



# Topological similarity of time-dependent objects

Chun-Xiao Nie

Received: 28 May 2022 / Accepted: 30 August 2022 / Published online: 19 September 2022  
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

**Abstract** Objects with time factor are important representations of data, such as time series, sequences generated by chaotic systems, and dynamic networks. This paper focuses on characterizing the topological similarity of such objects. First, we introduce time-dependent object (*TDO*), which is time-dependent datasets with metric structures. Second, we extend the concept of topological correlation coefficient so that it can analyze time-dependent objects. The generalized topological correlation coefficient (*GTCC*) can be well defined on *TDO* and characterize the topological similarity between datasets. Third, we analyze examples constructed from chaotic systems and temporal networks. Calculations show that there are non-trivial topological similarities in both datasets, where surrogate *TDOs* provide benchmark values for comparison. In particular, *GTCC* can identify patterns in the temporal network arising from human behavior. The analytical framework presented in this paper has the potential to be widely used in time series analysis, nonlinear dynamical systems, and dynamic networks.

**Keywords** Complex network · Time-dependent object · Topological correlation coefficient

## 1 Introduction

Detecting correlation and causality between time series is a long-standing topic. A large number of detection tools have been proposed to analyze the relationship between observations. For example, the classical Pearson correlation coefficient can detect a linear relationship between observations [1]. Since data sets extracted from complex systems often include nonlinear dependence and causality, previous studies have developed a large number of nonlinear detection tools. For example, *DCCA* (Detrended Cross-Correlation Analysis) provides a powerful analytical tool for analyzing nonlinear relationships between time series [2]. In addition, some studies have extended some important indicators based on *DFA* (Detrended Fluctuation Analysis) and *DCCA*, such as detrended cross-correlation coefficient, multifractal cross-correlation analysis and so on [3,4]. For high-dimensional data, Székely et al. proposed distance correlation to characterize the dependencies between random vectors in Euclidean space [5]. In recent years, some tools based on information theory focus on detecting nonlinear information flow and nonlinear relationship between variables [6–8]. In addition to classical statistical methods and methods based on information theory, some emerging methods have been proposed in recent years. For example, some studies have shown that complex network theory can be effectively applied to analyze time series. We can represent time series through a network and study the

---

C. Nie (✉)  
School of Statistics and Mathematics, Zhejiang Gongshang University, Xuezheng Street, Hangzhou 310018, Zhejiang, China  
e-mail: niechunxiao2009@163.com

characteristics of the original series. Common methods for connecting time series and networks include visibility graphs [9], recurrence networks [10], etc [11]. Furthermore, Nie constructed *TCC* (Topological Correlation Coefficient) to detect nonlinear dependencies between time series [12]. This method identifies nonlinear dependence by detecting structural similarities between  $k$ -nearest neighbor networks of time series, where the network sequence extracts the multi-scale structure of the distance matrix. Since *TCC* is essentially defined on the distance matrix, it can be generalized to a wider range of objects, not just time series. One of the topics of this study is to define *TCC* on time-dependent objects with distance matrices.

Time-dependent objects include not only time series, but also some other forms of datasets, such as dynamic networks or temporal networks. As a powerful tool for characterizing the dynamics of complex systems, dynamic networks usually consist of a series of network snapshots that describe the system [13]. In empirical analysis, such networks with dynamic features are often represented as some labeled static networks, thus formally represented as some network snapshots. Some researchers name such networks as snapshot networks [13], which are commonly used representations of temporal networks [14, 15]. Both snapshot networks and time series label data series based on timestamps, the difference being that the former has a more complex structure than the latter. For example, network structures exhibit multi-scale features, resulting in rich micro-, meso-, and macro-structures [16]. The multi-scale structure makes it a difficult task to study the time-varying features of dynamic networks.

Similar to analyzing time series, an important topic is to detect correlations or dependencies in dynamic networks. In recent years, some researchers have tried to construct detection tools from different perspectives. For example, Tang et al. [17] introduced the temporal-correlation coefficient of the dynamic network. Sugishita et al. [18] established the autocorrelation function on the dynamic network and introduced the corresponding Fourier transform. Valdano et al. [19] used Jaccard similarity to characterize the similarity between adjacent snapshots. Recently, Nie defined the Hurst exponent of dynamic networks through the distance matrix of snapshot networks [20]. These tools characterize persistence in dynamic networks from different perspectives, and thus are part of the broader topic, persistence of complex systems [21].

Both time series and dynamic networks include time factors, making it possible to study them from a general perspective. In this paper, we introduce a new concept, time-dependent object (*TDO*), to include these important special cases, and extend the topological correlation coefficient to be defined on *TDO* [12]. We organize this article as follows. First, we introduce the definition of time-dependent object and give several important examples. Second, we define *GTCC* on *TDO* and construct surrogate *TDO*. Third, in the results section, we analyze the data generated by the Hénon map and a temporal network dataset.

## 2 Method

### 2.1 Time-dependent object

Data sets with time factors include some objects with complex structure or high nonlinearity, such as dynamic networks and chaotic time series. In particular, these data are often equipped with suitable metrics so that the structure can be analyzed from the distance matrix of the dataset. This leads us to introduce the following definition.

We define a time-dependent object (*TDO*) as a collection of time-tagged sets  $O(\Theta, d_{\text{def}})$  ( $\Theta = \{\Theta_t\}$ ), and it is equipped with distance  $d_{\text{def}}$ . The relationship between these sets can be represented by a distance matrix  $D_{\Theta} = [d_{\text{def}}(\Theta_{t_1}, \Theta_{t_2})]$ . It should be noted that different *TDOs* can be defined on the same set by adjusting the definition of  $d_{\text{def}}$ . We will see later that on some complex objects, the definition of distance can lead to significant changes in the properties of *TDO*.

We can directly give the following three special cases. First, if  $\Theta_t \in R$ , and  $d_{\text{def}}(\Theta_{t_1}, \Theta_{t_2}) = \text{abs}(\Theta_{t_1} - \Theta_{t_2})$ , then  $O$  is a time series with classical Euclidean distance. Here, the symbol “*abs*” represents an absolute value function. Second, if  $\Theta$  is a sequence of vectors generated by a nonlinear discrete system, and the vectors are equipped with Euclidean distances, then  $O$  is the orbit generated by the nonlinear system. For example, for the phase space generated by the Hénon map [22],  $O$  is a two-dimensional vector set with Euclidean distance. Third, *TDO* can also be used to analyze some complex discrete objects. For example, snapshot networks are often used as representations of dynamic networks. A snapshot network is a sequence of static networks with time stamps. Network distance can char-

acterize the structural similarity between these static networks [23]. To sum up, a snapshot network with network distance is naturally a *TDO*.

