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

Lu Bai\textsuperscript{a, b}, Lixin Cui\textsuperscript{c, d}, Zihong Zhang\textsuperscript{a, e}, Lixiang Xu\textsuperscript{f}, Yue Wang\textsuperscript{g} and Edwin R. Hancock\textsuperscript{a, h, i, Fellow, IEEE}

Abstract—Network representations are powerful tools to modeling the dynamic time-varying financial complex systems consisting of multiple co-evolving financial time series, e.g., stock prices. 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 (CT) 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 CT 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 effectiveness of the proposed method.

I. INTRODUCTION

THERE has been a surge of research interest aimed at using machine learning algorithms for the analysis of dynamic networks. This is due to their potential applications in many practical systems, such as social systems, neural networks [1], uncertain systems [2], and financial market systems [3]. In this article, we focus in detail on financial markets that can be considered as a complex time-varying system consisting of multiple interacting financial components [4], e.g., the stock trade price. Due to the evolution of these financial variables with time, the system can be represented as multiple co-evolving financial time series generated from the original data.

The analysis of market crises or risk in the time-varying financial markets involves change point detection. This process has played a critical role to identify abrupt changes in time series properties [5]. Unfortunately, for a system consisting of complex interactions between constituent co-evolving time series detecting such crucial points remains challenging. The reason for this is that it is difficult to detect the changes since they cannot be easily observed [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], [8]. 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 the past \( W \) time steps ending at time step \( t \)). Each edge between a pair of co-evolving financial time series indicates the strength of their interactions (e.g., the absolute Pearson correlation or the Euclidean distance). 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], [9]–[12]. Under this scenario, network-based machine learning approaches can be directly employed for the structural analysis of financial risk, e.g.: 1) graph embedding methods [12]–[14]; and 2) graph kernels [15], [16].

In machine learning, graph kernels are usually defined in terms of similarity measures between graphs [17]. Unlike graph embedding algorithms that tend to approximate structural correlations of graphs in a low-dimensional pattern space and thus lead to information loss, graph kernels are effective approaches for characterizing graph structures in a high-dimensional Hilbert space [18] for structure analysis [15], [16], [19]. The aim of this article 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 a discrete entropy time series and using the classical dynamic time
warped 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 [6], [9]–[12], 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 [20] 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 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 [12]. 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: 1) derive network characteristics using connectivity structures, or statistics capturing connectivity structures and 2) characterize the networks using statistical physics. Proponents of the former approach focus on capturing network substructures using communities, hubs and clusters [21]–[23]. 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 [12], [24]–[27]. 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 [17]. 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: 1) walks; 2) subgraphs; and 3) subtrees. Representative R-convolution graph kernels based on substructures include the Weisfeiler–Lehman subtree kernel [28], the tree-based continuous attributed kernel [15], the aligned subtree kernel [29], the Jensen–Tsallis q-difference graph kernel [30], the optima assignment Weisfeiler–Lehman kernel [31], the core variants-based shortest path kernel [32], and the random walk graph kernel [33]. 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 preserve only the original edges indicating pairs of more interacted vertices. Under this scenario, Ye et al. [12], Silva et al. [6], and Wang et al. [34] 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 understand this weakness, consider the example of financial network shown in Fig. 2. Here, each vertex represents a stock and each edge corresponds to the correlation value between a pair of stocks. In the example, one can only preserve two 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 challenging.

B. Contributions

The objective of this article 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 [35], 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 [6], [12], 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 main objective is to compute the commute time (CT) matrices associated with their original weighted adjacency matrices, i.e., the absolute Pearson correlation-based matrices. The reason of using the CT 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. Thus, the CT 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 CT is robust under the perturbation of the network structure (e.g., the changes of edges or paths). As a result, the CT matrix can provide a more stable representation for the financial network structure that may accumulate a lot of noises over time. In summary, the CT 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 CT 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 CT 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 CT 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 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 CT 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. Article Outline

This article 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 article.

II. Preliminary Concepts

In this section, we briefly review preliminary concepts which will be utilized in this article. We first review the concept of the CT. 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. In order to capture a reliable and robust mutually correlated stock set, in this work, we propose to utilize the CT matrix associated with the original correlation matrix as the network representation.

