.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Embedding and trajectories of temporal networks

CHANON THONGPRAYOON[1], LORENZO LIVI[2], AND NAOKI MASUDA .[1,3,4]
[1]Department of Mathematics, State University of New York at Buffalo, Buffalo, NY, USA (e-mail: chanonth@buffalo.edu)
[2]Department of Computer Science, University of Manitoba, Winnipeg, Manitoba, Canada (e-mail: lorenz.livi@gmail.com)
[3] Computational and Data-Enabled Science and Engineering Program, State University of New York at Buffalo, Buffalo, NY, USA
[4]Faculty of Science and Engineering, Waseda University, Shinjuku, Tokyo, Japan

Corresponding author: Naoki Masuda (e-mail: naokimas@gmail.com).

ABSTRACT Temporal network data are increasingly available in various domains, and often represent highly complex systems with intricate structural and temporal evolutions. Due to the difficulty of processing such complex data, it may be useful to coarse grain temporal network data into a numeric trajectory embedded in a low-dimensional space. We refer to such a procedure as temporal network embedding, which is distinct from procedures that aim at embedding individual nodes. Temporal network embedding is a challenging task because we often have access only to discrete time-stamped events between node pairs, and, in general, the events occur with irregular intervals, making the construction of the network at a given time a nontrivial question already. We propose a method to generate trajectories of temporal networks embedded in a low-dimensional space given a sequence of time-stamped events as input. We realize this goal by combining the landmark multidimensional scaling, which is an out-of-sample extension of the well-known multidimensional scaling method, and the framework of tie-decay temporal networks. This combination enables us to obtain a continuous-time trajectory describing the evolution of temporal networks. We then study mathematical properties of the proposed temporal network embedding framework. Finally, we showcase the method with empirical data of social contacts to find temporal organization of contact events and loss of them over a single day and across different days.

INDEX TERMS Dimension reduction, dynamical systems, network theory (graphs)

I. INTRODUCTION

MANY complex systems can be modeled as networks that consist of nodes and edges connecting nodes [1], [2]. Network analysis has enabled us to extract useful information from data that represent relationships between elements constituting a system and provided concepts and methods with which to understand and intervene into phenomena occurring on networks such as epidemic spreading, cascading failures and synchronization. In fact, the structure of many types of networks changes over time on a timescale that is comparable with the timescale of dynamical processes occurring on networks such as epidemic processes. Examples of such networks, called temporal networks [3–6], include social contact networks whose structure evolves owing to mobility of individuals, e-mail communication networks among users, and air transportation networks in which direct commercial flights connecting a pair of airports are time-stamped and also vary over months. Research on temporal networks has revealed that the temporality of the networks changes various structural and dynamical properties of networks relative to those of static networks. For example, communities in temporal networks show much richer features than static-network counterparts, such as birth, death, split, and merge of communities, and heterogeneous distributions of lifetimes of individual communities, which are concepts that are relevant only to temporal networks [7], [8]. Such dynamic community structure of temporal networks has been useful for understanding evolution of scientific research fields [9] and that of groups of senators over the history in the US [10], to name a few.

Temporal networks are intimately related to time series data. However, it seems that temporal network data analyses that have been developed in the past two decades or so have not exploited a full potential of general time series analysis, which has a long history and provides an ample repertoire of methods. Important exceptions include change-point and anomaly detection in temporal networks [11–13] and extraction of the discrete state of the temporal network and its

transitory dynamics [14–17], which are clearly hinted on the same analyses of general time series data. We propose that this gap is because most methods for analyzing time series assume numeric data, whereas the temporal network data is network-valued by definition. Such methods include the Fourier transform, autocorrelation, curve fitting, fractal analysis, and the Lyapunov exponent, to name a few. Although one could transform a temporal network data into a numeric time series by, for example, vectorizing the adjacency matrix at each time, such a transformation is usually very high-dimensional because the number of entries of the adjacency matrix scales with $O(N^2)$, where N is the number of nodes. Furthermore, temporal network data are often provided in the form of time-stamped event sequences, and how one should transform such data into numeric vectors in a computationally feasible and informative manner is nontrivial.

To fill this knowledge gap, in the present study, we propose a method for generating a continuous-valued trajectory lying in a Euclidean space given temporal network data as input in the form of time-stamped event sequences; the proposed method is also directly applicable to the input given as a discrete-time sequence of adjacency matrices, which is another major representation of temporal networks [5].

Network embedding is commonly understood as embedding of nodes of a network into a low-dimensional space by preserving their pairwise similarity as much as possible [18–20]. Alternatively, an edge embedding method embeds edges of the given network into a low-dimensional space [21], [22]. An obvious application of network embedding is visualization of networks. In addition, if network embedding preserves sufficient information of the original network data, one may be able to efficiently carry out downstream data mining tasks by working on the network represented in the embedding space, possibly exploiting machine learning algorithms. Such downstream tasks include link prediction [23], visualization [24], and anomaly detection [25]. There are various node embedding methods for temporal networks, with which nodes move over time in the embedding space [26–30].

In contrast to such methods, the network embedding for temporal networks that we develop in this paper does not belong to the traditional classes of network embedding such as node or edge embedding. Instead, we aim at embedding temporal networks into a Euclidean space as a single trajectory such that the network at a given point in time is mapped to a point in the embedding space. The proposed method combines elements of the mathematical modeling framework of tie-decay temporal network [31] and the landmark multidimensional scaling (LMDS), which is a dimension reduction method for general data [32]. The LMDS is an out-of-sample extension of multidimensional scaling (MDS) [33]. This particular combination of methods enables us to generate continuous-time trajectories describing structural and dynamical features of temporal network data given in the form of list of time-stamped events on network edges.

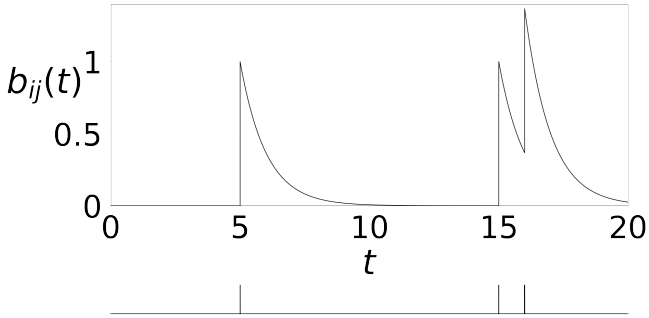

FIGURE 1: Time course of the edge weight, $b_{ij}(t)$, in a tie-decay network. The vertical bars on the timeline to the bottom of the figure represent the times of the events, $\tilde{t}_1 = 5, \tilde{t}_2 = 15, \tilde{t}_3 = 16$, on edge (i, j). We set $\alpha = 1$.

II. THE PROPOSED TEMPORAL NETWORK EMBEDDING ALGORITHM

A. TIE-DECAY NETWORKS

Consider a set of N nodes and a sequence of time-stamped contact events. A contact event, which we simply refer to as the event in the following text, is assumed to occur between a pair of individuals. Therefore, each event is a tuple (i_e, j_e, t_e), where i_e and j_e are the pair of nodes, and t_e is the time of the event. We allow the event to be directed from i_e to j_e. We seek to embed such a data set into a low-dimensional space as a network trajectory using tie-decay networks. A tie-decay network is a continuous-time temporal network model in which the weight of the edge (i.e., its tie) increases when an event occurs on the edge, and continuously decays over time [31]. Let $B(t) = [b_{ij}(t)]$, $t \in \mathbb{R}_{\geq 0}$ be a time-dependent $N \times N$ matrix, where $b_{ij}(t)$ represents the weight of directed edge (i, j) at time t. We set

$$b_{ij}(t) = \sum_{\ell} e^{-\alpha(t-\tilde{t}_\ell)} H(t - \tilde{t}_\ell), \tag{1}$$

where H is the Heaviside function, and $\tilde{t}_\ell$ is the time of the ℓth event on edge (i, j), with $0 \leq \tilde{t}_1 \leq \tilde{t}_2 \leq \cdots$. Equation (1) implies that an event increases the edge weight by 1 and that the edge weight exponentially decreases with rate α. We show in Fig. 1 an example time course of $b_{ij}(t)$.

Next, we consider a sequence of time steps $0 \leq t_1 < t_2 < \cdots$, where t_ℓ is the ℓth time when at least one event occurs in the entire network. We denote by A_ℓ the adjacency matrix of the network constructed from all the events occurring at time t_ℓ. The input data is then identified with a sequence of adjacency matrices $\{A_1, A_2, \ldots\}$ and the associated times $\{t_1, t_2, \ldots\}$. We obtain

$$B(t) = \sum_{\ell=1}^{\infty} e^{-\alpha(t-t_\ell)} A_\ell. \tag{2}$$

B. CLASSICAL AND LANDMARK MULTIDIMENSIONAL SCALING

The classical multidimensional scaling embeds a set of high-dimensional data points into a low-dimensional vector space

while preserving the dissimilarity between any pair of data points as much as possible [33–35]. We consider a set of M data points, denoted by $S = \{x_1, \ldots, x_M\}$ in a metric space (X, d), where d is a distance measure. The MDS embeds each $x_i \in S$ into $\mathbb{R}^k$, where k is relatively small, by carrying out the following steps. First, we construct the squared distance matrix $\Delta = [\Delta_{ij}]$, where $\Delta_{ij} = d^2(x_i, x_j)$. Second, we calculate the double mean-centered dot product matrix $D = -\frac{1}{2} H_M \Delta H_M$, where $H_M = I_M - \frac{1}{M} J_M$, matrix I_M is the $M \times M$ identity matrix, and J_M is the $M \times M$ matrix in which all the elements are equal to 1. Finally, we obtain the embedding coordinates in $\mathbb{R}^k$ for x_i as the ith column of the following $k \times M$ matrix:

$$L_k = \begin{bmatrix} \sqrt{\lambda_1} \boldsymbol{v}_1^\top \\ \vdots \\ \sqrt{\lambda_k} \boldsymbol{v}_k^\top \end{bmatrix}, \tag{3}$$

where λ_i is the ith largest positive eigenvalue of D, vector $\boldsymbol{v}_\ell$ is the associated right eigenvector normalized in terms of the ℓ^2-norm, and $^\top$ represents the transposition.

MDS has a time complexity of $O(M^3)$ because it requires eigenvalues and eigenvectors of D [36]. Therefore, it is difficult to apply MDS for large data sets. An out-of-sample extension of a dimension reduction method calculates the exact coordinates in the embedding space only for a small fraction of data points and approximates the coordinates for the remaining data points [37], [38]. The landmark multi-dimensional dimension scaling (LMDS) is an out-of-sample extension of the classical MDS [32], [39], [40]. The LMDS applies the MDS to a subset of the given data points called the landmarks and then uses the landmarks to calculate the embedding coordinates for the other data points. We use $S = \{x_1, \ldots, x_M\}$ as the landmarks and therefore first apply MDS to S. The embedding coordinates for S are given by (3). Then, we determine the embedding coordinates of the remaining data as follows. First, we define a $k \times M$ matrix by

$$L'_k = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} \boldsymbol{v}_1^\top \\ \vdots \\ \frac{1}{\sqrt{\lambda_k}} \boldsymbol{v}_k^\top \end{bmatrix}. \tag{4}$$

Second, we calculate

$$\boldsymbol{\delta}_\mu = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{\delta}_i, \tag{5}$$

where $\boldsymbol{\delta}_i$ is the ith column of Δ. Third, we define the embedding coordinate for an arbitrary out-of-sample data point $z \notin S$ by

$$\psi(z) = -\frac{1}{2} L'_k (\boldsymbol{\delta}_z - \boldsymbol{\delta}_\mu), \tag{6}$$

where

$$\boldsymbol{\delta}_z = \begin{bmatrix} d^2(z, x_1) \\ \vdots \\ d^2(z, x_M) \end{bmatrix}. \tag{7}$$

C. EMBEDDING OF TIE-DECAY NETWORKS BY LANDMARK MULTIDIMENSIONAL SCALING

We propose a method to embed tie-decay networks in continuous time into a low-dimensional Euclidean space using LMDS. The key idea is to use the networks at all the times at which the input event happens, i.e., $t \in \{t_1, t_2, \ldots\}$, as landmarks and embed the networks at all the other times, i.e., $t \notin \{t_1, t_2, \ldots\}$, using LDMS. The algorithm proceeds as follows. First, for a given value of α (i.e., rate of tie decay), we construct the tie-decay network from the input adjacency matrices, $\{A_1, \ldots, A_M\}$, using (2), where M is the number of the unique time points at which events occur on any edge. Second, we apply classical MDS to $\{B(t_1), \ldots, B(t_M)\}$. Third, we run LDMS with $\{B(t_1), \ldots, B(t_M)\}$ as landmarks as follows. We need to embed an arbitrary adjacency matrix $\{B(t) : t \geq 0\}$, which we call interpolating points when $t \notin \{t_1, \ldots, t_M\}$, into $\mathbb{R}^k$ to obtain a continuous trajectory. Using the exponential tie-decay property, we obtain

$$B(t) = e^{-\alpha(t - t_\ell)} B(t_\ell), \tag{8}$$

where $\ell \in \{1, \ldots, M\}$ satisfies $t_\ell \leq t < t_{\ell+1}$. Then, we use (6) to obtain the embedding coordinates for $B(t)$ for any t as follows:

$$\psi(B(t)) = -\frac{1}{2} L'_k (\boldsymbol{\delta}_B(t) - \boldsymbol{\delta}_\mu), \tag{9}$$

where

$$\boldsymbol{\delta}_B(t) = \begin{bmatrix} d^2(B(t), B(t_1)) \\ \vdots \\ d^2(B(t), B(t_M)) \end{bmatrix}. \tag{10}$$

D. DISTANCE MEASURES FOR NONNEGATIVE MATRICES

To calculate $\psi(B(t))$, one needs the squared distance between the matrix $B(t)$ and each of the M landmark matrix representations. Because $d(B(t), B(t_m)) = d(e^{-\alpha(t-t_\ell)} B(t_\ell), B(t_m))$, we should use a distance measure for networks for which we can easily calculate $d(cM_1, M_2)$ given $d(M_1, M_2)$, where M_1 and M_2 are general weighted adjacency matrices and $0 < c < 1$. Here we provide three eligible distance measures.

