



PAPER • OPEN ACCESS

## Time series synchronization in cross-recurrence networks: uncovering a homomorphic law across diverse complex systems

To cite this article: Shijia Song and Handong Li 2024 *New J. Phys.* **26** 013044

View the [article online](#) for updates and enhancements.

You may also like

- [Quantum homomorphic broadcast multi-signature based on homomorphic aggregation](#)

Xin Xu, , Ai-Han Yin et al.

- [Eikonal solutions for moment hierarchies of chemical reaction networks in the limits of large particle number](#)

Eric Smith and Supriya Krishnamurthy

- [Search for an Isotropic Gravitational-wave Background with the Parkes Pulsar Timing Array](#)

Daniel J. Reardon, Andrew Zic, Ryan M. Shannon et al.

**PAPER****OPEN ACCESS**RECEIVED  
2 August 2023REVISED  
18 December 2023ACCEPTED FOR PUBLICATION  
11 January 2024PUBLISHED  
24 January 2024Original content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.

# Time series synchronization in cross-recurrence networks: uncovering a homomorphic law across diverse complex systems

Shijia Song and Handong Li\* 

School of Systems Science, Beijing Normal University, Beijing 100875, People's Republic of China

\* Author to whom any correspondence should be addressed.

E-mail: [lhd@bnu.edu.cn](mailto:lhd@bnu.edu.cn)**Keywords:** time series analysis, cross-recurrence network, event synchronicity, complex systems**Abstract**

Exploring the synchronicity between time series, especially the similar patterns during extreme events, has been a focal point of research in academia. This is due to the fact that such special dependence occurring between pairs of time series often plays a crucial role in triggering emergent behaviors in the underlying systems and is closely related to systemic risks. In this paper, we investigate the relationship between the synchronicity of time series and the corresponding topological properties of the cross-recurrence network (CRN). We discover a positive linear relationship between the probability of pairwise time series event synchronicity and the corresponding CRN's clustering coefficient. We first provide theoretical proof, then demonstrate this relationship through simulation experiments by coupled map lattices. Finally, we empirically analyze three instances from financial systems, Earth's ecological systems, and human interactive behavioral systems to validate that this regularity is a homomorphic law in different complex systems. The discovered regularity holds significant potential for applications in monitoring financial system risks, extreme weather events, and more.

## 1. Introduction

Time series data mining has long been a widely researched topic [1–4]. One of its subfields focuses on evaluating the synchronization and dependence between time series, assessing the degree to which a given time series is similar or correlated with another. Considering individual time series as elements in a system, the dependence between elements often leads to emergent behaviors in the system [5–8]. Particularly, the dependence or synchronous patterns between the extreme values in time series deserve much more attention as it is often associated with revealing systemic risks [9–11]. Traditional methods for analyzing the relationship between two time series typically employ linear measures such as Euclidean distance and Pearson correlation coefficient. However, time series in real complex systems often exhibit nonlinear similarity and dependence, and the complex dynamical mechanisms underlying time series are difficult to capture solely based on values of a single observation. Therefore, there is a compelling demand for investigating more comprehensive nonlinear measures for time series in order to unveil the synchronicity between pairwise time series, particularly during occurrences of extreme events. Traditional linear measures exhibit notable constraints in this regard, whereas techniques such as recurrence plots (RPs) and complex networks serve as potent nonlinear approaches for elucidating the intricate dynamical mechanisms inherent in time series. The former enables the extraction of underlying dynamical properties in short or non-stationary time series, while the latter facilitates the mapping of time series onto networks and the representation of original sequence attributes using network statistics.

Eckmann *et al* first introduced the concept of RPs, highlighting the recurrence of states as a fundamental attribute of every dynamic system [12]. Based on this attribute, RP was developed as a two-dimensional visualization tool. Over the past two decades, RP has evolved into a nonlinear method for describing complex dynamics [13]. RP represents a graphical representation of a binary symmetric matrix, encoding the times when two states are very close (i.e. neighbors in phase space). The core idea is to reconstruct the

attractor from the time series using the delay embedding method [14]. If the metric distance between two points on the reconstructed attractor is less than or equal to a threshold, it is considered a state reoccurred. All recurrence of trajectory points can be represented by a two-dimensional graph, with each point indicating whether the corresponding trajectory point has a recurrence. RP provides abundant information about the underlying dynamical properties of the system. By analyzing the recurrence matrix, information about the system's dynamics can be extracted and quantified using techniques like recurrence quantification analysis (RQA) [15]. Additionally, RP-based techniques can analyze short and non-stationary data, making them highly applicable to studying real-world data [16]. Although there is not a precise definition in academia for the length of so-called 'short' time series, recurrence-based methods have found extensive application in fields facing data acquisition challenges like geology, paleoclimatology, physiology, among others. Research in these areas, employing simulation and empirical evidence, has demonstrated the effectiveness and robustness of recurrence-based methods when analyzing short time series with lengths greater than 100 but not exceeding 500 [12, 15, 17–19]. Therefore, this paper will also employ empirical studies using sample sequences within this range of lengths.

In recent years, significant progress has been made in developing methods for complex system analysis based on RP. However, since RP is typically used to analyze the recurrence patterns of a single time series, further extensions are required to analyze the similar patterns between pairwise time series. Marwan *et al* proposed the concept of cross-RPs (CRP) [13], which simultaneously embeds two time series into phase space and compares their dynamic behaviors [20]. The cross-RP (CRP) displays all times when the state of a dynamic system occurs simultaneously in the second dynamic system, providing a two-dimensional cross-recurrence matrix. In other words, CRP shows all the times when the phase space trajectory of the first system is roughly the same as the trajectory in the phase space of the second system. Similarly, quantification tools based on CRP, such as cross-RQA (CRQA), can measure how and to what extent two time series exhibit similar patterns. This analysis framework was initially developed and widely used in natural sciences, such as heart rate variability, seismology, and chemical fluctuations, among other fields [13, 20]. In psychology, it has found extensive applications in the field of motor control [21–23]. Richardson *et al* provided a broader context for the method in the domains of dynamical systems and psychology [24]. It has also been applied to capture dynamic patterns between individuals, for example, uncovering interaction behaviors during goal-oriented tasks [25] or conversations [26]. In these applications, the time series of interest can be factorial data describing body sway or eye movement states, as well as numerical data such as heart rate [21, 26, 27].

Due to the ability of complex networks to capture both local and global properties, they have become instrumental in understanding the complex relationships and information flow among different components in extended systems [16]. Complex networks have gained considerable popularity in analyzing complex, particularly spatially extended systems [28, 29]. Additionally, because the CRP plot provides an adjacency matrix, it can serve as a basis for constructing a complex network. Therefore, by utilizing the cross-RP as an intermediary, it becomes possible to map pairwise nonlinear time series onto a network, enabling the use of network analysis tools instead of traditional methods for time series analysis. Donner *et al* demonstrated the fundamental relationship between the topological properties of the recurrence network and the statistical properties of attractor densities in the underlying dynamical system [30]. This complements the existing RQA by incorporating network descriptions as new quantitative features of the dynamic complexity of time series. Recurrence networks and related statistical measurements have also become important tools for analyzing time series data [31]. However, to the best of our knowledge, statistical metrics of cross-recurrence networks (CRNs) have received little attention and have not been widely utilized to assist in the analysis of the original time series. As we are interested in capturing certain types of cross-correlation or dependence between pairwise time series, which are often not accurately captured by traditional linear analysis tools such as Pearson correlation coefficients, we construct CRNs based on the original sequences. We establish a connection between the statistical features of the CRN and the original time series, aiming to reveal the coupling relationships of the original sequences using statistical metrics derived from the established network.

Inspired by Wallot *et al* [27], we observe the statistical metrics of CRNs established from pairwise time series and the properties of the time series themselves. Through our theoretical derivations, we discover a significant positive linear correlation between the clustering coefficient of the CRN and the synchronicity of the pairwise time series. To validate this regularity, we consider multicomponent dynamical systems and employ the coupled map lattices (CML) technique to simulate multidimensional time series with coupling relationships. By varying the coupling strength between the sequences, we find that there exists a strong linear relationship between the synchronicity of pairwise sequences and the clustering coefficient of their corresponding CRNs. The simulation results confirm our findings. Furthermore, we conduct empirical analyses in financial systems, Earth systems, and human interactive systems. We find strong universality of this regularity in time series from different complex systems, indicating it as a homomorphic law. Due to the

relatively stable topological structure of recurrence networks and their applicability to short time series, the regularity uncovered in this study reveals the relationship between statistical metrics of complex networks derived from time series mapping and event synchronicity. Thus, it holds significant practical implications for real-time monitoring of future financial risks, natural crises, or human behavior using CRNs, particularly when faced with limited data.

The remaining sections are organized as follows: section 2 provides an introduction to the fundamental methods of RPs and cross-RPs (CRP), along with their corresponding statistical metrics. Additionally, it discusses the network metrics derived from RP and CRP. In section 3, a theoretical derivation is presented to establish the mathematical relationship between the clustering coefficient of the CRN and the probability of synchronicity. Section 4 validates this relationship by employing the CML technique to simulate coupled multidimensional time series. Section 5 presents empirical evidence using real-world financial time series, rainfall time series, and eye-tracking sequences during human interactions. These empirical analyses aim to demonstrate the homomorphic nature of the discovered regularity across diverse complex systems. Finally, section 6 provides a summary and conclusion of the paper, highlighting the key findings and contributions.

## 2. Method

A RP is a graphical representation of the recursive states of a dynamical system in its  $m$ -dimensional phase space. For all phase space vectors  $\vec{x}_i$  ( $i = 1, \dots, N$ ,  $\vec{x}_i \in \mathcal{R}^m$ ), a pairwise measurement based on distance is performed:

$$R_{i,j} = \Theta(\varepsilon - d(\vec{x}_i, \vec{x}_j)), \quad (1)$$

where  $\Theta(\cdot)$  is the Heaviside function,  $\varepsilon$  represents the threshold of closeness, and  $d(\vec{x}_i, \vec{x}_j)$  is the measure of closeness. Different measures can be used to quantify closeness, such as spatial distance, string metrics, or local rank orders [13, 32]. In most cases, spatial distance is considered using the Euclidean distance, where  $d(\vec{x}_i, \vec{x}_j) = \|\vec{x}_i - \vec{x}_j\|$ . In the binary recurrence matrix  $R$ , if the distance  $\|\vec{x}_i - \vec{x}_j\|$  is less than  $\varepsilon$ , the corresponding  $R_{i,j}$  is set to 1. The phase space trajectories can be reconstructed from the time series  $\{u_i\}_{i=1}^N$  using time-delay embedding [33]:

$$\vec{x}_i = (u_i, u_{i+\tau}, \dots, u_{i+\tau(m-1)}), \quad (2)$$

where  $m$  represents the embedding dimension and  $\tau$  represents the time delay. And  $\{u_i\}_{i=1}^N$  represents the observed values of the time series variable of interest, derived from our collected data samples. Optimal values for  $m$  and  $\tau$  can be determined by calculating the average mutual information (AMI) function and the false nearest neighbors (FNN) function, ensuring the coverage of all free parameters and avoiding autocorrelation effects [34]. Specifically, FNN can be computed for different embedding dimensions, and the optimal embedding dimension is chosen as the first local minimum or the dimension that corresponds to a smooth curve. Similarly, for time delay, AMI is calculated for different time lags, and the optimal time lag is selected as the first local minimum value of AMI. Let us illustrate the meaning of equation (2) with a simple example. Suppose the series  $\{\vec{x}_i\}_{i=1}^N$  represents the actual phase space vector series of a variable.

Considering  $\{u_i\}_{i=1}^N$  as observations from one dimension, unable to fully reflect the variable's true dynamical characteristics, we employ delaying and embedding for  $\{u_i\}_{i=1}^N$  for phase space reconstruction. Assuming the optimal embedding dimension  $m = 3$  and optimal delay  $\tau = 1$ , we aim to obtain the first-order lagged sequence of  $\{u_i\}_{i=1}^N$ ,  $\{u_{i+1}\}_{i=1}^N$ , as observations for the second dimension, and subsequently, the first-order lagged sequence of  $\{u_{i+1}\}_{i=1}^N$ ,  $\{u_{i+2}\}_{i=1}^N$ , as observations for the third dimension. After this embedding process,  $\{\vec{x}_i\}_{i=1}^N$  becomes a three-dimensional sequence, where each vector  $\vec{x}_i$  is obtained from observations in three dimensions:  $(u_i, u_{i+1}, u_{i+2})$  for  $i = 3, \dots, N$ . Due to the lagged operation, the number of points with observations in all three dimensions reduces from  $N$  to  $N - (m - 1)$ . Thus,  $d(\vec{x}_i, \vec{x}_j)$  represents the distance between points  $i$  and  $j$  in this three-dimensional space. If the reconstruction of phase space is not performed, i.e.  $m = 1$ ,  $d(\vec{x}_i, \vec{x}_j)$  signifies the absolute difference between  $u_i$  and  $u_j$  in the  $\{u_i\}_{i=1}^N$  sequence. We use a simple schematic diagram in figure 1 to illustrate the above process.

