy of each vertex associated with its probability belonging to the dominant correlated time series, based on Eq.(8). 3) Assume the value of the time interval w is 3, the third step computes the resulting dominant entropy time series of G_t as $\mathcal{S}_t$, where each rows corresponds to the dominant entropy characteristics vector of the network between time $t-2$ to t (i.e., the dominant entropy characteristics vectors of the recent 3 time steps ending to time step t). Here, each column of $\mathcal{S}_t$ corresponds to the entropy time series of the stock represented by the corresponding vertex. For instance the column marked by green color corresponds to the series of the stock represented by vertex v_2, and the red arrow indicates the time-varying direction.

Note that, if $a_i = 0$, we say that the i-th time series does not belong to $\mathbf{S}_1$ and we set $-a_i \log a_i = 0$. With the sub-entropies of all vertices to hand, we compute the dominant entropy characteristics for each network G_t at time step t as

$$E_t = \{\bar{H}_S(v_1), \ldots, \bar{H}_S(v_i), \ldots, \bar{H}_S(v_V)\}^T, \tag{9}$$

where

$$\bar{H}_S(v_i) = \begin{cases} -a_i \log a_i, & a_i > 0, \text{ i.e., } v_i \in \mathbf{S}_1 \\ 0, & a_i = 0, \text{ i.e., } v_i \in \mathbf{S}_2 \end{cases} \tag{10}$$

Eq.(10) indicates that we only compute the sub-entropies for the dominant correlated time series in $\mathbf{S}_1$, and do not consider the non-dominant correlated time series in $\mathbf{S}_2$.

With the dominant entropy characteristics to hand, we further characterize each network G_t as entropy time series. To this end, we move a time interval of w time steps over all the time-varying networks of the financial system to construct a time-varying dominant entropy time series for each network G_t at time step t. In this work, we set the value of w as 28. Specifically, for each network G_t, we compute its dominant entropy time series $\mathcal{S}_t$ associated with its time interval s as

$$\mathcal{S}_t = \{E_{t-w+1} | E_{t-w+2} | \ldots | E_s | \ldots | E_t\}, \tag{11}$$

where $s \in \{t - w + 1, t - w + 2, \ldots, t\}$, and each column E_s of $\mathcal{S}_t$ is the entropy characteristics vector of each network $G_s \in \mathbf{G}$ at time step s and is defined by Eq.(9). Clearly, the dominant entropy time series $\mathcal{S}_t$ of the network G_t encapsulates the w time-varying entropy characteristics vectors of the networks G_{t-w+1} at time $t - w + 1$ to G_t at time t. An instance of characterizing the time-varying network G_t at time step t is exhibited in Figure 3

Assume $G_p \in \mathbf{G}$ and $G_q \in \mathbf{G}$ are a pair of time-varying networks at time steps p and q respectively, and their associated entropy time series are

$$\mathcal{S}_p = \{E_{p-w+1}|E_{p-w+2}|\ldots|E_p\}$$

and

$$\mathcal{S}_q = \{E_{q-w+1}|E_{q-w+2}|\ldots|E_q\}.$$

We define the Entropic Dynamic Warping Kernel (EDTWK) between G_p and G_q as

$$\begin{aligned} k_{\mathbf{DCTE}}(G_p, G_q) &= k_{\mathbf{GA}}(\mathcal{S}_p, \mathcal{S}_q) \\ &= \sum_{\pi \in \mathcal{A}(w,w)} e^{-D_{\mathbf{p},\mathbf{q}}(\pi)}, \end{aligned} \quad (12)$$

where $k_{\mathbf{GA}}$ is the dynamic time warping inspired Global Alignment Kernel (GAK) defined in Eq.(3), π is the warping alignment between the entropy time series of G_p and G_q, $\mathcal{A}(w,w)$ is all possible alignments and $D_{\mathbf{p},\mathbf{q}}(\pi)$ refers to the alignment cost obtained via Eq.(4).

Remarks: Although the proposed EDTWK kernel is related to the general principles of the GAK kernel. However, the proposed kernel has two distinct theoretical differences. First, the original GAK kernel is only designed for vectorial time series and cannot capture intrinsic relationships between time series. In contrast, our proposed kernel is explicitly designed for time-varying financial networks that reflect correlations between pairs of time series. Second, only the proposed EDTWK kernel can identify the dominant correlated time series through the analysis over the commute time matrix. Based on financial risk theory [30], financial crises are usually caused by a set of the most correlated stock time series having less uncertainties. Therefore, only the proposed kernel is able to capture more reliable financial information.

In summary, the proposed kernel provides an effective way of incorporating the structural correlations between time series into the process of multiple co-evolving time series analysis.

C. Time Complexity

For a pair of networks, the proposed kernel $k_{\mathbf{DCTE}}$ requires time complexity $O(n^3 + w^2)$. The reasons are as follows. Assume a family of time-varying networks and each network has n vertices. Computing the dominant commute time entropy kernel k between a pair of networks associated with a time window of w steps requires time complexity $O(n^3 + w^2)$. Because computing the required entropy time series is based on the computation of the commute time. This computation relies on the spectral decomposition of the Laplician matrix and thus requires time complexity $O(n^3)$. Moreover, computing all possible warping alignments over w time steps requires

time complexity $O(w^2)$. Thus, the whole time complexity of the proposed kernel k is $O(n^3 + w^2)$.

D. Related Works to The Proposed Kernel

Comparing to some state-of-the-art approaches, the proposed EDTWK kernel has a number of advantages.

First, unlike the dynamic time warping inspired GAK kernel [33], the proposed kernel is developed for time-varying complex networks. Since the network encapsulates rich co-relationship between pairwise co-evolving time series, the proposed kernel can reflect richer correlated information than the classical dynamic time warping framework for original vectorial time series.

Second, the proposed kernel is based on the new dominant entropy time series that is computed through a quadratic programming method on the commute matrix to identify the most correlated time series subset. As a result, unlike the existing threshold-based approaches [6], [11], [29] that roughly select pairs of relatively more correlated time series, the proposed kernel can reflect reliable dominant correlated information between time series through the dominant entropy time series. Furthermore, the commute time encapsulates the integrated effectiveness of all possible paths between a pair of vertices. As a result, the dominant entropy time series computed through the commute time matrix can potentially encapsulate the weighted information over all edges, and overcome the shortcoming of information loss arising in the threshold-based approaches.

Third, as we have stated, the time-varying networks are usually complete weighted networks. Most existing graph kernels are designed based on the concept of R-convolution [22] and cannot directly accommodate such network structures and need to transform them into sparse structures [6]. Unfortunately, these sparse structures discard many weighted edges and certainly lead to information loss. By contrast, the commute time is computed through the Laplacian matrix that can directly accommodate complete weighted graphs. Thus, the proposed kernel encapsulates the whole structural information residing on all weighted edges. On the other hand, the discrete-time quantum walk kernel [32] can also accommodate complete weighted graphs through the commute time and has been demonstrated effective for time-varying financial network analysis. However, similar to the aforementioned approaches, this quantum kernel also needs to extract the minimum spanning trees over the commute time matrix as the sparser network structure. Since the number of the preserved edges is always equivalent to that of the vertices, this process cannot adaptively determine the most correlated time series. Thus, unlike the proposed kernel, this quantum kernel cannot reflect reliable characteristics of the financial network structures.