The main reasons of employing the CT matrix are threefold. First, the CT 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 CT can be considered as the enhanced correlation matrix. Second, since the CT 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 CT-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 CT is calculated through the Laplacian matrix of the original correlation-based weighted adjacency matrix. In Section III, we will show how the CT 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 CT. 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 \( \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \) is a \( |V| \times |V| \) diagonal matrix with ascending eigenvalues as elements, i.e., \( 0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{|V|} \), and \( \Phi \) is a \( |V| \times |V| \) matrix \( \Phi = (\phi_1, \phi_2, \ldots, \phi_{|V|}) \) with the corresponding ordered eigenvectors as columns. For \( G \), the hitting time \( H(u, v) \) between each pair of vertices \( u \) and \( v \) is computed as the expected number of steps taken by a classical random walk commencing from \( u \) and ending at \( v \). Likewise, the CT \( C(u, v) \) 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 CT \( C(u, v) \) can be calculated through the unnormalized Laplacian eigenvalues and eigenvectors [36] as

\[
C(u, v) = \sum_{v \in V} D(u, u) \sum_{j=2}^{\lvert V \rvert} \frac{1}{\lambda_j} (\phi_j(u) - \phi_j(v))^2.
\]

**Remarks:** The CT has been proven to be a powerful tool to extract rich characteristics from complete weighted graphs. In previous works, Bai et al. [8] have employed the CT 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 CT matrix. For a pair of complete weighted graphs to be compared, the resulting quantum kernel is defined by measuring the similarity between their associated CT 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 [6], [12], [34]. This is because the weights of the preserved edges on spanning tree structures correspond to the CT values between corresponding pairs of vertices, and the CT values integrates the effectiveness over all possible paths residing on the original weighted edges. However, similar to these threshold-based approaches [6], [12], [34], the quantum kernel [8] 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 CT matrix.

### B. Dynamic Time Warping Framework

We review the global alignment kernel that is defined through the classical dynamic time warping framework [37]. Assume \( T \) is a set of discrete time series that takes values in a space \( \mathcal{X} \). For each pair of discrete time series \( P = (p_1, \ldots, p_m) \in T \) and \( Q = (q_1, \ldots, q_n) \in T \) with lengths \( m \) and \( n \) respectively, the alignment \( \pi \) between \( P \) and \( Q \) is computed as a pair of increasing integral vectors \( (\pi_p, \pi_q) \) 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 \( (\pi_p, \pi_q) \) is assumed to possess unitary increments and no simultaneous repetitions. For \( P \) and \( Q \), 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 \( \pi = (\pi_p, \pi_q) \), that is,

\[
\begin{bmatrix}
\pi_p(i+1) - \pi_p(i) \\
\pi_q(i+1) - \pi_q(i)
\end{bmatrix} \in \left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.
\]

Within the framework of the classical dynamic time warping [37], the coordinates \( \pi_p \) and \( \pi_q \) of the alignment \( \pi \) define the warping function. Assume that \( A(m, n) \) corresponds to a set of all possible alignments between \( P \) and \( Q \). Cuturi [37] has proposed a dynamic time warping inspired kernel, namely the Global Alignment Kernel, by considering all the possible alignments in \( A(m, n) \). The kernel is defined as

\[
k_{GA}(P, Q) = \sum_{\pi \in A(m, n)} e^{-D_{PQ}(\pi)}
\]

where \( D_{PQ}(\pi) \) is the alignment cost given by

\[
D_{PQ}(\pi) = \sum_{i=1}^{\lvert \pi \rvert} \phi(p_{\pi_p(i)}, q_{\pi_q(i)})
\]

and is defined through a local divergence \( \phi \) that quantifies the discrepancy between each pair of elements \( p_i \in P \) and \( q_i \in Q \). In general, \( \phi \) is defined as the squared Euclidean distance [37]. Note that, the kernel \( k_{GA} \) measures the quality of both the optimal alignment and all other alignments \( \pi \in A(m, n) \), thus it is positive definite. Moreover, \( k_{GA} \) provides richer statistical measures of similarity by encapsulating the overall spectrum of the alignment costs \( D_{PQ}(\pi), \pi \in A(m, n) \).

**Remarks:** The dynamic time warping-based global alignment kernel \( k_{GA} \) has been proven to be a powerful tool of analyzing vectorial time series [37]. To extend the effective
kernel $k_{GA}$ for time series analysis domain into the graph kernel domain. Bai et al. [38] 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 [33], [39], [40] 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. 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 CT 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 [37].

A. Identifying Dominant Correlated Time Series

We identify a set of the most mutually correlated time series for each time-varying financial network. Let $G = \{G_1, G_2, \ldots, G_T\}$ be a family of time-varying financial networks extracted from a complex system $S$, 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 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 $S$ has a fixing number of components co-evolving with time.