However, our study primarily focuses on identifying cross-recurrence patterns among pairwise time series variables in multivariate systems, necessitating equal lengths. Due to variations in the optimal embedding dimensions and delays for each time series, ensuring equal lengths for any pairwise sequences after phase space reconstruction is challenging. Therefore, for simplicity, we do not perform embedding of the original series  $\{u_i\}_{i=1}^N$  by defaulting to  $m = 1$ . Consequently, the equation (2) can be written as  $\vec{x}_i = u_i$  in our case, and the process shown in figure 1 is not involved. This practice aligns with common approaches

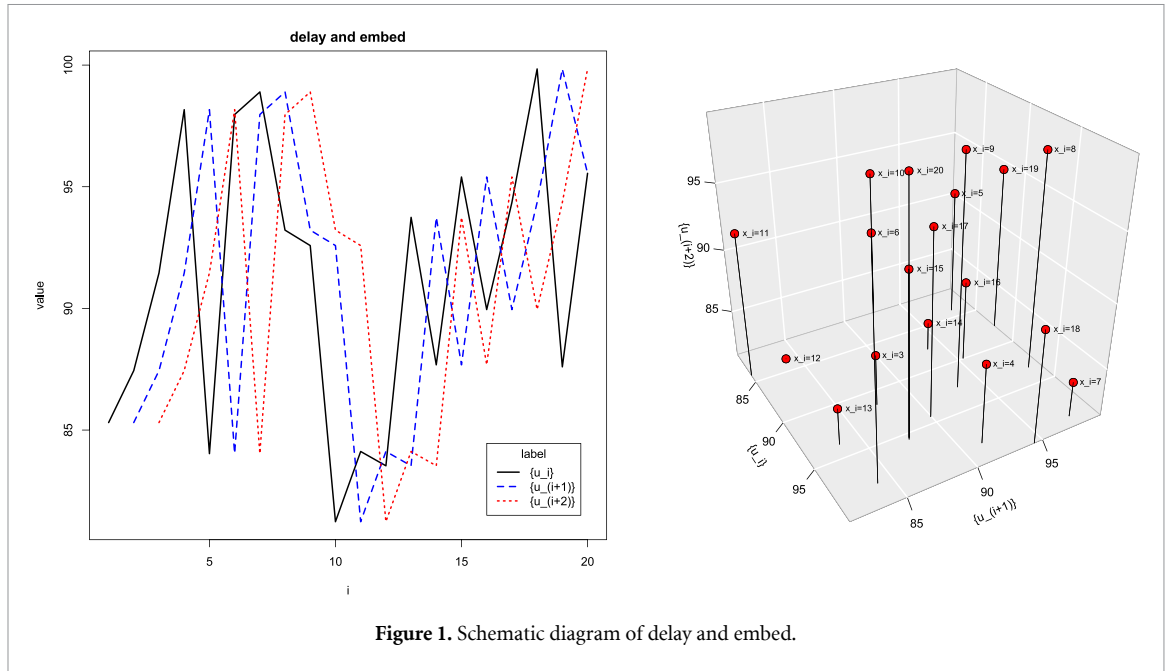


Figure 1. Schematic diagram of delay and embed.

in studying multivariate systems where multivariate analysis itself offers a better description of a system’s evolutionary features compared to univariate analysis [35].

Clearly, in RP, there is an evident diagonal line representing the recurrence of each point with itself. If spatial distance is used as the criterion for recurrence, RP is symmetric. Small-scale features in RP can be observed through diagonal and vertical lines, and the morphology of these special lines reflects the dynamics of the system. Following a heuristic approach, Zbilut and Webber [15] introduced a quantitative description of RP based on these line structures, known as RQA. RQA defines measures such as diagonal line length, recurrence rate (RR), determinism (DET), average length of diagonal structures, and entropy to characterize the diagonal segments in the RP. Table 1 presents these measures and their definitions.

The RP is extended to the CRP, which compares the dynamic behavior of two time series simultaneously embedded in phase space [13]. Specifically, for each point in the first trajectory  $\vec{x}_i$  ( $i = 1, \dots, N, \vec{x}_i \in \mathcal{R}^m$ ), and each point in the second trajectory  $\vec{y}_j$  ( $j = 1, \dots, N, \vec{y}_j \in \mathcal{R}^m$ ), the distance measure  $d(\vec{x}_i, \vec{y}_j)$  is calculated, resulting in an  $N \times N$  matrix indicating their closeness:

$$CR_{i,j} = \Theta(\varepsilon - d(\vec{x}_i, \vec{y}_j)) = \Theta(\varepsilon - \|\vec{x}_i - \vec{y}_j\|) . \tag{3}$$

CRP is the two-dimensional plot generated from the binary cross-recurrence matrix CR. In CRP, long diagonal line structures reveal synchronization between the two time series in phase space. Corresponding RQA measures are redefined, where RR, DET, and mean diagonal line length (MEAN\_DL) are functions of the distance from the main diagonal [13]. This indicates that CRP can explore the dynamical similarity of the two sequences in the presence of time delays. For example, in CRP, the diagonal-wise RR focuses only on the recurrence along the diagonal and is defined as:

$$RR_k = \frac{1}{N-k} \sum_{j-i=k} CR_{i,j}, \tag{4}$$

where  $k$  represents the time delay between the second trajectory  $\vec{y}$  and the first trajectory  $\vec{x}_i$ , and  $k$  can be positive or negative. In practical research, the optimal time delay,  $k$ , which maximizes  $RR_k$ , indicates the best delay for exhibiting the highest similarity between the two sequences.

The graphical properties of RP and CRP reflect the dynamic patterns of the sequences. When transformed into complex networks, the adjacency matrices derived from the recurrence matrix R and the cross-recurrence matrix CR can further reveal the self-similarity and mutual similarity patterns of the sequences. Various metrics commonly used to measure the topological structure of complex networks are listed in table 2.

These metrics provide insights into the properties of recurrence networks, capturing characteristics such as clustering, connectivity, average degree, average path length, and centrality measures. They help quantify the structure and importance of vertices within the recurrence network. Donner *et al* [30] established a connection between the properties of recurrence networks and the phase space topology of dynamic systems

**Table 1.** Main qualification analysis metrics of recurrence plots.

Measurement	Definition	Explanation
The number of recurrence points (RP_N)	The number of recurrent points in the RP plot	
Recurrence rate (RR)	The proportion of recurrent points in the RP plot	It represents the ratio of recurrent points to all points, indicating the probability of specific recurrent patterns.
Determinism (DET)	The percentage of recurrent points forming diagonal lines in the RP plot, given a minimum diagonal line length threshold	It reflects the extent to which deterministic behavior dominates the system, as random behavior leads to shorter diagonal lines while deterministic behavior results in longer diagonal lines. Thus, the ratio of recurrent points forming diagonal structures to all recurrent points serves as an indicator of the determinism of a system.
Mean diagonal line length (MEAN_DL)	The average length of diagonal lines in the RP plot	Diagonal structures in the RP plot indicate segments of trajectories that are close to each other at different time points. The length of diagonal line represents the duration of their closeness.
Entropy of diagonal line length distribution (ENT_DL)	The entropy of the distribution of diagonal line lengths	It measures the Shannon entropy of the histogram of diagonal line lengths, reflecting the complexity of deterministic structures in the system.
Laminarity (LAM)	The percentage of recurrent points forming vertical lines in the RP plot, given a minimum vertical line length threshold	It measures the extent of deterministic behavior in the system by evaluating the ratio of recurrent points forming vertical structures to all recurrent points.
Trapping time (TT)	The average length of vertical lines in the RP plot	Vertical line structures indicate segments of trajectories that stay close to specific points in another trajectory. The trapping time represents the duration of their closeness.

**Table 2.** Main metrics of recurrence networks.

Measurement	Explanation
Clustering coefficient (CC)	It describes the extent to which vertices in a network form clusters
Edge density (Edge_Density)	It characterizes the density of connections between vertices in the network
Average degree (MEAN_DEG)	It represents the average number of edges per vertex
Average path length (MEAN_PL)	It represents the average shortest path length between all pairs of vertices in the network
Graph-level betweenness centralization (CENTAL_BET)	It measures the importance of vertices at the graph level based on the number of shortest paths passing through them
Graph-level degree centralization (Central_DEG)	It measures the importance of vertices at the graph level based on their degrees

represented by RPs. However, their work primarily focused on demonstrating the fundamental relationship between the statistical properties of the underlying dynamical system and the topological properties of the corresponding recurrence network. In contrast, this paper emphasizes the relationship between the statistical metrics of CRNs and the probability of pairwise time series experiencing synchronized ‘events’. The objective is to utilize the network’s topological structure to reveal the likelihood of synchronization risks or states occurring between two time series. This approach offers an alternative perspective for understanding and mitigating synchronization risks among multiple entities.



### 3. Theoretical derivation

In our study, we observe the common statistical metrics of CRN, as listed in table 2, in relation to the proportion of synchronized states between pairwise time series. We find a deterministic linear relationship between the clustering coefficient and the occurrence of synchronized events (i.e. relatively extreme states). We will present further details and characteristics of the CRN in this section, and then derive the theoretical basis for the pattern we have discovered. To simplify the process of determining the threshold for recurrence when constructing the recurrence network, we transformed the numerical time series into categorical time series based on the relative magnitudes of their observed values. Specifically, assuming equal probability of the occurrence of each state, by increasing percentile levels at fixed increments from 0% to 100%, we can obtain  $n + 1$  percentiles from the pseudo-sample distribution of the numerical time series of length  $N$ . Consequently, this partition yields  $n$  ranges, where the sample values falling within these ranges are classified into  $n$  states. These  $n$  states can be ordered according to the relative sizes of their corresponding ranges. Hence, the final state space can be designated as  $(a_1, a_2, \dots, a_n)$  ( $1 \leq n \leq N$ ), facilitating the identification of so-called extreme states due to its inherent orderability. The choice of  $n$  should neither be too small nor too large; if it is too small, the probability of occurrences of  $a_1$  or  $a_n$  will be significantly higher than that of extreme events. Conversely, if it is too large, each state will occur with lower frequency, resulting in very few edges in the CRN and making it challenging to extract useful information. Considering the application of recurrence-based methods on shorter time series, an empirical range for  $n$  could be between [10, 20].

Consider a two-dimensional time series  $\{X_t, Y_t; t \geq 0\}$ , where the components have the same state space  $(a_1, a_2, \dots, a_n)$  ( $1 \leq n \leq N$ ). With a set of observation samples of length  $N$ ,  $(x_1, y_1), \dots, (x_N, y_N)$ , we can construct a CRN for these two time series variables. Specifically, we designate the  $N$  moments as network vertices, denoted as  $V = (1, 2, \dots, N)$ . The cross-recurrence matrix  $R = \{A_{ij}\}_{N \times N}$  is constructed based on whether the respective states at two moments, one from each time series, are the same or not, defined as follows:

$$A_{ij} = \begin{cases} 1 & x_i = y_j \\ 0 & x_i \neq y_j \end{cases} \quad i, j = 1, \dots, N. \quad (5)$$

Clearly, this matrix is not symmetric; generally,  $A_{ij} \neq A_{ji}$ . As defined above, the determination of states is done from the perspective of  $X_t$ , where the state of each moment in  $X_t$  is compared sequentially to that of every moment in  $Y_t$ . If the matrix is constructed from the perspective of  $Y_t$ , the comparison is done similarly, and the resulting cross-recurrence matrix is denoted as  $R^* = \{B_{ji}\}_{N \times N}$ , where:

$$B_{ji} = \begin{cases} 1 & y_j = x_i \\ 0 & y_j \neq x_i \end{cases} \quad i, j = 1, \dots, N. \quad (6)$$

Hence, we have  $A_{ij} = B_{ji}$ , establishing the relationship between  $R$  and  $R^*$  as transpose matrices, i.e.,  $R^T = R^*$ . This property is demonstrated graphically in figure 2.