Fourth, Cuturi11 [33] has shown that the dynamic time warping inspired GAK kernel is both faster and more efficient than other kernel approaches based on the dynamic time warping framework. Since the proposed kernel is based on measuring the GAK kernel between the entropic time series for a pair of time-varying networks, the proposed kernel can
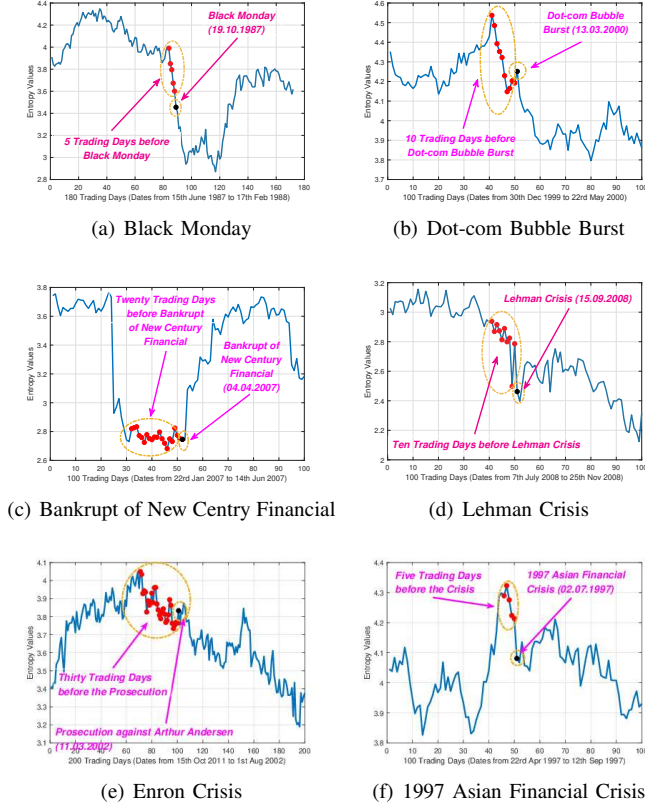
Fig. 4. Evaluation of Dominant Shannon Entropy.

naturally inherit these advantages of the GAK kernel. On the other hand, with the entropic time series to hand, any other state-of-the-art kernel approach designed for vectorial time series can be directly employed for the analysis of time-varying networks, e.g., the time series cluster kernel [39]. However, it has been shown that the cluster kernel is associated with a lot of parameters. Selecting a set of suitable parameters usually tends to be a complicated problem if there is no priori knowledge. By contrast, the associated GAK kernel of our approach only needs little parameters and is more applicable for time-varying network analysis.

In summary, the proposed kernel bridges the gap between state-of-the-art graph kernels and the classical dynamic time warping framework for time-varying networks, providing a new alternative way for analyzing time series more effectively.

IV. EXPERIMENTS OF TIME SERIES ANALYSIS

We empirically validate the effectiveness of the proposed kernel approach on a family of time-varying financial networks extracted from the New York Stock Exchange (NYSE) dataset [6], [11]. The NYSE dataset consists of 347 stocks associated with their daily closing prices over 6004 transaction days starting from January 1986 to February 2011. These prices are all collected from the public financial dataset on Yahoo (http://finance.yahoo.com). To abstract the time-varying financial network structures, we employ a time window W of fixed size (i.e., 28 days) that has been stated in Section I. We slide this fixed sized window along time to derive a sequence from the 29th trading day to the 6004th trading day, where

each temporal time window encapsulates a set of 347 co-evolving daily stock price time series of the 347 stocks over 28 days. We characterize the trades between various stocks as a network structure with each stock as the vertex. Specifically, for each time window we calculate the absolute value of the Pearson correlation between the time series for pairwise stocks as their edge weight. This in turn generates a family of time-varying financial network with a fixed number of 347 vertices and varying edge weights for the 5976 trading days. The aim of this study is to investigate whether the proposed kernel approach can be used to detect fluctuations in trading network structure due to global political or economic events.

A. Evaluation of The Entropy Time Series

We commence by exploring whether the dominant entropy time series can significantly characterize the time-varying financial networks, since these new developed time series play an important role for the proposed kernel. Specially, we investigate the evolutionary behavior of the NYSE stock market through calculating the dominant Shannon entropy on the time-varying financial networks at each time step, i.e., we investigate how the sum of the dominant sub-entropies of the network varies with increasing time t. We exhibit the results in Fig.4, where the x-axis corresponds to the date (time) and the y-axis corresponds to the dominant Shannon entropy values. Fig.4 shows that the dominant Shannon entropy is sensitive to different financial crises (i.e., Black Monday [40], Dot-com Bubble Burst [41], Bankrupt of New Centry Financial, Lehman Crisis in Sub-prime Crisis Period [42], Enron Crisis, and 1997 Asian Financial Crisis), and the entropy values usually lead to a rapid decrease even many days before the significant financial event. In other words, each significant fluctuation of the dominant Shannon entropy values corresponds to a financial crisis, and provides early warning before the crisis occurs. The reason for the effectiveness is that the financial networks are constructed by computing the correlation between pairwise stock time series and the dominant Shannon entropy is computed based on the dominant correlated time series subset that is identified through the commute time matrix. Based on the financial risk theory stated by [30], the financial crisis is usually caused by a set of the most correlated stocks having less uncertainties. Thus, the dominant Shannon entropy, that characterizes the the dominant correlated stocks, tends to significantly drop down before a financial crisis. The experiments demonstrates that the proposed dominant entropy time series through the commute time can capture significant financial information, satisfying financial theory.