1) Frobenius distance

Let $\langle \cdot, \cdot \rangle$ be the Frobenius inner product for real-valued matrices given by

$$\langle P, Q \rangle = \sum_{i=1}^{a} \sum_{j=1}^{b} p_{ij} q_{ij}, \tag{11}$$

where $P = [p_{ij}]$ and $Q = [q_{ij}]$ are $a \times b$ matrices. We define the induced Frobenius norm $\| \cdot \|$ by

$$\|P\| = \sqrt{\langle P, P \rangle}. \tag{12}$$

We obtain the Frobenius distance by $d_F(B(t_\ell), B(t_{\ell'})) = \|B(t_\ell) - B(t_{\ell'})\|$. For $\ell, \ell' \in \{1, \ldots, M\}$, we obtain

$$d_F^2(B(t_\ell), B(t_{\ell'})) = \|B(t_{\ell'})\|^2 - 2\langle B(t_\ell), B(t_{\ell'}) \rangle + \|B(t_{\ell'})\|^2. \tag{13}$$

Using (13), we obtain for any $t \in [t_\ell, t_{\ell+1})$

$$d_F(e^{-\alpha(t-t_\ell)}B(t_\ell), B(t_{\ell'})) = (e^{-\alpha(t-t_\ell)}(e^{-\alpha(t-t_\ell)}-1)\|B(t_\ell)\|^2+(1-e^{-\alpha(t-t_\ell)})\|B(t_{\ell'})\|^2+ e^{-\alpha(t-t_\ell)}d_F^2(B(t_\ell), B(t_{\ell'})))^{\frac{1}{2}} . \quad (14)$$

Therefore, we can easily compute $\psi(B(t))$ for any t by precomputing $\|B(t_\ell)\|^2$ for $\ell \in \{1, \ldots, M\}$ in addition to matrix Δ.

2) Laplacian distance

In this and next subsections, we assume undirected networks. The Laplacian matrix $L(t)$ of a tie-decay network at time t is defined by $L(t) \equiv D^*(t) - B(t)$, where $D^*(t)$ is the diagonal matrix whose ith diagonal entry is given by $\sum_{j=1}^{N} b_{ij}(t)$. The unnormalized Laplacian distance between two Laplacian matrices, which we refer to as the Laplacian distance for short in the following text, is defined by

$$d_L(L(t_\ell), L(t_{\ell'})) = \sqrt{\sum_{k=1}^{N} [\rho_k(L(t_\ell)) - \rho_k(L(t_{\ell'}))]^2} , \quad (15)$$

where $\rho_k(P)$ is the kth smallest eigenvalue of symmetric matrix P [14], [41], [42]. It should be noted that one can replace $\sum_{k=1}^{N}$ in (15) by $\sum_{k=2}^{N}$ because the smallest eigenvalue of a Laplacian matrix is equal to 0. For any t satisfying $t_\ell \le t < t_{\ell+1}$, we obtain $L(t) = e^{-\alpha(t-t_\ell)}L(t_\ell)$ and therefore

$$d_L(e^{-\alpha(t-t_\ell)}L(t_\ell), L(t_{\ell'})) = \sqrt{\sum_{k=2}^{N} \left[e^{-\alpha(t-t_\ell)}\rho_k(L(t_\ell)) - \rho_k(L(t_{\ell'}))\right]^2}. \quad (16)$$

3) Bregman distance

For a pair of networks G_1 and G_2 each with N nodes, we consider the distance measure given by

$$F(G_1, G_2) = \sum_{k=1}^{N}\sum_{k'=1}^{N} f(\rho_k(L_1), \rho_{k'}(L_2)) \, |\langle \boldsymbol{u}_k(L_1), \boldsymbol{u}_{k'}(L_2)\rangle_E|^m , \quad (17)$$

where $f : \mathbb{R}^2 \to \mathbb{R}_{\ge 0}$ is symmetric with respect to the two arguments and satisfies the triangle inequality; L_i (with $i \in \{1, 2\}$) is the Laplacian matrix of G_i; $\boldsymbol{u}_k(L_i)$ is an eigenvector of L_i that is normalized in terms of the ℓ^2-norm and associated with eigenvalue $\rho_k(L_i)$; $\langle \cdot, \cdot \rangle_E$ is the Euclidean inner product of two vectors; m is a positive integer [43]. Using the tie-decay networks at times t and $t_{\ell'}$ as the input to F, where $t_\ell \le t < t_{\ell+1}$, we obtain

$$F(B(t), B(t_{\ell'})) = \sum_{k=1}^{N}\sum_{k'=1}^{N} f(e^{-\alpha(t-t_\ell)}\rho_k(L(t_\ell)), \rho_{k'}(L(t_{\ell'}))) \, |\langle \boldsymbol{u}_k(L(t_\ell), \boldsymbol{u}_{k'}(L(t_{\ell'}))\rangle_E|^m . \quad (18)$$

We assume $m = 2$, which is known to make (17) independent of the choice of the eigenvectors [43]. An example of F is given through the Bregman divergence as follows. Let $h : \Omega \to \mathbb{R}$ be a continuously differentiable and strictly convex function, where $\Omega \subseteq \mathbb{R}^n$. We define the Bregman divergence associated with h, which we denote by d_h, by

$$d_h(\boldsymbol{x}, \boldsymbol{y}) = h(\boldsymbol{x}) - h(\boldsymbol{y}) - \langle \boldsymbol{x} - \boldsymbol{y}, \nabla h(\boldsymbol{y})\rangle_E , \quad (19)$$

where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, and $\nabla h(\boldsymbol{y})$ denotes the gradient of $h(\boldsymbol{y})$. If $h(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{x}\rangle_E$, then d_h reduces to the squared ℓ^2-norm because

$$\begin{aligned} d_{\langle\cdot,\cdot\rangle_E}(\boldsymbol{x}, \boldsymbol{y}) &= \langle \boldsymbol{x}, \boldsymbol{x}\rangle_E - \langle \boldsymbol{y}, \boldsymbol{y}\rangle_E - \langle \boldsymbol{x} - \boldsymbol{y}, \nabla\langle \boldsymbol{y}, \boldsymbol{y}\rangle_E\rangle_E \\ &= \langle \boldsymbol{x}, \boldsymbol{x}\rangle_E - \langle \boldsymbol{y}, \boldsymbol{y}\rangle_E - \langle \boldsymbol{x} - \boldsymbol{y}, 2\boldsymbol{y}\rangle_E \\ &= |\boldsymbol{x} - \boldsymbol{y}|_2^2 , \end{aligned} \quad (20)$$

where $|\cdot|_2$ denotes the ℓ^2-norm [44]. By setting $f = d_{\langle\cdot,\cdot\rangle_E}$, $n = 1$, and $m = 2$ in (18), we obtain the Bregman distance, denoted by d_B, given by

$$d_B(B(t), B(t_{\ell'})) = \sum_{k=1}^{N}\sum_{k'=1}^{N} \left[e^{-\alpha(t-t_\ell)}\rho_k(L(t_\ell)) - \rho_{k'}(L(t_{\ell'}))\right]^2 \langle \boldsymbol{u}_k(L(t_\ell)), \boldsymbol{u}_{k'}(L(t_{\ell'}))\rangle_E^2 , \quad (21)$$

where $t_\ell \le t < t_{\ell+1}$. A modification of the Bregman distance that better captures some global dissimilarities between two networks such as connectivity is given by [43]

$$d_{B'}(B(t), B(t_{\ell'})) = \sum_{k=1}^{N}\sum_{k'=1}^{N} \frac{\left[e^{-\alpha(t-t_\ell)}\rho_k(L(t_\ell)) - \rho_{k'}(L(t_{\ell'}))\right]^2}{e^{-\alpha(t-t_\ell)}\rho_k(L(t_\ell)) + \rho_{k'}(L(t_{\ell'}))} \langle \boldsymbol{u}_k(L(t_\ell)), \boldsymbol{u}_{k'}(L(t_{\ell'}))\rangle_E^2 . \quad (22)$$

E. GOODNESS OF TEMPORAL NETWORK EMBEDDING

MDS better preserves the pairwise dissimilarity between data points if the embedding dimension, k, is larger. Furthermore, the quality of the embedding produced by MDS may depend on the distance measure. To select a distance measure and k for MDS, we use the following goodness of fit criteria.

The first criterion uses the fact that embedding into the Euclidean space preserves the pairwise dissimilarity for all pairs of data points if and only if matrix $D = -\frac{1}{2}H_M\Delta H_M$ is positive semidefinite [45, 46]. The existence and magnitude of negative eigenvalues indicate the deviation from the

Euclidean behavior. Therefore, as goodness of fit, we use the ratio of the k most positive eigenvalues over the sum of the absolute values of all eigenvalues [47]. Specifically, we assume that D has at least k positive eigenvalues and define

$$g = \frac{\sum_{\ell=1}^{k} \lambda_\ell}{\sum_{\ell'=1}^{M} |\lambda_{\ell'}|}, \tag{23}$$

where $\lambda_1 > \cdots > \lambda_M$ are the eigenvalues of D. Index g ranges between 0 and 1 and it is close to 1 when the embedding is quasi-Euclidean.

Another goodness of fit criterion for MDS is given by the normalized stress function [34], [35], [48], denoted by σ, which is defined as

$$\sigma = \sqrt{\frac{\sum_{\ell=1}^{M} \sum_{\ell'=1}^{\ell-1} (d(x_{\ell'}, x_\ell) - |\psi(x_{\ell'}) - \psi(x_\ell)|_2)^2}{\sum_{\ell=1}^{M} \sum_{\ell'=1}^{\ell-1} d^2(x_{\ell'}, x_\ell)}}. \tag{24}$$

Note that $|\psi(x_{\ell'}) - \psi(x_\ell)|_2$ is the distance between the two data points in the embedding space. A small σ value indicates that MDS is accurate at representing the dissimilarity structure of the original data set. Values of σ less than 0.15 have been suggested to be acceptable [34].

F. TIME COMPLEXITY OF THE TEMPORAL NETWORK EMBEDDING

Given a squared distance matrix Δ, the computation of the embedding coordinate, given by (9), consists of computing the double mean-centered dot product matrix D, which requires $O(M^3)$ time, computing the eigenvalues and eigenvectors of D, which requires $O(M^3)$ time, and then computing L'_k, which requires $O(kM)$ time, computing $\boldsymbol{\delta}_\mu$, which requires $O(M^2)$ time, computing $\boldsymbol{\delta}_B(t) - \boldsymbol{\delta}_\mu$, which requires $O(M)$ time, and computing $L'_k(\boldsymbol{\delta}_B(t) - \boldsymbol{\delta}_\mu)$, which requires $O(kM)$ time. The total time complexity of these computations combined is $O(M^3)$. The time complexity for calculating $\boldsymbol{\delta}_B(t)$ depends on the choice of the network distance measure. The time complexity for embedding $B(t)$ for any $t \in [t_{\tilde{\ell}}, t_{\tilde{\ell}+1})$ is as follows.

In the case of the Frobenius distance, the time complexity for computing the distance between two landmarks is $O(N^2)$. Therefore, constructing the $M \times M$ squared distance matrix takes $O(M^2N^2)$ time. By precalculating $\|B(t_\ell)\| \, \forall \ell \in \{1, \ldots, M\}$, which takes $O(MN^2)$ time, and storing it and the pairwise distance between the landmarks, we can compute $\boldsymbol{\delta}_B(t)$ in $O(M)$ time. It follows that the total time complexity for embedding $B(t)$ is $O(M^2N^2 + M^3)$.

In the case of the Laplacian distance, the time complexity for computing the Laplacian matrices of all landmarks is $O(MN^2)$. To compute the distance between two landmarks, we need the eigenvalues of $L(t_\ell) \forall \ell \in \{1, \ldots, M\}$. Because the time complexity for calculating the eigenvalues is $O(N^3)$, it takes $O(MN^3)$ to compute $\rho_k(L(t_\ell)) \; \forall k \in \{2, \ldots, N\}, \forall \ell \in \{1, \ldots, M\}$. Once we store $\rho_k(L(t_\ell)) \; \forall k \in \{2, \ldots, N\}, \forall \ell \in \{1, \ldots, M\}$, it takes $O(M^2N)$ time to compute the squared distance matrix using (15) and $O(MN)$ time to compute $\boldsymbol{\delta}_B(t)$. Therefore, the total time complexity for embedding $B(t)$ is $O(MN^3 + M^2N + M^3)$.

Finally, we consider the case of the Bregman distance. In addition to the time complexity of $O(MN^2)$ for computing the Laplacian matrices of all landmarks, it takes $O(MN^3)$ time to compute all the eigenvalues and eigenvectors of all the Laplacian matrices, i.e., $\rho_k(L(t_\ell))$ and $\boldsymbol{u}_k(L(t_\ell))$, $\forall k$, $\forall t_\ell$. Given them, it takes $O(M^2N^3)$ time to compute the squared Bregman distance matrix using (21) or the squared modified Bregman distance using (22). Then, the time complexity for computing $\boldsymbol{\delta}_B(t)$ is $O(MN^3)$. Therefore, the total time complexity of embedding $B(t)$ is $O(M^2N^3 + M^3)$.