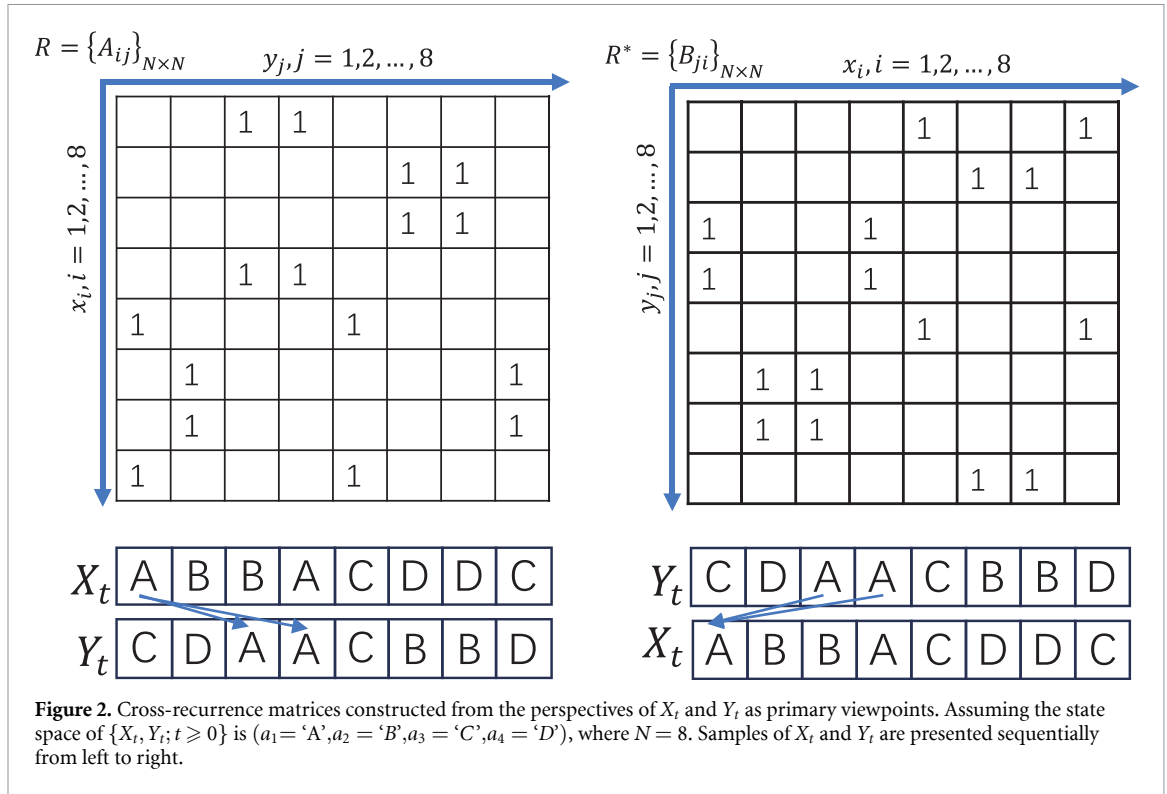
When establishing the CRNs, since our primary interest lies in determining whether there is a connection between nodes rather than focusing on the directionality or weight of these connections, we obtain symmetric adjacency matrices,  $R'$  and  $R'^*$ , based on  $R$  and  $R^*$ . And then the CRNs are constructed by these symmetric adjacency matrices. Consequently, the resulted network is undirected and unweighted. Moreover, this network includes self-loops, denoted as  $A_{ii} = 1$  when  $x_i = y_{j=i} = a_k$ , ( $i = 1, \dots, N; k = 1, \dots, n$ ).

Within the constructed CRN, we observed a particular property: the clustering coefficient of this network is positively correlated with the probability of synchronous states between the two time series, i.e. the synchronization probability. This relationship is expressed as:

$$C = P(X = Y), \quad (7)$$

where  $C$  represents the global clustering coefficient of the CRN, and  $(X = Y)$  signifies the simultaneous occurrence of identical states in both time series. The proof of this relationship is as follows:

Let  $v$  be any vertex in the CRN, and  $C_v$  be the local clustering coefficient of that vertex. Since the adjacency matrix corresponding to the CRN is symmetric, when analyzing the actual meaning of the network's edges  $e_{ij}$  ( $i, j = 1, 2, \dots, N$ ), it is reasonable to consider either  $A_{ij} = 1$  or  $A_{ji} = 1$ . Therefore, the definition formula for the local clustering coefficient can be written as:



$$\begin{aligned}
 C_\nu &= P(A_{ij} = 1 | A_{vi} = 1, A_{vj} = 1) \\
 &= P(x_i = y_j | x_v = y_i, x_v = y_j) \\
 &= \sum_{k=1}^n P(x_i = y_j = a_k | x_v = y_i = y_j) \\
 &= \sum_{k=1}^n P(x_i = y_i = a_k | x_v = y_i = y_j) \\
 &= P(x_i = y_i | x_v = y_i, x_v = y_j).
 \end{aligned} \tag{8}$$

This equation demonstrates that the local clustering coefficient of CRN is equal to the conditional probability of synchronization between two time series. Particularly, when the states of the two-dimensional time series at different moments are mutually independent, we have:

$$C_\nu = P(x_i = y_i | x_v = y_i, x_v = y_j) = P(x_i = y_i) = P(X = Y) . \tag{9}$$

Under this condition, the local clustering coefficient equals the unconditional synchronization probability of the two time series. However, assuming the independence of the states of the two-dimensional time series at different moments is a strong assumption, indicating that this two-dimensional time series is independently and identically distributed (i.i.d.) sequences. In reality, when the auto-correlation and cross-correlation functions of the two time series decay exponentially, i.e. there's only short-term correlation within and between the time series, we can derive an approximate relationship:  $C_\nu \approx P(X = Y)$ .

It can be observed that the local clustering coefficient of any vertex in the network is determined by a same term. Consequently, the global clustering coefficient  $C$  can be obtained as the average of all local clustering coefficients:

$$C = \frac{1}{N} \sum_{v=1}^N C_\nu = P(x_i = y_i | x_v = y_i, x_v = y_j) \tag{10}$$

$C = P(X = Y)$  holds if the strong independence assumption is met, and  $C \approx P(X = Y)$  when the correlations are weak or decay rapidly over time. Under the strong independence assumption, if we are more



concerned about the synchronization probability of extreme events, that is, the probability of the co-occurrence of the extreme states in the two series, the above equation can be further decomposed as:

$$\begin{aligned} C &= C_\nu = P(X = Y) \\ &= \sum_{k=1}^{n-1} P(x_i = y_i = a_k) + P(x_i = y_i = a_n) \\ &= P_0 + P_1, \end{aligned} \quad (11)$$

where  $P_0 = \sum_{k=1}^{n-1} P(x_i = y_i = a_k)$ , and  $P_1 = P(x_i = y_i = a_n)$ .  $P_1$  represents the probability of extreme event synchronization, while  $P_0$  denotes the probability of the synchronization of other states. Clearly, while maintaining  $P_0$  constant, the variation in  $P_1$  is directly proportional to  $C$ . This derivation demonstrates two key points: firstly, the clustering coefficient of a CRN established between two time series is equal to the synchronization probability when the states of the two-dimensional time series at different moments are independent; secondly, the clustering coefficient is linearly correlated to the synchronization probability of extreme event, whether the strong independence assumption is satisfied or not.

From a physical perspective, the synchronization of two time series is determined by the coupling strength between their respective attractors. In general, if the coupling strength is increased, the synchronization between the time series is enhanced, leading to a higher clustering coefficient in the CRN of the two time series. Although Chen *et al* [36] also emphasized that the relationship between coupling strength and synchronization can be complex and nonlinear, and it might depend on the specific characteristics of the coupled systems. The coupling strength remains a crucial variable that plays a significant role in altering the synchronization between time series. In the next section, we will employ simulation experiments to further demonstrate the findings mentioned above.

#### 4. Simulation experiments

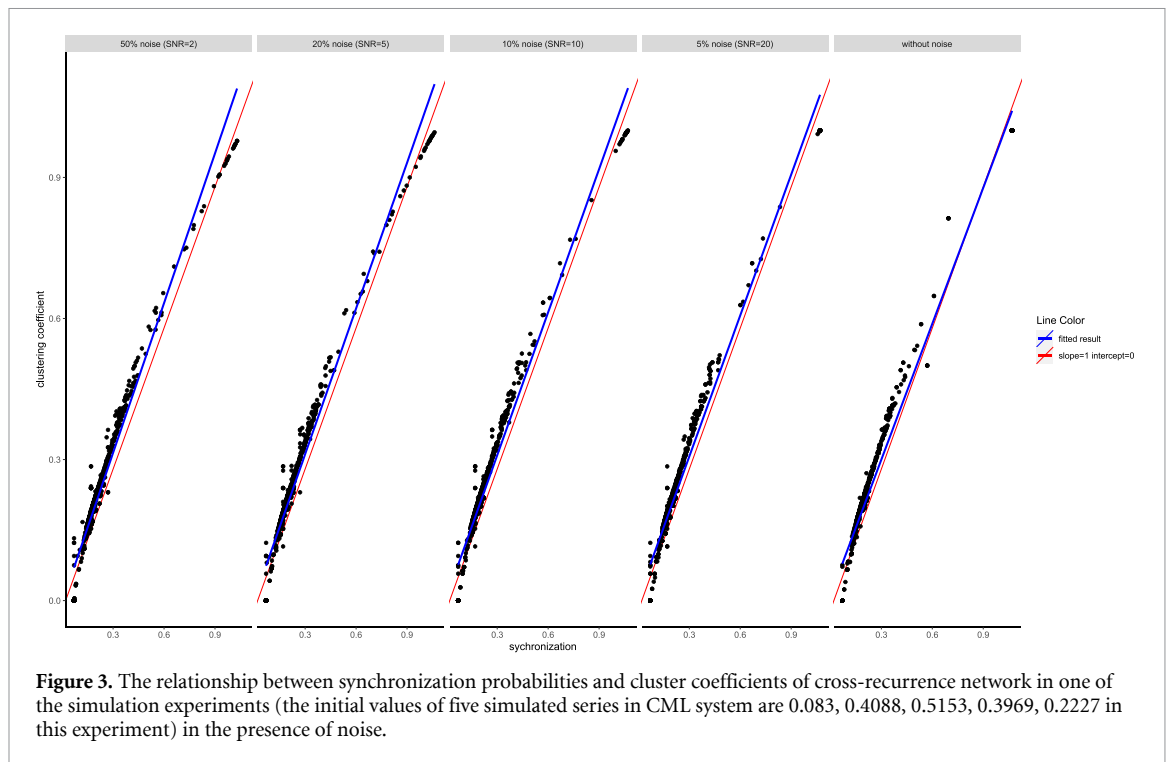
From a physical perspective, the synchronization of two time series can be understood as the effect of coupling between their respective attractors, manifested as the dependence between the states of the time series. In this paper, we consider CMLs, which are simple spatiotemporal chaotic models widely used for modeling complex spatiotemporal dynamics. Generally, their form is given by:

$$x_{t+1}^{[\kappa]} = (1 - \zeta)f(x_t^{[\kappa]}) + \zeta h_t^{[\kappa]}, \quad (12)$$

where  $x_t^{[\kappa]}$  can be understood as the value of the  $\kappa^{\text{th}}$  sequence at time  $t$  ( $\kappa = 1, 2, \dots, M$ ,  $M$  denotes the dimension of the multivariate time series), and  $h_t^{[\kappa]}$  represents the interaction of the elements from other time series at time  $t$  on the elements of the  $\kappa^{\text{th}}$  sequence. The first term on the right side represents the internal chaotic dynamics determined by the nonlinear mapping function  $f(x)$ , while the second term represents the mutual coupling effect generated by the coupling parameter  $\zeta$  ( $0 < \zeta < 1$ ). Generally, CMLs have three types of coupling: local, direct-neighbor coupling, where  $h_t^{[\kappa]} = f(x_t^{[\kappa+1]})$  or  $h_t^{[\kappa]} = f(x_t^{[\kappa-1]})$ ; global coupling, where  $h_t^{[\kappa]} = \frac{1}{M} \sum_{\kappa'} f(x_t^{[\kappa']})$ ; and intermediate-range coupling, where  $h_t^{[\kappa]} = \frac{1}{2K+1} \sum_{k=-K}^K f(x_t^{[\kappa+k]})$ . In this experiment, following the approach of Lacasa *et al* [37] and Eroglu *et al* [35], we consider the  $M$ -dimensional time series as  $M$  points on a ring model. With this approach, the dynamic evolution of each point  $x^{[\kappa]}$  is determined by the internal chaotic evolution and the average coupling effect between neighboring points, resembling a form of direct-neighbor coupling:

$$x_{t+1}^{[\kappa]} = (1 - \zeta)f(x_t^{[\kappa]}) + \frac{\zeta}{2} \left[ f(x_t^{[\kappa-1]}) + f(x_t^{[\kappa+1]}) \right] \quad (\kappa = 1, \dots, M) . \quad (13)$$

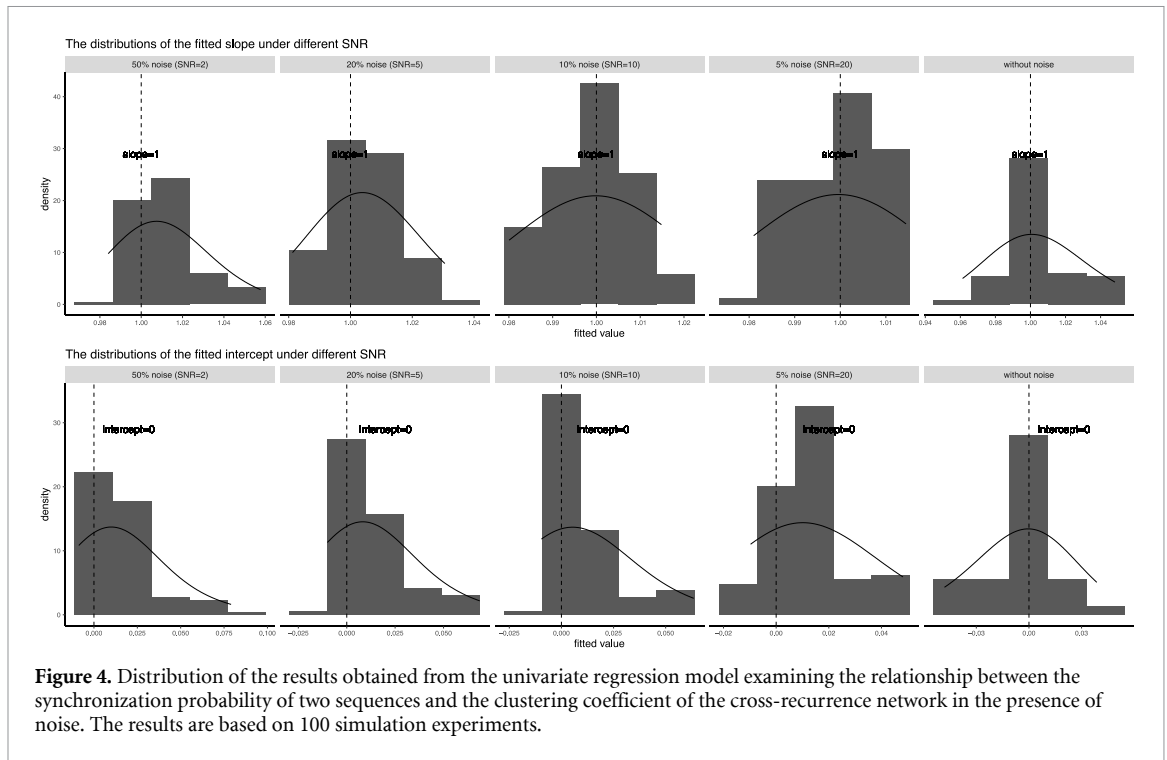
This coupling mechanism ensures that the simulated sequences only exhibit short-range correlations, as the single simulated value only couple with its neighbors. As  $\zeta$  varies, the system is trapped in different attractors, leading to different degrees of synchronization and dynamical phases. Although  $\zeta$  does not directly reflect the synchronization between time series, it acts as a coupling strength parameter that alters the synchronization between time series. Thus, we can observe the relationship between the synchronization probability of each pair of time series and the clustering coefficient of the corresponding CRN under different values of  $\zeta$  in the system.