For each network $G_t$, we first compute its CT matrix as $C_t$ associated with its original absolute Pearson correlation-based adjacency matrix. As we have stated previously, the CT 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 CT matrix $C_t$ can be seen as a reliable enhanced absolute Pearson correlation matrix for $G_t$. In other words, the CT matrix provides a stable representation to further characterize the dynamic network $G_t$ associated with time-varying correlations between vertices.

With the CT 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. [41]. 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. See Fig. 2 for an example. 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 in red. As a result, the time series specified by DS can be 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 [41]. More specifically, associated with $C_t$, we compute the solution $a$ of the following quadratic program problem [41]:

$$ \arg \max_{a} \frac{1}{2} a^T C_t a $$

subject to $a \in \mathbb{R}^{|V_t|}$, $a_i \geq 0$ and $\sum_{i=1}^{|V_t|} a_i = 1$. The solution vector $a$ of (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 $a$. Based on the definition of Pavan and Pelillo [41], we can solve the local maximum of $f(a)$ by

$$ a_i(k + 1) = a_i(k) \frac{(C_t a(k))_i}{a(k)^T C_t a(k)} $$

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 $a$, all-time series represented by the vertices $v_1, \ldots, v_{|V_t|}$ fall into
two disjoint subsets, i.e.,

\[ S_1(a) = \{ v_i \in V_t \mid a_i > 0 \} \]

and

\[ S_2(a) = \{ v_j \in V_t \mid a_j = 0 \} \].

Clearly, the set \( S_1 \) 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 \( a \) also corresponds to a probability distribution of the time series belonging to the dominant set \( S_1 \), i.e., each element \( a_i \) corresponds to the probability of the \( i \)th time series belonging to \( S_1 \).

### B. 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 \( G \) at time step \( t \), we first compute the associated CT matrix \( C_t \). Moreover, by solving the quadratic program problem [41] on the CT matrix \( C_t \), we identify the set of dominant correlated time series \( S_1 \) and compute the associated probability distribution \( a \) of the time series belonging to \( S_1 \). Based on Section III-A, the remaining nondominant correlated time series are included in the set \( S_2 \). With the probability distribution \( a \) to hand, based on (42) we compute a dominant Shannon entropy for \( G_t \) as

\[
H_S(a) = -\sum_{i=1}^{|V_t|} a_i \log a_i
\]

where \( a_i \) is the probability of the \( i \)th time series represented by vertex \( v_i \in V_t \). Equation (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 subentropy \( H_S(v_i) \) of the \( i \)th time series represented by \( v_i \), that is,

\[
H_S(v_i) = -a_i \log a_i.
\]

Note that, if \( a_i = 0 \), we say that the \( i \)th time series does not belong to \( S_1 \) 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

\[
E_t = (H_S(v_1), \ldots, H_S(v_i), \ldots, H_S(v_{V_t}))^T
\]

where

\[
H_S(v_i) = \begin{cases} -a_i \log a_i, & a_i > 0, \ \text{i.e., } v_i \in S_1; \\ 0, & a_i = 0, \ \text{i.e., } v_i \in S_2. \end{cases}
\]

Equation (10) indicates that we only compute the sub-entropies for the dominant correlated time series in \( S_1 \), and do not consider the nondominant correlated time series in \( S_2 \).

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 \( S_t \) associated with its time interval \( s \) as

\[
S_t = \{ E_t-w+1 | E_t-w+2 | \ldots | E_t \} \]

where \( s \in [t-w+1, t-w+2, \ldots, t] \), and each column \( E_s \) of \( S_t \) is the entropy characteristics vector of each network \( G_t \) at time step \( s \) and is defined by (9). Clearly, the dominant entropy time series \( S_t \) 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 \( G_t \). An instance of characterizing the time-varying network \( G_t \) at time step \( t \) is illustrated in Fig. 3.

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

\[
S_p = \{ E_{p-w+1} | E_{p-w+2} | \ldots | E_p \}
\]

and

\[
S_q = \{ E_{q-w+1} | E_{q-w+2} | \ldots | E_q \}.
\]
We define the Entropic Dynamic Warping Kernel (EDTWK) between $G_p$ and $G_q$ as

$$k_{DCTE}(G_p, G_q) = k_{GA}(S_p, S_q) = \sum_{\pi \in \mathcal{A}(w, w)} e^{-D_{p,q}(\pi)}$$

(12)