### 2.2 Graph distance

We choose two special graph distances here, Jaccard distance and Laplace spectral distance [23]. The two distances characterize the difference between the two networks from the local structure and the global structure, respectively. Below, we consider networks  $W_1(V, e_1)$  and  $W_2(V, e_2)$ , where the networks share a set of nodes ( $V = \{v_i, i = 1, 2, \dots, n\}$ ). The edge sets are  $e_1$  and  $e_2$  respectively, then the Jaccard distance is as in Eq. (1) [24]. Here, the symbol ‘‘Card’’ represents the cardinality. The part  $\frac{Card(e_1 \cap e_2)}{Card(e_1 \cup e_2)}$  in this definition is the Jaccard similarity, and the value is in the interval  $[0, 1]$ .

$$d_{Jaccard}(W_1, W_2) = 1 - \frac{Card(e_1 \cap e_2)}{Card(e_1 \cup e_2)} \tag{1}$$

If the normal Laplace spectrum of  $W_1$  and  $W_2$  are  $\lambda^1 = \{\lambda^1(1) \geq \lambda^1(2) \geq \dots \geq \lambda^1(n)\}$  and  $\lambda^2 = \{\lambda^2(1) \geq \lambda^2(2) \geq \dots \geq \lambda^2(n)\}$ , respectively, the  $l_p$  spectral distance is defined as Eq. (2) [23]. Here, the symbol  $p$  is a parameter for adjusting the distance. In this study, we define  $p = 2$  and the spectral distance is the classical Euclidean distance. Since the spectrum of the network is an isomorphic invariant, the distance between isomorphic networks is 0. In particular, it is not necessary to consider whether there is a correspondence between the node sets of  $W_1(V, e_1)$  and  $W_2(V, e_2)$  when defining the spectral distance.

$$d_{Laplace}(W_1, W_2) = \left[ \sum_{k=1}^n abs(\lambda^1(k) - \lambda^2(k))^p \right]^{\frac{1}{p}} \tag{2}$$

### 2.3 *kNN* filtering

Below, we extend the steps proposed by previous studies and define a *kNN* (*k*-nearest neighbour) filtering sequence for *TDO* [12]. For a specified  $k$ , we can calculate the  $k$ -nearest neighbor network of *TDO* via the distance matrix [25]. Here, each set  $\Theta_t$  corresponds to a node  $v_t$ . From the distance  $d_{def}$ , the set of neighbors of each node  $v_t$  can be calculated. We denote the neighbor set of node  $v_t$  as  $N_t^k$ . First, we construct a network including  $n$  nodes. Then, if  $v_i \in N_t^k$ , a

directed link  $v_t \rightarrow v_i$  is added between  $v_t$  and  $v_i$ . Finally, for each node, directed links are added between it and its neighbors. We adjust  $k$  to generate a network sequence  $\{W^\Theta(V, E_j), j = 1, 2, \dots, n - 1\}$  ( $V = \{v_i, i = 1, 2, \dots, n\}$ ). As  $k$  increases, the set  $N_t^k$  includes more neighbors, resulting in Eq. (3).

$$E_1 \subseteq E_2 \subseteq \dots \subseteq E_j \dots \subseteq E_{n-1} \tag{3}$$

The *kNN* network is constructed by the relationship between the neighbors, making it have an important invariance. For  $d_{def1}$  and  $d_{def2}$ , if the elements of the distance matrix  $D_\Theta^2 = [d_{def2}(\Theta_{t_1}, \Theta_{t_2})]$  can be generated by the elements of  $D_\Theta^1 = [d_{def1}(\Theta_{t_1}, \Theta_{t_2})]$  and a monotonically increasing function  $f(d_{def2}(\Theta_{t_1}, \Theta_{t_2}) = f(d_{def1}(\Theta_{t_1}, \Theta_{t_2})))$ , then the two distance definitions correspond to the same filtering sequence.

### 2.4 *GTCC*

Similar to time series analysis, an important topic is to study the relationship between a pair of *TDOs*. In linear time series analysis, patterns in the data can be well captured by models such as *ARMA*, so that theoretical relationships between model-driven series can be well analyzed. If the set  $\Theta$  is represented by a sequence of sparse matrices or high-dimensional vectors, such as a snapshot network, it is difficult to construct a model to capture the dynamics in *TDO*. In particular, it is also difficult to construct a suitable model for general *TDO* as in linear time series analysis. However, by extracting the neighbor relations in the distance matrix, we can characterize the topological similarity between *TDOs*.

Below, we define general topological correlation coefficient (*GTCC*) between *TDOs* by *kNN* filtering sequences [12]. Except that a  $d_{def}$  needs to be specified, the steps of *GTCC* are the same as those of calculating *TCC* of time series. We assume there are two *TDOs* ( $O(\Theta_1, d_{def,1})$ ,  $O(\Theta_2, d_{def,2})$ ) and  $card(\Theta_1) = card(\Theta_2) = n$ . If *kNN* filtering is implemented for each *TDO*, two network sequences are generated, denoted as  $W^{\Theta_1} = \{W_k^{\Theta_1}(V_1, E_k^{\Theta_1})\}$  and  $W^{\Theta_2} = \{W_k^{\Theta_2}(V_2, E_k^{\Theta_2})\}$ , respectively. For each pair of networks  $W_k^{\Theta_1}(V_1, E_k^{\Theta_1})$  and  $W_k^{\Theta_2}(V_2, E_k^{\Theta_2})$ , we can calculate the Jaccard similarity  $J(k) = \frac{Card(E_k^{\Theta_1} \cap E_k^{\Theta_2})}{Card(E_k^{\Theta_1} \cup E_k^{\Theta_2})}$  to characterize the structural similarity between them. In this way, there are  $n - 1$  Jaccard similarity values between a pair of filtered sequences.

The  $TCC$  of  $TDO$  is defined as the mean of the Jaccard similarity between two sequences:  $\frac{1}{n-1} \sum_{k=1}^{n-1} J(k)$ . Since the  $kNN$  filtering sequence extracts the nearest neighbors of  $\Theta$  from multiple scales, it represents the topological structure of the dataset. In this way,  $GTCC$  characterizes the topological similarity between two  $TDOs$ .