**Figure 3.** The relationship between synchronization probabilities and cluster coefficients of cross-recurrence network in one of the simulation experiments (the initial values of five simulated series in CML system are 0.083, 0.4088, 0.5153, 0.3969, 0.2227 in this experiment) in the presence of noise.

The classic logistic mapping function  $f(x) = 4x(1 - x)$  is adopted as the internal chaotic dynamics mechanism to represent the underlying dynamics of the system and the system is assumed to be in a five-dimensional phase space ( $M = 5$ ). The coupling parameter  $\zeta$  is set within the range  $[0, 0.4]$ , with an increment of  $\Delta \zeta = 0.005$  to investigate the effect of different coupling strengths. Due to the sensitivity of chaotic systems to initial conditions, small variations in the initial values can lead to significantly different system structures. To capture the prominent avalanche effect, we iterate the system 15 000 times and use the last 10 000 simulation values for analysis. Although ample previous research confirms the robustness of recurrence-based methods in the presence of noise [19, 31, 38], to demonstrate the stability of the identified homomorphic regularity across different signal-to-noise ratios (SNRs), we will add varying levels of noise to the above simulated series. During each experiment, we introduce white noise (following a Gaussian distribution with a standard deviation of 1 and mean of 0) into the five simulated sequences within the CML system, setting SNRs at 2, 5, 10, and 20, representing 50%, 20%, 10%, and 5% noise addition, respectively. Then we factorize each time series into ten states based on the relative magnitudes of their values. We establish CRNs for the updated series and compute the clustering coefficients and synchronization rates. With multiple sample results under the coupling parameter  $\zeta$ , we perform linear fitting on the sample series of clustering coefficients and synchronization probabilities, yielding parameter estimates (slope and intercept) for first-order linear regression. Through 100 iterations of the aforementioned setup, we derive pseudo-sample distributions of slopes and intercepts for different random initial values. Based on these pseudo-distributions, we establish 95% confidence intervals for the slope and intercept estimates. The criterion for assessing the strong linear correlation between CRN clustering coefficients and synchronization probabilities relies on whether the slope 1 and intercept 0 fall within the respective confidence intervals. It is worth noting that under certain  $\zeta$  values, the simulation sequences may evolve into a few fixed values, leading to a reduced number of distinct states, possibly fewer than ten. In such cases, when computing the synchronization probability and constructing the CRN for any two time series, we adopt the number of states with the least partition from both time series as the reference. This ensures an equal number of distinct states in each series of a pair.

Figure 3 illustrates the relationship between CRN clustering coefficients and synchronization probabilities at various SNRs, including a noise-free scenario for comparison. Remarkably, even at a SNR of 2, representing a 50% noise addition, the robust linear relationship between CRN clustering coefficients and synchronization probabilities persists. The fitted slope tends closer to 1 when noise is minimal or absent. Meanwhile, figure 4 presents the pseudo-sample distributions of parameter estimates from the fitted linear models at different SNRs. As noise increases, the pseudo-sample distribution of slopes gradually skews right; however, a significant portion of slope estimates remains within proximity to 1. Similarly, the distribution pattern of the intercept did not exhibit systematic changes with alterations in the noise. Table 3 displays the



**Figure 4.** Distribution of the results obtained from the univariate regression model examining the relationship between the synchronization probability of two sequences and the clustering coefficient of the cross-recurrence network in the presence of noise. The results are based on 100 simulation experiments.

**Table 3.** The limits of 95% confidence intervals for slope and intercept estimates across different scenarios.

Simulated series	Slope		Intercept	
	Lower limit of confidence interval	Upper limit of confidence interval	Lower limit of confidence interval	Upper limit of confidence interval
50% noise (SNR = 2)	0.9900	1.0489	-0.0075	0.0657
20% noise (SNR = 5)	0.9847	1.0287	-0.0096	0.0597
10% noise (SNR = 10)	0.9813	1.0144	-0.0090	0.0578
5% noise (SNR = 20)	0.9823	1.0135	-0.0079	0.0456
Without noise	0.9681	1.0472	-0.0472	0.0319

95% confidence intervals for slope and intercept estimates across different scenarios, showcasing that the slopes 1 and intercepts 0 fall within the corresponding intervals irrespective of the noise levels. This observation suggests the insensitivity of the relationship between CRN clustering coefficients and synchronization probabilities to some extent, regardless of the noise present. Therefore, we have valid reasons to believe that the observed regularity remains effective even in empirical series with varying degrees of noise.

### 5. Empirical analysis

In the empirical study, we will focus on verifying the homogeneity of the linear relationship revealed by equation (11) in different systems, as event synchronization is often a crucial factor in triggering systemic risks. We will also examine the performance of the equivalence between the clustering coefficient in CRN and the unconditional synchronization probability in actual data. We select price time series from the financial system, rainfall data from the Earth’s ecosystem, and eye-tracking sequences from human interactive behavior as empirical objects. For each system, we establish CRNs for all pairwise time series and calculate their global clustering coefficients. Subsequently, we obtain the occurrence ratio of synchronous extreme events for each pair of sequences, aiming to verify whether there exists a significant positive linear correlation between the global clustering coefficient and the ratio of event synchronization.

#### 5.1. Financial system

##### 5.1.1. Data

The stock data from the two major trading markets in China, namely, the Shanghai Stock Exchange and the Shenzhen Stock Exchange is considered empirical objects. We select constituent stocks from the Shanghai Composite Index (SH000001) and the Shenzhen Component Index (SZ399001) based on their market capitalization weights arranged in descending order. Our focus is on choosing stocks with complete trading

data to ensure data integrity and reliability. Ultimately, we identify 56 stocks from the Shanghai Composite Index and 34 stocks from the Shenzhen Component Index that have complete trading data from 4 January 2015, to 30 October 2020. The data is divided into two parts: the training set (from 2015 to 2019) and the test set (2020). The data length of each time series is denoted as  $N$ , resulting in an  $N \times N$  recurrence matrix when establishing the CRNs or CRPs. It is worth noting that a large  $N$  is not suitable for constructing CRNs or CRPs due to the computational complexity. Considering that financial returns often exhibit yearly cyclical changes, we will use the daily return data from 2015 to 2019 to build CRNs for each year separately. And the length of the sample data of each year falls within the range of 200–240. Subsequently, we will calculate the ratio of daily coexceedances of pair stocks within a year to verify the robustness of the relationship between the topological structure of CRNs and the ratio of coexceedances of daily returns across different years. Exceedances here refer to positive returns exceeding the 95th percentile and negative returns falling below the 5th percentile. And coexceedances specifically refer to exceedances in the pair stocks that occur at the same moment or at the fixed lead-lag moments. Table A1 and A2 present the specific stock codes for the selected stocks in the SH000001 and SZ399001, respectively.

### 5.1.2. Parameter setting

As described in section 3, establishing CRNs based on numerical data involves the selection of the radius parameter. However, determining a suitable fixed radius is challenging, and there is currently no reliable literature supporting a specific choice. In this context, our primary interest lies in the frequency of occurrences of joint extreme returns between pairs of stocks. Extreme values represent absolute values of significantly positive and negative returns. We then discretize returns into ten equidistant states, ensuring a uniform distribution of returns across these states. By converting the numerical variable into a factor variable, recurrence is only considered to have occurred when the states are entirely consistent (radius = 0), thereby avoiding the need for optimal parameter selection. Consequently, CRNs are established based on factor variables. Since we consider the data of different constituent stocks as descriptions of different dimensions within the financial system, we will no longer perform the reconstruction of the phase space in the empirical analysis. Therefore, the embedding dimension is set to 1, and the time-lag value does not affect the results when the embedding dimension equals 1.

### 5.1.3. Optimal delay between time series

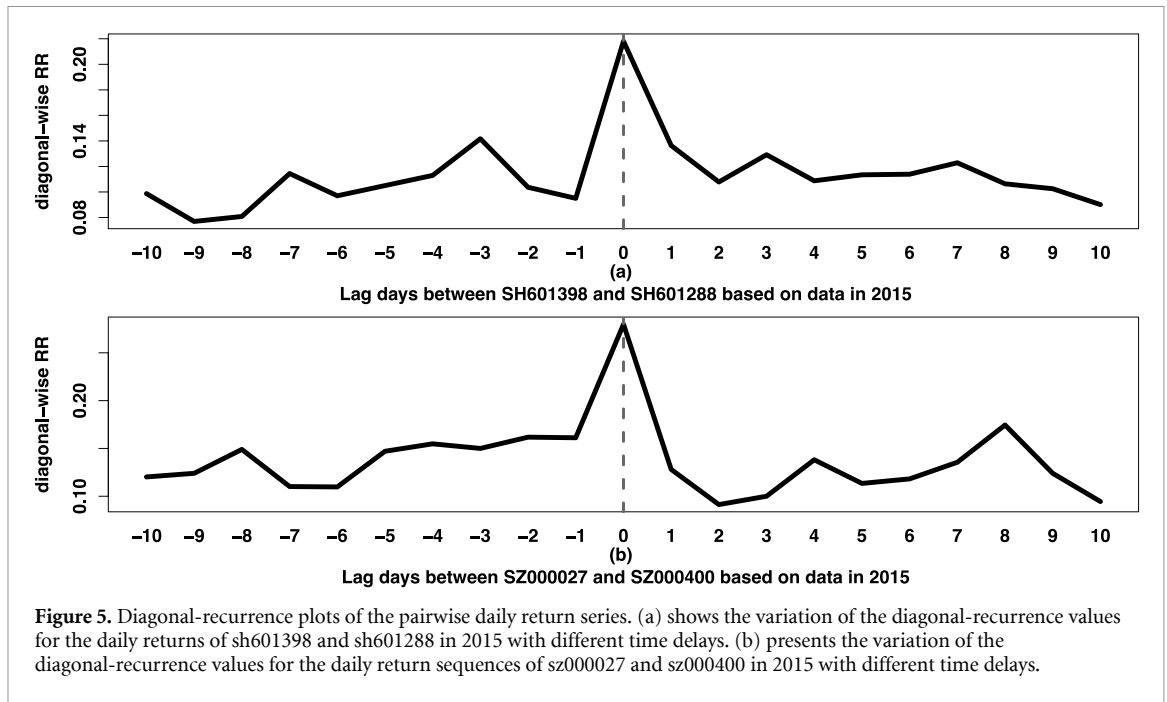
In cross-RQA (CRQA), we can construct CRPs for all possible lags between pairs of sequences and extract corresponding indicators, such as diagonal-wise RR, to explore the maximum ratio of state recurrence between the two sequences at different time lags. This helps us investigate the delayed similarity patterns between pairs of stock return sequences within the same financial system. If there exists a time delay that maximizes the RR between two return sequences, we can reasonably assume that there are statistically significant differences in their reaction speeds to the same financial events in the market. This information can also guide us in establishing rules for identifying coexceedances, which means detecting exceedances in two sequences either with a time delay or simultaneously.

Based on the data from each year in the training set, with a maximum delay of 10 ( $k_{\max} = 10$ ), we construct all possible CRPs for any pair of factorized return series and observe the optimal delay that maximizes the diagonal-wise RR. The results show that the optimal delay for any pair of sequences is 0, indicating that the highest RR occurs when there is no lead-lag relationship in the pairwise daily return series. Taking the daily returns of sh601398 and sh601288 in the Shanghai market and sz000027 and sz000400 in the Shenzhen market in 2015 as examples, figure 5 shows the variation of diagonal-wise RR  $RR_k$  within the range of  $\{k = \pm i\}_{i=1}^{10}$ . The  $RR_k$  curves for any other pair of sequences exhibit a similar pattern to that in figure 5, where  $RR_k$  reaches its highest relative value at  $k = 0$ . In other words, when there is no delay between pairwise stock return series, there are more occurrences of the same state on the same day. Based on this observation, when determining coexceedances, if exceedances occur in different assets on the same day, we consider it as a case of coexceedance.