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

Remarks: 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 CT matrix. Based on financial risk theory [35], 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 $k_{DCTE}$ 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 CT 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 CT. This computation relies on the spectral decomposition of the Laplacian 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 $k_{DCTE}$ 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 [37], 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 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], [12], [34] 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 CT encapsulates the integrated effectiveness of all possible paths between a pair of vertices. As a result, the dominant entropy time series computed through the CT 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 [17]. They thus cannot directly accommodate such network structures and instead need to transform them into sparser structures [6]. Unfortunately, these sparse structures discard many salient weighted edges and certainly lead to information loss. By contrast, the CT is computed through the Laplacian matrix that can directly accommodate complete weighted graphs. Thus, the proposed kernel encapsulates the entire set of structural information residing on all weighted edges. On the other hand, the discrete-time quantum walk kernel [8] can also accommodate complete weighted graphs through the CT 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 CT 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, Cuturi [37] 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 naturally inherit the 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 [43]. 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 prior 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 NYSE data set [6], [12]. The NYSE data set 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
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 [44], Dot-com Bubble Burst [45], Bankrupt of New Centry Financial, Lehman Crisis in Sub-prime Crisis Period [46], 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 CT matrix. Based on the financial risk theory stated by [35], 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 dominant correlated stocks, tends to significantly drop down before a financial crisis. The experiments demonstrate that the proposed dominant entropy time series through the CT 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.

B. Kernel Embeddings of Financial Networks From \( kPCA \)

In this section, we evaluate the performance of the proposed EDTWK kernel on time-varying networks of the NYSE data set and explore whether the proposed kernel can distinguish the network evolution with time. Specifically, with the proposed kernel, we perform kernel Principle Component Analysis (kPCA) [47] 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 (GC) methods, that is, the Shannon entropy associated with the classical steady state random walk [48], the von Neumann entropy associated with the normalized Laplacian matrix [49], 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 in Fig. 5(b).

Moreover, we also compare the proposed kernel with five state-of-the-art kernel methods, that is, the dynamic time warping inspired global alignment kernel (GAK) for original time series [37] and four graph kernels for time-varying financial networks. Specifically, the graph kernels include the
Fig. 5. Path of financial networks over all trading days. (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 QK. (h) Path for EDTWKO.

Weisfeiler–Lehman subtree kernel (WLSK) [39], the quantum Jensen–Shannon kernel (QJSK) [48], the feature space Laplacian graph kernel (FLGK) [50], and the quantum-inspired kernel (QK) [8]. For the GAK kernel, we also adopt a time window of 28 days for each trading day. For the alternative graph kernels, we propose to employ the Euclidean distance to replace the Pearson correlation between pairwise time series as the edge weights. Because, we have demonstrated that these graph kernels can achieve better performance on the distance-based time-varying financial networks in the previous work [8]. 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 QK kernel, we accommodate the original weighted adjacency matrices of financial networks through the minimum spanning tree over the CT matrices, following the same setting of our previous work [8]. For all the five 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)–(g), respectively.

Finally, as we previously stated, the CT 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 CT matrix plays a significant role to determine the effectiveness of the proposed kernel. To further validate its effectiveness, we also compare the proposed DETWK kernel associated with CT matrices with that associated with the original weighted adjacency matrices (DETWKO), and the result is displayed 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-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

$$DS = \sum_{t} \frac{\|x_t - x_{t-1}\|^2}{\sum_{t} \|x_t - x_m\|^2}$$  \hspace{1cm} (13)$$

where $t = 2, 3, \ldots, n$, $x_t$ is the network embedding vector at time step $t$, and $x_m$ 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

| Methods | EDTWK | GC | GAK | WLSK |
|---------|-------|----|-----|------|
| Distance Stress | 1.1931 | 8.9918 | 2.9677 | 5.7053 |

| Methods | QJSK | FLK | QK | EDTWKO |
|---------|------|-----|----|--------|
| Distance Stress | 2.8953 | 2.9297 | 2.9167 | 2.1330 |

Table I: Distance Stress of the Network Embeddings
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 shown 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 CT matrix has a better performance than that associated with the original absolute Pearson correlation matrix, demonstrating the effectiveness of the CT integrated in the proposed approach. To further reveal the effectiveness of adopting the CT matrix, we randomly select a financial network. For this network, we perform multidimensional on both its CT 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 CT or the Pearson correlation matrix), then the centered similarity matrix is given by

\[
K = -\frac{1}{2}(I - \frac{J}{n}) C (I - \frac{J}{n})
\]  

(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 eigendecomposition of the kernel matrix in terms of the diagonal matrix of ordered eigenvalues \( \Lambda \) and the correspondingly ordered matrix of column eigenvectors \( \Phi \), 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 CT 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 CT-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 indicate that the proposed EDTWK kernel associated with the CT matrix can better understand the structural evolutions of the financial networks with time than the remaining methods.