For large values of  $n$ , we can set  $\rho_k = \frac{k}{n-1}$  and replace  $J(k)$  with  $J(\rho_k)$ . The  $TCC$  value can be approximated by an integral (Eq. 4), where  $J(\rho)$  can be replaced by a fitted polynomial function. In addition, we can choose  $l$  values at equal intervals within  $[1, n-1]$  and calculate the average value of  $\{J(k), k = 1, 1+2l, \dots, ml\}$ . Here, the positive  $m$  is equal to  $[(n-1)/l]$ , and  $l$  is a parameter that controls the length of the interval. Later, if we use the approximate calculation method, the value of  $l$  is set. Since the Jaccard similarity value is in the interval  $[0, 1]$ , in Eq. (4),  $J(\rho) \leq 1$ . Furthermore, due to the relation Eq. (3), the function  $J(\rho)$  is a monotonically increasing function and takes the value 1 when  $k = n-1$ . These properties of  $J(\rho)$  allow it to be estimated well and do not need to consider convergence.

$$TCC_{OTD} = \int_0^1 J(\rho) d\rho \quad (4)$$

$TCC$  is a special case of  $GTCC$ , i.e.  $\Theta_t$  is a univariate time series and  $d_{\text{def}}$  is the Euclidean distance defined over real numbers. In time series analysis, autocorrelation is an important analysis indicator, such as it can be used to identify the order of moving average process [26]. For a  $TDO$ , we can also detect the time-delay similarity of its structure through  $TCC$ , where the  $TCC$  with time lag  $\tau$  is defined on  $O_1(\{\Theta_t, t = 1, 2, \dots, n-\tau\}, d_{\text{def}})$  and  $O_2(\{\Theta_t, t = \tau+1, \dots, n\}, d_{\text{def}})$ . Next, we use  $TCC_\tau$  to denote the delay similarity.

## 2.5 Surrogate $TDO$

$TDO$  includes a time factor, and in some cases  $\Theta_t$  has a complex structure, eg.  $\Theta_t$  is a graph or network. In order to analyze the characteristics of  $TDO$  in detail, we introduce surrogate  $TDO$  for constructing significance test.

Based on previous studies, we can construct surrogate  $TDOs$  for some special cases. For time series,

there have been some classical methods to construct surrogate time series [27]. These methods remove or maintain the dynamics inherent in time series from different perspectives. For example, the random permutation ( $RP$ ) method scrambles the observations, thereby providing a way to completely eliminate the original dynamics [28]. In previous studies, the  $RP$  method has been effectively applied to the significance analysis of  $TCC$  [12].

Similar to time series, there are various methods for generating benchmark networks in temporal network theory [29]. These methods provide multiple benchmarks of the original network by disrupting the network structure or timeline. A common approach is sequence shufflings, which destroy temporal correlations by shuffling the order of network snapshots [29]. This method randomly shuffles the timestamps of the temporal network, so that the structural relationship between adjacent snapshots is destroyed. It can be found that this method has formal similarities with the  $RP$  method for time series.

In general, we can use the shuffling method to construct a  $TDO$  for comparison. This is a generalized extension of the  $RP$  method and the sequence shuffling method. For a  $TDO$ , we scramble the timestamp  $t$  of  $\Theta = \{\Theta_t\}$ , and transform the distance matrix  $D_\Theta = [d_{\text{def}}(\Theta_{t_1}, \Theta_{t_2})]$  accordingly. Here, the transformation does not change the elements of the matrix, but only shuffles the positions, and the main diagonal elements are equal to 0, and it is a symmetric matrix. The shuffling method only changes the order of timestamps, but does not change  $\Theta$ . To sum up, for a  $TDO$ , this method can be used to test whether it is time-dependent, so the null hypothesis is that the time-tagged set  $(\{\Theta_t\})$  is time-independent. In particular, this null hypothesis corresponds to the distance  $d_{\text{def}}$ .

## 3 Results

Since previous studies have systematically analyzed various types of time series data using  $TCC$  [12], here we only analyze the  $TDOs$  of chaotic systems and temporal networks.

### 3.1 TDO generated by Hénon map

#### 3.1.1 Structural similarity analysis between TDOs generated by Hénon map

We use the Hénon map to generate two sets of observations with parameters  $a = 1.4$  and  $b = 0.3$  (Eq. 5) [22]. We set the initial values to  $(0, 0)$  and  $(0.01, 0.01)$  to generate two TDOs, and mark them with  $H_1$  and  $H_2$  respectively. Each set of observations consists of 1000 two-dimensional vectors and is assigned a Euclidean distance, resulting in two TDOs ( $\Theta_t = (x_t, y_t)$ ). The TCC value between the two TDOs is 0.3871.

$$\begin{aligned} x_{t+1} &= -ax_t^2 + y_t + 1, \\ y_{t+1} &= bx_t. \end{aligned} \tag{5}$$

Next, we test the significance of TCC [12]. We generate a surrogate TDO from each original TDO, calculate the TCC value and denote it as  $TCC_{rand}$  ( $l = 2$ ). Here, we repeat the above steps 600 times and generate 600 TCC values. Figure 1 is a frequency histogram of  $TCC_{rand}$  values, including four descriptive statistics. We then perform a Jarque-Bera test on these observations. The test report shows that the statistic  $\chi = 5.3224$  and the  $p$ -value = 0.06987, indicating that the null hypothesis cannot be rejected at the significance level  $\alpha = 0.05$ . The test results support that this set of observations comes from a normal population. Finally, we calculate the Z-score,  $Z = \frac{0.3871 - 0.3869}{0.0014} = 0.1645$ . A Z-score less than 1.96 indicates that the TCC value is within the interval  $[ave - 1.96std, ave + 1.96std]$  ( $\alpha = 0.05$ ) and thus not significant. The TCC value is not significantly distinguishable from the baseline values, so there is no significant time-dependent structural similarity between the two TDOs. In this example, a chaotic system with the same parameters generates two vector sequences with no significant time-dependent topological similarity, thus demonstrating the sensitivity of TCC to initial values.

#### 3.1.2 Time-delay similarity of TDO

Below, we calculate the  $TCC_\tau$  values for each TDO, where the order  $\tau$  of the lags is chosen from 1 to 20. In addition, we generated surrogate TDOs, denoted by  $H_1^s$  and  $H_2^s$ , respectively, and then calculated  $TCC_\tau$  values. Here, we use an approximate calculation method, where the parameter  $l = 2$ . Figure 2

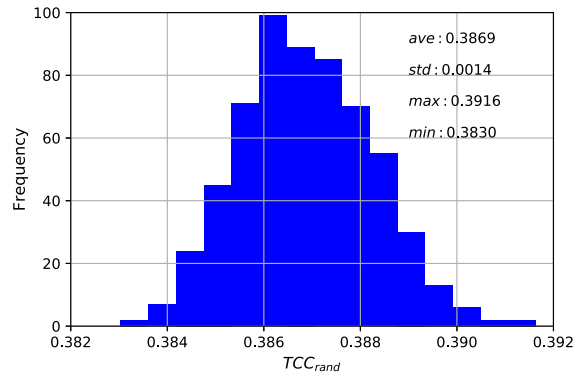


Fig. 1 Frequency histogram of  $TCC_{rand}$  values. The symbol ‘ave’ represents the average value