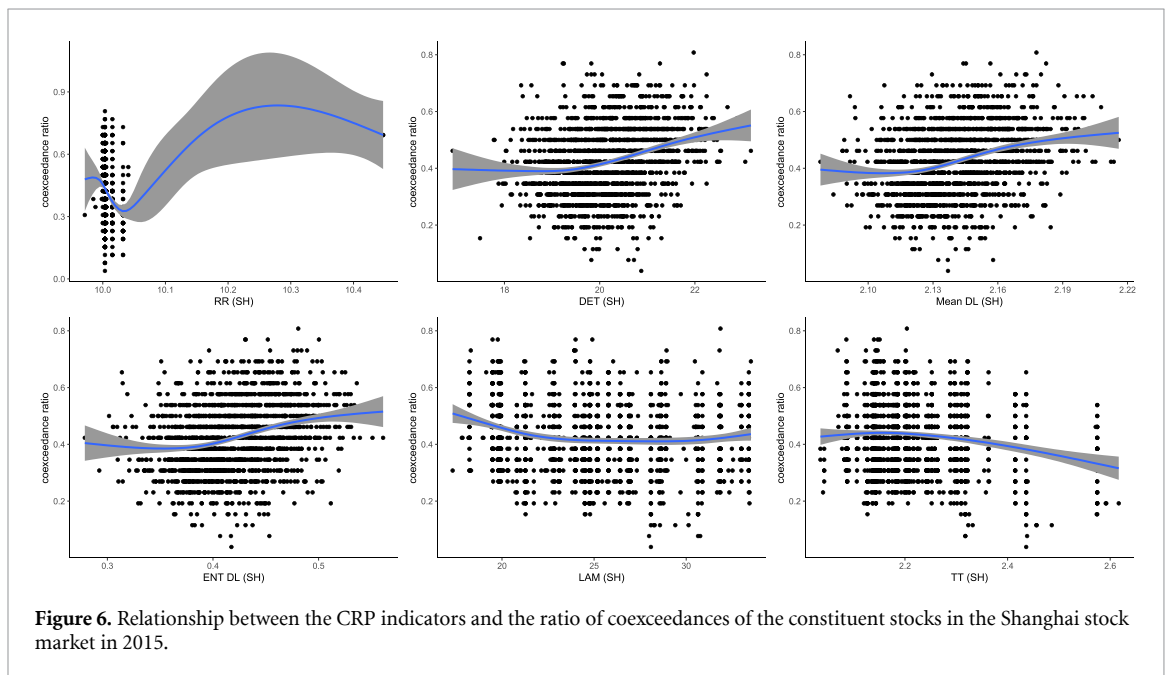
### 5.1.4. Relationship between CRP and CRN indicators and the ratio of coexceedances

In the absence of delay in the pairwise series, we construct the CRP and CRN separately and obtain the commonly analyzed indicators in tables 1 and 2. Taking the results of 2015 as an example, we show the scatter plots of any pair of stocks in the Shanghai market corresponding to the CRP indicators versus the ratio of coexceedances in figure 6, and the CRN indicators versus the ratio of coexceedances in figure 7. The corresponding plots figures A1 and A2 for the Shenzhen market are shown in the appendix.

In figure 6, there is no clear linear or nonlinear relationship observed between the CRP indicators and the proportion of coexceedances. However, in figure 7, it can be observed that there is a positive linear relationship between the average shortest path in the complex network established based on pairwise returns



**Figure 5.** Diagonal-recurrence plots of the pairwise daily return series. (a) shows the variation of the diagonal-recurrence values for the daily returns of sh601398 and sh601288 in 2015 with different time delays. (b) presents the variation of the diagonal-recurrence values for the daily return sequences of sz000027 and sz000400 in 2015 with different time delays.



**Figure 6.** Relationship between the CRP indicators and the ratio of coexceedances of the constituent stocks in the Shanghai stock market in 2015.

and the proportion of coexceedances between the returns of the two stocks. Additionally, there is also a positive linear relationship between the network’s clustering coefficient and the ratio of coexceedances.

We validate these two relationships using the 5 year training set data separately. As shown in figure 8, the linear relationship between the average shortest path and the ratio of coexceedances in individual stocks’ CRN networks in the Shanghai Composite Index is not consistently significant, such as in 2016 and 2017. However, for each year’s data, there is a clear positive linear relationship between the clustering coefficient of stocks’ CRN and the ratio of coexceedances. Figure 9 illustrates the corresponding relationships in the Shenzhen Composite Index. It can also be observed that the linear relationship between the average shortest path and the coexceedances is weaker than the linear relationship between the clustering coefficient and the coexceedances. To further investigate these linear relationships, we conduct Pearson correlation tests and use the Wilcoxon rank-sum test to verify if the clustering coefficient and the ratio of coexceedances follow the same distribution. The results are presented in table 4. As indicated in the table, for constituent stocks in the Shanghai market, the clustering coefficient of CRN from 2015 to 2019 is significantly and positively correlated with the corresponding ratio of coexceedances. This strong correlation is evident as all the linear

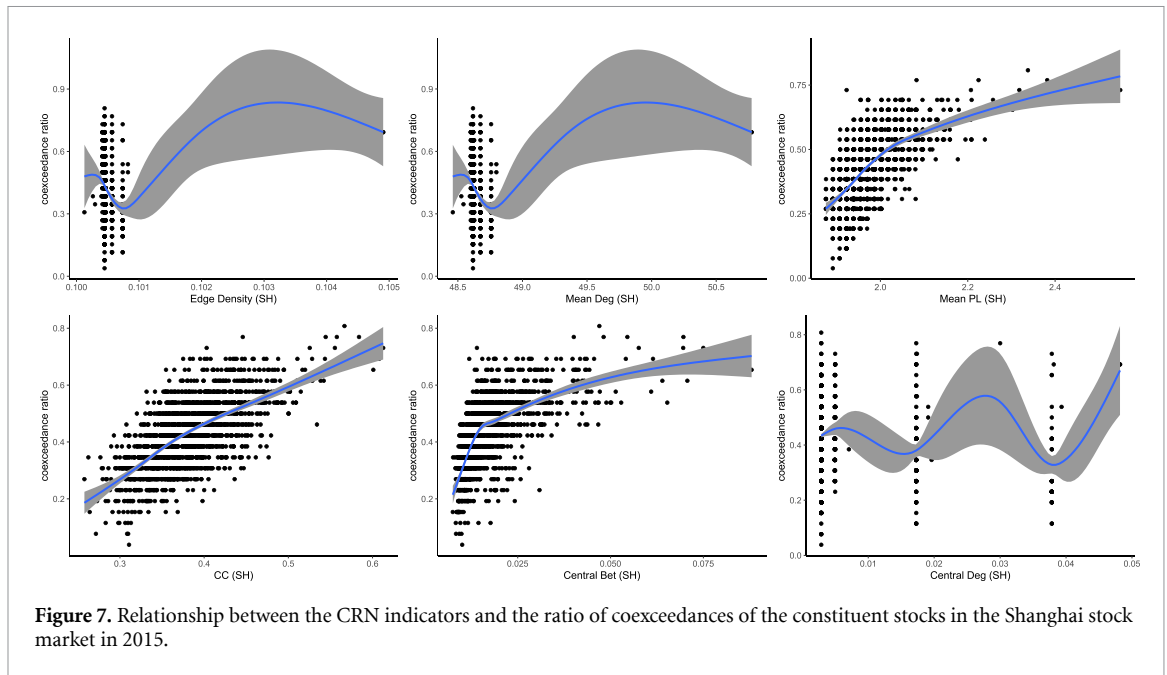


Figure 7. Relationship between the CRN indicators and the ratio of coexceedances of the constituent stocks in the Shanghai stock market in 2015.

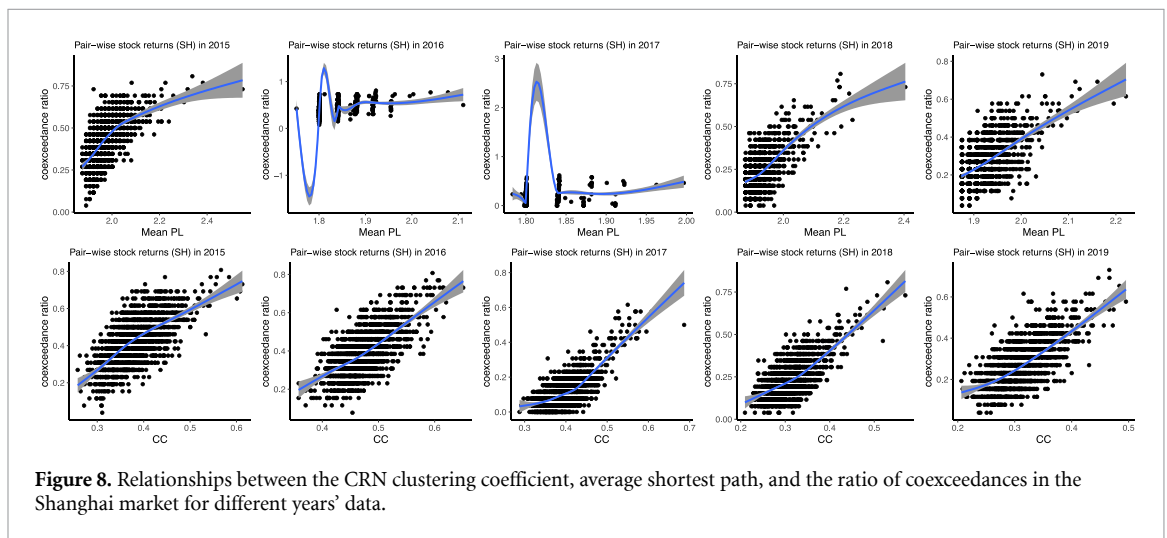


Figure 8. Relationships between the CRN clustering coefficient, average shortest path, and the ratio of coexceedances in the Shanghai market for different years' data.

correlation coefficients for each year are greater than 0.7. For constituent stocks in the Shenzhen market, there is also a significant positive linear relationship between the CRN clustering coefficient and the ratio of coexceedances from 2015 to 2019, but the strength of the relationship is weaker than that in the Shanghai market. The correlation coefficients between the average shortest path of CRN and the ratio of coexceedances in 2017 and 2019 are lower than 0.5.

Based on these findings, we conclude that only the clustering coefficient of the CRN for pairwise daily returns shows a robust and strong positive linear relationship with the ratio of coexceedances. The Wilcoxon rank-sum test results suggest that both in the Shanghai and Shenzhen markets, and for all years, the clustering coefficient and the average shortest path of individual stocks' CRN do not belong to the same distribution as the corresponding proportion of coexceedances.

As mentioned earlier, mapping time series onto complex networks allows us to discover not only correlations but also richer information and specific patterns. Besides extreme values in financial systems, the relationships between extreme values in time series of other complex systems have been of interest to researchers. To emphasize the universality of this pattern, in the following section, we will conduct separate analyses on rainfall data from different regions in the climate system and eye-tracking data from human interaction behavior.

Financial returns exhibit several stylized facts, one of which is long memory, characterized by long-range autocorrelation in the absolute values of returns. This characteristic may interfere with the equivalence of  $C$  and  $P(X = Y)$ . Taking the example of data from the Shenzhen market, we segment the return series into 10



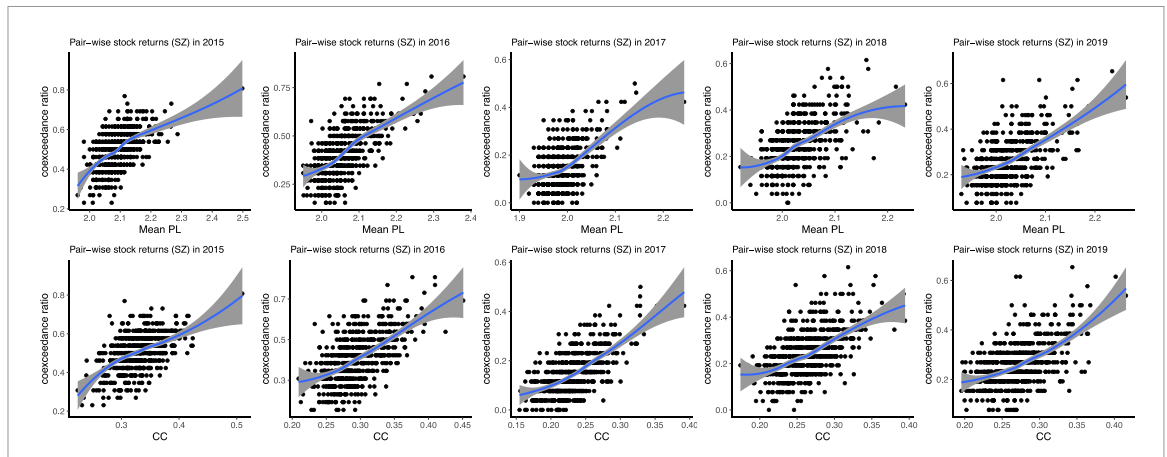


Figure 9. Relationships between the CRN clustering coefficient, average shortest path, and the ratio of coexceedances in the Shenzhen market for different years' data.