III. THEORETICAL RESULTS

In this section, we mathematically investigate two properties of the proposed embedding algorithm for temporal networks: periodicity and boundedness.

A. PERIODIC INPUT EVENTS

We first examine the convergence of the trajectory in the embedding space when the input events periodically arrive over time. Let $\{A_\ell\}_{\ell=1}^{\infty}$ be an infinite sequence of adjacency matrices with each A_ℓ encoding the new network structure at time t_ℓ. We assume that $\{A_\ell\}_{\ell=1}^{\infty}$ is periodic such that $A_{qT+k} = \tilde{A}_k$ and $t_{qT+k} = q\tilde{T} + t_k$ for each $q \in \mathbb{N} \cup \{0\}$ and $k \in \{1, \ldots, T\}$, where $T \in \mathbb{N}$ is the number of time points in one cycle at which at least one event occurs, and $\tilde{T}$ is the period in terms of absolute time.

Lemma III.1. *For any $t \in [0, \infty)$, write $t = q\tilde{T} + t'$ such that $t_k \leq t' < t_{k+1}$ for unique values of $q \in \mathbb{N} \cup \{0\}$ and $k \in \{1, \ldots, T\}$. In the case of $k = T$, we interpret $t_{k+1} = \tilde{T} + t_1$. Then, we obtain $B(t) = \sum_{\ell=1}^{T} c_\ell(t) A_\ell$, where*

$$c_\ell(t) = \begin{cases} \dfrac{1 - e^{-\alpha(q+1)\tilde{T}}}{1 - e^{-\alpha\tilde{T}}} e^{-\alpha(t - q\tilde{T} - t_\ell)} & \text{if } 1 \leq \ell \leq k, \\ \dfrac{1 - e^{-\alpha q\tilde{T}}}{1 - e^{-\alpha\tilde{T}}} e^{-\alpha(t - (q-1)\tilde{T} - t_\ell)} & \text{if } k+1 \leq \ell \leq T. \end{cases} \tag{25}$$

In particular, the tie-decay network with the periodic input converges to the limit cycle of period $\tilde{T}$ given by

$$X(t') \equiv \sum_{\ell=1}^{T} \frac{e^{-\alpha(t' - t_\ell)}}{1 - e^{-\alpha\tilde{T}}} A_\ell, \tag{26}$$

where $0 \leq t' < \tilde{T}$.

Proof. Using (2), we obtain

$$\begin{aligned}B(t) &= \sum_{\ell=1}^{qT+k} e^{-\alpha(t-t_\ell)} A_\ell \\ &= \sum_{\ell=1}^{k}\sum_{j=0}^{q} e^{-\alpha\left[t-(j\tilde{T}+t_\ell)\right]} A_\ell \\ &\quad + \sum_{\ell=k+1}^{T}\sum_{j=0}^{q-1} e^{-\alpha\left[t-(j\tilde{T}+t_\ell)\right]} A_\ell \\ &= \sum_{\ell=1}^{k} e^{-\alpha(t'-t_\ell)} A_\ell \sum_{j=0}^{q} e^{-\alpha(q-j)\tilde{T}} \\ &\quad + \sum_{\ell=k+1}^{T} e^{-\alpha(\tilde{T}+t'-t_\ell)} A_\ell \sum_{j=0}^{q-1} e^{-\alpha(q-1-j)\tilde{T}} \\ &= \sum_{\ell=1}^{k} \frac{1-e^{-\alpha(q+1)\tilde{T}}}{1-e^{-\alpha\tilde{T}}} e^{-\alpha(t'-t_\ell)} A_\ell \\ &\quad + \sum_{\ell=k+1}^{T} \frac{1-e^{-\alpha q\tilde{T}}}{1-e^{-\alpha\tilde{T}}} e^{-\alpha(\tilde{T}+t'-t_\ell)} A_\ell \\ &= \sum_{\ell=1}^{T} c_\ell(t) A_\ell. \end{aligned} \tag{27}$$

Equation (25) implies that

$$\lim_{q\to\infty} c_\ell(q\tilde{T}+t') = \frac{e^{-\alpha(t'-t_\ell)}}{1-e^{-\alpha\tilde{T}}} \tag{28}$$

for any $\ell \in \{1,\ldots,T\}$ and any $t' \in [0,\tilde{T})$. Therefore, we obtain (26). □

Theorem III.2. *The embedding coordinates of tie-decay networks, given by* (9)*, with periodic input converge to a limit cycle.*

Proof. Under the periodic input, Lemma III.1 guarantees that $\lim_{q\to\infty} B(q\tilde{T}+t')$ converges to the limit cycle $X(t')$. Furthermore, because of the continuity of the distance metric function, d, the $\lim_{q\to\infty}$ operation and d are interchangeable. Therefore, we obtain

$$\begin{aligned}\lim_{q\to\infty} d(B(q\tilde{T}+t'),B(t_\ell)) &= d(\lim_{q\to\infty} B(q\tilde{T}+t'),B(t_\ell)) \\ &= d(X(t'),B(t_\ell)) \end{aligned} \tag{29}$$

for any $\ell \in \{1,\ldots,T\}$. Using (29), we obtain

$$\lim_{q\to\infty} \boldsymbol{\delta}_B(q\tilde{T}+t') = \lim_{q\to\infty} \begin{bmatrix} d(B(q\tilde{T}+t'),B(t_1)) \\ \vdots \\ d(B(q\tilde{T}+t'),B(t_T)) \end{bmatrix} = \boldsymbol{\delta}_X(t'). \tag{30}$$

By combining (9) and (30), we obtain

$$\begin{aligned}\lim_{q\to\infty} \psi(B(q\tilde{T}+t')) &= -\lim_{q\to\infty} \frac{1}{2} L'(\boldsymbol{\delta}_B(q\tilde{T}+t')-\boldsymbol{\delta}_\mu) \\ &= -\frac{1}{2} L'(\boldsymbol{\delta}_X(t')-\boldsymbol{\delta}_\mu) = \psi(X(t')). \end{aligned} \tag{31}$$

Therefore, the limit cycle in the embedding space is given by $\psi(X(t'))$, where $0 \le t' < \tilde{T}$. □

To illustrate Theorem III.2 by an example, we consider a temporal network on $N = 3$ nodes. We give a periodic input with $T = 4$ input matrices defined by

$$\tilde{A}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \tilde{A}_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
$$\tilde{A}_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad \tilde{A}_4 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}. \tag{32}$$

We set $t_\ell = \ell$ for each $\ell \in \mathbb{N}$ and $\alpha = 0.15$. The network trajectory in the two-dimensional embedding space is shown in Fig. 2. Figs. 2(a) and 2(b) show a trajectory obtained with the Frobenius and the Laplacian distance measure, respectively. As expected, both trajectories approach a limit cycle with period 4.

Under the Frobenius distance measure and $\alpha = 0.15$, we obtain $g = 0.617$ and $\sigma = 0.32$ with $k = 1$, $g = 0.951$ and $\sigma = 0.06$ with $k = 2$, and $g > 0.999$ and $\sigma = 9.35 \times 10^{-15}$ with $k = 3$. We remind that g and σ are goodness of embedding measures defined by (23) and (24), respectively. Under the Laplacian distance measure, D has only two positive eigenvalues. Therefore, we consider g and σ, when $k = 1, 2$. When $\alpha = 0.15$, we obtain $g = 0.820$ and $\sigma = 0.26$ with $k = 1$ and $g > 0.999$ and $\sigma = 5.34 \times 10^{-15}$ with $k = 2$. Therefore, for both distance measures, we conclude that $k = 2$ is sufficient for embedding this tie-decay network.

B. BOUNDEDNESS OF TRAJECTORIES

Next we investigate boundedness of the network trajectory, i.e., the property of the trajectory that it remains within a particular bounded domain. We assume that adjacency matrices $\{A_\ell\}_{\ell=1}^{\infty}$ arrive with regular intervals, i.e., $t_\ell = \ell\tau$, where $\tau(>0)$ is a constant. However, the extension of the following results to the case of aperiodic inputs, with the minimal interval between two consecutive input adjacency matrices being equal to τ, is straightforward. We also assume that each adjacency matrix A_ℓ is selected from a finite set, which we denote by $\{\tilde{A}_m\}_{m=1}^{M}$. In the periodic input case discussed in the previous section, we obtain $M = T$ and $A_\ell = \tilde{A}_{((\ell-1) \bmod T)+1}$.

Lemma III.3. *For any sequence of input adjacency matrices* $\{A_\ell\}_{\ell=1}^{\infty}$*, where each* $A_\ell \in \{\tilde{A}_1,\ldots,\tilde{A}_M\}$*, let* $B(t)$ *be the corresponding tie-decay matrix, which we write as*

$$B(t) = \sum_{m=1}^{M} \tilde{c}_m(t)\tilde{A}_m \tag{33}$$

with $\tilde{c}_m(t) \ge 0$*. Then, each* $\tilde{c}_m(t)$ *is bounded.*

Proof. We first show that $\sum_{m=1}^{M} \tilde{c}_m(t)$ is bounded for any input sequence, hence so is each $\tilde{c}_m(t)$. Consider an arbitrary $t \in$

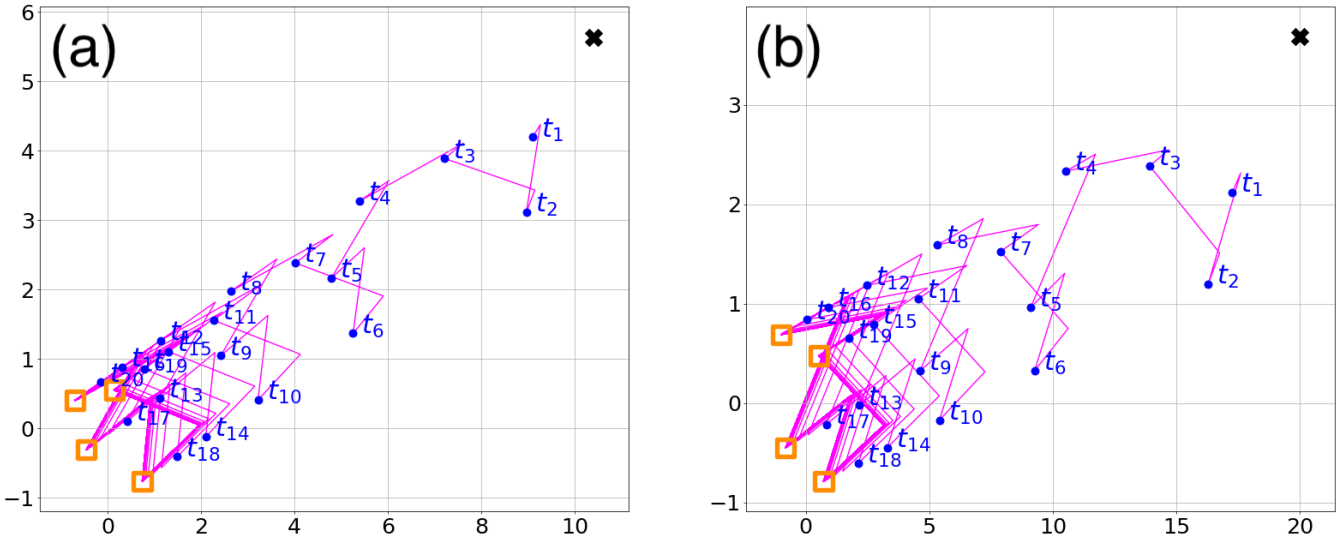

FIGURE 2: Network trajectories with periodic input. (a) A trajectory under the Frobenius distance measure. (b) A trajectory under the Laplacian distance measure. The input adjacency matrices are periodic with four adjacency matrices. We set $\alpha = 0.15$. The squares indicate the coordinates of the limit cycle right after an input adjacency matrix arrives. The crosses indicate the coordinate of the empty network.

$[t_\ell, t_{\ell+1})$, for which we can write $t = t_\ell + \Delta t$ with $\Delta t \in [0, \tau)$. For each $m \in \{1, \ldots, M\}$, we obtain

$$\begin{aligned} 0 \le \tilde{c}_m(t) &\le \sum_{m'=1}^{M} \tilde{c}_{m'}(t) \\ &= \sum_{j=1}^{\ell} e^{-\alpha(t-t_j)} \\ &= e^{-\alpha\Delta t} \sum_{j=0}^{\ell-1} e^{-\alpha j\tau} \\ &\le e^{-\alpha\Delta t} \frac{1-e^{-\alpha\ell\tau}}{1-e^{-\alpha\tau}} \\ &< \frac{1}{1-e^{-\alpha\tau}} . \end{aligned} \tag{34}$$

□

Now we show a boundedness of the network trajectory constructed with the Frobenius distance.