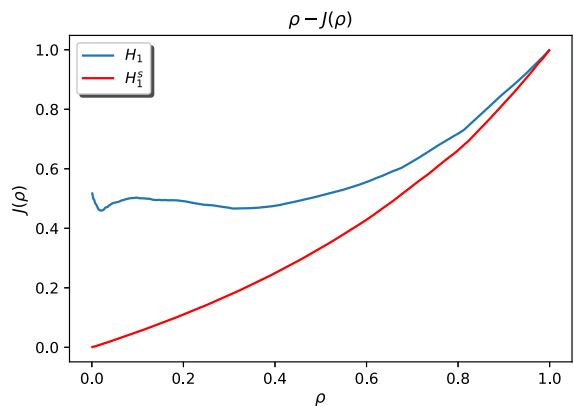
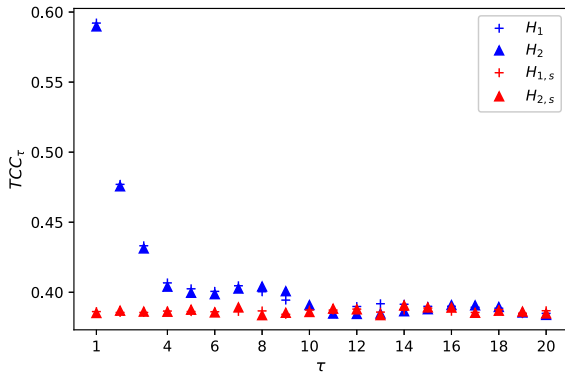


Fig. 2 Most of the  $J(\tau)$  values of  $H_1$  are greater than those of  $H_1^s$ , and both sequences of  $J(\tau)$  values tend to 1 as  $\tau$  increases

shows the  $J(\tau)$  curves for  $H_1$  and  $H_1^s$ , where  $\tau = 1$ . The two curves can be distinguished significantly and result in a difference between the  $TCC_1$  values. The  $TCC_\tau$  of  $H_1$  is 0.5921 and that of  $H_1^s$  is 0.3862, indicating that the dataset generated by Eq. (5) includes structural similarity and is significantly different from the benchmark dataset.

Figure 3 shows the sequence of TCC values for these two TDOs, where the abscissa corresponds to the lag order. As  $\tau$  increases, the  $TCC_\tau$  value decays, and the  $TCC_\tau$  values of the same lag order are always close to each other. In addition, we generated surrogate TDOs, denoted by  $H_1^s$  and  $H_2^s$ , respectively, and then calculated  $TCC_\tau$  values. In Fig. 3, we also identify the lagged TCC values of the TDO constructed based on the shuffling method. Table 1 lists the four  $TCC_\tau$  value sequences. Comparing the  $TCC_\tau$  values of  $H_1$  and  $H_1^s$ , it can be found that there is a significant differ-



**Fig. 3**  $TCC_\tau$  sequences of the two  $TDO$ s generated by Hénon map ( $\tau = 1, 2, \dots, 20$ )

**Table 1** The sequence of  $\tau$  values for  $TDO$  generated by the Hénon map

$\tau$	$H_1$	$H_1^s$	$H_2$	$H_2^s$
1	0.5921	0.3862	0.5897	0.3852
2	0.4770	0.3856	0.4756	0.3868
3	0.4332	0.3856	0.4313	0.3862
4	0.4067	0.3866	0.4040	0.3861
5	0.4025	0.3862	0.3996	0.3875
6	0.4006	0.3861	0.3986	0.3856
7	0.4046	0.3863	0.4026	0.3892
8	0.4004	0.3867	0.4041	0.3836
9	0.3944	0.3843	0.4007	0.3853
10	0.3882	0.3873	0.3909	0.3859
11	0.3855	0.3873	0.3848	0.3883
12	0.3898	0.3880	0.3845	0.3878
13	0.3918	0.3858	0.3848	0.3837
14	0.3915	0.3872	0.3864	0.3906
15	0.3900	0.3888	0.3879	0.3892
16	0.3900	0.3864	0.3909	0.3890
17	0.3893	0.3856	0.3906	0.3853
18	0.3886	0.3876	0.3895	0.3868
19	0.3848	0.3850	0.3855	0.3862
20	0.3849	0.3868	0.3838	0.3849

ence between the two  $TCC_1$  values, and the difference between the  $TCC_\tau$  values of  $H_1$  and  $H_1^s$  is small when  $\tau \geq 9$ . Similar results can be observed for  $H_2$  versus  $H_2^s$ .

Table 1 shows that  $GTCC$  can detect hidden linear and nonlinear relationships in orbital sequences generated by Hénon system. Here, the variables  $x_t$  and

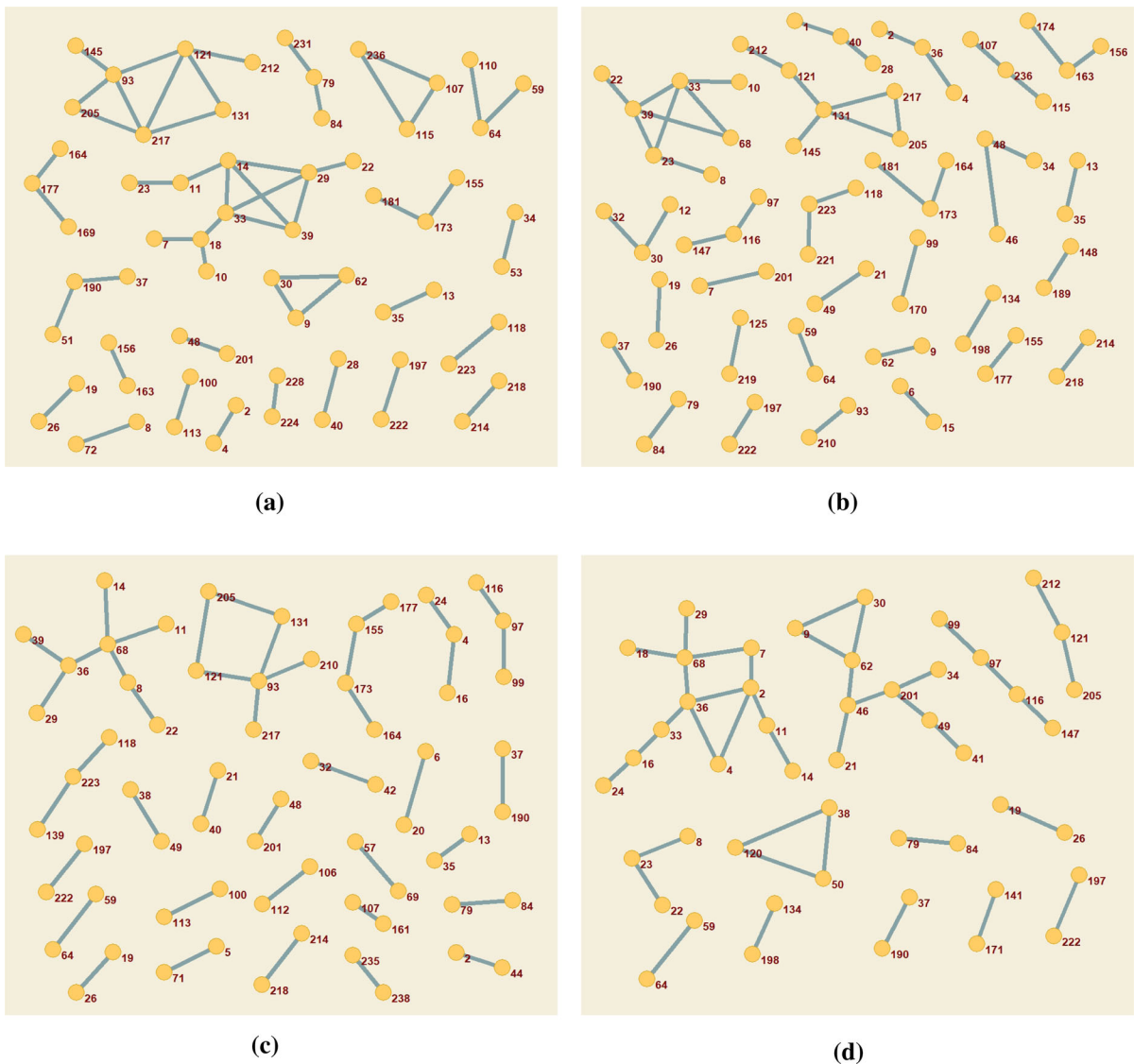
$y_t$  are represented by a quadratic function and a linear function, respectively, and dependencies defined by functions are broken in surrogate  $TDO$ . Furthermore, when  $\tau$  is large enough, the level of dependence decreases near the baseline. In particular, although the  $TCC$  between the two  $TDO$ s was insignificant, the decay rates of their  $TCC_\tau$  values exhibited similar patterns.