Note that, although Fig.4 indicates that the dominant Shannon entropy is effective for identifying the extreme financial events in the evolution of the time-varying financial networks, the entropy measure can only represent the network characteristics in an one-dimensional pattern space and thus ignores the information regarding specific changes in the network structure. In contrast, our proposed kernel associated with the dominant entropy is able to map the network structures into a high dimensional Hilbert space via kernelizing the entropies. In other words, our kernels should better preserve structural information contained in the network structures.
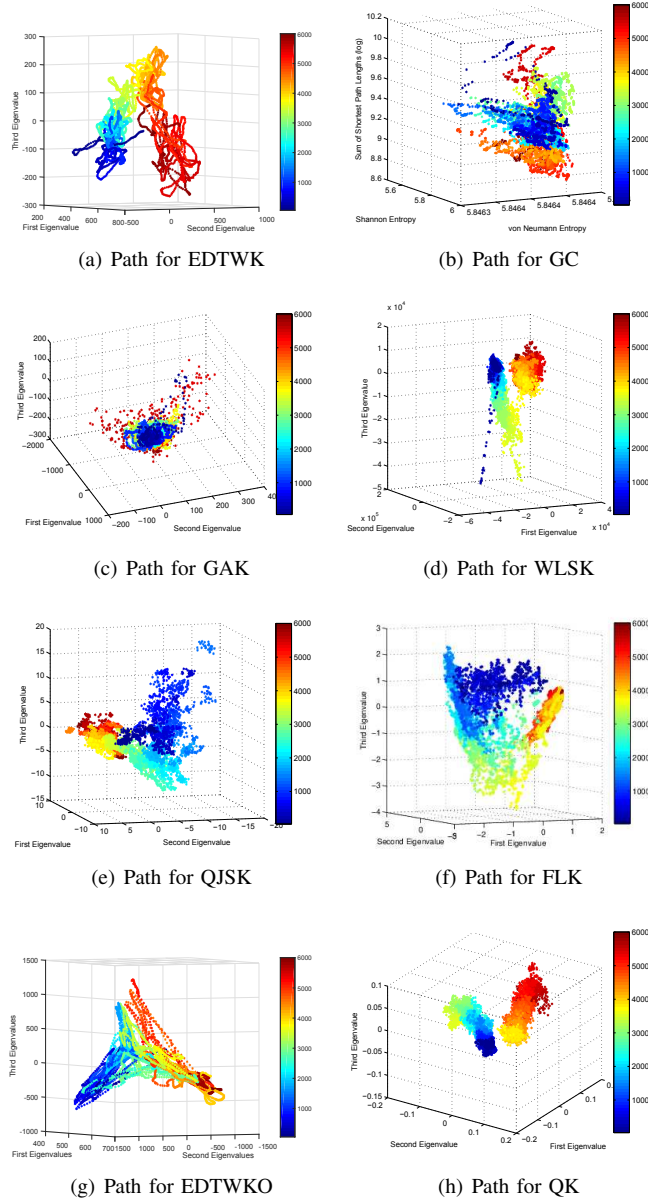
(a) Path for EDTWK

(b) Path for GC

(c) Path for GAK

(d) Path for WLSK

(e) Path for QJSK

(f) Path for FLK

(g) Path for EDTWKO

(h) Path for QK

Fig. 5. Path of financial networks over all trading days.

B. Kernel Embeddings of Financial Networks from kPCA

In this subsection, we evaluate the performance of the proposed EDTWK kernel on time-varying networks of the NYSE dataset, and explore whether the proposed kernel can distinguish the network evolution with time. Specifically, associated with the proposed kernel, we perform kernel Principle Component Analysis (kPCA) [43] on the kernel matrix of the financial networks and embed the networks in a vector space. We visualize the embedding results through the first three principal components in Fig.5(a). In addition, we compare the proposed kernel to three classical graph characterization methods (GC), that is, the Shannon entropy associated with the classical steady state random walk [44], the von Neumann entropy associated with the normalized Laplacian matrix [45], and the average length of the shortest path over all pairs of vertices [6]. All the three network characterization methods can accommodate complete weighted graphs. The visualization spanned by the three graph characterizations are shown

TABLE I
THE DISTANCE STRESS OF THE NETWORK EMBEDDINGS

Methods	EDTWK	GC	GAK	WLSK
Distance Stress	1.1931	5.5215	2.9677	4.0045
Methods	QJSK	FLK	EDTWKO	QK
Distance Stress	2.7220	2.9297	2.133	2.9167

in Fig.5(b). Finally, we also compare the proposed kernel with four state-of-the-art kernel methods, that is, the dynamic time warping inspired global alignment kernel (GAK) for original time series [33] and three graph kernels for time-varying financial networks. Specifically, the graph kernels include the Weisfeiler-Lehman subtree kernel (WLSK) [35], the quantum Jensen-Shannon kernel (QJSK) [44], and the feature space Laplacian graph kernel (FLGK) [46]. For the GAK kernel, we also adopt a time window of 28 days for each trading day. For the WLSK kernel, since it cannot accommodate a complete weighted graph and the edge weight, we transform each original network into a minimum spanning tree and ignore the edge weights. Note that, although one can also adopt the aforementioned threshold-based strategy [6], [11] to abstracted sparse network structures for the WLSK kernel, it has been shown that the minimum spanning tree abstracted from the commute time matrix has more advantages than that abstracted from the original adjacency matrix [32]. Thus, in this work we do not consider the threshold-based strategy for the WLSK kernel. For the QJSK and FLGK kernels, since they can accommodate edge weights, we directly apply these kernels to the original financial networks. For the four kernel methods, we also perform kPCA on the resulting kernel matrices and embed the original time series or the time-varying networks into a vector space, and the embedding results are exhibited in Fig.5(c), 5(d), 5(e) and 5(f), respectively.

As we previously stated, the commute time matrix computed from the original absolute Pearson correlation based adjacency matrix of each financial network can be seen as the enhanced Pearson correlation matrix for the proposed DETWK kernel. Thus, the commute time matrix plays a significant role to determine the effectiveness of the proposed kernel. To further validate its effectiveness, we also compare the proposed DE-TWK kernel associated with commute time matrices with that associated with the original weighted adjacency matrices (DE-TWKO), and the result is displayed in Fig 5(g). In addition, we compare the proposed kernel with the quantum-inspired kernel (QK) [32], since the QK kernel can also accommodate the original Pearson correlation based adjacency matrices of financial networks through the commute time matrices. The result of the QK kernel is exhibited in Fig 5(h).

Fig.5 exhibits the traces of the time-varying financial networks (or the original time series) in the different kernel spaces together with the classical graph characterization pattern space over all trading days. The color bar of each subfigure indicates the specific date over time. We observe that the embedding from the proposed EDTWK kernel exhibits a better manifold structure. Moreover, the embedding resulting from the EDTWK, WLSK and QK kernels are better distributed than those defined by the remaining approaches for comparisons. Another interesting phenomenon exhibited in Fig.5 is that only the proposed EDTWK kernel produces a clear time-
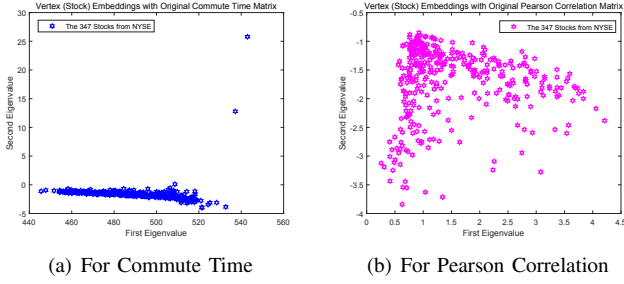
(a) For Commute Time   (b) For Pearson Correlation

Fig. 6. Vertex (Stock) Embedding Comparisons.