Theorem III.4. *If we use the Frobenius distance for MDS, the embedding coordinate $\psi(B(t))$ for any $t \in [0, \infty)$ is bounded as follows:*

$$|\psi(B(t))|_2 < \frac{1}{2}\left[\sqrt{M}\left(\frac{\sum_{m=1}^{M}\left\|\tilde{A}_m\right\|}{1-e^{-\alpha\tau}}\right)^2 \sqrt{\sum_{\ell=1}^{k}\frac{1}{\lambda_\ell}} + |L_k'\boldsymbol{\delta}_\mu|_2\right]. \tag{35}$$

Proof. Using (33) and (34), we obtain

$$\begin{aligned} d_F(B(t), B(t_\ell)) &= \|B(t) - B(t_\ell)\| \\ &= \left\|\sum_{m=1}^{M}\left[\tilde{c}_m(t) - \tilde{c}_m(t_\ell)\right]\tilde{A}_m\right\| \\ &\le \sum_{m=1}^{M}\left|\tilde{c}_m(t) - \tilde{c}_m(t_\ell)\right| \left\|\tilde{A}_m\right\| \\ &< \frac{1}{1-e^{-\alpha\tau}} \sum_{m=1}^{M}\left\|\tilde{A}_m\right\| . \end{aligned} \tag{36}$$

Next, we evaluate the distance of the embedding coordinates of $B(t)$ from the origin. Using (9), we obtain

$$\begin{aligned} |\psi(B(t))|_2 &= \left|-\frac{1}{2}L_k'(\boldsymbol{\delta}_B(t) - \boldsymbol{\delta}_\mu)\right|_2 \\ &\le \frac{1}{2}\left(|L_k'\boldsymbol{\delta}_B(t)|_2 + |L_k'\boldsymbol{\delta}_\mu|_2\right) . \end{aligned} \tag{37}$$

We obtain

$$|L_k'\boldsymbol{\delta}_B(t)|_2^2 = \sum_{\ell=1}^{k}\langle\frac{1}{\sqrt{\lambda_\ell}}\boldsymbol{v}_\ell, \boldsymbol{\delta}_B(t)\rangle_E^2. \tag{38}$$

By applying Cauchy-Schwarz inequality to (38) and using

the fact that the eigenvectors of D are normalized, we obtain

$$\begin{aligned}|L'_k\boldsymbol{\delta}_B(t)|_2^2 &\leq \sum_{\ell=1}^{k}\frac{1}{\lambda_\ell}|\boldsymbol{v}_\ell|_2^2\,|\boldsymbol{\delta}_B(t)|_2^2\\ &= \sum_{\ell=1}^{k}\frac{1}{\lambda_\ell}|\boldsymbol{\delta}_B(t)|_2^2\\ &= \sum_{\ell=1}^{k}\frac{1}{\lambda_\ell}\sum_{\ell'=1}^{M}d_F^4(B(t),B(t_{\ell'}))\end{aligned} \tag{39}$$

By combing (36) and (39), we obtain

$$\begin{aligned}|L'_k\boldsymbol{\delta}_B(t)|_2^2 &< \sum_{\ell=1}^{k}\frac{1}{\lambda_\ell}\sum_{\ell'=1}^{M}\left[\frac{1}{(1-e^{-\alpha\tau})^4}\sum_{m=1}^{M}\left(\left\|\tilde{A}_m\right\|\right)^4\right]\\ &= \sum_{\ell=1}^{k}\frac{1}{\lambda_\ell}\left[\frac{M}{(1-e^{-\alpha\tau})^4}\sum_{m=1}^{M}\left(\left\|\tilde{A}_m\right\|\right)^4\right].\end{aligned} \tag{40}$$

By substituting (40) in (37), we obtain (35).

□

Next we show the boundedness when we use the Laplacian distance for the MDS. Recall that we let $L(t)$ denote the Laplacian matrix of the tie-decay matrix $B(t)$.

Lemma III.5. *Let M be an $N\times N$ matrix, and L_M be the Laplacian matrix induced by M. Then, the map $M\mapsto L_M$ is a linear map.*

Proof. Let M_1 and M_2 be $N\times N$ matrices and let $D^*_{M_1}$, $D^*_{M_2}$, and $D^*_{cM_1+M_2}$ be the degree matrices of M_1, M_2 and cM_1+M_2, respectively, where c is a constant. One obtains

$$D^*_{cM_1+M_2} = cD^*_{M_1}+D^*_{M_2}\,. \tag{41}$$

Using (41), we obtain

$$\begin{aligned}L_{cM_1+M_2} &= D^*_{cM_1+M_2}-(cM_1+M_2)\\ &= cD^*_{M_1}+D^*_{M_2}-(cM_1+M_2)\\ &= c(D^*_{M_1}-M_1)+(D^*_{M_2}-M_2)\\ &= cL_{M_1}+L_{M_2}\,,\end{aligned} \tag{42}$$

which proves the linearity of the map. □

By applying Lemma III.5 to the tie-decay adjacency matrix $B(t)=\sum_{m=0}^{M}\tilde{c}_m(t)\tilde{A}_m$, we express $L(t)$ as follows:

$$L(t)=\tilde{c}_1(t)L_{\tilde{A}_1}+\cdots+\tilde{c}_M(t)L_{\tilde{A}_M}\,, \tag{43}$$

where $L_{\tilde{A}_\ell}$ is the Laplacian matrix of $\tilde{A}_\ell$. In the next theorem, we use the same notations as those defined in section II-D.

Theorem III.6. *If we use the Laplacian distance for the MDS, the embedding coordinate for any $t\in[0,\infty)$ is bounded as follows:*

$$|\psi(B(t))|_2 < \frac{1}{2}\left[\sqrt{M}\left(\frac{2}{1-e^{-\alpha\tau}}\right)^2\left(\sum_{m=1}^{M}\left\|\tilde{A}_m\right\|_1\right)^2\sqrt{\sum_{\ell=1}^{k}\frac{1}{\lambda_\ell}}+|L'_k\boldsymbol{\delta}_\mu|_2\right], \tag{44}$$

where $\|A\|_1=\sum_{i=1}^{N}\sum_{j=1}^{N}|a_{ij}|$ for any $N\times N$ matrix $A=[a_{ij}]$.

Proof. We obtain

$$\begin{aligned}d_L(L(t),L(t_\ell)) &= \sqrt{\sum_{j=1}^{N}[\rho_j(L(t))-\rho_j(L(t_\ell))]^2}\\ &\leq \sum_{j=1}^{N}|\rho_j(L(t))-\rho_j(L(t_\ell))|\\ &\leq \sum_{j=1}^{N}|\rho_j(L(t))|+\sum_{j=1}^{N}|\rho_j(L(t_\ell))|\,.\end{aligned} \tag{45}$$

Because any eigenvalue of a Laplacian matrix is non-negative, we obtain

$$\begin{aligned}\sum_{j=1}^{N}|\rho_j(L(t))| &= \sum_{j=1}^{N}\rho_j(L(t))\\ &= \mathrm{tr}(L(t))\\ &= \tilde{c}_1(t)\mathrm{tr}\left(L_{\tilde{A}_1}\right)+\cdots+\tilde{c}_M(t)\mathrm{tr}\left(L_{\tilde{A}_M}\right).\end{aligned} \tag{46}$$

Similarly, we obtain

$$\sum_{j=1}^{N}|\rho_j(L(t_\ell))| = \tilde{c}_1(t_\ell)\mathrm{tr}\left(L_{\tilde{A}_1}\right)+\cdots+\tilde{c}_M(t_\ell)\mathrm{tr}\left(L_{\tilde{A}_M}\right). \tag{47}$$

By combining (45), (46), (47), and $\mathrm{tr}(L_{\tilde{A}_\ell})=\sum_{i=1}^{N}\sum_{j=1}^{N}a_{ij}^{(\ell)}=\left\|\tilde{A}_\ell\right\|_1$, we obtain

$$\begin{aligned}d_L(L(t),L(t_\ell)) &\leq \sum_{m=1}^{M}[\tilde{c}_m(t)+\tilde{c}_m(t_\ell)]\left\|\tilde{A}_m\right\|_1\\ &< \frac{2}{1-e^{-\alpha\tau}}\sum_{m=1}^{M}\left\|\tilde{A}_m\right\|_1.\end{aligned} \tag{48}$$

Similarly to the derivation of (40), we obtain

$$\begin{aligned}|L'_k\boldsymbol{\delta}_B(t)|_2^2 &\leq \sum_{\ell=1}^{k}\frac{1}{\lambda_\ell}\sum_{\ell'=1}^{M}d_L^4(B(t),B(t_{\ell'}))\\ &< M\left(\frac{2\sum_{m=1}^{M}\left\|\tilde{A}_m\right\|_1}{1-e^{-\alpha\tau}}\right)^4\sum_{\ell=1}^{k}\frac{1}{\lambda_\ell}.\end{aligned} \tag{49}$$

By substituting (49) in (37), we obtain (44). □

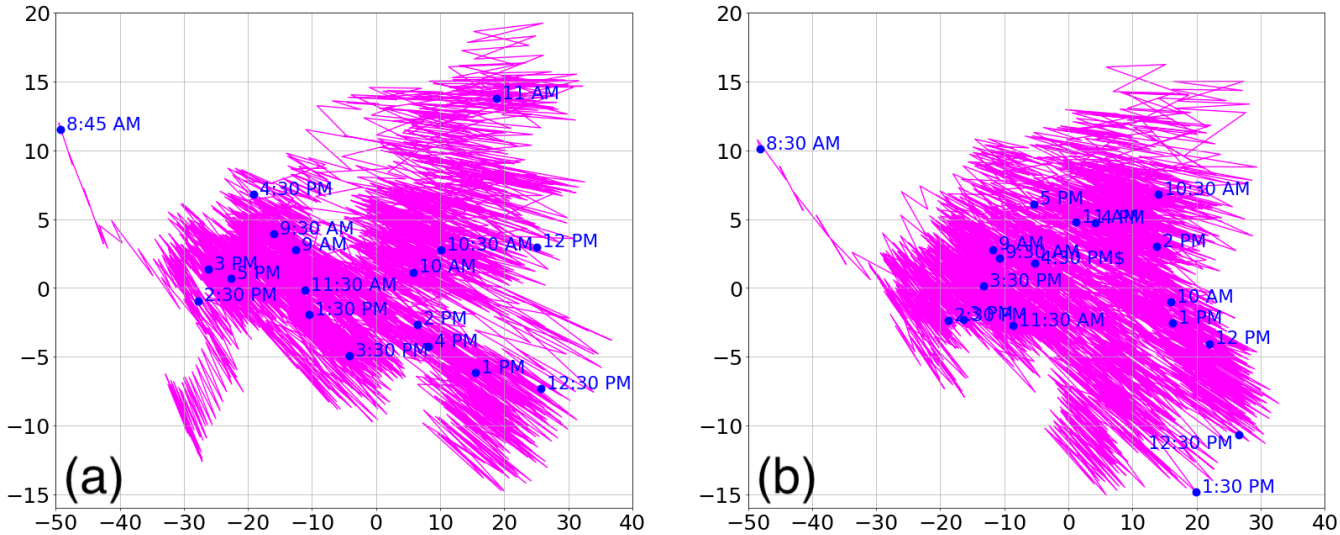

FIGURE 3: Trajectories of the Primary School temporal networks over two days in the two-dimensional embedding space. (a) Day 1. (b) Day 2. We set $\alpha = 10^{-2}$. The dots indicate the location of the trajectory every 30 minutes.

IV. TRAJECTORIES OF EMPIRICAL CONTACT NETWORKS

In this section, we apply the proposed temporal network embedding method to two empirical data sets of time-stamped proximity events between human individuals. Both data sets are provided by the Sociopatterns project. The recording was made every 20 seconds. We use the Laplacian distance measure for computing the distance between networks. We set $\alpha = 10^{-2}$. The results of the two data sets with $\alpha = 10^{-3}$, shown in Appendix A, are similar to those with $\alpha = 10^{-2}$, supporting the robustness of our method with respect to the α value.

A. PRIMARY SCHOOL

The Primary School data set is a temporal contact network recorded from students and teachers in a primary school in Lyon, France over two days, i.e., 10/1/2009 and 10/2/2009 [49], [50]. The data spans from 8:45 AM to 5:20 PM on the first day and from 8:30 AM to 5:05 PM on the second day. We consider the time-stamped contact events between pairs of 242 individuals; there are 8,317 edges and 125,773 contacts. For each day, there are three breaks that allow students to interact with peers in other classes, which are the morning break for 20–25 minutes starting at 10:30 AM, the lunch break between 12 PM and 2 PM, and the afternoon break for 20–25 minutes starting at 3:30 PM.

We show in Figs. 3(a) and 3(b) the trajectories in the two-dimensional embedding space for the first and the second day, respectively. We set $\alpha = 10^{-2}$. In each day, the tie-decay network is initially empty, which corresponds to $(x, y) \approx (-51.42, 14.93)$ in the embedding space. In both days, the trajectory resides in the first quadrant during the morning break (see the points labeled 10:30 AM in Fig. 3) and in the fourth quadrant during the lunch break (i.e., 12 PM to 2 PM). The segments of the trajectories during these time windows are relatively, while not completely, distinguishable from those for the rest of the day.

To compare between the two days, we calculate the average coordinate of the trajectory in the embedding space in each of the consecutive 30-minute windows. We show the average coordinate for the two days in Fig. 4(a). The point labeled 10 AM, for example, represents the (x, y) coordinate averaged between 10 AM and 10:30 AM. Fig. 4(a) suggests that the two networks at the same point of time on the two days tend to be located close to each other in the embedding space. For example, the networks at 10:30 AM, corresponding to the morning break, on the two days are both located at (x, y) with large x and y values. As another example, the networks during lunch breaks (i.e., 12 PM to 2 PM) tend to linger around (x, y) with $-10 \leq x \leq 20$ and $-3 \leq y \leq 0$. To quantitatively examine the day-to-day consistency of the network trajectory, we compute the distance between the 30-minute average of the network at the same time between the two days, which we show in Fig. 4(b). First, we observe that the networks during the lunch and afternoon breaks are in particular similar between the two days compared to other times of the daytimes. Second, the distance between the two networks at the same time of the day tends to decrease over the course of the day. Third, the coordinate at the same time of the day is in fact close between the two days. Specifically, the average distance between an arbitrarily chosen pair of times indicated by the dots in Fig. 4(a) is equal to 16.87, which is substantially larger than the distance between the two networks at the same time of the day shown in Fig. 4(b), except at 8:30 AM.