Table 4. Results of correlation tests between the CRN indicators and the proportion of coexceedances.

Year	SH				SZ			
	Pearson's correlation test		Wilcoxon rank sum test		Pearson's correlation test		Wilcoxon rank sum test	
	Cojump ratio and Mean PL	Cojump ratio and CC	Cojump ratio and Mean PL	Cojump ratio and CC	Cojump ratio and Mean PL	Cojump ratio and CC	Cojump ratio and Mean PL	Cojump ratio and CC
2015	0.6229 (***)	0.7731 (***)	0 (***)	1607 693 (***)	0.5567 (***)	0.5053 (***)	0 (***)	336 326 (***)
2016	0.3672 (**)	0.8479 (***)	0 (***)	576 740 (***)	0.5781 (***)	0.5539 (***)	0 (***)	290 161 (***)
2017	0.3325 (**)	0.8100 (***)	0 (***)	56 871 (***)	0.4234 (**)	0.6025 (***)	0 (***)	64 367 (***)
2018	0.6431 (***)	0.7079 (***)	0 (***)	535 347 (***)	0.5055 (***)	0.5303 (***)	0 (***)	138 777 (***)
2019	0.6032 (***)	0.7476 (***)	0 (***)	803 536 (***)	0.4941 (***)	0.5059 (***)	0 (***)	147 341 (***)

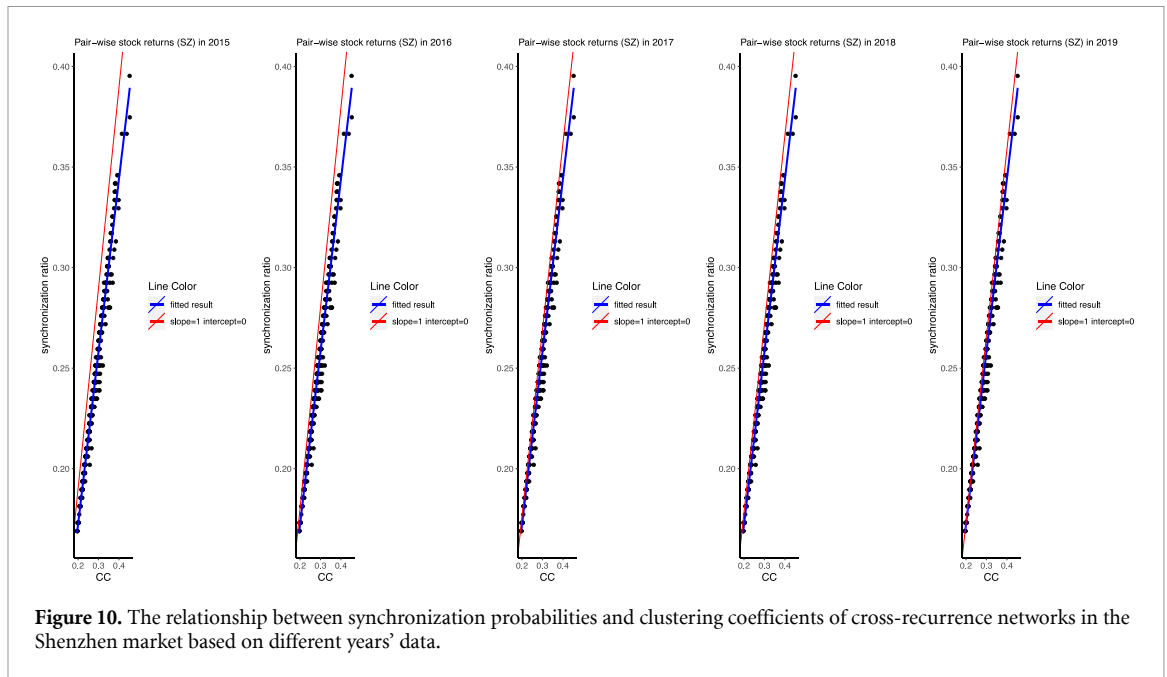
Note: the values corresponding to the Pearson's correlation test are correlation coefficients, and the values corresponding to the Wilcoxon rank sum test are test statistics. \* in parentheses indicate the significance level of the corresponding P-value. \*, \*\*, and \*\*\* represent statistical significance levels of 5%, 1%, and 0.1%, respectively.

states and calculate the synchronization probability of pairwise state sequences. Then, using the clustering coefficient series obtained above, we plot the relationship between the clustering coefficient of CRN and the synchronization probability of states of returns for each year from 2015 to 2019 in figure 10.

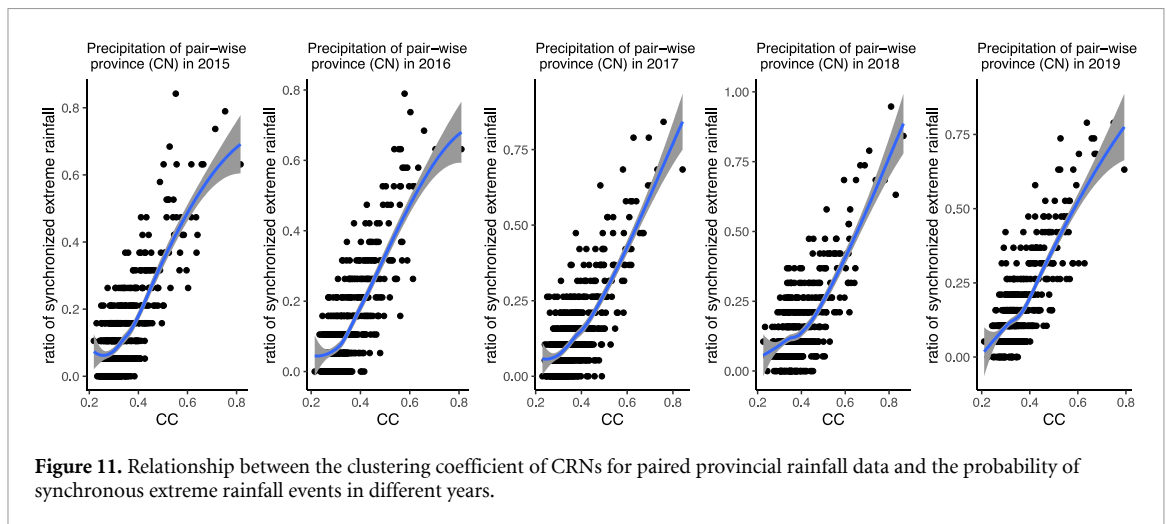
From figure 10, it is evident that the fitted curve does not align perfectly with the straight line of slope 1 and intercept 0. However, the deviation is small and it also suggests a strong linear correlation between the CRN clustering coefficient and the synchronization probability. This indicates that the long memory of returns has minimal impact on the conclusion of  $C \approx P(X = Y)$ . This could be attributed to the fact that the length of the data is just one year, comprising just over two hundred values, where the long memory of returns is not notably pronounced in such relatively short time series.

### 5.2. Earth's ecosystems

In this study, we obtain the China's ground climate daily data set (V3.0) from the National Tibetan Plateau Data Center (<https://data.tpdc.ac.cn/home>). The original data provides daily observations from various weather stations. We first interpolate the original data into grid data, covering  $500 \times 500$  grids over China, with each grid size of  $0.123\ 1924$  (longitude degrees)  $\times$   $0.099\ 4549$  (latitude degrees). By regionally averaging, we calculate the daily average precipitation for 34 provinces in China for the period from 2015 to 2019. The length of the sample data of each year is about 365. The unit of precipitation is 0.1 mm, and the total length of each province's precipitation sequence is 1826 d.



**Figure 10.** The relationship between synchronization probabilities and clustering coefficients of cross-recurrence networks in the Shenzhen market based on different years' data.



**Figure 11.** Relationship between the clustering coefficient of CRNs for paired provincial rainfall data and the probability of synchronous extreme rainfall events in different years.

For analysis, we consider extreme rainfall events as those exceeding the 95th percentile of daily precipitation in each year. The occurrence of extreme rainfall in two provinces on the same day is defined as synchronous rainfall. We calculate the probability of synchronous extreme rainfall events between pairs of provinces. Then, for each pair of provinces, we construct a CRN based on their respective rainfall sequences, following similar procedures and parameter choices as described in section 5.1.2. The clustering coefficient of each recurrence network is calculated.

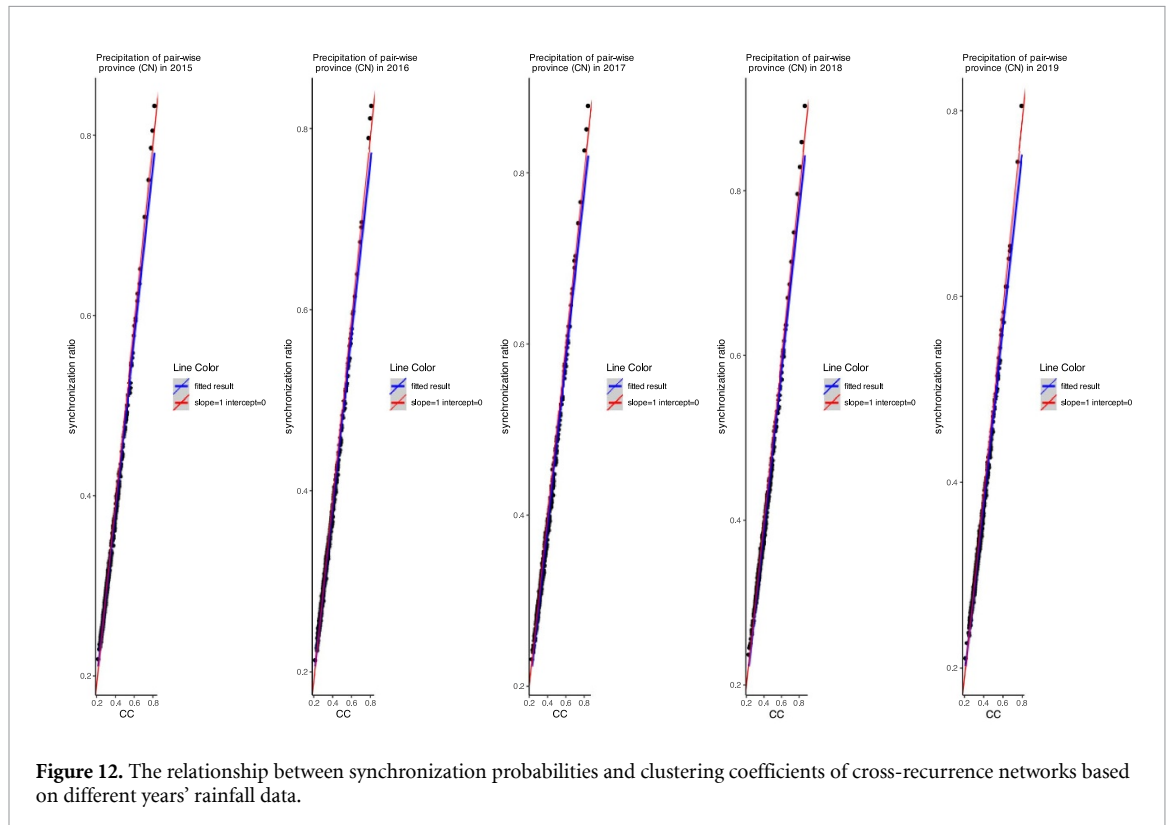
Figure 11 presents the relationship between the clustering coefficient of the CRNs and the probability of synchronous extreme rainfall events between pairs of provinces from 2015 to 2019. It tells that this relationship is positively linear and exhibits a certain level of robustness. The results from the correlation tests in table 5 also confirm the statistical significance of this linear relationship. Similarly, even though these two variables show a linear correlation, the results from the rank sum test indicate that they do not come from the same distribution.

We are also interested in investigating whether the equivalence between  $C$  and  $P(X = Y)$  approximately holds in rainfall data samples. Similarly, we partition rainfall sequences into 10 states based on their respective percentiles and compute the synchronization probability between pairwise rainfall sequences. Figure 12 depicts the relationship between the clustering coefficient of CRN and the synchronization rate for each year from 2015 to 2019. It can be seen that the fitted curve of  $C$  and  $P(X = Y)$  is closely aligned with the straight line of slope 1 and intercept 0, indicating a more obvious equivalence between  $C$  and  $P(X = Y)$  in

**Table 5.** Results of correlation tests between the precipitation CRNs clustering coefficients and the proportion of synchronized rainfall events.

Year	2015	2016	2017	2018	2019
Pearson's correlation test	0.7947 (***)	0.8086 (***)	0.7780 (***)	0.7077 (***)	0.7844 (***)
Wilcoxon rank sum test	29 425 (***)	29 655 (***)	27 766 (***)	22 185 (***)	34 844 (***)

*Note:* the values corresponding to the Pearson's correlation test are correlation coefficients, and the values corresponding to the Wilcoxon rank sum test are test statistics. \* in parentheses indicate the significance level of the corresponding  $P$ -value. \*, \*\*, and \*\*\* represent statistical significance levels of 5%, 1%, and 0.1%, respectively.