(a) Black Monday for EDTWK   (b) Black Monday for GC

(c) Black Monday for GAK   (d) Black Monday for WLSK

(e) Black Monday for QJSK   (f) Black Monday for FLK

(g) Black Monday for EDTWKO   (h) Black Monday for QK

Fig. 7. The 3D embeddings of Black Monday.

varying trajectory associated with the financial networks of consecutive time steps, i.e., the embedding of each time-varying financial network on the current time step is closely near to that on the last time step in the embedding space. By contrast, the alternative methods hardly result in such a trajectory and the associated embeddings tend to distribute as clusters. To further demonstrate this effectiveness of the proposed EDTWK kernel, we compute the distance stress of the financial network embeddings based on different methods. Specifically, the distance stress DS is defined as

$$\mathrm{DS} = \frac{\sum_t \parallel x_t - x_{t-1} \parallel^2}{\sum_t \parallel x_t - x_{\mathrm{tn}} \parallel^2}, \qquad (13)$$

where $t = 2, 3, \ldots, n$, x_t is the network embedding vector at time step t, and x_{tn} is the nearest network embedding vector of x_t. For each embedding vector x_t at time step t, if the nearest embedding vector is always the embedding vector at last time step (i.e., x_{t-1}), the value of the distance stress **DS** will be 1. In other words, the distance stress value nearer to 1 indicates the better performance of the embeddings to form a clear time-varying trajectory. The distance stress values of the financial network embeddings based on different methods are show in Table I. It is clear that only the distance stress value of the proposed EDTWK kernel is nearer to 1. This further indicates that the proposed method can better preserve the ordinal arrangement of the time-varying financial networks.

Finally, although the EDTWK kernel (i.e., the proposed kernel associated with the original weighted adjacency matrix) can also form a clear time-varying trajectory, its embedding points of the financial networks are not well distributed. This reveals that the proposed approach associated with the commute time matrix has better performance than that associated with the original absolute Pearson correlation matrix, demonstrating the effectiveness of the commute time integrated in the proposed approach. To further reveal the effectiveness of adopting the commute time matrix, we randomly select a financial network. For this network we perform multidimensional on both its commute time matrix and the Pearson correlation matrix, to embed each vertex (i.e., the 347 stocks from NYSE) into a 2-dimensional pattern space. Suppose the affinity matrix in this question is C (i.e., the commute time or the Pearson correlation matrix), then the centred similarity matrix is given by

$$K = -1/2(I - J/n)C(I - J/n), \qquad (14)$$

where n is the number of vertices, I is the $n \times n$ identity matrix, and J is the $n \times n$ all-ones matrix. If $K = \Phi\Lambda\Phi^T$ is the eigendecompositon of the kernel matrix in terms of
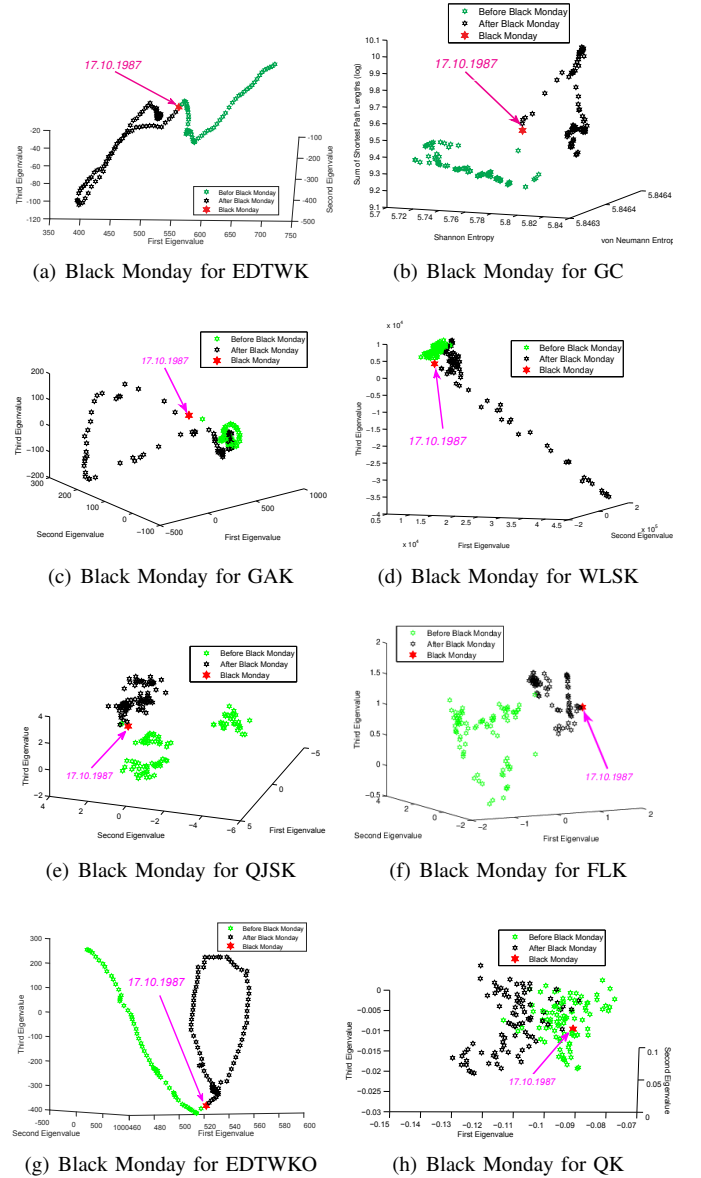
the diagonal matrix of ordered eigenvalues Λ and the correspondingly ordered matrix of column eigenevtors Φ, then the matrix with the embedding co-ordinates as column vectors is $X = \sqrt{\Lambda}$. Here we discard rows corresponding to negative eigenvalues and take the leading two rows. The results are shown in Fig.6. It is interesting that the stock embedding points through the commute time matrix distribute well and form an approximately linear manifold structure. By contrast, the stock embedding points through the Pearson correlation matrix disperse over a larger volume of the space. Note that we will observe similar results for most of our financial networks. This reveals that when compared to the Pearson correlation matrix, the commute time based vertex affinity matrix offers the advantage of capturing more reliable relationships between the stocks which reside in an approximately linear subspace.

Overall, the above observations indicates that the proposed EDTWK kernel associated with the commute time matrix can better understand the structural evolutions of the financial
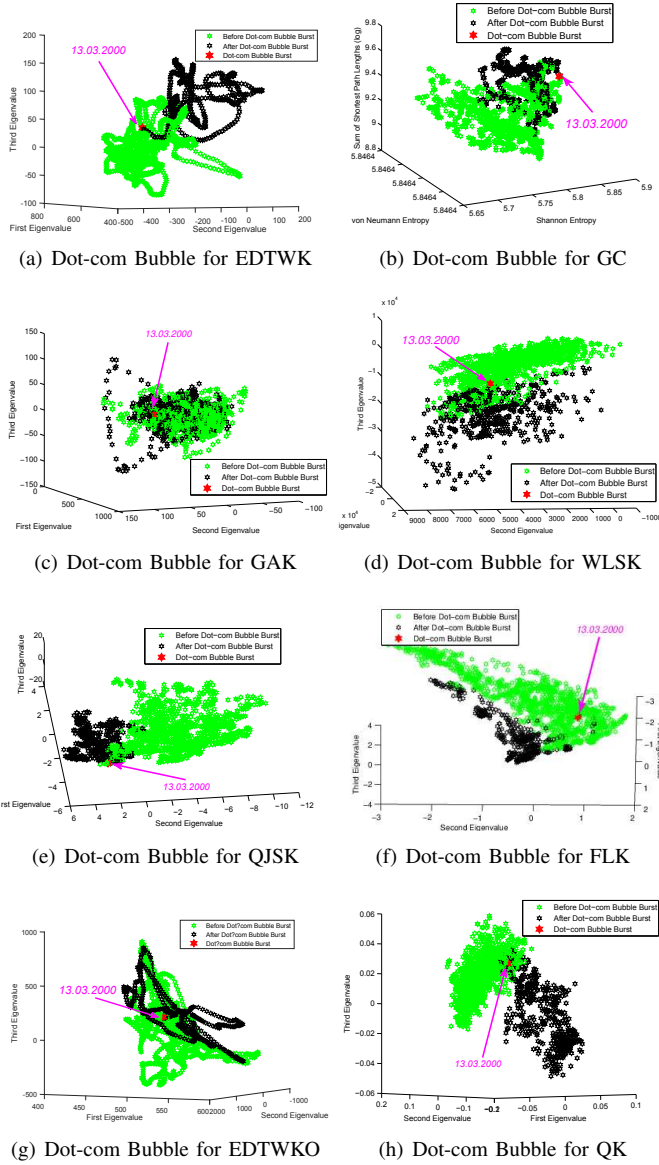
(a) Dot-com Bubble for EDTWK (b) Dot-com Bubble for GC

