# Fractal Conditional Correlation Dimension Infers Complex Causal Networks

Özge Canlı Usta [1,2,3,*,†] and Erik M. Bollt [1,2,*,†]

1    Department of Electrical and Computer Engineering, Clarkson University, 8 Clarkson Ave.,
     Potsdam, NY 13699, USA
2    Clarkson Center for Complex Systems Science, Clarkson University, 8 Clarkson Ave.,
     Potsdam, NY 13699, USA
3    Department of Electrical and Electronics Engineering, Dokuz Eylül University, Izmir 35390, Turkey
*    Correspondence: ocanlius@clarkson.edu (Ö.C.U.); bolltem@clarkson.edu (E.M.B.)
†    These authors contributed equally to this work.

**Abstract:** Determining causal inference has become popular in physical and engineering applications. While the problem has immense challenges, it provides a way to model the complex networks by observing the time series. In this paper, we present the optimal conditional correlation dimensional geometric information flow principle (*oGeoC*) that can reveal direct and indirect causal relations in a network through geometric interpretations. We introduce two algorithms that utilize the *oGeoC* principle to discover the direct links and then remove indirect links. The algorithms are evaluated using coupled logistic networks. The results indicate that when the number of observations is sufficient, the proposed algorithms are highly accurate in identifying direct causal links and have a low false positive rate.

**Keywords:** causal inference; correlation dimension; geometric information flow

## 1. Introduction

Causal inference has attracted attention in various scientific fields, from engineering [1] to climate science [2,3] and from neuroscience [4] to ecological systems [5]. The problem is reconstructing the causal relations from the observed time series of a complex network. However, the underlying dynamics of the networks are often unknown, and the observations can be limited. Hence, the ability to model the networks and infer causal relationships among the systems can be quite challenging.

We have written that [6,7] "a basic question when defining the concept of information flow is to contrast versions of reality for a dynamical system. Either a subcomponent is closed or alternatively, there is an outside influence due to another component". Claude Granger's Nobel prize [8]-winning work leading to Granger causality (see also Wiener [9]) formulates causal inference as a concept of quality of forecasts. That is, we ask, does system $X$ provide sufficient information regarding forecasts of future states of system $X$, or are there improved forecasts with observations from system $Y$? We declare that $X$ is not closed, as it is receiving influence (or information) from system $Y$, when data from $Y$ improve forecasts of $X$, and this is called Weiner–Granger causality, WGC. In Granger's original test for causality (GC) between two time series, a time series $Y$ has a causal inference on a second time series $X$ if the future of $X$ includes information from past terms of $Y$ [10] as decided by forecasting $X$ in two different ways with linear models with and without considering the information from $Y$. GC deals with the identification of causality in stochastic and linear systems, and its extensions have been introduced to tackle the problem of detecting causation in separability between multivariate time series and nonlinear models [11–13]. Other variations on the concepts of WGC exist based on other concepts of forecasting.

Cross-mapping (CM) techniques, which use the predictions of one system based only on the past observations from the other system, are also employed in detecting causal inference problems [14,15]. Rulkov et al. have studied the connections of two unidirectional coupled systems and the detection of synchronization using the CM-based technique in [14]. The authors have also focused on the connections of unidirectional coupled systems and applied the mutual nonlinear prediction method to neuroscience [15]. Several methods have been proposed to infer causal relationships and synchronization using CM techniques [16–19]. Following the line of CM techniques-based research works, Sugihara et al. have also proposed convergent cross-mapping (CCM) that utilizes a state–space reconstruction technique [5]. CCM can identify causality in weakly coupled networks and find causal links in complex ecosystems. Although CCM requires a large amount of data and fails in case of strong coupling or synchrony, CCM and its alternatives have also been widely used in recent years [20–22].