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 different methods, Fig. 7 corresponds to the Black Monday period (from June 15, 1987 to February 17, 1988), Fig. 8 corresponds to the Dot-com Bubble period (from January 3, 1995 to December 31, 2001), and Fig. 9 corresponds to the Enron Incident period (the red points, from October 16, 2001 to March 11, 2002). These figures show that the Black Monday (October 17, 1987), the Dot-com Bubble Burst (March 13, 2000), and the Enron Incident period (December 2, 2001 to March 11, 2002) are all extreme financial events. The embedding points of the proposed EDTWK kernel before and after these events can be observed in the figures.

![Fig. 6. Vertex (Stock) embedding comparisons. (a) For CT. (b) For Pearson correlation.](image)

![Fig. 7. 3-D embeddings of Black Monday. (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 QK. (h) Black Monday for EDTWKO.](image)
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 (November 3, 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.

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 January 2, 2006 to July 1, 2009) in more detail. The financial events for comparisons are the Newcentury Financial Bankruptcy (April 4, 2007) and the Lehman Brothers Bankruptcy (September 15, 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 color
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 dis-
tributions 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 networks belonging to the Newcentury crisis period and
the Lehman crisis period, as well as that belonging to all the
6004 trading data of the NYSE data set. 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

$$k_n(G_p, G_q) = \frac{k(G_p, G_q)}{\sqrt{k(G_p, G_p)k(G_q, G_q)}}$$ (15)

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 color bar beside each subfigure of Fig. 11 represents the kernel value of the kernel
matrix. From 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 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, and over the entire financial crisis
period. On the other hand, we show that only the kernel
embeddings associated with the proposed EDTWK kernel give a clear trajectory in terms of the embedding distributions with time. In other words, 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 help us understand how the local structural changes of the time-varying network gradually evolves with time, i.e., it measures the fine gradations in the structure of the time-varying financial network as it gradually changes 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 time intervals during each specific financial crisis period. The crisis periods used for evaluation 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 was not a single event crisis but continued for multiple trading days, we do not use it for the classification evaluation. For each of these selected financial crisis periods, we utilize 100 continuous trading days around the particular day on which the financial event occurred, 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 ten short-time intervals and each interval contains ten continuous trading days, i.e., the financial networks of the 100 trading days are separated into ten 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 tenfold cross-validation using the C-support vector machine (C-SVM) to compute the classification accuracies, using LIBSVM [51]. We use nine samples for training and one for testing. All the C-SVMs were performed along with their parameters optimized on each data set. We repeat the whole experiment ten 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 [8], [39], as well as the QJSK kernel [48], and the evaluations associated with these kernels follow the same experimental setup as that of the proposed kernel. The experimental 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 evolving within a short-time interval, but also discriminates the main changes that evolve over a long time interval across more than one stages. This demonstrates the better ability of the proposed kernel to measure how the structure of the financial network gradually evolves with time.

Note that, in this evaluation one can also classify the time-varying financial networks from each crisis period into a different number of classes, e.g., two classes, three classes, etc. If the number of the classes is 2, the classification evaluation tends to explore the global discriminative ability of the changes in the financial network before and after a specific dominant financial event. The more classes used to describe the time series, the more the classification focuses on the local discriminative ability of the network changes with time. The proposed kernel tends to significantly outperform the remaining kernels, when we divide the financial networks into more classes. However, it is not meaningful to divide the financial networks into more than 20 classes, since less than five financial networks of each local stage are insufficient to represent the main common characteristics at each stage.

F. Discussions of the Experimental Advantages

All of the above experiments demonstrate the effectiveness of the proposed kernel. The reasons for this are fourfold. First, unlike the WLSK kernel, the proposed kernel can directly accommodate those 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. Thus, it better preserves the network characteristics. Third, unlike the GAK for the original time series, the proposed kernel for time-varying networks is based on the dominant entropy time series and reflects richer correlation 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 dynamic time-varying financial networks.