To justify our choice of the dimension of the embedding space (i.e., $k = 2$) and the distance measure (i.e., Laplacian distance measure), we carry out the following analysis.

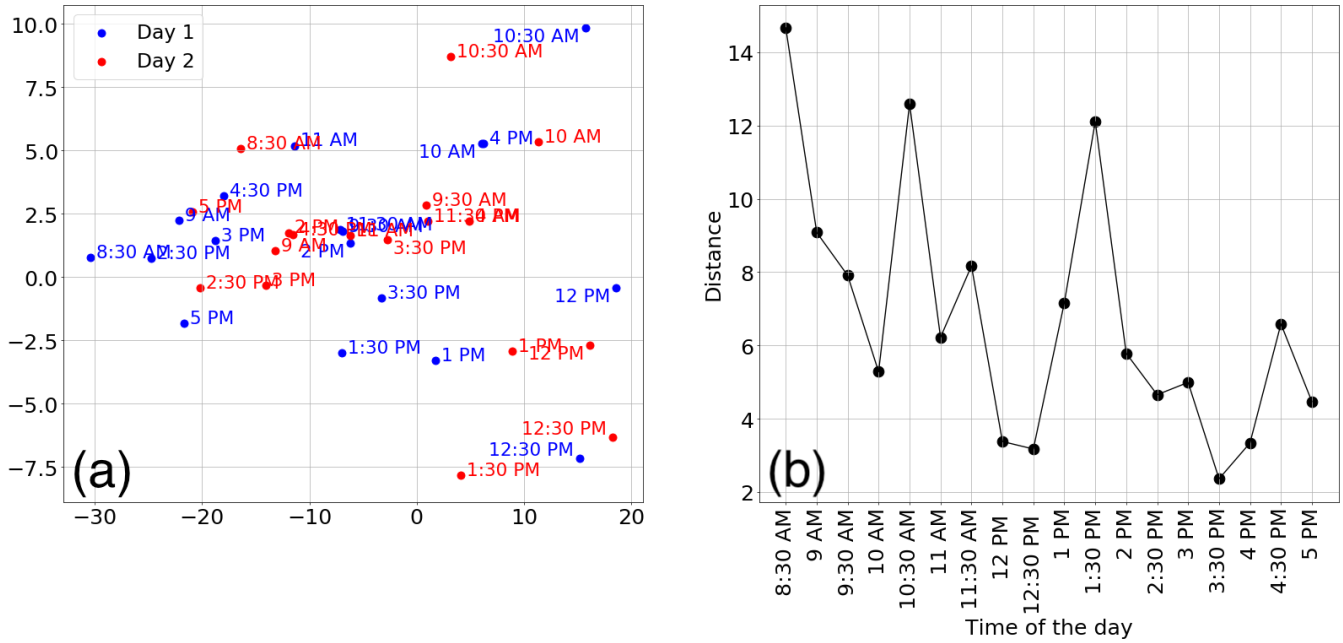

FIGURE 4: Comparison of the trajectory of the Primary School temporal network between the two days. (a) Average coordinate of the tie-decay network in the embedding space over every 30-minute window. (b) Day-to-day distance between the networks at the same time of the day. We set $\alpha = 10^{-2}$.

We show in Table 1 the values of g and σ for $k = 1$, 2, and 3 for the two distance measures. We find that the Laplacian distance measure yields much larger g values and much smaller σ values than the Frobenius distance measure does. This result justifies the use of the Laplacian distance. Moreover, Table 1 shows that, with the Laplacian distance, the embedding dimension of $k = 2$ is enough, but not $k = 1$. Note that the σ value with $k = 2$ is equal to 0.06, which is smaller than a suggested criterion value of 0.15 [34]. With $k = 3$, the goodness of embedding further improves compared with the case of $k = 2$. However, the gain with $k = 3$ is not large because the embedding is already good with $k = 2$. Therefore, we conclude that the combination of the Laplacian distance measure and $k = 2$ is a reasonable choice for this data set.

B. HOSPITAL

The second data set is time-stamped contact network collected from a hospital ward in Lyon, France [51]. We refer to this data set as Hospital. The network is composed of 75 nodes, of which 11 nodes represent medical doctors, 35 nurses, and 29 patients. The network has $1,139$ edges and $32,424$ contacts. The data span the weekdays of a week, from 1 PM on Monday, 12/6/2010 to 2 PM on Friday, 12/10/2010.

To compare the network trajectory across days, we show in Fig. 5(a) the hourly average of the embedding coordinate. For example, the coordinate labeled 10 AM represents the coordinate of the trajectory averaged between 10 AM and 11 AM. Fig. 5(a) apparently indicates that, similar to the case of the Primary School data, the network trajectory for the different days tends to visit a similar location at the same time of the day. To quantitatively investigate this property, we first compute the distance between the coordinate at a certain time and day and the coordinate at the same time of a different day. Then, given the time of the day, we average the obtained distance value over all possible pairs of the day of the week. For example, there are four tie-decay networks at 10 AM, corresponding to Tuesday, Wednesday, Thursday, and Friday. Therefore, we average the pairwise distance over the 6 possible pairs of days. We show the average distance at each time of the day in Fig. 5(b). The figure suggests that the tie-decay networks at the same time of the day are apparently close to each other across the days in the morning, but less so in the afternoon, which is opposite to the case of the Primary School data (see Fig. 4(b)). We note that all the day-to-day distance values shown in Fig. 5(b) are substantially less than 7.61, which is the average distance between an arbitrary pair of hours shown in Fig. 5(a). Therefore, we conclude that the tie-decay network in different days recurs to a similar network at the same time of the day, similar to the case of the Primary School data.

The previous study showed that contacts among nurses and patients are more abundant in the morning than the afternoon and that those among medical doctors are similar in frequency between the morning and the afternoon [51]. We verify this observation in Fig. 6(a), which shows the normalized number of contacts of different types in each hour. For normalization, we first divided the number of contacts of each type in each hourly window by the number of the days available. For example, data between 8 AM and

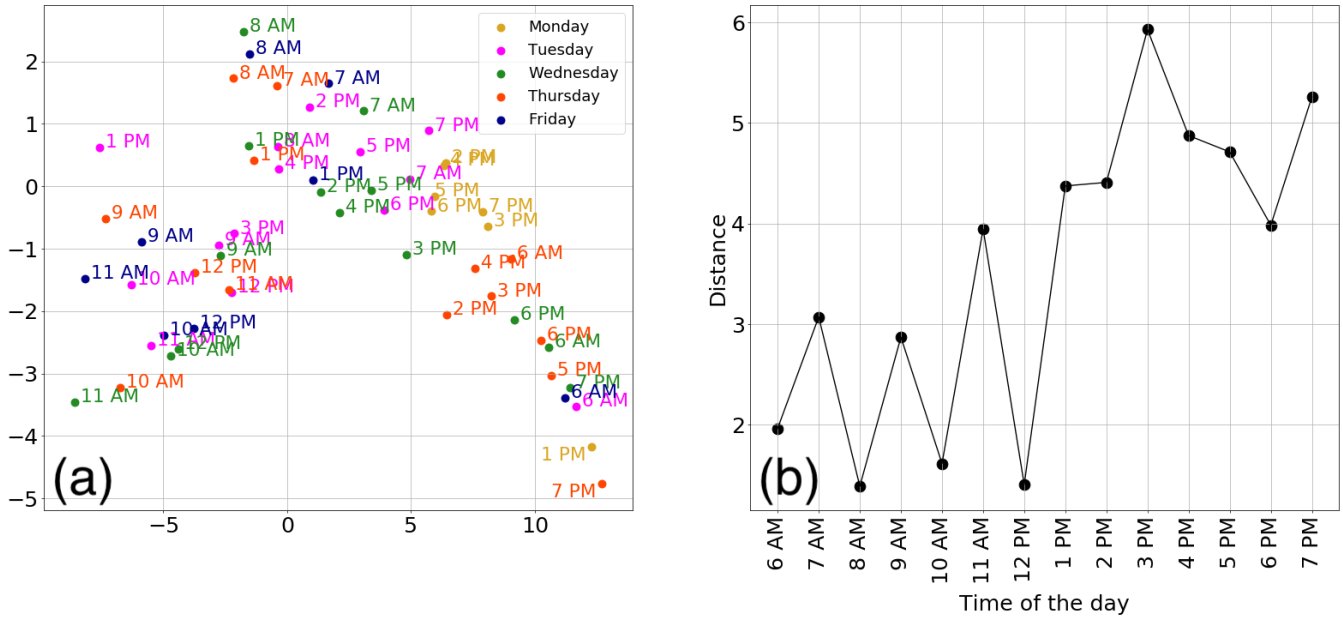

FIGURE 5: Comparison of the trajectory of the Hospital temporal network across the five days. (a) Average coordinate of the tie-decay network in the embedding space over every hourly time window. (b) Day-to-day distance between the networks at the same time of the day, which is averaged over all pairs of days. We omitted the tie-decay networks before 6 AM and after 8 PM because most of the contacts (i.e., 95.6%) occur between 6 AM and 8 PM. We set $\alpha = 10^{-2}$.

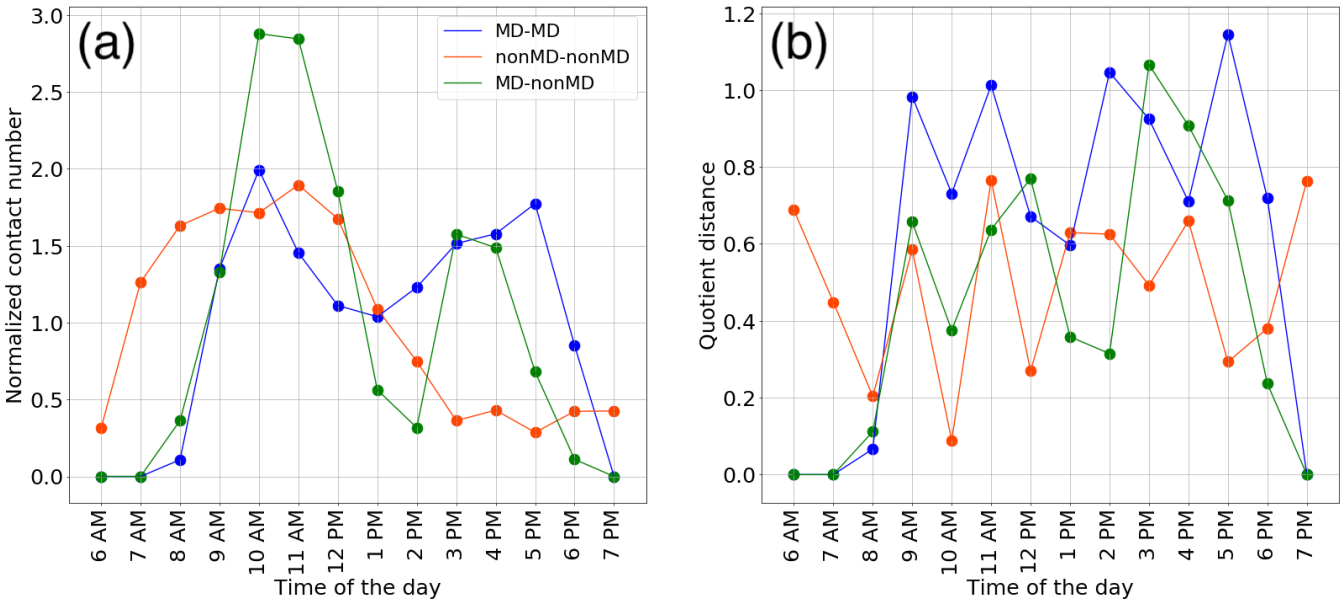

FIGURE 6: Comparison among temporal subnetworks in the Hospital data. (a) Normalized number of contacts over time. (b) Normalized day-to-day distance at the same time of the day for the different subnetworks. In both (a) and (b), we separately analyzed the trajectories of MD-MD, MD-nonMD, and nonMD-nonMD subnetworks.

9AM was recorded from Tuesday to Friday. Therefore, we divided the number of contacts between 8 AM and 9 AM by four. Second, we further normalized the obtained count so that the average over the hours of the day is equal to 1, separately for each type of contact. Similar to the previous study [51], we consider three types of contacts, i.e., contacts between medical doctors, labeled MD-MD, those between a medical doctor and a nondoctor, i.e., nurse or patient, which we label MD-nonMD, and those between individuals that are not medical doctors, labeled nonMD-nonMD. Fig. 6(a)

indicates that the number of MD-MD contacts is relatively stable over the course of the day. In contrast, the MD-nonMD and nonMD-nonMD contacts are roughly more abundant in the morning than the afternoon.

Because the entire network trajectory is more diverse and the MD-MD contacts are relatively more frequent in the afternoon than in the morning, we hypothesize that the temporal subnetwork composed only of the MD-MD contacts varies more considerably across the days of the week than the MD-nonMD and nonMD-nonMD subnetworks. To examine this hypothesis, we analyze the trajectory of the MD-MD, MD-nonMD, and nonMD-nonMD temporal subnetworks. The union of the sets of contacts from the three subnetworks is the set of all contacts in the original data set. For each subnetwork, we embed the corresponding tie-decay network and compute its hourly average coordinate in the embedding space. Then, for each subnetwork, similar to the calculation for Fig. 5(b), we compute the average distance between pairs of coordinates at the same time of the day. Finally, we divide this number by the average distance over all the possible pairs of hours. If this normalized distance is smaller than 1, then the trajectory of the subnetwork tends to visit similar locations at the same time of the day across the days.