On the other hand, information–theoretic approaches are implemented to solve the causal inference problem in many applications due to being model-free, including transfer entropy [23–26]. Of particular interest to us here is the more nuanced concept of direct information flow, which considers if $X$ causes $Y$ conditioned on considering intermediaries $Z$: that is, if $X$ flows through $Z$ to influence $Y$, but perhaps there is no direct influence from $X$ to $Y$. In a particular study, Sun et al. have suggested the optimal causation entropy principle (oCSE), an algorithm that reveals the causal relations in a complex network by using causation entropy [25], to learn direct and indirect influences. The principle is based on the idea that the causal parents of a node in the network contain the minimal set of nodes that maximize causation entropy. The oCSE principle allows us to differentiate causal parents of a node from indirect influences of a node by using discovery and removal algorithms.

The idea of understanding connections between geometric information flow and causal inference has been investigated in recent decades [7,27–30]. An index has been proposed in [27] where it was called the dynamic complexity coherence measure. The index is the ratio of the sum of the correlation dimensions of individual subsystems to the correlation dimension of the coupled dynamical system. If the two systems are independent, the sum of the correlation dimensions of individual subsystems equals the correlation dimension of the concatenated dynamical system. However, if two systems are coupled, the sum of the correlation dimensions of individual subsystems is greater than the correlation dimension of the concatenated dynamical system. The index can determine the degree of synchronization and the presence of coupling [27]. The authors in [28] have also shown that the correlation dimension can reveal the presence and the direction of coupling. The synchronization of coupled systems can be determined by using this method.

Krakovská has used the correlation dimension to detect causality [30]. The study investigates the causal relevance between or within two systems for the different coupling strengths using the correlation dimension in the reconstructed state space. It has been emphasized that the correlation dimension in causal analysis can be a promising method between and within systems. Furthermore, the study states that correlation-based methods provide some advantages in finding causal relations in dynamical systems when we have sufficiently long observations and the states are observable.

Surasinghe and Bollt have suggested the correlation dimension geometric information flow measure to quantify causal inference between two related systems in the geometric sense [7]. The authors have proposed a new measure, i.e., geometric information flow $GeoC_{.\to.}$, based on the conditional correlation dimension, which enables the identification of causal relations between two related systems by geometric terms. They found that $GeoC_{.\to.}$ provides us with geometric interpretable results when detecting causality in synthetic and real examples.

Conversely, Cummins et al. have established a theoretical model that builds on Takens' theorem [31] for recovering dynamic interactions between weakly coupled or moderately coupled dynamical systems. The authors have examined the limitations of the state–space reconstruction methods. The manifold of the systems from one single

coordinate observation function has been reconstructed. Then, the approach seeks to identify the reconstructions that have mutual driving, one-directional driving, or are completely independent. The approach fails to recover self-loops, and it cannot differentiate between mutual and unidirectional dynamical driving in connected components [32]. Although the study in [30] claims that the correlation dimension reveals the causal relations in the reconstructed space, it can fail in some cases. If two uncoupled systems $(X, Y)$ are driven by a hidden common driver $Z$, $X$ and $Z$ cannot distinguish, and it implies a directional link from $X$ to $Y$ when $X$ and $Z$ are synchronized.

Although the correlation dimension and measures based on it have been explored for detecting causality and synchronization, they have not been studied extensively to reveal connections in the network. The existing methods in the literature are particularly interested in detecting synchronization. In this paper, we focus on detecting causality in the networks, assuming the networks are not synchronized.

Additionally, the previously discussed CM or state–space methods reconstruct the phase space from a single observation of a node. In contrast, we observe all states of the subsystems in the network. Even if we analyze the correlation dimension in the reconstructed state–space or observe all states, using only the correlation dimension may be insufficient to determine the direct and indirect influences of the network. However, we have utilized the conditional correlation dimensions-based measure in [7] to infer direct and indirect influences in this paper.

The main goal of this paper is to quantify the causal inference between the subsystems in a network in the geometric sense. Unlike previous studies, this paper extends the analysis of causal inference problems using only geometric interpretation to detect causal links in the networks. Expanding upon the fractal geometric concepts of the consequence of information flow in [7], we develop the optimal conditional correlation dimensional geometric information flow principle (*GeoC*) that resembles the oCSE principle previously proposed by Sun et al. [25]. We present two algorithms to detect the causal links and remove indirect links using the correlation dimension geometric information flow *GeoC*. The performance of the *oGeoC* algorithm is investigated in coupled logistic networks.