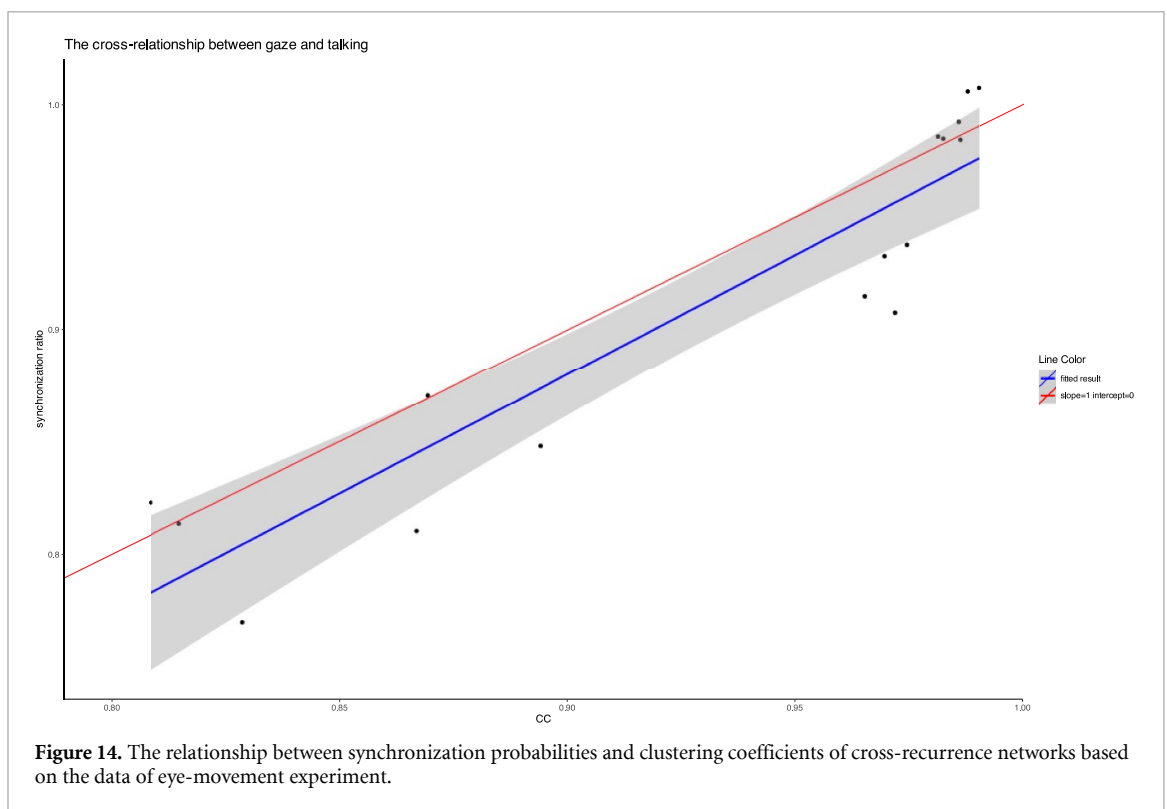
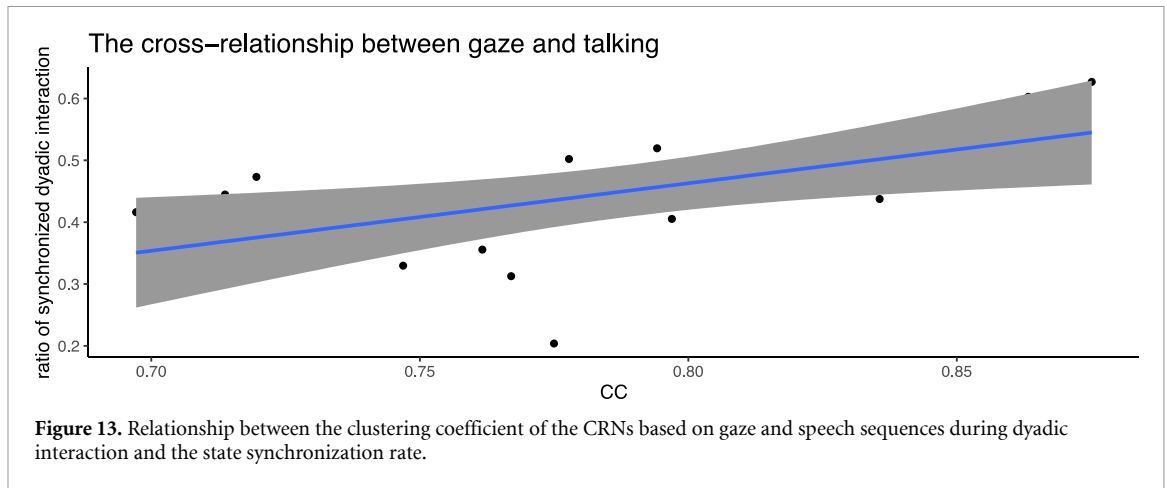
**Figure 12.** The relationship between synchronization probabilities and clustering coefficients of cross-recurrence networks based on different years' rainfall data.

this case. The differences between figures 10 and 12 also indicate that rainfall sequences exhibit stronger independence compared to return sequences.

### 5.3. Human interaction behavior system

Ho *et al* [39] conducted a study to investigate the intercorrelation between gaze and speech during social interaction games. Specifically, they organized 20 pairs of participants in small groups, where each pair played two social guessing games, and their eye movements were tracked during the interaction. The length of each pair of sample sequences is around 100. Through cross-correlation analysis of the gaze and speech signals between the two participants, they found that speakers often end their turn by directly gazing at the listener, signaling the listener to respond, resulting in a lagged synchronization of the listener's speech state with the speaker's direct gaze behavior. In other words, during social interaction, the direct gaze state of participant A exhibits a certain level of synchrony with the speech state of participant B, even though the latter's speech state consistently lags behind the former's gaze state.

We process the data from one of these social games and explore the relationship between the synchronization rate of one participant's gaze state with the other participant's speech state and the clustering coefficient of the corresponding CRN established from their respective behavioral sequences. In this experimental data, both the gaze time series and speech time series are binary sequences, where 1 represents the presence of gazing or speaking, and 0 indicates looking away or the end of speech. Treating both gaze and speech as events, the event synchronization rate between pairs of participants during the interaction activities shows a positive linear relationship with the clustering coefficient of their respective



CRNs, as depicted in figure 13. And the p-value of the corresponding Pearson's correlation test is 0.03 ( $<0.05$ ), which indicates that the linear relationship is significant at 5% confidence level.

Similarly, figure 14 illustrates the relationship between the CRN clustering coefficient and the state synchronization probability between pairs of participants in this scenario. The fitted curve of  $C$  and  $P(X = Y)$  shows a slope approaching 1, and while the intercept is not equal to 0, the confidence interval of the fitted curve still contains the line with a slope of 1 and an intercept of 0. Therefore, this also indicates an approximate equivalence between  $C$  and  $P(X = Y)$ .

In conclusion, there exists a significant positive linear relationship between the synchronization rate of events occurring in pairs of objects and the clustering coefficient of the CRNs established based on their respective time series. This pattern is consistently observed across various complex systems, including but not limited to financial systems, Earth's ecological systems, and human interactive behaviors, demonstrating strong universality. Moreover, data from these three different systems all support the approximate equivalence between the clustering coefficient of CRN and the unconditional synchronization probability, as long as the weak independence assumption is satisfied.

## 6. Discussion and conclusion

This paper focuses on the similarity patterns or synchronicity of extreme events occurring in pairs of time series. By leveraging cross-recurrence analysis to capture the complex dynamics of time series and the relative stability of network topology, we map the two time series into a complex network based on cross-recurrence, aiming to establish a relationship between the network's statistical metrics and the probability of synchronized events occurring in the two time series. Our analysis reveals a positive linear correlation between the clustering coefficient of the CRN and the event synchronization rate of the two time series. We conduct simulation experiments by CMLs and observe that the synchronization probability and the clustering coefficient of CRNs tend to become approximately equivalent. We also conduct empirical analyses in financial systems, Earth's ecological systems, and human interactive systems, and find that this pattern is universally applicable across different complex systems. However, when dealing with limited data, there are inherent limitations in observing synchronized events between two time series. In contrast, the CRP is particularly suitable for non-stationary and short time series, with relatively stable statistical features of the CRN. This context provides a practical scenario for the discovered regularity. For instance, in financial systems, systemic risk is often represented by index volatility. By identifying stocks that exhibit similar patterns to the index, i.e. with higher event synchronization rates, we may monitor or even predict systemic risk using this kind of constituent stock. When only short historical data is available and observing extremes is rare, it becomes challenging to compute event synchronization rates to identify stocks with high synchronization with the index. However, due to the positive linear relationship between the clustering coefficient of the CRN and event synchronicity, we can achieve this goal by constructing CRNs for short time series.

Similarly, this method can be applied to predict critical events such as extreme rainfall and earthquakes in specific regions. Therefore, the discovered homomorphic regularity in this study bears significant practical implications in revealing systemic risk. In the future, we plan to implement and expand the aforementioned applications based on this regularity.

### Data availability statement

The data cannot be made publicly available upon publication because the cost of preparing, depositing and hosting the data would be prohibitive within the terms of this research project. The data that support the findings of this study are available upon reasonable request from the authors.

## Appendix

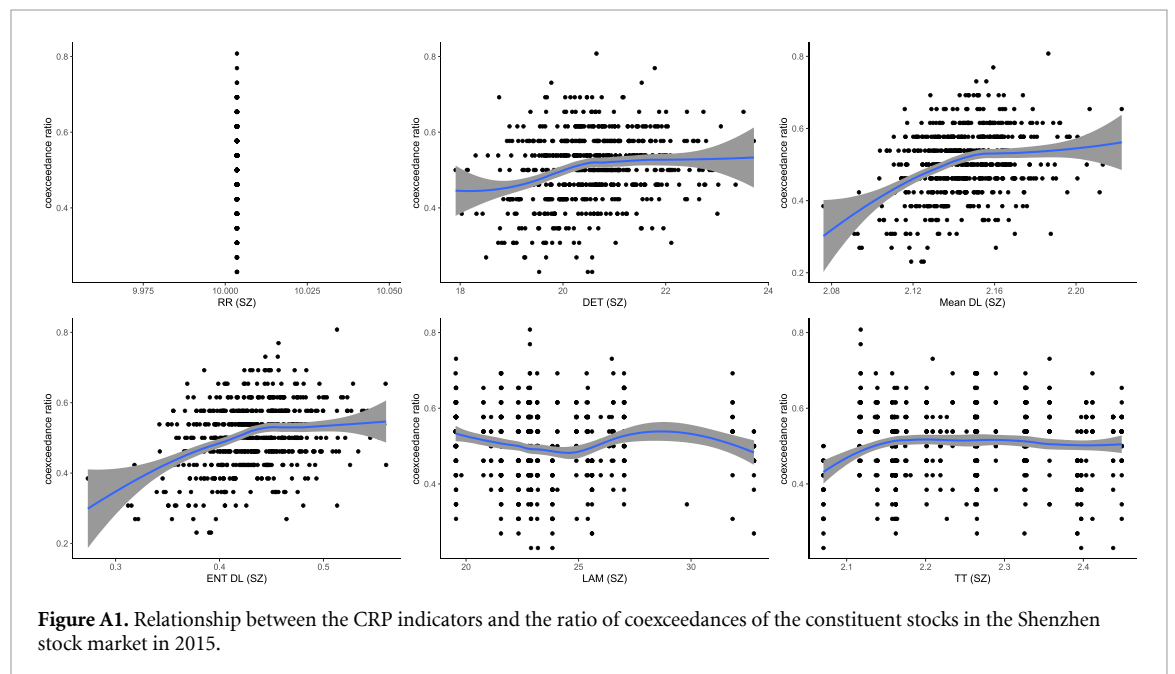
**Table A1.** List of selected stocks in the Shanghai stock market.