(c) Dot-com Bubble for GAK (d) Dot-com Bubble for WLSK

(e) Dot-com Bubble for QJSK (f) Dot-com Bubble for FLK

(g) Dot-com Bubble for EDTWKO (h) Dot-com Bubble for QK

Fig. 8. The 3D embedding of Dot-com Bubble Burst.

networks with time than the remaining methods.

(a) Enron Incident for EDTWK (b) Enron Incident for GC

(c) Enron Incident for GAK (d) Enron Incident for WLSK

(e) Enron Incident for QJSK (f) Enron Incident for FLK

(g) Enron Incident for EDTWKO (h) Enron Incident for QK

Fig. 9. The 3D embedding of Enron Incident.

C. Kernel Embeddings for Financial Crisis Analysis

To take our study one step further, in this subsection we exhibit more details of the kernel embedding results for three different financial crisis periods that have been preliminarily explored in Section IV-A. Specifically, for differen methods, Fig.7 corresponds to the Black Monday period (*from 15th June 1987 to 17th February 1988*), Fig.8 corresponds to the Dot-com Bubble period (*from 3rd January 1995 to 31st December 2001*), and Fig.9 corresponds to the Enron Incident period (*the red points, from 16th October 2001 to 11th March 2002*). These figures show that the Black Monday (*17th October, 1987*), the Dot-com Bubble Burst (*13rd March, 2000*) and the Enron Incident period (*from 2nd December 2001 to 11th March 2002*) are all extreme financial events. The embedding points of the proposed EDTWK kernel before and after these events can be better separated into independent clusters, and the points representing the extreme
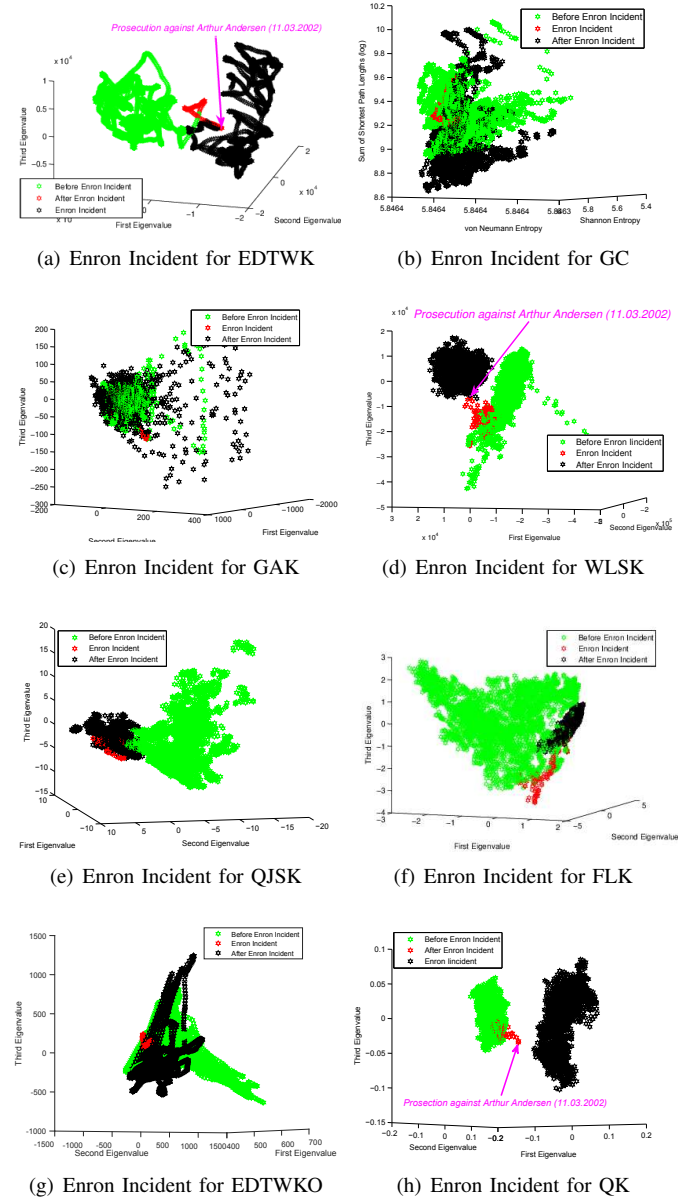
financial events are in the middle of the corresponding clusters. Another interesting phenomenon from Fig.9 is that the embedding points of the networks between year 1986 and year 2011 are distinctly divided by the Prosecution against Arthur Andersen (*3rd November, 2002*). Since the prosecution symbolizes the end of the Enron Incident, the Enron Incident can be considered as a watershed at the beginning of the 21st century, which significantly distinguishes the financial networks of the 21st century from those of the 20th century. These observations again indicate that the proposed kernel can well understand and detect the abrupt financial incident significantly changing the network structures. Although some methods are competitive, only the proposed kernel produces a clearer trajectory in terms of the embedding distributions, i.e., the proposed kernel can better reflect the time-varying transition with time. Note that, we can observe similar results if we explore the proposed kernel and the remaining methods on the alternative financial crisis mentioned in Section IV-A.

(a) Newcentury for EDTWK

(b) Newcentury for WLSK

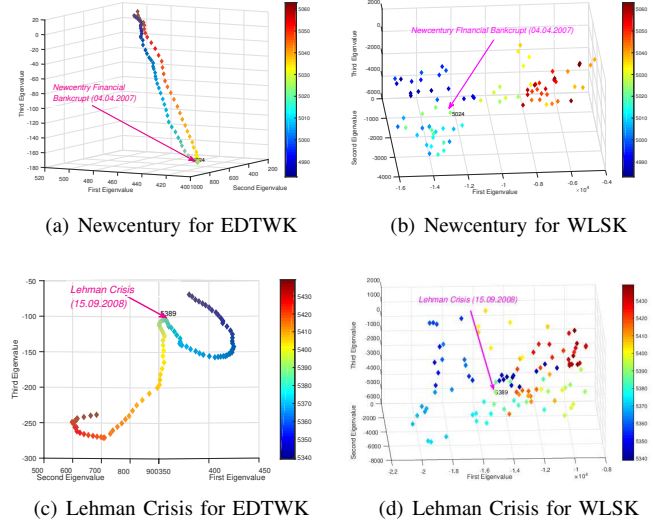(c) Lehman Crisis for EDTWK

(d) Lehman Crisis for WLSK

Fig. 10. Time-varying Trajectory Comparisons of the EDTWK and WLSK Kernels for Sub-prime Crisis.