We show the normalized distance for the three subnetworks from 6 AM to 8 PM in Fig. 6(b). We find that the day-to-day difference of the MD-MD subnetwork at the same time of the day is not particularly small, with the normalized day-to-day distance value being close to 1 across the different hours of the day (shown in blue in Fig. 6(b)). The day-to-day difference of the MD-nonMD and nonMD-nonMD subnetworks at the same time of the day is smaller than that of the MD-MD subnetwork in most times of the day. In particular, the normalized day-to-day distance for the nonMD-nonMD subnetwork is subsantially smaller than 1 throughout the day.

Fig. 6(a) implies that the fraction of the MD-MD contacts among all contacts is higher in the afternoon than the morning. Fig. 6(b) indicates that the medical doctors form more dissimilar networks across the days than nurses and patients. These two phenomena in combination explain why the day-to-day variation in the entire tie-decay network at the same time of the day is larger in the afternoon than the morning.

We end this section by justifying the choice of the distance measure (i.e., Laplacian distance) and embedding dimension (i.e., $k = 2$). We show the g_λ and σ values for the two distance measures and three values of k in Table 1. Similar to the case of the Primary School data, we find that the Laplacian distance is much superior to the Frobenius distance, that $k = 2$ is necessary, and that the accuracy of the embedding does not improve much with $k = 3$ compared to $k = 2$. The combination of the Laplacian distance and $k = 2$ is also satisfactory for embedding the three subnetworks, as we show in Table 2.

TABLE 1: Goodness of temporal network embedding. We set $\alpha = 10^{-2}$.

Data	Distance measure	Index	$k=1$	$k=2$	$k=3$
Primary School	Frobenius	g	7.73%	13.70%	17.39%
		σ	0.81	0.76	0.73
	Laplacian	g	84.76%	95.16%	97.13%
		σ	0.18	0.06	0.03
Hospital	Frobenius	g	6.03%	10.51%	14.52%
		σ	0.84	0.74	0.68
	Laplacian	g	78.84%	90.69%	95.13%
		σ	0.22	0.09	0.05

TABLE 2: Goodness of temporal network embedding for the subnetworks of the Hospital's temporal network. We set $\alpha = 10^{-2}$. We use the Laplacian network distance.

Subnetwork	Index	$k=2$
MD-MD	g	93.23%
	σ	0.07
nonMD-nonMD	g	91.04%
	σ	0.09
MD-nonMD	g	91.99%
	σ	0.08

V. GRAPH CLASSIFICATION

To demonstrate the utility of our embedding method, we apply the Support Vector Machine (SVM) to the embedding coordinates of the Primary School data to carry out a graph classification task [52], [53]. We define four groups of the time of the school day depending on the activities in the school: morning break, lunch break, afternoon break, and class time. We apply the SVM to the set of the embedding coordinates of tie-decay networks averaged over each 30-minute time window (see Fig. 4) to estimate the label of the group to which each 30-minute network has. There are 36 such averaged coordinates, 18 from each of the two days. Then, we compare the classification results to those obtained by applying the SVM to the set of embedding coordinates calculated with baseline methods, i.e., principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE) [24], and MDS. In the PCA and t-SNE, we compute an adjacency matrix by aggregating the events in each 30-minute time window. We then vectorize the upper-diagonal part of the adjacency matrices into $N(N-1)/2$-dimensional vectors and apply the PCA or t-SNE, obtaining the embedding coordinates in a two-dimensional vector space. In the MDS, we use the same set of aggregated adjacency matrices to compute the distance between each pair of the aggregated adjacency matrices, on which we run the MDS to generate a two-dimensional representation of the adjacency matrices. To learn the SVM, we first randomly choose 80% of the coordinates in the embedding space as the training set. The remaining 20% of the coordinates are the test set. Then, we measure the accuracy of the classifier as the number of the correctly classified 30-minute networks divided by the total number of the 30-minute networks in the test set. After repeating this procedure 10^4 times, we compute the average and the standard deviation of the accuracy.

We show the accuracy values for the tie-decay networks with $\alpha = 10^{-2}$ and 10^{-3}, PCA, t-SNE, and MDS in Table 3. The mean accuracy for the different classifiers is between 0.8

TABLE 3: Mean and standard deviation of accuracy of the different embedding methods in the graph classification task. We set the embedding dimension to two. We calculated the mean and standard deviation on the basis of 10^4 iterations of the graph classification task.

Method	Mean	Standard deviation
Tie-decay with $\alpha = 10^{-2}$	0.88	0.10
Tie-decay with $\alpha = 10^{-3}$	0.81	0.12
PCA	0.85	0.12
t-SNE	0.65	0.15
MDS	0.89	0.10

and 0.9 except for the t-SNE, for which the mean accuracy is substantially lower (i.e., 0.65) than for the other classifiers. The accuracy for the tie-decay networks with $\alpha = 10^{-3}$ (i.e., 0.81) is also notably lower than that for the tie-decay networks with $\alpha = 10^{-2}$, PCA, and MDS. These results remain similar when the embedding dimension is three, as we show in Appendix B. Then, for the embedding dimension equal to two, we ran a repeated measures ANOVA to test the null hypothesis that the mean accuracy is not different among the remaining three embedding methods, i.e., tie-decay networks with $\alpha = 10^{-2}$, PCA, and MDS. The critical value of the F-statistics corresponding to the significant level of $p = 0.05$ is approximately 1.03. Our test yielded a F-statistics value of 0.99, which is less than 1.03. Therefore, we accept the null hypothesis to conclude that our method based on the tie-decay networks with $\alpha = 10^{-2}$ performs on par with PCA and MDS in this graph classification task.

VI. DISCUSSION

In the present study, we have developed a method to embed temporal network data into an embedding space as a continuous trajectory. To this end, we combined the framework of tie-decay networks and the LMDS. We then proved some mathematical properties of the proposed embedding method and applied the method to two empirical data sets of temporal social contact networks among human individuals. For both data sets, we find that the networks at the same time of the day are relatively similar across the different days compared to those at different times of the day. In addition, for the Primary School data set, the coordinates of the morning break and the lunch break in the embedding space are in particular distinguishable from the rest of the day. For the Hospital data set, we have found that the day-to-day variation of the network at the same time of the day is larger in the afternoon than the morning and that this owes to the networking behavior of the medical doctors rather than the nurses and patients.

The network embedding problem that we have addressed in this study is distinct from the more popular node embedding, which aims at mapping individual nodes into a low-dimensional space such that their pairwise similarity is preserved as much as possible. Examples of node embedding include DeepWalk [20], node2vec [21], and LINE [54]. Node embedding methods for temporal networks are also available [55–57]. See [58], [59] for reviews of node embedding for static and temporal networks. With node embedding for temporal networks, each node moves in the embedding space as the input network varies over time. In contrast, we have proposed to embed the entire network structure at any given time into a single point in the embedding space. Thus, the sequence of networks observed over time generates a continuous-time trajectory lying in the embedding space. Node embedding for temporal networks and the embedding of a temporal network into a trajectory in the embedding space have different goals. With the latter, we drastically coarse grain the original temporal network by representing the snapshot network at each time point, $B(t)$, by a single point in the embedding space. In this manner, we are not focusing on the individual nodes and edges, or their relationships. Rather, we aim at capturing both structural and dynamical features of the evolving networks. By construction, the distance between two points on the trajectory reflects how different the networks at two times are. This property is potentially useful for downstream tasks such as identifying macroscopic states and dynamics of the system state of temporal networks [14], [60], change-point analysis [11], [12], [61], anomaly detection [62], systematic identification of recurrence using recurrence quantification analysis [60], [63], and link prediction [23], [55].

The LMDS is one among many out-of-sample extension methods for dimension reduction, and it can be regarded as an application of the Nyström approximation of the eigenvalues and eigenvectors of matrices [40]. The Nyström approximation is useful for our temporal network embedding framework because, with a suitably chosen network distance measure, we can efficiently approximate the elements of the distance matrix that are unobserved. We specifically exploited the exponentially decaying nature of $B(t)$ in the absence of input to calculate the distance between pairs of networks at arbitrary continuous times in a manner compatible with the Nyström approximation. In fact, the LMDS is not the only Nyström-type out-of-sample extension of the MDS, and MDS is not the only dimension reduction method that allows the Nyström algorithm [40]. Therefore, our method should also be generalizable to other Nyström-type out-of-sample extensions. Examples include the landmark Isomap algorithm [37], [39] (see [64] for the Nyström algorithm for the Isomap) and the landmark diffusion map based on the Nyström approximation [65].

Our temporal network embedding method has arbitrary parameters and allows methodological choices. The tie-decay rate, α, is defined by the user. The embedding results were qualitatively the same between $\alpha = 10^{-2}$ and $\alpha = 10^{-3}$ for the two empirical temporal networks taken into account. If one uses a large α value, the tie-decay network decays fast such that, the network trajectory would quickly approach and linger near the coordinate to which the null network is mapped. In contrast, if α is small, the tie-decay network at time t reflects the input events that are far in the past. Then, the network trajectory may change too slowly and be

TABLE 4: Goodness of temporal network embedding. We set $\alpha = 10^{-3}$.

Data	Distance measure	Index	$k=1$	$k=2$	$k=3$
Primary School	Frobenius	g	18.59%	33.48%	40.63%
		σ	0.69	0.60	0.56
	Laplacian	g	82.95%	96.16%	97.77%
		σ	0.18	0.05	0.03
Hospital	Frobenius	g	11.86%	20.70%	28.16%
		σ	0.77	0.64	0.54
	Laplacian	g	81.33%	92.79%	96.70%
		σ	0.21	0.08	0.04

TABLE 5: Goodness of temporal network embedding for the subnetworks of the Hospital's temporal network with $\alpha = 10^{-3}$. We use the Laplacian network distance.

Subnetwork	Index	$k=2$
MD-MD	g	94.45%
	σ	0.07
nonMD-nonMD	g	93.85%
	σ	0.08
MD-nonMD	g	91.81%
	σ	0.09

insensitive to new input events, potentially compromising the interpretation of the results and performance of downstream tasks. Automatic determination of the α value warrants future work. The choice of the embedding dimension, k, and the network distance measure plays an important role. To this end, we evaluated two indices to assess the embedding quality, concluding that the combination of $k = 2$ and the Laplacian network distance is our recommendation. Generalizing this result to a larger set of goodness of fit indices [66], network distance measures [5], [41], [67], [68], and various data of time-stamped contact events also warrants future work.

Although we focused on contact network data, one can apply the proposed framework and methods of temporal network embedding to various temporal network data. For example, dynamic functional connectivity [69], [70] and chronnectome [71] aim to construct and interpret temporal networks from time-dependent correlational measurements of brain activity. Our temporal network embedding method and its possible extensions are expected to give new insights for such functional network data in neuroscience. In particular, the location and dynamics of the network trajectory in the embedding space may give us a compact summary of brain dynamics without the need of explicitly defining and detecting discrete dynamical states of the brain network or their transient dynamics [15–17]. State-transition dynamics has also been applied to data of cell cycle dynamics [72]. In future works, it could be interesting to apply our methods to these data to characterize periodicity and fluctuations around periodic dynamics in cell cycles.

.

APPENDIX A RESULTS FOR $\alpha = 10^{-3}$

To assess the robustness of network trajectories in terms of the α value, we repeated the same numerical experiments with $\alpha = 10^{-3}$. Table 4 shows the values of g and σ for the Primary School and Hospital data sets, with the two network distance measures and embedding dimensions $k = 1, 2$, and 3. The results are qualitatively the same as those with $\alpha = 10^{-2}$. Specifically, the Laplacian distance is by far superior to the Frobenius distance, and $k = 2$ is necessary and sufficient in practice. The values of these indices for the subnetworks of the Hospital temporal network, shown in Table 5, indicate that the combination of the Laplacian measure and $k = 2$ is also sufficiently good for these subnetworks. Therefore, in the rest of this section, we use the Laplacian distance and $k = 2$. We show in Fig. 7 the trajectory of the tie-decay network constructed from the Primary School data in the embedding space, with each panel representing each day. As expected, the trajectory is more slowly changing than in the case of $\alpha = 10^{-2}$ (see Fig. 3) because a smaller α value used here (i.e., $\alpha = 10^{-3}$) implies a stronger memory effect. In both days, the trajectory resides in the third quadrant during the morning break (i.e., 10:30 AM) and, except for 12 PM, in the second quadrant during the lunch break (i.e., from 12.30 PM to 2 PM). In addition, the location of the segments of the trajectory in these time windows is relatively distinguishable from the rest of the day. These results are similar to those with $\alpha = 10^{-2}$ (see Fig. 3). We show in Fig. 8(a) the 30-minute average of the embedding coordinate over the two days. The figure suggests that the coordinate at the same time of the day tends to be close between the two days. We show in Fig. 8(b) the distance between the time-averaged tie-decay networks at the same time of the two days in the embedding space. We find that, at any time of the day, the distance is substantially smaller than 156.07, which is the average distance between all the pairs of 30-minute windows. These results are similar to those with $\alpha = 10^{-2}$ (see Fig. 4). We carried out the same analysis for the Hospital temporal network, with the window size for averaging being an hour. The results shown in Fig. 9 are similar to those with $\alpha = 10^{-2}$ (see Fig. 5). In particular, the average day-to-day distance between the tie-decay networks at the same time of the day, shown in Fig. 9(b), is substantially smaller than the average distance between an arbitrary pair of networks (i.e., 61.01), and this tendency is stronger in the morning than in the afternoon. The variability of the day-to-day distance over the course of the day for each subnetwork of the Hospital temporal network is shown in Fig. 10. The results are similar to those with $\alpha = 10^{-2}$ (see Fig. 6). We conclude that our temporal network embedding method yields qualitatively similar trajectories between $\alpha = 10^{-3}$ and $\alpha = 10^{-2}$.