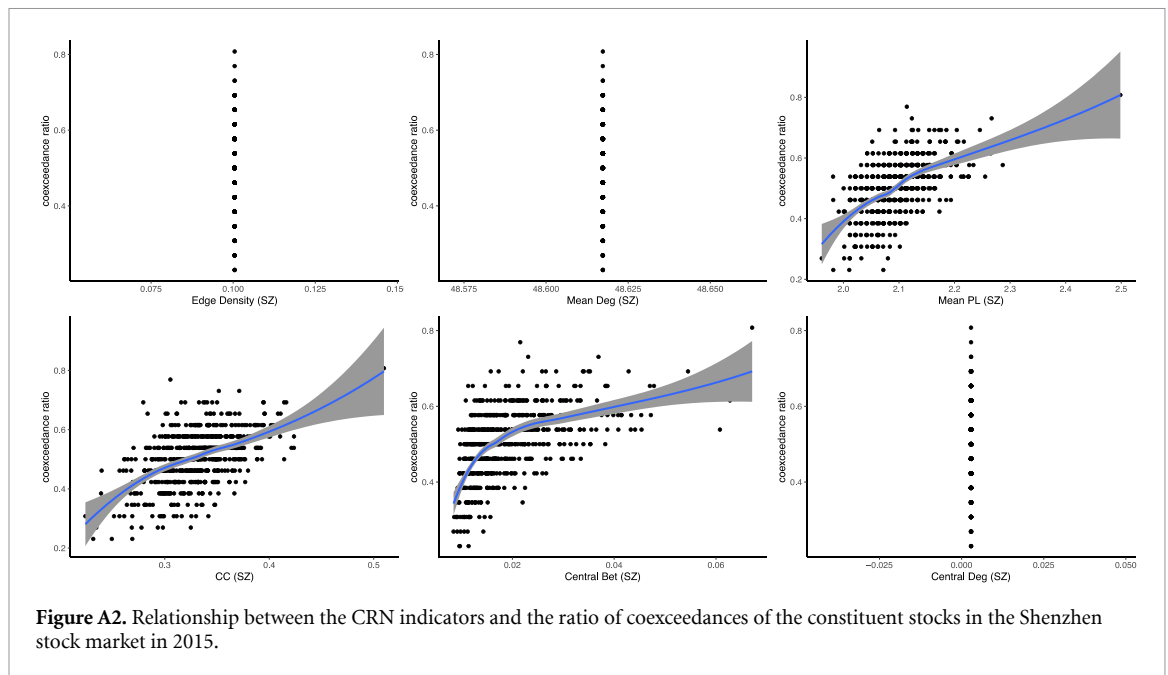
Stock index	Name	Stock index	Name
sh600519	Kweichow Moutai Co Ltd	sh600885	Hongfa Technology Co., Ltd
sh601398	Industrial and Commercial Bank of China Ltd	sh600183	Shengyi Technology Co., Ltd
sh601288	Agricultural Bank of China Co Ltd	sh600674	Sichuan Chuantou Energy Co Ltd
sh601857	PetroChina Co Ltd	sh600023	Zhejiang Zheneng Electric Power Co., Ltd
sh601988	Bank of China Ltd	sh600085	Beijing Tongrentang Co Ltd
sh601318	Ping An Insurance (Group) Company of China Ltd	sh600298	Angel Yeast Co Ltd
sh603288	Foshan Haitian Flavouring and Food Company Ltd	sh600038	AVIC Helicopter Co., Ltd
sh600028	China Petroleum & Chemical Corporation	sh601808	China Oilfield Services Limited
sh600111	China Northern Rare Earth (Group) High-Tech Co., Ltd	sh600549	Xiamen Tungsten Co Ltd
sh601668	China State Construction Engineering Co Ltd	sh601607	Shanghai Pharmaceuticals Holding Co., Ltd
sh601601	China Pacific Insurance (Group) Co Ltd	sh600456	Baoji Titanium Industry Co Ltd
sh601328	Bank of Communications Co LTD	sh600688	Sinopec Shanghai Petrochemical Co Ltd
sh600585	Anhui Conch Cement Co Ltd	sh600377	Jiangsu Expressway Co Ltd
sh600016	China Minsheng Banking Corp Ltd	sh600118	China Spacesat Co Ltd
sh601225	Shaanxi Coal Industry Company Limited	sh600642	Shenergy Co Ltd
sh601688	Huatai Securities Co Ltd	sh601179	China XD Electric Co Ltd
sh600660	Fuyao Glass Industry Group Co., Ltd	sh600316	Jiangxi Hongdu Aviation Industry Co Ltd
sh601006	Daqin Railway Co Ltd	sh600027	Huadian Power International Corporation Ltd
sh600015	Hua Xia Bank Co Ltd	sh600348	Shan Xi Hua Yang Group New Energy Co., Ltd
sh601186	China Railway Construction Co Ltd	sh600066	Zhengzhou Yutong Bus Co Ltd
sh600009	Shanghai International Airport Co Ltd	sh600004	Guangzhou Baiyun International Airport Co Ltd
sh601800	China Communications Construction Company Limited	sh601677	Henan Mingtai AL.Industrial Co Ltd
sh600600	Tsingtao Brewery Co Ltd	sh601958	Jinduicheng Molybdenum Co Ltd
sh600132	Chongqing Brewery Co Ltd	sh600808	Maanshan Iron and Steel Co Ltd
sh600011	Huaneng Power International Inc	sh600535	TASLY PHARMACEUTICAL GROUP CO., Ltd
sh600563	Xiamen Faratronic Co., Ltd	sh601965	China Automotive Engineering Research Institute Co., Ltd
sh601898	China Coal Energy Co Ltd	sh600583	Offshore Oil Engineering Co Ltd
sh601939	China Construction Bank	sh601666	Pingdingshan Tianan Coal Mining Co Ltd



**Table A2.** List of selected stocks in the Shenzhen stock market.

Stock index	Name	Stock index	Name
sz000001	Ping An Bank Co., Ltd	sz000983	Shanxi Coking Coal Energy Group Co., Ltd
sz000027	Shenzhen Energy Group Co., Ltd	sz000999	China Resources Sanjiu Medical & Pharmaceutical Co., Ltd
sz000400	XJ Electric Co., Ltd	sz002028	Sieyuan Electric Co., Ltd
sz000402	Financial Street Group Co., Ltd	sz002179	AVIC Jonhon Optronic Technology Co., Ltd
sz000423	Dong-E-E-Jiao Co., Ltd	sz002262	Jiangsu Nhwa Pharmaceutical Co., Ltd
sz000528	Guangxi Liugong Machinery Co., Ltd	sz002271	Beijing Oriental Yuhong Waterproof Technology Co., Ltd
sz000596	Anhui Gujing Distillery Co., Ltd	sz002273	Zhejiang Crystal-Optech Co., Ltd
sz000598	Chengdu Xingrong Environment Co., Ltd	sz002294	Shenzhen Salubris pharmaceuticals Co., Ltd
sz000729	Beijing Yanjing Brewery Co., Ltd	sz002304	Jiangsu Yanghe Brewery Joint-Stock Co., Ltd
sz000776	GF Securities Co., Ltd	sz002311	Guangdong Haid Group Co., Ltd
sz000789	Jiangxi Wannianqing Cement Co., Ltd	sz002372	Zhejiang Weixing New Building Materials Co., Ltd
sz000830	Luxi Chemical Group Co., Ltd	sz002419	Rainbow Digital commercial Co., Ltd
sz000848	Cheng De LoLo Co., Ltd	sz002424	Guizhou Bailing Group Pharmaceutical Co., Ltd
sz000869	Yantai Changyu Pioneer Wine Co., Ltd	sz002463	Wus Printed Circuit (Kunshan) Co., Ltd
sz000921	Hisense Home Appliances Group Co., Ltd	sz002595	Himile Mechanical Science And Technology (Shandong) Co., Ltd
sz000937	Jizhong Energy Resources Co., Ltd	sz002737	Sunflower Pharmaceutical Group Co., Ltd
sz000951	Sinotruk Jinan Truck Co., Ltd	sz300024	Siasun Robot&Automation Co., Ltd





## ORCID iD

Handong Li  <https://orcid.org/0000-0003-3613-7327>

## References

- [1] Fu T 2011 A review on time series data mining *Eng. Appl. Artif. Intell.* **24** 164–81
- [2] Keogh E and Kasetty S 2003 On the need for time series data mining benchmarks: a survey and empirical demonstration *Data Min. Knowl. Discov.* **7** 349–71
- [3] Mirshahi S and Novák V 2021 A fuzzy method for evaluating similar behavior between assets *Soft Comput.* **25** 7813–23
- [4] Warren Liao T 2005 Clustering of time series data—a survey *Pattern Recogn.* **38** 1857–74
- [5] Atay F M 2011 Synchronization and emergence in complex systems *Pramana J. Phys.* **77** 855–63
- [6] Gualdi S, Cimini G, Primicerio K, Di Clemente R and Challet D 2016 Statistically validated network of portfolio overlaps and systemic risk *Sci. Rep.* **6** 39467
- [7] Lu J and Ho D W C 2008 Local and global synchronization in general complex dynamical networks with delay coupling *Chaos Solitons Fractals* **37** 1497–510
- [8] Villegas P, Moretti P and Muñoz M A 2014 Frustrated hierarchical synchronization and emergent complexity in the human connectome network *Sci. Rep.* **4** 5990
- [9] Donges J F, Schleussner C-F, Siegmund J F and Donner R V 2016 Event coincidence analysis for quantifying statistical interrelationships between event time series *Eur. Phys. J.: Spec. Top.* **225** 471–87
- [10] Tajvidi N, Kiatsupaibul S, Tirapat S and Panyangam C 2015 Behavior of extreme dependence between stock markets when the regime shifts *J. Appl. Stat.* **16** 21–40
- [11] Zscheischler J *et al* 2018 Future climate risk from compound events *Nat. Clim. Change* **8** 469–77
- [12] Eckmann J-P, Kamphorst S O and Ruelle D 1987 Recurrence plots of dynamical systems *Europhys. Lett.* **4** 973
- [13] Marwan N, Carmen Romano M, Thiel M and Kurths J 2007 Recurrence plots for the analysis of complex systems *Phys. Rep.* **438** 237–329
- [14] Grassberger P and Procaccia I 1983 Measuring the strangeness of strange attractors *Physica D* **9** 189–208
- [15] Zbilut J P and Webber C L 1992 Embeddings and delays as derived from quantification of recurrence plots *Phys. Lett. A* **171** 199–203
- [16] Marwan N, Donges J F, Zou Y, Donner R V and Kurths J 2009 Complex network approach for recurrence analysis of time series *Phys. Lett. A* **373** 4246–54
- [17] Webber C and Zbilut J 1994 Dynamical assessment of physiological systems and states using recurrence plot strategies *J. Appl. Physiol.* **76** 965–73
- [18] Zimatore G *et al* 2021 Recurrence quantification analysis of heart rate variability to detect both ventilatory thresholds *PLoS One* **16** e0249504
- [19] Bianciardi M, Sirabella P, Hagberg G E, Giuliani A, Zbilut J P and Colosimo A 2007 Model-free analysis of brain fMRI data by recurrence quantification *NeuroImage* **37** 489–503
- [20] Marwan N 2008 A historical review of recurrence plots *Eur. Phys. J.: Spec. Top.* **164** 3–12
- [21] Richardson D C and Dale R 2005 Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension *Cogn. Sci.* **29** 1045–60
- [22] Shockley K and Turvey M T 2005 Encoding and retrieval during bimanual rhythmic coordination *J. Exp. Psychol.* **31** 980–90
- [23] Stephen D G, Dixon J A and Isenhower R W 2009 Dynamics of representational change: entropy, action, and cognition *J. Exp. Psychol.* **35** 1811–32

- [24] Richardson M J, Dale R and Marsh K L 2014 Complex dynamical systems in social and personality psychology: theory, modeling, and analysis *Handbook of Research Methods in Social and Personality Psychology* 2nd edn, ed C M Judd and H T Reis (Cambridge University Press) pp 253–82
- [25] Balasubramaniam R, Riley M A and Turvey M T 2000 Specificity of postural sway to the demands of a precision task *Gait Posture* **11** 12–24
- [26] Shockley K, Santana M-V and Fowler C A 2003 Mutual interpersonal postural constraints are involved in cooperative conversation *J. Exp. Psychol.* **29** 326–32
- [27] Wallot S, Fusaroli R, Tylén K and Jegindø E-M 2013 Using complexity metrics with R-R intervals and BPM heart rate measures *Front. Physiol.* **4** 211
- [28] Arenas A, Díaz-Guilera A, Kurths J, Moreno Y and Zhou C 2008 Synchronization in complex networks *Phys. Rep.* **469** 93–153
- [29] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D-U 2006 Complex networks: structure and dynamics *Phys. Rep.* **424** 175–308
- [30] Donner R V, Zou Y, Donges J F, Marwan N and Kurths J 2010 Recurrence networks—a novel paradigm for nonlinear time series analysis *New J. Phys.* **12** 033025
- [31] Jacob R, Harikrishnan K P, Misra R and Ambika G 2018 Recurrence network measures for hypothesis testing using surrogate data: application to black hole light curves *Commun. Nonlinear Sci. Numer. Simul.* **54** 84–99
- [32] Bandt C *et al* 2008 Analysis of bivariate coupling by means of recurrence *Mathematical Methods in Signal Processing and Digital Image Analysis* pp 153–82
- [33] Packard N H, Crutchfield J P, Farmer J D and Shaw R S 1980 Geometry from a time series *Phys. Rev. Lett.* **45** 712–6
- [34] Kantz H and Schreiber T 2003 *Nonlinear Time Series Analysis* 2nd edn (Cambridge University Press)
- [35] Eroglu D, Marwan N, Stebich M and Kurths J 2018 Multiplex recurrence networks *Phys. Rev. E* **97** 012312
- [36] Chen J Y, Wong K W, Zheng H Y and Shuai J W 2002 The coupling of dynamics in coupled map lattices *Discrete Dyn. Nat. Soc.* **7** 157–60
- [37] Lacasa L, Nicosia V and Latora V 2015 Network structure of multivariate time series *Sci. Rep.* **5** 15508
- [38] Rusakov A V, Medvinsky A V and Nurieva N I 2018 Analysis of the recurrence of noisy time series *Biophysics* **63** 590–5
- [39] Ho S, Foulsham T and Kingstone A 2015 Speaking and listening with the eyes: gaze signaling during dyadic interactions *PLoS One* **10** e0136905