The above experimental results indicate that the WLSK kernel as well as the QK kernel are the most competitive alternative methods to the proposed EDTWK kernel. Since the embedding results of these two kernels can also be better divided before and after a financial event. To further demonstrate the effectiveness of the proposed kernel, we compare the three methods on two financial events happened in Sub-Prime Crisis period (*from 2nd January 2006 to 1st July 2009*) in more details. The financial events for comparisons are the Newcentury Financial Bankruptcy (*4th April 2007*) and the Lehman Brothers Bankruptcy (*15th September 2008*), both having significant impacts in the world finance history. Specifically, for each method, we visualize the set of points that indicate the path of the kPCA embeddings with time over about 90 trading days (i.e., 90 points) around each of the two events. The results are shown in Fig.10, and the colour bar beside each subfigure represents the data in time series. Note that, we only show the result from the WLSK kernel, since we will observe the similar phenomenon from both the WLSK and QK kernel. For the proposed kernel, we observe that the point distribution forms a clear trajectory with time, and the trajectory around each of the financial events usually undergoes significant changes, i.e., the trajectory starting from the financial event will tempestuously change the distribution direction during a short time period. By contrast, the point distributions from the WLSK kernel are chaos and cannot observe any significant phenomenon. This demonstrates that only the proposed kernel has better capability to both characterize and distinguish financial crises.

D. Evaluations of the Kernel Matrix

To further reveal the effectiveness of the proposed EDTWK kernel, in this subsection we also draw the kernel matrices of both the proposed EDTWK kernel and the competitive alternative WLSK kernel. Note that, for another competitive QK kernel, we can observe the similar phenomenon with the WLSK kernel, thus we only exhibit the experiment with the WLSK kernel. The kernel matrices are computed between the

(a) For EDTWK on All Days

(b) For WLSK on All Days

(c) For EDTWK on Newcentry

(d) For WLSK on Newcentry
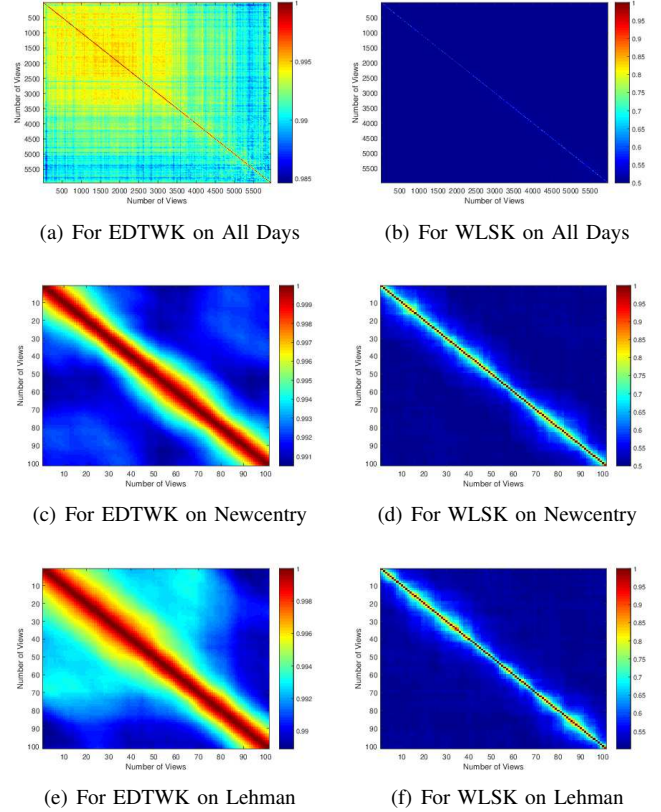

(e) For EDTWK on Lehman

(f) For WLSK on Lehman

Fig. 11. Kernel Matrices Visualizations for EDTWK and WLSK Kernels.

networks belonging to the Newcentury crisis period and the Lehman crisis period, as well as that belonging to all the 6004 trading dats of the NYSE dataset. The kernel matrix visualization results are shown in Fig.11, where both the x-axis and y-axis represent the time steps. Note that, to compare the two kernels in the same scaled Hilbert space, we consider the normalized version of both the kernels as

$$k_n(G_p, G_q) = \frac{k(G_p, G_q)}{\sqrt{k(G_p, G_p)k(G_q, G_q)}},$$

where k_n is the normalized kernel, and k is either the EDTWK kernel or the WLSK kernel. As a result, the kernel values are all bounded between 0 to 1, and the colour bar beside each subfigure of Fig.11 represents the kernel value of the kernel matrix. Through Fig.11, we observe that the kernel values tend to decrease when the elements of the kernel matrix are far away from the trace of the matrix. This is because such elements of the kernel matrix are computed between time-varying financial networks having long time spans, there are more structure changes when the network evolves with a long time variation. Thus, both the EDTWK and the WLSK kernels can reflect the structural evolution of the financial networks with time. However, on the other hand, we observe that the kernel values associated with the WLSK kernel tend to suddenly drop down when the element is a little far from the trace. By contrast, the kernel values associated with the proposed EDTWK kernel tend to gradually decrease when the element gets far away from the trace. This observation reveal the reason why only the kPCA embeddings through the proposed EDTWK kernel can form clear trajectory with time

TABLE II
CLASSIFYING TIME-VARYING NETWORKS THROUGH DIFFERENT KERNELS ASSOCIATED WITH C-SVMS.

Datasets	Black Monday	Newcentury	Lehman	Asia 1997	Dot-com
EDTWK	**91.47** $\pm$ 0.54	**88.79** $\pm$ 0.63	**88.00** $\pm$ 0.65	**90.82** $\pm$ 0.57	**91.03** $\pm$ 0.61
QJSK	89.71 $\pm$ 0.78	88.32 $\pm$ 0.74	87.85 $\pm$ 0.81	89.88 $\pm$ 1.11	89.17 $\pm$ 0.72
WLSK	88.70 $\pm$ 0.51	88.30 $\pm$ 0.80	86.36 $\pm$ 0.88	85.44 $\pm$ 0.91	88.03 $\pm$ 0.71
QK	90.10 $\pm$ 0.51	88.50 $\pm$ 0.80	87.69 $\pm$ 0.88	90.21 $\pm$ 0.91	89.79 $\pm$ 0.71

variation in Fig.10. Therefore, only the proposed EDTWK kernel can well distinguish and understand the structural changes of the network structures evolving with a long time period. Note that, we can observe similar results if we explore the proposed kernel and the remaining methods on the alternative financial crisis mentioned in Section IV-A.