V. CONCLUSION

In this article, we have proposed a new dynamic time warping framework inspired kernel, namely the EDTWKS 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 CT matrix on each of the network structures and have 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

| Datasets | Black Monday | Newcentury | Lehman | Asia 1997 | Dot-com |
|----------|--------------|------------|--------|-----------|---------|
| EDTWK    | 91.47 ± 0.54 | 88.79 ± 0.63 | 88.00 ± 0.65 | 90.82 ± 0.57 | 91.03 ± 0.61 |
| QJSK     | 89.71 ± 0.78 | 88.32 ± 0.74 | 87.85 ± 0.81 | 89.88 ± 1.11 | 89.17 ± 0.72 |
| WLSK     | 88.70 ± 0.51 | 88.30 ± 0.80 | 86.36 ± 0.88 | 85.44 ± 0.91 | 88.03 ± 0.73 |
| QK       | 90.10 ± 0.51 | 85.50 ± 0.80 | 87.69 ± 0.88 | 90.21 ± 0.91 | 89.79 ± 0.71 |
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 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. CT implicitly averages over all paths connecting a pair of stocks in the network, and not just the first-order nearest neighbor 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 CT to define the groups of stock and develop a hypergraph kernel to measure the relationships between groups. Moreover, in recent years, there have been increasing interests to develop novel graph convolutional networks (GCN) for graph data analysis [52]–[55]. It will be interesting to develop new GCN models for dynamic financial network analysis.

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. Learn. Syst., vol. 25, no. 7, pp. 1229–1262, Jul. 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. Learn. Syst., vol. 29, no. 5, pp. 1637–1651, May 2018.

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

[4] Y. Yin, P. Shang, and J. Xia, “Compositional segmentation of time series in the financial markets,” Appl. Math. Comput., vol. 268, pp. 399–412, Oct. 2015.

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

[6] F. N. Silva et al., “Modular dynamics of financial market networks,” 2015, arXiv:1501.05040. [Online]. Available: http://arxiv.org/abs/1501.05040

[7] H. Bunke, P. J. Dickinson, A. Humm, C. Irmiger, 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. Berlin, Germany: Springer, 2007, pp. 227–245.

[8] L. Bai, L. Rossi, L. Cui, J. Cheng, and E. R. Hancock, “A quantum-inspired similarity measure for the analysis of complete weighted graphs,” IEEE Trans. Cybern., vol. 50, no. 3, pp. 1264–1277, Mar. 2020.

[9] J. Zhang and M. Small, “Complex network from pseudoperiodic time series: Topology versus dynamics,” Phys. Rev. Lett., vol. 96, no. 23, Jun. 2006, Art. no. 238701.

[10] G. Nicolis, A. G. Cantu, and C. Nicolis, “Dynamical aspects of interaction networks,” Int. J. Bifurcation Chaos, vol. 15, no. 11, pp. 3467–3480, 2005.

[11] Y. Shimada, T. Kimura, and T. Ikeguchi, “Analysis of chaotic dynamics using measures of the complex network theory,” in Proc. Int. Conf. Artif. Neural Netw. (ICANN), 2009, pp. 61–70.

[12] C. Ye et al., “Thermodynamic characterization of networks using graph polynomials,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 92, no. 3, Sep. 2015, Art. no. 032810.

[13] L. Wang, L. Liu, and L. Zhou, “A graph-embedding approach to hierarchical visual word mergence,” IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 2, pp. 308–320, Feb. 2017.

[14] A. Dutta and H. Sahbi, “Stochastic graphlet embedding,” IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 8, pp. 2369–2382, Aug. 2019.

[15] G. Da San Martino, N. Navarin, and A. Sperti, “Tree-based kernel for graphs with continuous attributes,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 7, pp. 3270–3276, Jul. 2018.

[16] L. Oneto et al., “Learning with kernels: A local Rademacher complexity-based analysis with application to graph kernels,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 10, pp. 4660–4671, Oct. 2018.

[17] D. Haussler, “Convolutional kernels on discrete structures,” Dept. Comput. Sci., UC Santa Cruz, Santa Cruz, CA, USA, Tech. Rep. USCC-CRL-99-10, 1999.

[18] X. Bai, C. Yan, H. Yang, L. Bai, J. Zhou, and E. R. Hancock, “Adaptive hash retrieval with kernel based similarity,” Pattern Recognition, vol. 75, pp. 136–148, Mar. 2018.

[19] L. Bai, Z. Zhang, C. Wang, X. Bai, and E. R. Hancock, “A graph kernel based on the Jensen-Shannon representation alignment,” in Proc. 24th Int. Conf. Intert. Infr. (IJCAI), 2015, pp. 3322–3328.

[20] E. Bullmore and O. Sporns, “Complex brain networks: Graph theoretical analysis of structural and functional systems,” Nature Rev. Neurosci., vol. 10, no. 3, pp. 186–198, Mar. 2009.

[21] D. P. Feldman and J. P. Crutchfield, “Measures of statistical complexity: Why?” Phys. Lett. A, vol. 238, nos. 4–5, pp. 244–252, Feb. 1998.