### 3.2 Temporal network

The temporal network analyzed in this study is represented by a snapshot network, where the time interval for extracting snapshots is 20s [30,31]. Each snapshot is constructed from face-to-face data between 232 students and 10 teachers. The dataset includes data for two periods, October 1, 2009 and October 2, 2009, and are labeled  $P_1$  and  $P_2$  below [30,31]. We specify the  $d_{def}$  of  $TDO$  using two definitions: Jaccard distance, spectral distance. Moreover, we use  $P_{1,s}$  and  $P_{2,s}$  to denote the surrogate  $TDO$  generated from the two periods of data, respectively.

#### 3.2.1 Time dependencies of snapshot network structures

Below, we show several snapshots of this temporal network, as shown in Fig. 4. Figure 4a–d correspond to snapshots No. 214–No. 217 on the first day, respectively. The four snapshots only include some small-sized subgraphs, and most nodes have no links to other nodes. We only show the nodes with links here, where the four networks contain 64, 74, 65 and 48 nodes respectively. Some links appeared in different snapshots, such as  $v_{173} - v_{181}$  in Fig. 4a, b, and  $v_{36} - v_{68}$  in Fig. 4c, d. However, most of the links in one snapshot were eliminated in other snapshots. The Jaccard distance matrix (Eq. 6) exhibits the similarity between the four snapshots, where all off-diagonal elements are greater than 0.7. The Jaccard distance depends on a one-to-one correspondence between node sets of different snapshots. Spectral distance captures the differences between networks globally, eliminating the need to specify this correspondence. There is a significant difference between the calculated results of the local structure and the global structure. Equation (7) shows the spectral distance between the four snapshots.



**Fig. 4** Four snapshots of the temporal network, where a–b correspond to No. 214–No. 217, respectively

The two distance definitions lead to different comparison results. For example,  $d_{Jaccard}(1, 2) < d_{Jaccard}(1, 3)$  and  $d_{Laplace}(1, 3) < d_{Laplace}(1, 2)$ . The analysis here is consistent with previous studies that the dynamics in temporal networks can be observed from different structural scales [20,23]. To sum up, the comparative analysis shows that there are significant differences between the local and global structures of network

snapshots, so that  $d_{def}$  needs to be specified when calculating  $TCC$ .

$$d_{Jaccard} = \begin{pmatrix} 0 & 0.7625 & 0.8333 & 0.8590 \\ 0.7625 & 0 & 0.8312 & 0.8267 \\ 0.8333 & 0.8312 & 0 & 0.8732 \\ 0.8590 & 0.8267 & 0.8732 & 0 \end{pmatrix} \quad (6)$$

$$d_{Laplace} = \begin{pmatrix} 0 & 2.3525 & 1.5559 & 2.6413 \\ 2.3525 & 0 & 1.3966 & 4.2158 \\ 1.5559 & 1.3966 & 0 & 3.6260 \\ 2.6413 & 4.2158 & 3.6260 & 0 \end{pmatrix} \quad (7)$$

Next, we first calculate the  $TCC_\tau$  values when  $d_{\text{def}}$  is the Jaccard distance ( $l = 2$ ). Figure 5a plots the  $TCC_\tau$  values for the two periods, where the difference between the two series is small, suggesting a time-dependent similarity between the structures of the two  $TDOs$ . Figure 5a also plots the  $TCC_\tau$  value of  $P_{i,s}$  ( $i = 1, 2$ ). For a  $\tau$  value, we find that the  $TCC_\tau$  values of the two periods are close to each other. However, all  $TCC_\tau$  values were greater than the baseline values generated by the surrogate  $TDOs$ . The table in the appendix lists the  $TCC_\tau$  values for the two periods. There are only minor differences between the  $TCC_\tau$  series of  $P_1$  and  $P_2$ . In particular,  $TCC_\tau$  values calculated from real data are always significantly larger than the baseline values obtained from surrogates.

The similarity of decaying features suggests a similarity between the dynamics hidden in the two temporal networks. Since the Jaccard distance is defined by the local structure, the similarity in the calculated results stems from the fact that human behavior exhibits similar patterns in the two periods.

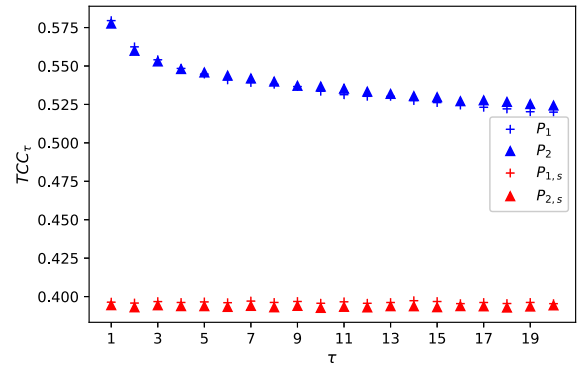
Second, we calculate the  $TCC_\tau$  value when  $d_{\text{def}}$  is the spectral distance ( $l = 2$ ). Figure 5b plots the  $TCC_\tau$  values of  $TDO$  with  $d_{\text{def}}$  defined by Laplace spectral distance. We find a significant difference between the  $TCC_\tau$  values of the two periods. The table in the appendix lists the  $TCC_\tau$  series. For the case where  $d_{\text{def}}$  is the spectral distance, the result is similar to the case of the Jaccard distance. For each  $\tau$  value, the  $TCC_\tau$  value also deviates significantly from the baseline value.