APPENDIX B RESULTS FOR THE GRAPH CLASSIFICATION WITH THE THREE-DIMENSIONAL EMBEDDING SPACE

We compare in Table 6 the accuracy of the different embedding methods in the graph classification task when the

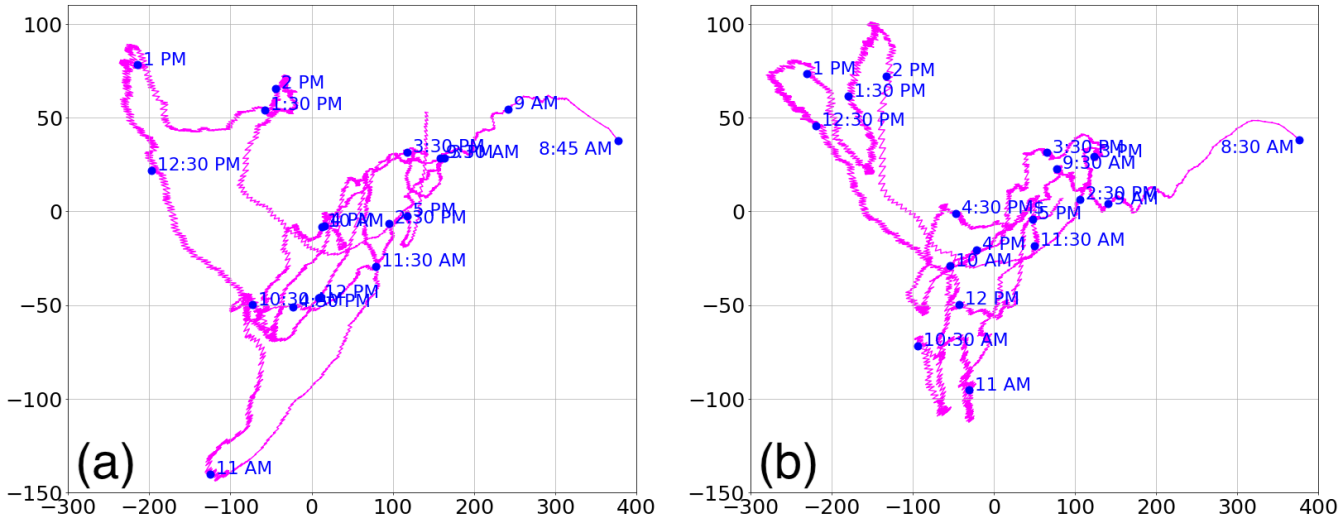

FIGURE 7: Trajectory of the tie-decay network for the Primary School data. (a) Day 1. (b) Day 2. We set $\alpha = 10^{-3}$. The dots represent the location of the trajectory every 30 minutes.

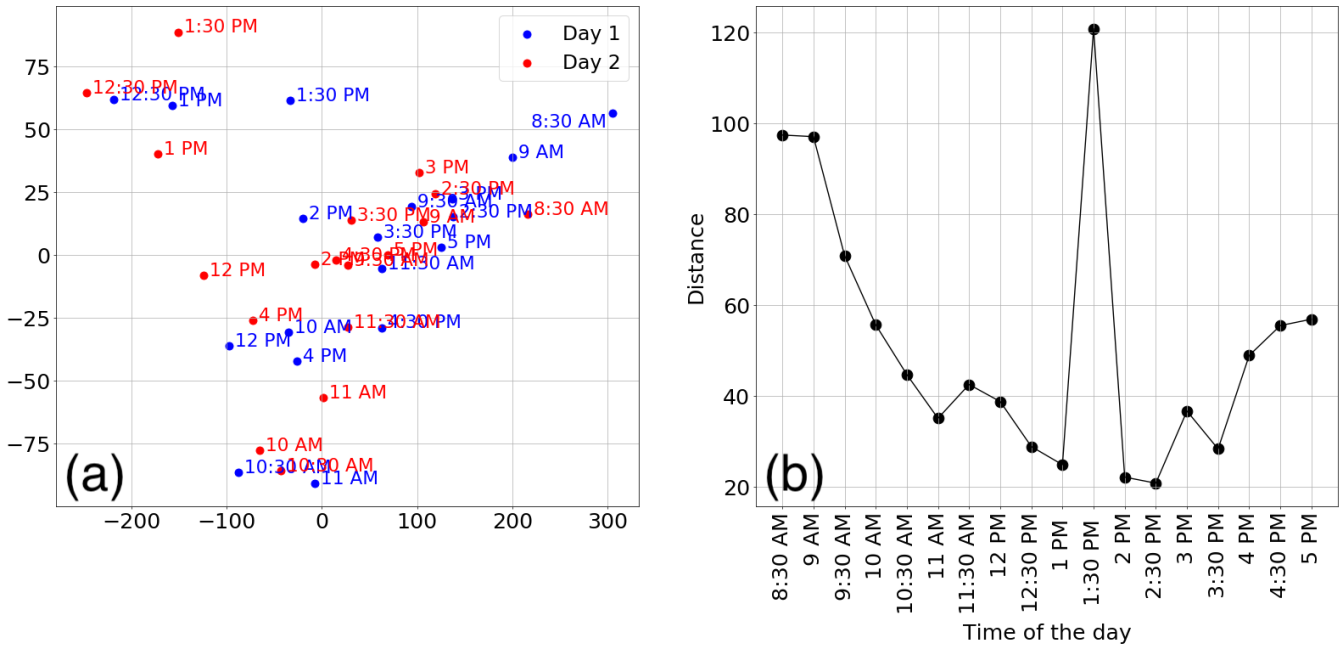

FIGURE 8: Comparison of the trajectory of the Primary School temporal network between the two days, with $\alpha = 10^{-3}$. (a) Average coordinate of the tie-decay network in the embedding space over every 30-minute time window. (b) Day-to-day distance between the networks at the same time of the day.

embedding dimension is three. The results are similar to those with the two-dimensional embedding space, shown in Table 3.

ACKNOWLEDGMENT

N.M. thanks to AFOSR European Office (under Grant No. FA9550-19-1-7024), National Science Foundation or NSF (under Grant No. DMS-2052720), the Sumitomo Foundation, and the Japan Science and Technology Agency (JST) for the financial support (under Grant No. JPMJMS2021). L.L. thanks the Canada Research Chairs program. We also thank SocioPatterns organization for providing the primary school and hospital ward data sets.

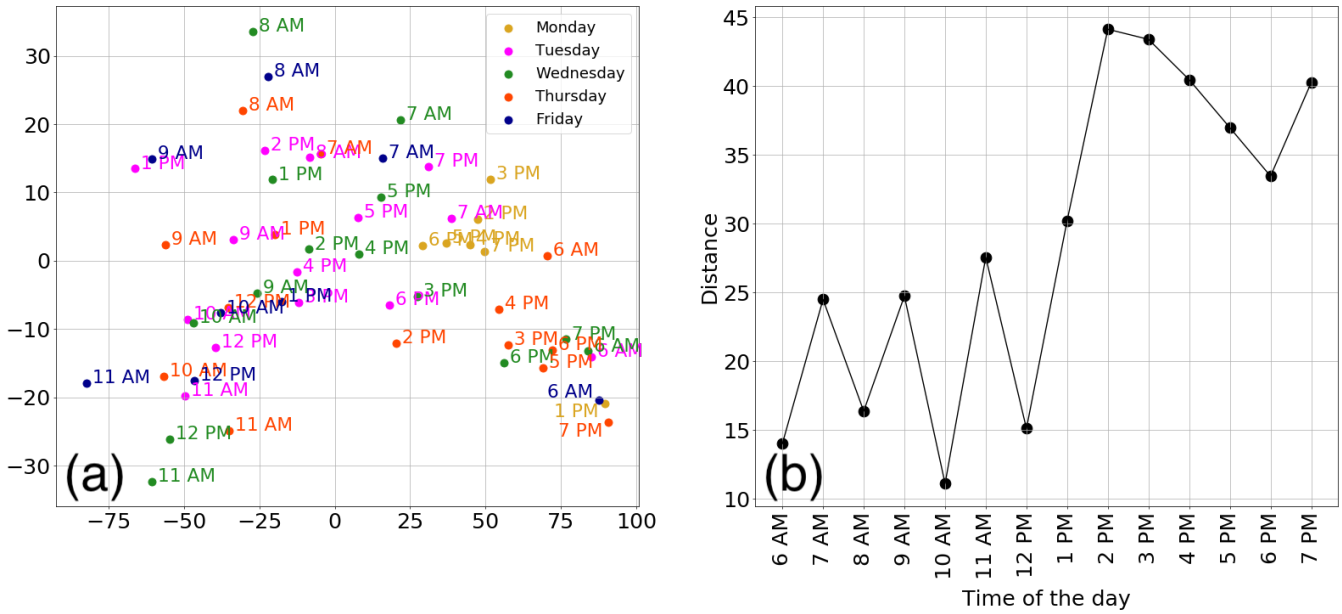

FIGURE 9: Comparison of the trajectory of the Hospital temporal network across the five days, with $\alpha = 10^{-3}$. (a) Average coordinate of the tie-decay network in the embedding space over every hourly time window. (b) Day-to-day distance between the networks at the same time of the day. We omit tie-decay networks before 6 AM and after 8 PM.

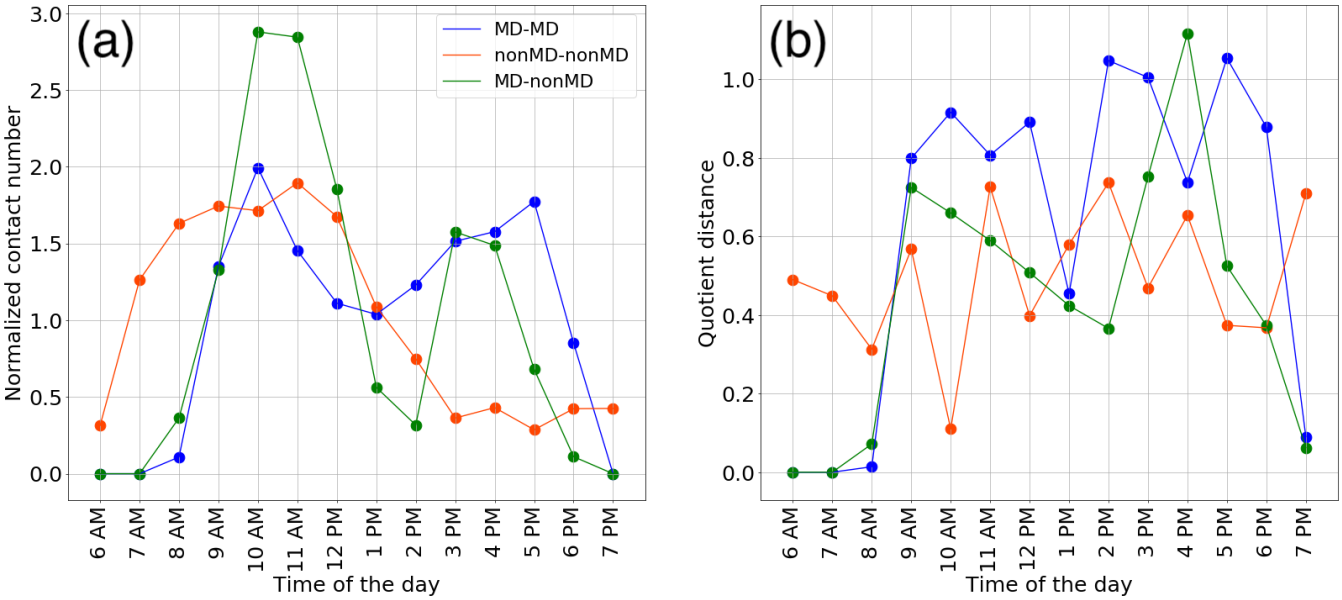

FIGURE 10: Comparison among temporal Hospital subnetworks with $\alpha = 10^{-3}$. (a) Normalized number of contacts over time, (b) Normalized day-to-day distance at the same time of the day for the different subnetworks. In (a), we use the same normalization as that used in Fig. 6(a).

TABLE 6: Mean and standard deviation of accuracy of the different embedding methods in the graph classification task when the embedding dimension is three. We calculated the mean and standard deviation on the basis of 10^4 iterations of the graph classification task.

Method	Mean	Standard deviation
Tie-decay with $\alpha = 10^{-2}$	0.88	0.10
Tie-decay with $\alpha = 10^{-3}$	0.77	0.14
PCA	0.84	0.13
t-SNE	0.67	0.15
MDS	0.86	0.11

REFERENCES

[1] M. Newman, *Networks*. Oxford Univ. Press, Oxford, UK, 2018.

[2] A. Barabási, *Network Science*. Cambridge Univ. Press, 2016.

[3] P. Holme and J. Saramäki, "Temporal networks," *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, 2012.

[4] P. Holme, "Modern temporal network theory: a colloquium," *Eur. Phys. J. B*, vol. 88, no. 9, pp. 1–30, 2015.

[5] N. Masuda and R. Lambiotte, *A Guide to Temporal Networks*. Europe: World Scientific, 2020.