E. Classifications from C-SVMs

The above experiments have demonstrated that the proposed EDTWK kernel can better discriminate the global structural changes of the dynamic time-varying financial networks before and after a financial event, over the entire financial crisis period. On the other hand, we show that only the kernel embeddings associated with the proposed EDTWK kernel can form a clear trajectory in terms of the embedding distributions with time, i.e., the proposed EDTWK kernel can better preserve the ordinal arrangement of the time-varying financial networks. Intuitively, this observation indicates that the proposed kernel may have a good ability to understand the local structural changes of the time-varying network gradually evolved with time, i.e., the structure of the time-varying financial network tends to gradually change with time.

To demonstrate this, we validate the effectiveness of the proposed EDTWK kernel on classification tasks. Specifically, we explore whether the proposed kernel can be used to correctly classify the time-varying financial networks into corresponding short-term stages during each specific financial crisis period. The crisis periods for evaluations include the Black Monday period, the Dot-com Bubble period, the Newcentury Financial crisis period, the Lehman Crisis period and the 1997 Asia Financial crisis period. Since the Enron crisis is not an emergency crisis and continued for many trading days, we do not perform the classification evaluation on this crisis. For each of these selected financial crisis periods, we utilize 100 continuous trading days around the particular day when the financial event happens, i.e., we respectively select 50 days before and after the crisis event. For each crisis, we sequentially divide the 100 continuous trading days into 10 short-term stages and each stage contains 10 continuous trading days, i.e., the financial networks of the 100 trading days are separated into 10 classes. For each financial crisis period, we calculate the kernel matrix between the financial networks of the trading days. Moreover, for the proposed kernel, we perform 10-fold cross-validation using the C-Support Vector Machine (C-SVM) to compute the classification accuracies, using LIBSVM [47]. We use nine samples for training and one for testing. All the C-SVMs were performed along with their parameters optimized on each dataset. We repeat the whole experiment 10 times and report the average classification accuracies and standard errors in Table.II. We also compare the proposed EDTWK kernel with the competitive WLSK and QK kernels [35], [32], as well as

the QJSK kernel [44], and the evaluations associated with these kernels follow the same experimental setup of the proposed kernel. The experiment results are also reported in Table.II. It is clear that the proposed EDTWK kernel outperforms the alternative graph kernels, and the proposed kernel can better classify each financial network into a correct local stage. This evaluation indicates that the proposed kernel not only captures the main common characteristics of the financial network evolved within a short-term stage, but also discriminates the main changes of that evolved over a long time period across more than one stages, demonstrating the better ability of the proposed kernel to understand how the structure of the financial network gradually evolve with time.

Note that, in this evaluation one can also classify the time-varying financial networks from each crisis period into other number of classes, e.g., 2 classes, 3 classes, etc. If the number of the classes is 2, the classification evaluation tends to explore the global discriminative ability of the financial network changes before and after a specific financial event. The more classes we divided the financial networks into, the more the classification evaluation focuses on the local discriminative ability of the network changes with time during an entire financial crisis period. The proposed kernel tends to significantly outperform the remaining kernels, when we divide the financial networks into more classes. However, it is not significant to divide the financial networks into more than 20 classes, since less than 5 financial networks of each local stage are not enough to represent the main common characteristics of such a stage.

F. Discussions of the Experimental Advantages

All the above experiments demonstrate the effectiveness of the proposed kernel. The reasons for the effectiveness are fourfold. First, unlike the WLSK kernel, the proposed kernel can directly accommodate the time-varying networks that are complete weighted graphs. Second, unlike the GC method that computes vectorial network characteristics and tends to approximate the time-varying networks in a low dimensional pattern space, the proposed kernel can represent the network structures in a high dimensional Hilbert space and thus better preserve the network characteristics. Third, unlike the GAK for original time series, the proposed kernel for time-varying networks is based on the dominant entropy time series and reflects richer correlated information than the original time series. Fourth, unlike the QJSK and FLK kernels, only the proposed kernel can capture reliable financial information through the dominant entropy time series. Overall, the proposed kernel provides an effective way to analyze the financial crisis associated with the dynamic time-varying financial networks.

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a new dynamic time warping framework inspired kernel, namely the Entropic Dynamic Time Warping Kernels between time-varying financial networks for multiple co-evolving financial time series analysis. Specifically, for a family of time-varying financial networks with each vertex representing the individual time series of a stock and each edge between pairwise series representing the correlation, we have computed the commute time matrix on each of the network structures and shown how this matrix allows us to identify a dominant correlated stock set as well as the associated dominant probability distribution of these stocks belonging to this set. Based on the probability distribution, we have represented each original network as dominant Shannon entropy time series. With the dominant entropy time series for each pair of financial networks to hand, the proposed kernel has been defined through the dynamic time warping based global alignment kernel between the entropy time series. We have shown that the proposed kernel bridges the gap between graph kernels and the classical dynamic time warping framework for time series analysis. Experiments on time-varying networks extracted from New York Stock Exchange (NYSE) database demonstrate the effectiveness.

In this work, we have identified the most strongly correlated stock trading patterns based on the compute-time between pairs of stocks in a network of trading relationships. Commute time implicity averages over all paths connecting a pair of stocks in the network, and not just the first-order nearest neighbour relations. This renders it robust to missed or erroneously inferred correlation relations in the time series of the stock closing prices. In the future, we will explore the use of hypergraph representations employing the relationships between multiple stocks. Here we will use the commute time to define the groups of stock and develop a hypergraph kernel to measure the relationships between groups.

ACKNOWLEDGMENTS

REFERENCES

[1] H. Zhang, Z. Wang, and D. Liu, "A comprehensive review of stability analysis of continuous-time recurrent neural networks," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 25, no. 7, pp. 1229–1262, 2014.

[2] Q. Zhang and Q. Yao, "Dynamic uncertain causality graph for knowledge representation and reasoning: Utilization of statistical data and domain knowledge in complex cases," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 29, no. 5, pp. 1637–1651, 2018.

[3] Y. Cao, Y. Li, S. A. Coleman, A. Belatreche, and T. M. McGinnity, "Detecting wash trade in financial market using digraphs and dynamic programming," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 27, no. 11, pp. 2351–2363, 2016.

[4] Y. Yin, P. Shang, and J. Xia, "Compositional segmentation of time series in the financial markets," *Applied Mathematics and Computation*, vol. 268, pp. 399–412, 2015.

[5] S. Liu, M. Yamada, N. Collier, and M. Sugiyama, "Change-point detection in time-series data by relative density-ratio estimation," *Neural Networks*, vol. 43, pp. 72–83, 2013.

[6] F. N. Silva, C. H. Comin, T. K. Peron, F. A. Rodrigues, C. Ye, R. C. Wilson, E. R. Hancock, and L. da F. Costa, "Modular dynamics of financial market networks," *arXiv preprint arXiv:1501.05040*, 2015.

[7] H. Bunke, P. J. Dickinson, A. Humm, C. Irniger, and M. Kraetzl, "Graph sequence visualisation and its application to computer network monitoring and abnormal event detection," in *Applied Graph Theory in Computer Vision and Pattern Recognition*, 2007, pp. 227–245.