Comparing Fig. 4 with Fig. 5, we find that the analysis results using the perspective of the global structure are significantly different from those of the local structure. Figures 4 and 5 suggest that there is a high similarity between the underlying local structural dynamics of the two periods, while there are significant differences between the global structural dynamics.

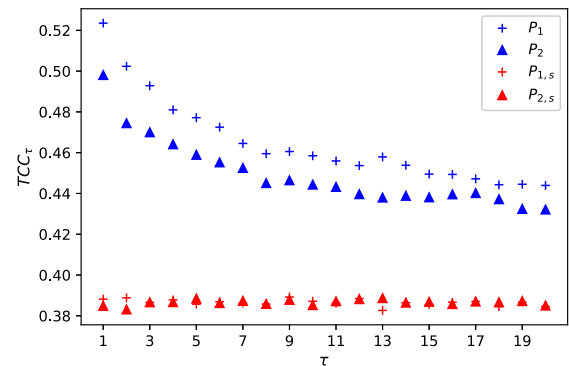
In this example, we found that adjusting distance on the same set  $\Theta$  can cause the analysis results to vary significantly. For complex networks, the structure is multi-scale, such as including local, meso and macroscopic structures [16]. Different distance definitions lead to changes in calculation results is essentially a manifestation of this feature.

### 3.2.2 $TCC$ -based dynamics of temporal networks

The changes of the local structure and the global structure are described by the Jaccard distance matrix



(a)



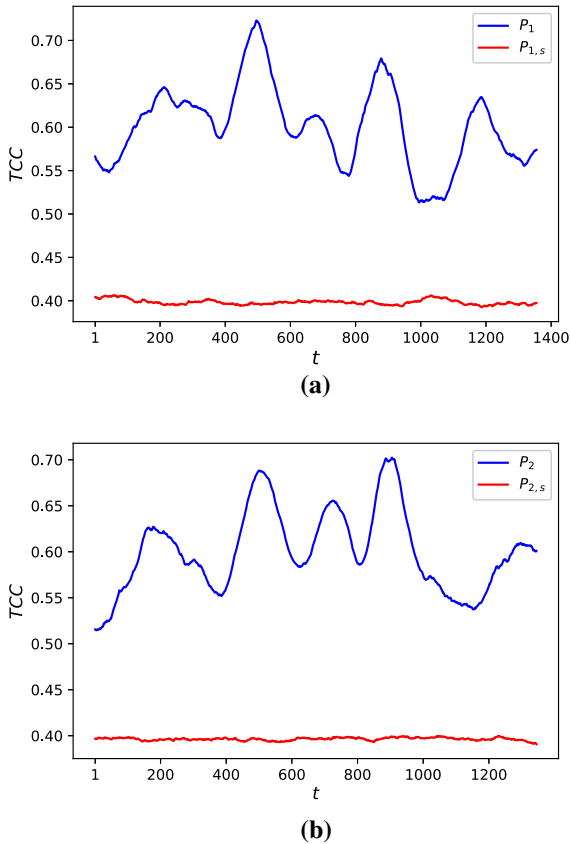
(b)

**Fig. 5**  $TCC_\tau$  series for two periods of the temporal network. Panel a plots the  $TCC_\tau$  series with  $d_{\text{def}}$  as the Jaccard distance, where there is only a slight difference between the series of the two periods. Panel b plots  $TCC_\tau$  series with  $d_{\text{def}}$  as the spectral distance, and there is a significant difference between  $P_1$  and  $P_2$

and the spectral distance matrix, respectively.  $TCC$  extracts the neighbor relations hidden in the distance matrix, so that the structural dynamics of snapshot networks can be studied by sliding calculation window. In order to observe the structural dynamics of  $TDO$  in detail, below, we set the calculation window and slide the window to calculate the sequence of  $TCC_\tau$  values, where  $\tau = 1$ . For window  $w$ , we extract two  $TDOs$ ,  $O_1(\{\Theta_t, t = i, i + 1, \dots, i + w - 1\}, d_{\text{def}})$  and  $O_2(\{\Theta_t, t = i + 1, i + 2, \dots, i + w\}, d_{\text{def}})$ , and detect the  $TCC_\tau$  value between them. Here, we choose  $w = 200$ . From this method, we can observe the details of network dynamics by changes in topological similarity.

Figure 6 shows the  $TCC_1$  sequence with  $d_{\text{def}}$  as the Jaccard distance (blue line). Some peaks and valleys can be clearly observed, implying that the topological similarity varies with time. These peaks and valleys





**Fig. 6** The  $TCC_\tau$  series calculated by the sliding calculation window (Jaccard distance), where the window length is 200 and  $\tau = 1$ . Figures **a** and **b** correspond to the  $P_1$  and  $P_2$  periods, respectively

suggest that the temporal network has structural critical times. This is confirmed by the benchmark value calculated based on the surrogate  $TDO$  (red line). Figure 6 shows that the baseline value changes slightly and is significantly smaller than the  $TCC_1$  value of the original temporal network. Furthermore, the blue lines in Fig. 6a, b exhibit similar patterns. For example, both periods included 5 peaks, suggesting similarities between the underlying dynamics.

Table 2 lists the descriptive statistics of the four  $TCC_1$  series calculated by sliding windows. The mean values of  $TCC_1$  series at both  $P_1$  and  $P_2$  were significantly larger than the mean values of baseline series. Furthermore, the minimum values are all larger than the maximum values of the baseline series. In particular, the standard deviation of each series is greater than ten times the standard deviation of the benchmark series. The significant difference in standard deviations

**Table 2** Descriptive statistics for the  $TCC_1$  series (Jaccard distance)

$TDO$	<i>Ave</i>	<i>Std</i>	<i>Max</i>	<i>Min</i>
$P_1$	0.6013	0.0470	0.7228	0.5133
$P_2$	0.6008	0.0444	0.7024	0.5146
$P_1^s$	0.3984	0.0030	0.4065	0.3925
$P_2^s$	0.3964	0.0016	0.4000	0.3907