[6] S. Bansal, J. Read, B. Pourbohloul, and L. A. Meyers, "The dynamic nature of contact networks in infectious disease epidemiology," *J. Biol. Dyn.*, vol. 4, no. 5, pp. 478–489, 2010.

[7] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying social group evolution," *Nature*, vol. 446, no. 7136, pp. 664–667, 2007.

[8] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3-5, pp. 75–174, 2010.

[9] M. Rosvall and C. T. Bergstrom, "Mapping change in large networks," *PLoS ONE*, vol. 5, no. 1, p. e8694, 2010.

[10] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," *Science*, vol. 328, no. 5980, pp. 876–878, 2010.

[11] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, 2009.

[12] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," *Data Min. Knowl. Discov.*, vol. 29, no. 3, pp. 626–688, 2015.

[13] D. Zambon, C. Alippi, and L. Livi, "Concept Drift and Anomaly Detection in Graph Streams," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5592–5605, Nov. 2018.

[14] N. Masuda and P. Holme, "Detecting sequences of system states in temporal networks," *Sci. Rep.*, vol. 9, no. 1, pp. 1–11, 2019.

[15] A. P. Baker, M. J. Brookes, I. A. Rezek, S. M. Smith, T. Behrens, P. J. P. Smith, and M. Woolrich, "Fast transient networks in spontaneous human brain activity," *eLife*, vol. 3, p. e01867, 2014.

[16] S. Ryali, K. Supekar, T. Chen, J. Kochalka, W. Cai, J. Nicholas, A. Padmanabhan, and V. Menon, "Temporal dynamics and developmental maturation of salience, default and central-executive network interactions revealed by variational bayes hidden markov modeling," *PLoS Comput. Biol.*, vol. 12, no. 12, p. e1005138, 12 2016.

[17] D. Vidaurre, R. Abeysuriya, R. Becker, A. J. Quinn, F. Alfaro-Almagro, S. M. Smith, and M. W. Woolrich, "Discovering dynamic brain networks from big data in rest and task," *NeuroImage*, vol. 180, pp. 646–656, 2018, brain Connectivity Dynamics.

[18] S. Cao, W. Lu, and Q. Xu, "Grarep: Learning graph representations with global structural information," in *Proc. the 24th ACM Int. Conf. Inf. Knowl. Manag.*, 2015, pp. 891–900.

[19] S. Cavallari, V. W. Zheng, H. Cai, K. C.-C. Chang, and E. Cambria, "Learning community embedding with community detection and node embedding on graphs," in *Proc. the ACM Conf. Inf. Knowl. Manag.*, 2017, pp. 377–386.

[20] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proc. the 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2014, pp. 701–710.

[21] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. the 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 855–864.

[22] C. Wang, C. Wang, Z. Wang, X. Ye, and P. S. Yu, "Edge2vec: Edge-based social network embedding," *ACM Trans. Knowl. Discov. Data*, vol. 14, no. 4, pp. 1–24, 2020.

[23] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *Proc. Int. Conf. Mach. Learn.* PMLR, 2016, pp. 2071–2080.

[24] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *J. Mach. Learn. Res.*, vol. 9, no. 11, 2008.

[25] A. Agovic, A. Banerjee, A. Ganguly, and V. Protopopescu, "Anomaly detection using manifold embedding and its applications in transportation corridors," *Intell. Data Anal.*, vol. 13, no. 3, pp. 435–455, 2009.

[26] J. Li, H. Dani, X. Hu, J. Tang, Y. Chang, and H. Liu, "Attributed network embedding for learning in a dynamic environment," in *Proc. the ACM Conf. Inf. Knowl. Manag.*, 2017, pp. 387–396.

[27] T. Li, J. Zhang, S. Y. Philip, Y. Zhang, and Y. Yan, "Deep dynamic network embedding for link prediction," *IEEE Access*, vol. 6, pp. 29 219–29 230, 2018.

[28] G. H. Nguyen, J. B. Lee, R. A. Rossi, N. K. Ahmed, E. Koh, and S. Kim, "Continuous-time dynamic network embeddings," in *Proc. TheWebConf*, 2018, pp. 969–976.

[29] P. Goyal, N. Kamra, X. He, and Y. Liu, "Dyngem: Deep embedding method for dynamic graphs," 2018, arXiv:1805.11273.

[30] T. K. Saha, T. Williams, M. A. Hasan, S. Joty, and N. K. Varberg, "Models for capturing temporal smoothness in evolving networks for learning latent representation of nodes," 2018, arXiv:1804.05816.

[31] W. Ahmad, M. A. Porter, and M. Beguerisse-Díaz, "Tie-decay networks in continuous time and eigenvector-based centralities," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1759–1771, 2021.

[32] V. De Silva and J. B. Tenenbaum, "Sparse multidimensional scaling using landmark points," Stanford Univer-

sity, Tech. Rep., 2004.
[33] S. L. France and J. D. Carroll, "Two-way multidimensional scaling: A review," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 5, pp. 644–661, 2010.
[34] I. Borg and P. J. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer, 2005.
[35] R. P. Duin and E. Pekalska, *Dissimilarity Representation for Pattern Recognition, Foundations and Applications*. World Scientific, 2005, vol. 64.
[36] A. Gisbrecht and F.-M. Schleif, "Metric and non-metric proximity transformations at linear costs," *Neurocomputing*, vol. 167, pp. 643–657, 2015.
[37] Y. Bengio, J. F. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet, "Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 177–184.
[38] L. Van Der Maaten, E. Postma, J. Van den Herik *et al.*, "Dimensionality reduction: a comparative review," *J. Mach. Learn. Res.*, vol. 10, no. 66-71, p. 13, 2009.
[39] V. De Silva and J. B. Tenenbaum, "Global versus local methods in nonlinear dimensionality reduction," in *Proc. Adv. Neural Inf. Process. Syst.*, 2003, pp. 721–728.
[40] J. C. Platt, "FastMap, MetricMap, and Landmark MDS are all Nyström algorithms," *Proc. the 10th Int. Workshop Artif. Intell. Stat.*, pp. 261–268, 2005.
[41] C. Donnat and S. Holmes, "Tracking network dynamics: A survey using graph distances," *Ann. Math. Stat.*, vol. 12, no. 2, pp. 971–1012, 2018.
[42] R. C. Wilson and P. Zhu, "A study of graph spectra for comparing graphs and trees," *Pattern Recognit.*, vol. 41, no. 9, pp. 2833–2841, 2008.
[43] N. Wicker, C. H. Nguyen, and H. Mamitsuka, "A new dissimilarity measure for comparing labeled graphs," *Linear Algebra Its Appl.*, vol. 438, no. 5, pp. 2331–2338, 2013.
[44] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with bregman divergences," *J. Mach. Learn. Res.*, vol. 6, pp. 1705–1749, 2005.
[45] T. Graepel, R. Herbrich, P. Bollmann-Sdorra, and K. Obermayer, "Classification on pairwise proximity data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 11. MIT Press, 1998, pp. 438–444.
[46] E. Pekalska, P. Paclik, and R. P. Duin, "A generalized kernel approach to dissimilarity-based classification," *J. Mach. Learn. Res.*, vol. 2, pp. 175–211, 2001.
[47] J. Graffelman, "Goodness-of-fit filtering in classical metric multidimensional scaling with large datasets," *J. Appl. Stat.*, vol. 47, no. 11, pp. 2011–2024, 2020.
[48] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.
[49] V. Gemmetto, A. Barrat, and C. Cattuto, "Mitigation of infectious disease at school: targeted class closure vs school closure," *BMC Infect. Dis.*, vol. 14, no. 1, p. 695, 2014.
[50] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina *et al.*, "High-resolution measurements of face-to-face contact patterns in a primary school," *PLoS ONE*, vol. 6, no. 8, p. e23176, 2011.
[51] P. Vanhems, A. Barrat, C. Cattuto, J.-F. Pinton, N. Khanafer, C. Régis, B.-a. Kim, B. Comte, and N. Voirin, "Estimating potential infection transmission routes in hospital wards using wearable proximity sensors," *PLoS ONE*, vol. 8, no. 9, p. e73970, 2013.
[52] S. M. Kazemi, R. Goel, K. Jain, I. Kobyzev, A. Sethi, P. Forsyth, and P. Poupart, "Representation learning for dynamic graphs: A survey." *J. Mach. Learn. Res.*, vol. 21, no. 70, pp. 1–73, 2020.
[53] A. Taheri, K. Gimpel, and T. Berger-Wolf, "Learning to represent the evolution of dynamic graphs with recurrent models," in *Proc. TheWebConf*, 2019, pp. 301–307.
[54] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," in *Proc. TheWebConf*, 2015, pp. 1067–1077.
[55] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Assoc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, 2007.
[56] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," *Physica A*, vol. 390, no. 6, pp. 1150–1170, 2011.
[57] M. A. Hasan and M. J. Zaki, "A survey of link prediction in social networks," in *Social Network Data Analytics*. Springer, Boston, MA, 2011, pp. 243–275.
[58] D. Zhang, J. Yin, X. Zhu, and C. Zhang, "Network representation learning: A survey," *IEEE Trans. Big Data*, vol. 6, no. 1, pp. 3–28, 2018.
[59] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," *Knowl.-Based Syst.*, vol. 151, pp. 78–94, 2018.
[60] K. Sugishita and N. Masuda, "Recurrence in the evolution of air transport networks," *Sci. Rep.*, vol. 11, no. 1, pp. 1–15, 2021.
[61] T. P. Peixoto and L. Gauvin, "Change points, memory and epidemic spreading in temporal networks," *Sci. Rep.*, vol. 8, p. 15511, 2018.
[62] S. Ranshous, S. Shen, D. Koutra, S. Harenberg, C. Faloutsos, and N. F. Samatova, "Anomaly detection in dynamic networks: a survey," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 7, no. 3, pp. 223–247, 2015.
[63] M. A. Lopes, J. Zhang, D. Krzemiński, K. Hamandi, Q. Chen, L. Livi, and N. Masuda, "Recurrence quantification analysis of dynamic brain networks," *Eur. J. Neurosci.*, vol. 53, no. 4, pp. 1040–1059, 2021.
[64] A. Talwalkar, S. Kumar, and H. Rowley, "Large-scale manifold learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
[65] A. W. Long and A. L. Ferguson, "Landmark diffusion maps (L-dMaps): Accelerated manifold learning out-

of-sample extension," *Appl. Comput. Harmon. Anal.*, vol. 47, pp. 190–211, 2019.

[66] L. Chennuru Vankadara and U. von Luxburg, "Measures of distortion for machine learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.

[67] H. Hartle, B. Klein, S. McCabe, A. Daniels, G. St-Onge, C. Murphy, and L. Hébert-Dufresne, "Network comparison and the within-ensemble graph distance," in *Proc. R. Soc. A*, vol. 476, no. 2243, 2020, p. 20190744.

[68] P. Wills and F. G. Meyer, "Metrics for graph comparison: a practitioner's guide," *PLoS ONE*, vol. 15, no. 2, p. e0228728, 2020.

[69] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo *et al.*, "Dynamic functional connectivity: promise, issues, and interpretations," *NeuroImage*, vol. 80, pp. 360–378, 2013.

[70] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cereb. Cortex*, vol. 24, no. 3, pp. 663–676, 11 2012.

[71] V. D. Calhoun, R. Miller, G. Pearlson, and T. Adalı, "The chronnectome: time-varying connectivity networks as the next frontier in fmri data discovery," *Neuron*, vol. 84, no. 2, pp. 262–274, 2014.

[72] M. Lucas *et al.*, "Inferring cell cycle phases from a partially temporal network of protein interactions," *Cell Rep. Methods*, vol. 3, no. 2, p. 100397.

CHANON THONGPRAYOON received a Bachelor of Science in Mathematics from Chulalongkorn University in June 2019 and a Master of Arts in Mathematics from State University of New York, Buffalo in February 2023. He is currently a Ph.D. student in the Department of Mathematics at State University of New York, Buffalo. His research interests lie on the area of network science, particularly, network embedding.

LORENZO LIVI received the B.Sc. degree and M.Sc. degree from the Department of Computer Science, Sapienza University of Rome, Italy, in 2007 and 2010, respectively, and the Ph.D. degree from the Department of Information Engineering, Electronics, and Telecommunications at Sapienza University of Rome, in 2014. He has been with the ICT industry during his studies. From January 2014 until April 2016, he was a Post-Doctoral Fellow at Ryerson University, Toronto, Canada. From May 2016 until September 2016, he was a Post-Doctoral Fellow at the Politecnico di Milano, Italy and Universita' della Svizzera Italiana, Lugano, Switzerland. Currently, he is an Associate Professor with the Department of Computer Science at the University of Manitoba, Canada. He is also an Honorary Senior Lecturer in Data Science at the Department of Computer Science, University of Exeter, United Kingdom. In November 2018, Dr. Livi was awarded the prestigious Tier 2 Canada Research Chair in Complex Data. He is a Senior (Associate) Editor of the IEEE Transactions on Neural Networks and Learning Systems and a regular reviewer for several international journals and conferences. His research interests include Machine Learning, Time Series Analysis and Complex Dynamical Systems, with focused applications in Systems Biology and Computational Neuroscience.

NAOKI MASUDA received his PhD in 2002 from the University of Tokyo. He worked as Lecturer and then Associate Professor at the University of Tokyo, Japan, between 2006 and 2014. Then, he worked as Senior Lecturer and Associate Professor at the University of Bristol, UK, between 2014 and 2019. He moved to Department of Mathematics at University at Buffalo in 2019 as Associate Professor and has been full Professor since 2021. His research interests include network science and mathematical biology.