[22] K. Anand, G. Bianconi, and S. Severini, “Shannon and von Neumann entropy of random networks with heterogeneous expected degree,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 83, no. 3, Mar. 2011, Art. no. 036109.

[23] K. Anand, D. Krioukov, and G. Bianconi, “Entropy distribution and combination in random networks with a given degree distribution,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 89, no. 6, Jun. 2014, Art. no. 062807.

[24] K. Huang, Statistical Mechanics. New York, NY, USA: Wiley, 1987.

[25] M. A. Javarone and G. Armano, “Quantum-classical transitions in complex networks,” J. Stat. Mech., Theory Exp., vol. 2013, no. 4, p. 04019, 2013.

[26] J.-C. Delvenne and A.-S. Libert, “Centrality measures and thermodynamic formalism for complex networks,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 83, no. 4, Apr. 2011, Art. no. 046117.

[27] A. Fronczak, P. Fronczak, and J. A. Holyst, “Thermodynamic forces, bond order and Onsager coefficients in complex networks,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 76, no. 6, Dec. 2007, Art. no. 061106.

[28] N. Shervashidze, P. Schweitzer, E. J. Van Leeuwen, K. Mehlhorn, and K. M. Borgwardt, “Weisfeiler-Lehman graph kernels,” J. Mach. Learn. Res., vol. 12, pp. 2539–2561, 2011.

[29] L. Bai, L. Rossi, Z. Zhang, and E. R. Hancock, “An aligned subtree kernel for weighted graphs,” in Proc. Int. Conf. Mach. Learn. (ICML), 2015, pp. 30–39.

[30] L. Bai, L. Rossi, H. Bunke, and E. R. Hancock, “Attributed graph kernels using the Jensen-Tsallis q-differences,” in Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases (ECML-PKDD), 2014, pp. 99–114.

[31] N. M. Kriege, P. Giscard, and R. C. Wilson, “On valid optimal assignment kernels and applications to graph classification,” in Proc. Neural Inf. Process. Syst. (NIPS), 2016, pp. 1615–1623.

[32] G. Nikolentzos, P. Meladianos, S. Limnios, and M. Vazirgiannis, “A degeneracy framework for graph similarity,” in Proc. 27th Int. Joint Conf. Intert. Infr. (IJCAI), Jul. 2018, pp. 2595–2601.

[33] H. Kashima, K. Tsuda, and A. Inokuchi, “Marginalized kernels between labeled graphs,” in Proc. 20th Int. Conf. Mach. Learn. (ICML), 2003, pp. 321–328.

[34] J. Wang, R. C. Wilson, and E. R. Hancock, “Directed and undirected network evolution from Euler–Lagrange dynamics,” Pattern Recognit. Lett., vol. 134, pp. 135–144, 2020.

[35] J. G. Haubrich and A. W. Lo, Quantifying Systemic Risk. USA: Univ. of Chicago Press, 2013.

[36] K. Huang, Statistical Mechanics. New York, NY, USA: Wiley, 1987.
[41] M. Pavan and M. Pelillo, “Dominant sets and pairwise clustering.” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 1, pp. 167–172, Jan. 2007.

[42] L. Bai and E. R. Hancock, “Graph kernels from the Jensen-Shannon divergence.” J. Math. Imag. Vis., vol. 47, nos. 1–2, pp. 60–69, Sep. 2013.

[43] K. O. Mikalsen, F. M. Bianchi, C. Soguero-Ruiz, and R. Jønsen, “Time series cluster kernel for learning similarities between multivariate time series with missing data,” Pattern Recognit., vol. 76, pp. 569–581, Apr. 2018.

[44] E. S. Browning, “Exorcising ghosts of Octobers past,” Wall Street J., vol. 15, pp. C1–C2, Oct. 15, 2007.

[45] K. Anderson, C. Brooks, and A. Katsaris, “Speculative bubbles in the S&P 500: Was the tech bubble confined to the tech sector?” J. Empirical Finance, vol. 17, no. 3, pp. 345–361, 2010.

[46] C. Mollenkamp, S. Craig, S. Ng, and A. Lucchetti, “Lehman files for bankruptcy, merrill sold, AIG seeks cash,” Wall Street J., Sep. 16, 2008. [Online]. Available: http://online.wsj.com/article/SB122145492097035549.html

[47] I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques. San Mateo, CA, USA: Morgan Kaufmann, 2011.

[48] L. Bai, L. Rossi, A. Torsello, and E. R. Hancock, “A quantum Jensen–Shannon graph kernel for unattributed graphs,” Pattern Recognit., vol. 48, no. 2, pp. 344–355, Feb. 2015.