*Ave* average, *Std* standard deviation

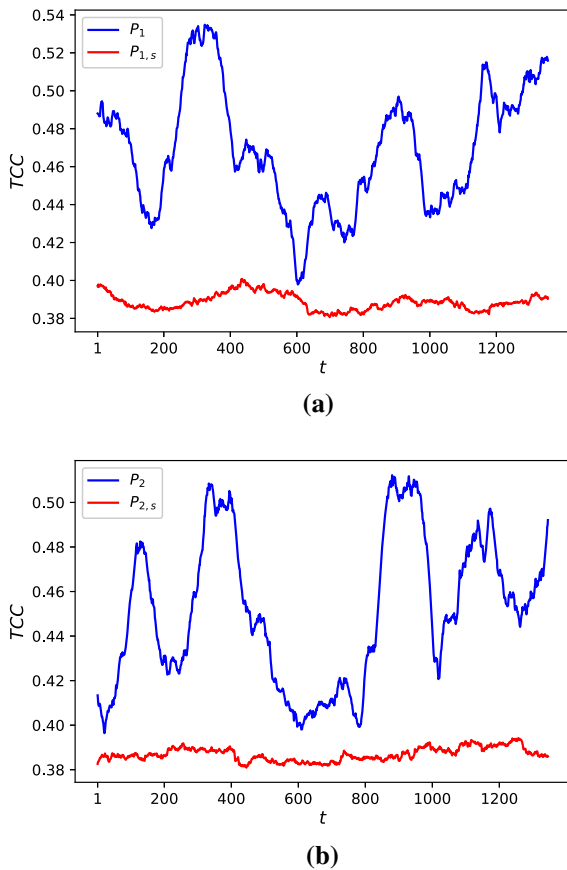
**Table 3** Descriptive statistics for the  $TCC_1$  series (spectral distance)

$TDO$	<i>Ave</i>	<i>Std</i>	<i>Max</i>	<i>Min</i>
$P_1$	0.4689	0.0317	0.5348	0.3979
$P_2$	0.4513	0.0332	0.5122	0.3964
$P_1^s$	0.3886	0.0043	0.4008	0.3806
$P_2^s$	0.3870	0.0030	0.3942	0.3809

is essentially due to the presence of non-trivial dynamics in  $TDOs$ .

We also calculated  $TCC_1$  series with  $d_{def}$  as spectral distance (Fig. 7). All  $TCC_1$  values of the original  $TDO$  are greater than those of the surrogate  $TDO$ . Comparing Fig. 6 with Fig. 7, it can be observed that the difference in the definition of distance results in significant changes in the sequence of  $TCC_1$  values. Similar to Fig. 6, in Fig. 7, the  $TCC_1$  series of  $P_i$  ( $i = 1, 2$ ) also includes some local peaks and valleys. This implies that critical times of dramatic changes are still included in the global structural dynamics. Table 3 shows the descriptive statistics calculated based on the sliding window, in which the mean values corresponding to  $P_1$  and  $P_2$  are significantly smaller than those in Table 2. In addition, the standard deviations corresponding to  $P_i^s$  ( $i = 1, 2$ ) are also significantly smaller than those of  $P_i$  ( $i = 1, 2$ ), thus supporting that the global structural dynamics is non-trivial.

In this example, we analyze the structural dynamics of two periods by  $TCC$  and find that the structural similarity exhibits complex patterns when  $\{\Theta_t\}$  is a sequence of network snapshots. Such complexity requires equipping different distances ( $d_{def}$ ) for analyzing structural dynamics from multiple perspectives.



**Fig. 7** The figure shows the  $TCC_\tau$  sequence where  $d_{\text{def}}$  is the spectral distance. Figures **a** and **b** correspond to  $P_1$  and  $P_2$  respectively, the calculation window length is 200, and  $\tau = 1$

## 4 Discussion and conclusions

### 4.1 Discussion

For special objects, such as time series or temporal networks, there are several ways to construct a surrogate  $TDO$ . For example, autoregressive ( $AR$ ) surrogates in nonlinear time series analysis assume that the data is described by an  $AR$  model [27]. Fourier transform ( $FT$ ) surrogates are used to test whether a data set has nonlinear structure from the perspective of power spectrum [27]. For a general  $TDO$ , we can check whether its structure is time-dependent by shuffling method. A related topic is how to construct other surrogates to examine the structure of  $TDO$  in detail.

### 4.2 Conclusion

We introduced  $TDO$  for generalized analysis of objects with time factors.  $TDO$  makes it possible to develop tools to analyze data objects from different research fields, such as time series, chaos dynamics, and dynamic networks. The definition of  $TDO$  includes two elements, time and distance. By combining these two elements, we can calculate the time-dependent structural similarity between  $TDOs$  by the  $kNN$  filtering method. The  $kNN$  filter sequence extracts the hidden multi-scale topology in the dataset through the distance matrix of  $TDO$ .  $GTCC$  captures the difference between two  $kNN$  filter sequences globally, thereby characterizing the topological similarity of two  $TDOs$ .

To demonstrate the effectiveness of  $GTCC$ , we analyzed a chaotic system and a temporal network using  $GTCC$ . The examples demonstrate that nonlinear dependencies in chaotic systems can be identified by  $GTCC$ , as well as topological similarity in temporal networks. In particular, by  $GTCC$ , we found rich dynamics in  $TDOs$  constructed from networks. For example, we observed local peaks and valleys in all  $TCC_1$  series, and two  $TCC_1$  sequences defined by the same  $d_{\text{def}}$  had similar patterns. To sum up, this study provides an analytical framework for insight into datasets with time factors.

**Acknowledgements** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Data availability** All data generated or analysed during this study, if not included in this published article, are available from the corresponding author on reasonable request.

### Declarations

**Conflict of interest** The author declares that he have no conflict of interest. The author has no relevant financial or non-financial interests to disclose.

## Appendix

Tables 4 and 5 show the calculation results of  $TCC_\tau$  of temporal network. The values in the columns corresponding to  $P_1^s$  and  $P_2^s$  are calculated based on surrogates. The  $TCC_\tau$  values of the real data decrease as  $\tau$  increases and are all significantly larger than the baseline values.

**Table 4** The series of  $TCC_\tau$  values for the temporal network, where  $d_{def}$  is defined by the Jaccard distance

$\tau$	$P_1$	$P_1^s$	$P_2$	$P_2^s$
1	0.5795	0.3964	0.5776	0.3944
2	0.5625	0.3958	0.5599	0.3930
3	0.5540	0.3967	0.5529	0.3943
4	0.5484	0.3962	0.5480	0.3938
5	0.5448	0.3965	0.5458	0.3937
6	0.5410	0.3960	0.5437	0.3933
7	0.5392	0.3971	0.5419	0.3939
8	0.5381	0.3962	0.5398	0.3930
9	0.5366	0.3968	0.5371	0.3939
10	0.5336	0.3957	0.5366	0.3926
11	0.5313	0.3966	0.5352	0.3933
12	0.5303	0.3956	0.5333	0.3929
13	0.5302	0.3962	0.5318	0.3937
14	0.5275	0.3973	0.5304	0.3937
15	0.5262	0.3968	0.5297	0.3931
16	0.5246	0.3954	0.5271	0.3938
17	0.5231	0.3962	0.5277	0.3936
18	0.5221	0.3955	0.5266	0.3928
19	0.5202	0.3962	0.5252	0.3935
20	0.5200	0.3954	0.5243	0.3943