[8] J. Zhang and M. Small, "Complex network from pseudoperiodic time series: Topology versus dynamics," *Physical Review Letters*, vol. 96, p. 238701, 2006.

[9] G. Nicolis, A. G. Cantu, and C. Nicolis, "Dynamical aspects of interaction networks," *International Journal of Bifurcation and Chaos*, vol. 15, p. 3467, 2005.

[10] Y. Shimada, T. Kimura, and T. Ikeguchi, "Analysis of chaotic dynamics using measures of the complex network theory," in *Proceedings of ICANN*, 2008, pp. 61–70.

[11] C. Ye, C. H. Comin, T. K. Peron, F. N. Silva, F. A. Rodrigues, L. da F. Costa, A. Torsello, and E. R. Hancock, "Thermodynamic characterization of networks using graph polynomials," *Physical Review E*, vol. 92, no. 3, p. 032810, 2015.

[12] L. Bai, Z. Zhang, C. Wang, X. Bai, and E. R. Hancock, "A graph kernel based on the jensen-shannon representation alignment," in *Proceedings of IJCAI*, 2015, pp. 3322–3328.

[13] G. D. S. Martino, N. Navarin, and A. Sperduti, "Tree-based kernel for graphs with continuous attributes," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 29, no. 7, pp. 3270–3276, 2018.

[14] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.

[15] D. P. Feldman and J. P. Crutchfield, "Measures of statistical complexity: Why?" *Physics Letters A*, vol. 238, no. 4, pp. 244–252, 1998.

[16] K. Anand, G. Bianconi, and S. Severini, "Shannon and von neumann entropy of random networks with heterogeneous expected degree," *Physical Review E*, vol. 83, no. 3, p. 036109, 2011.

[17] K. Anand, D. Krioukov, and G. Bianconi, "Entropy distribution and condensation in random networks with a given degree distribution," *Physical Review E*, vol. 89, no. 6, p. 062807, 2014.

[18] K. Huang, *Statistical Mechanic*. Wiley, New York, 1987.

[19] M. A. Javarone and G. Armano, "Quantum-classical transitions in complex networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2013, no. 04, p. 04019, 2013.

[20] J.-C. Delvenne and A.-S. Libert, "Centrality measures and thermodynamic formalism for complex networks," *Physical Review E*, vol. 83, no. 4, p. 046117, 2011.

[21] A. Fronczak, P. Fronczak, and J. A. Hołyst, "Thermodynamic forces, flows, and onsager coefficients in complex networks," *Physical Review E*, vol. 76, no. 6, p. 061106, 2007.

[22] D. Haussler, "Convolution kernels on discrete structures," in *Technical Report UCS-CRL-99-10*, Santa Cruz, CA, USA, 1999.

[23] N. Shervashidze, P. Schweitzer, E. J. van Leeuwen, K. Mehlhorn, and K. M. Borgwardt, "Weisfeiler-lehman graph kernels," *Journal of Machine Learning Research*, vol. 1, pp. 1–48, 2010.

[24] L. Bai, L. Rossi, Z. Zhang, and E. R. Hancock, "An aligned subtree kernel for weighted graphs," in *Proceedings of ICML*, 2015, pp. 30–39.

[25] L. Bai, L. Rossi, H. Bunke, and E. R. Hancock, "Attributed graph kernels using the jensen-tsallis q-differences," in *Proceedings of ECML-PKDD*, 2014, pp. 99–114.

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

[27] G. Nikolentzos, P. Meladianos, S. Limnios, and M. Vazirgiannis, "A degeneracy framework for graph similarity," in *Proceedings of IJCAI*, 2018, pp. 2595–2601.

[28] H. Kashima, K. Tsuda, and A. Inokuchi, "Marginalized kernels between labeled graphs," in *Proceedings of ICML*, 2003, pp. 321–328.

[29] J. Wang, R. C. Wilson, and E. R. Hancock, "Directed and undirected network evolution from eulerclagrange dynamics," To Appear.

[30] J. G. Haubrich and A. W. Lo, *Quantifying Systemic Risk*. The University of Chicago Press, 2013.

[31] H. Qiu and E. R. Hancock, "Clustering and embedding using commute times," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1873–1890, 2007.

[32] L. Bai, L. Rossi, L. Cui, J. Cheng, Y. Wang, and E. R. Hancock, "A quantum-inspired similarity measure for the analysis of complete weighted graphs," *IEEE Transactions on Cybernetics*, vol. 50, no. 3, pp. 1264 – 1277, 2020.

[33] M. Cuturi, "Fast global alignment kernels," in *Proceedings of ICML*, 2011, pp. 929–936.

[34] L. Bai, L. Cui, L. Rossi, L. Xu, X. Bai, and E. R. Hancock, "Local-global nested graph kernels using nested complexity traces," in *Pattern Recognition Letters*, To Appear.

[35] N. Shervashidze, S. Vishwanathan, K. M. T. Petri, and K. M. Borgwardt, "Efficient graphlet kernels for large graph comparison," *Journal of Machine Learning Research*, vol. 5, pp. 488–495, 2009.

[36] F. D. Johansson, V. Jethava, D. P. Dubhashi, and C. Bhattacharyya, "Global graph kernels using geometric embeddings," in *Proceedings of ICML*, 2014, pp. 694–702.

[37] M. Pavan and M. Pelillo, "Dominant sets and pairwise clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 167–172, 2007.

[38] L. Bai and E. R. Hancock, "Graph kernels from the jensen-shannon divergence," *Journal of Mathematical Imaging and Vision*, vol. 47, no. 1-2, pp. 60–69, 2013.

[39] K. Ø. Mikalsen, F. M. Bianchi, C. Soguero-Ruíz, and R. Jenssen, "Time series cluster kernel for learning similarities between multivariate time series with missing data," *Pattern Recognit.*, vol. 76, pp. 569–581, 2018.

[40] E. S. Browning, "Exorcising ghosts of octobers past," *The Wall Street Journal*, pp. C1–C2, 2007.

[41] K. Anderson, C. Brooks, and A. Katsaris, "Speculative bubbles in the s&p 500: Was the tech bubble confined to the tech sector?" *Journal of empirical finance*, vol. 17, no. 3, pp. 345–361, 2010.

[42] C. Mollenkamp, S. Craig, S. Ng, and A. Lucchetti, "Lehman files for bankruptcy, merrill sold, aig seeks cash," *The Wall Street Journal*, 2008.

[43] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2011.

[44] L. Bai, L. Rossi, A. Torsello, and E. R. Hancock, "A quantum jensen-shannon graph kernel for unattributed graphs," *Pattern Recognition*, vol. 48, no. 2, pp. 344–355, 2015.

[45] M. Dehmer and A. Mowshowitz, "A history of graph entropy measures," *Information Science*, vol. 181, no. 1, pp. 57–78, 2011.

[46] R. Kondor and H. Pan, "The multiscale laplacian graph kernel," in *Proceedings of NIPS*, 2016, pp. 2982–2990.

[47] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm*, 2011.