[49] M. Dehmer and A. Mowshowitz, “A history of graph entropy measures,” Proc. Nat. Acad. Sci., vol. 116, no. 15, pp. 7015–7020, Apr. 2019.

[50] R. Kondor and H. Pan, “The multiscale Laplacian graph kernel,” in Proc. NIPS, 2016, pp. 2982–2990.

[51] C.-C. Chang and C.-J. Lin, “LIBSVM: A Library for Support Vector Machines.” [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[52] Z. Zhang, D. Chen, J. Wang, L. Bai, and E. R. Hancock, “Quantum-based subgraph convolutional neural networks,” Pattern Recognit., vol. 88, pp. 38–49, Apr. 2019.

[53] Y. Chen et al., “Graph convolutional network with structure pooling and joint-wise channel attention for action recognition,” Pattern Recognit., vol. 103, Jul. 2020, Art. no. 107321.

[54] F. Manessi, A. Rozza, and M. Manzo, “Dynamic graph convolutional networks,” Pattern Recognit., vol. 97, Jan. 2020, Art. no. 107000.

[55] L. Bai, Y. Jiao, L. Cui, and E. R. Hancock, “Learning aligned-spatial graph convolutional networks for graph classification,” in Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases, Würzburg, Germany, Sep. 2019, pp. 464–482.

Lu Bai received the B.Sc. and M.Sc. degrees from the Faculty of Information Technology, Macau University of Science and Technology, Macau, China, in 2007 and 2009, respectively, and the Ph.D. degree from the University of York, York, U.K., in 2015. He is currently an Associate Professor with the Central University of Finance and Economics, Beijing, China. His current research interests include pattern recognition, machine learning, and financial data analysis.

Dr. Bai was a recipient of the National Award for Outstanding Self-Financed Chinese Students Study Abroad by China Scholarship Council in 2015, and the Best Paper Awards of the International Conferences ICIAP 2015 (Eduardo Caianello Award) and ICPPR 2018 organized by IAPR. He is a leading Guest Editor of Pattern Recognition Journal.

Lixin Cui received the B.Sc. and M.Sc. degrees from Tianjin University, Tianjin, China, in 2007 and 2009, respectively, and the Ph.D. degree from the University of Hong Kong, Hong Kong, in 2013. She is currently an Associate Professor with the School of Information, Central University of Finance and Economics, Beijing, China. Her current research interests include machine learning, optimization techniques, deep learning, and their applications in FinTech problems.

Dr. Cui was nominated as the recipient of the Best Paper Awards of the International Conference IEEE IEEM 2019. She is currently an Associate Editor of Pattern Recognition Journal.

Zhihong Zhang received the B.Sc. degree in computer science from the University of Ulster, Ulster, U.K., in 2009, and the Ph.D. degree in computer science from the University of York, York, U.K., in 2013.

He is currently an Associate Professor with the Software School, Xiamen University, Fujian, China. His research interests are wide-reaching but mainly involve the areas of pattern recognition and machine learning, particularly problems involving graphs and networks.

Lixiang Xu received the Ph.D. degree in computer application technology from Anhui University, Anhui, China, in 2016.

He worked at Huawei Technologies Company Ltd., Hefei, China, in 2008. He was with Hefei University, Anhui. His current research interests include structural pattern recognition, machine learning, graph spectral analysis, image and graph matching, especially in kernel methods and complexity analysis on graphs and networks.

Yue Wang received the B.Sc. and M.Sc. degrees from the Hefei University of Technology, Anhui, China, in 2004 and 2007, respectively, and the Ph.D. degree from Sichuan University, Sichuan, China, in 2010.

He was a Post-Doctoral Researcher with Peking University, Beijing, China. He is currently an Associate Professor with the School of Information, Central University of Finance and Economics, Beijing. His current research interests include data mining and machine learning.

Edwin R. Hancock (Fellow, IEEE) received the B.Sc., Ph.D., and D.Sc. degrees from the University of Durham, Durham, U.K., in 1977, 1981, and 2008, respectively.

He is currently an Emeritus Professor with the Department of Computer Science, University of York, York, U.K. He has authored or coauthored nearly 200 journal articles and 550 conference articles.

Dr. Hancock is a fellow of the IAPR, the Institute of Physics, the Institute of Engineering and Technology, and the British Computer Society. He was a recipient of the Royal Society Wolfson Research Merit Award in 2009, the Pattern Recognition Society Medal in 1991, and the BMVA Distinguished Fellowship in 2016. He was the second Vice President of IAPR. He is currently the Editor-in-Chief of Pattern Recognition Journal.