**Table 5** The series of  $TCC_\tau$  values for the temporal network, where  $d_{def}$  is defined by the spectral distance

$\tau$	$P_1$	$P_1^s$	$P_2$	$P_2^s$
1	0.5235	0.3882	0.4981	0.3849
2	0.5024	0.3888	0.4745	0.3831
3	0.4928	0.3866	0.4700	0.3866
4	0.4810	0.3878	0.4641	0.3867
5	0.4772	0.3857	0.4590	0.3884
6	0.4725	0.3869	0.4553	0.3862
7	0.4645	0.3863	0.4525	0.3874
8	0.4595	0.3858	0.4451	0.3858
9	0.4606	0.3892	0.4465	0.3878
10	0.4585	0.3871	0.4443	0.3853
11	0.4559	0.3862	0.4432	0.3871
12	0.4537	0.3885	0.4397	0.3882
13	0.4579	0.3826	0.4380	0.3887
14	0.4538	0.3864	0.4389	0.3865
15	0.4495	0.3855	0.4381	0.3869
16	0.4493	0.3867	0.4396	0.3858
17	0.4471	0.3870	0.4402	0.3870
18	0.4443	0.3845	0.4372	0.3867
19	0.4445	0.3867	0.4324	0.3872
20	0.4439	0.3845	0.4321	0.3851

**References**

- Asuero, A.G., Sayago, A., González, A.G.: The correlation coefficient: an overview. *Crit. Rev. Anal. Chem.* **36**(1), 41–59 (2006)
- Podobnik, B., Stanley, H.E.: Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. *Phys. Rev. Lett.* **100**(8), 084102 (2008)
- Oświęcimka, P., Drożdż, S., Forczek, M., Jadach, S., Kwapiień, J.: Detrended cross-correlation analysis consistently extended to multifractality. *Phys. Rev. E* **89**(2), 023305 (2014)
- Kwapiień, J., Oświęcimka, P., Drożdż, S.: Detrended fluctuation analysis made flexible to detect range of cross-correlated fluctuations. *Phys. Rev. E* **92**(5), 052815 (2015)
- Székely, G.J., Rizzo, M.L., Bakirov, N.K.: Measuring and testing dependence by correlation of distances. *Ann. Stat.* **35**(6), 2769–2794 (2007)
- Schreiber, T.: Measuring information transfer. *Phys. Rev. Lett.* **85**(2), 461–464 (2000)
- Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M., Sabeti, P.C.: Detecting novel associations in large data sets. *Science* **334**(6062), 1518–1524 (2011)
- Berret, T.B., Samworth, R.J.: Nonparametric independence testing via mutual information. *Biometrika* **106**(3), 547–566 (2019)
- Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuno, J.C.: From time series to complex networks: the visibility graph. *Proc. Natl. Acad. Sci.* **105**(13), 4972–4975 (2008)
- Marwan, N., Romano, M.C., Thiel, M., Kurths, J.: Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**(5–6), 237–329 (2007)
- Zou, Y., Donner, R.V., Marwan, N., Donges, J.F., Kurths, J.: Complex network approaches to nonlinear time series analysis. *Phys. Rep.* **787**, 1–97 (2019)
- Nie, C.-X.: Nonlinear correlation analysis of time series based on complex network similarity. *Int. J. Bifurc. Chaos* **30**(15), 2050225 (2020)
- Rossetti, G., Cazabet, R.: Community discovery in dynamic networks: a survey. *ACM Comput. Surv. (CSUR)* **51**(2), 1–37 (2018)
- Holme, P.: Modern temporal network theory: a colloquium. *Eur. Phys. J. B* **88**(9), 1–30 (2015)
- Holme, P., Saramäki, J.: 1. In: Holme, P., Saramäki, J. (eds.) *Temporal network theory*, vol. 2, pp. 1–24. Springer, Gewerbestrasse 11, 6330 Cham, Switzerland (2019)
- Zanin, M., Papo, D., Sousa, P.A., Menasalvas, E., Nicchi, A., Kubik, E., Boccaletti, S.: Combining complex networks and data mining: why and how. *Phys. Rep.* **635**, 1–44 (2016)
- Tang, J., Scellato, S., Musolesi, M., Mascolo, C., Latora, V.: Small-world behavior in time-varying graphs. *Phys. Rev. E* **81**(5), 055101 (2010)
- Sugishita, K., Masuda, N.: Recurrence in the evolution of air transport networks. *Sci. Rep.* **11**(1), 1–15 (2021)
- Valdano, E., Poletto, C., Giovannini, A., Palma, D., Savini, L., Colizza, V.: Predicting epidemic risk from past temporal contact data. *PLoS Comput. Biol.* **11**(3), 1004152 (2015)
- Nie, C.-X.: Hurst analysis of dynamic networks. *Chaos Interdiscipl. J. Nonlinear Sci.* **32**(2), 023130 (2022)

21. Salcedo-Sanz, S., Casillas-Pérez, D., Del Ser, J., Casanova-Mateo, C., Cuadra, L., Piles, M., Camps-Valls, G.: Persistence in complex systems. *Phys. Rep.* **957**, 1–73 (2022)
22. Hénon, M.: A two-dimensional mapping with a strange attractor. *Commun. Math. Phys.* **50**, 69–77 (1976)
23. Donnat, C., Holmes, S.: Tracking network dynamics: a survey using graph distances. *Ann. Appl. Stat.* **12**(2), 971–1012 (2018)
24. Levandowsky, M., Winter, D.: Distance between sets. *Nature* **234**(5323), 34–35 (1971)
25. Hautamäki, V., Kärkkäinen, I., Fränti, P.: Outlier detection using k-nearest neighbour graph. In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3, pp. 430–433 (2004). IEEE
26. Chatfield, C., Xing, H.: *The Analysis of Time Series?: An Introduction with R*, 7th edn., p. 49. CRC Press, Boca Raton (2019)
27. Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V., Stefanovska, A.: Surrogate data for hypothesis testing of physical systems. *Phys. Rep.* **748**, 1–60 (2018)
28. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J.D.: Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena* **58**(1–4), 77–94 (1992)
29. Gauvin, L., Génois, M., Karsai, M., Kivela, M., Takaguchi, T., Valdano, E., Vestergaard, C.L.: Randomized reference models for temporal networks. [arXiv:1806.04032](https://arxiv.org/abs/1806.04032) (2018)
30. Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.-F., Quaghiotto, M., Van den Broeck, W., Régis, C., Lina, B.: High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE* **6**(8), 23176 (2011)
31. Gemmetto, V., Barrat, A., Cattuto, C.: Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC Infect. Dis.* **14**(1), 1–10 (2014)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.