## 2. Problem Statement

### 2.1. Preliminaries and Basic Definitions

In this section, we present the notation and the basic definitions. A graph, $\mathcal{G} \equiv (\mathcal{V}, \mathcal{E})$ is defined by the set of vertices (nodes), $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$, and the set of edges (links), $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. If $\forall\, (v_i, v_j) \in \mathcal{E} \implies (v_j, v_i) \in \mathcal{E}$, the graph is undirected. Otherwise, it is defined as a directed graph. The set of $\mathcal{N}_i = \{v_j \in \mathcal{V} | (v_i, v_j) \in \mathcal{E}\}$ is denoted as the parents of the $i^{th}$ node. In short, we denote $v_i$ simply as $i$. The graphs can also be represented by their adjacency matrix $\mathbf{A}$. The elements of $\mathbf{A}$ are $a_{ij} = 1$ if there is an edge from $j$ to $i$. Otherwise, $a_{ij} = 0$.

Consider a discrete-time dynamical system in $\mathbb{R}^d$ expressed as

$$\mathbf{x}_{n+1} = f(\mathbf{x}_n) \tag{1}$$

where $\mathbf{x}_n \in \mathbb{R}^d$ is the state variable at time step $n$ and $f(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is the local dynamics. We also consider a discrete-time dynamical network consisting of $N$ identical components

$$\mathbf{x}_{n+1}^{(i)} = f(\mathbf{x}_n^{(i)}) + \sigma \sum_{\substack{i=1 \\ i \neq j}}^{N} a_{ij} \kappa g(\mathbf{x}_n^{(i)}, \mathbf{x}_n^{(j)}) \quad i = 1, 2, \ldots, N. \tag{2}$$

Here, $\mathbf{x}_n^{(i)} \in \mathbb{R}^d$ is the state variable of node $i$ at time step $n$, $f(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is the local dynamic, $\sigma$ is the coupling strength, $a_{ij}$ represents the coupling from node $j$ to node $i$ and it can be expressed in matrix form $\mathbf{A} \in \mathbb{R}^{N \times N}$, $\kappa \in \mathbb{R}^{d \times d}$ is the inner coupling matrix and $g(\mathbf{x}_n^{(i)}, \mathbf{x}_n^{(j)}) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is the coupling function. To simplify the notation, we define the next step of $\mathbf{x}_n^{(i)}$ as

$$\mathbf{x}^{'(i)} = \mathbf{x}_{n+1}^{(i)}, \tag{3}$$

where the $'$ denotes the next time step as an alternative notation to explicitly indexing time. Let $\left\{\mathbf{x}_n^{(i)}\right\}_{n=1}^{T}$ and $\left\{\mathbf{x}_n^{(j)}\right\}_{n=1}^{T}$ represent sets of measurements from a network in (2). Assume that a manifold of observations of $(\mathbf{x}_n^{(i)}, \mathbf{x}_n^{(j)}, \mathbf{x}_n^{'(i)}) \in (\mathbf{X}^{(i)} \times \mathbf{X}^{(j)} \times \mathbf{X}^{'(i)})$ and $(\mathbf{x}_n^{(i)}, \mathbf{x}_n^{'(i)}) \in (\mathbf{X}^{(i)} \times \mathbf{X}^{'(i)})$ are defined as $\mathcal{M}$ and $\Omega_1$, respectively. Based on how these manifolds lie in the space provides crucial information about whether $\mathbf{x}_n^{'(i)}$ depends only on $\mathbf{x}_n^{(i)}$ or on $(\mathbf{x}_n^{(i)}, \mathbf{x}_n^{(j)})$. Thus, using the dimensions of the manifold of the subsystems can be decisive in determining causal inference between systems [7]. First, the conditional correlation dimensional geometric information flow is defined as follows:

**Definition 1** (Conditional Correlation Dimensional Geometric Information Flow [7]). *Assume that $\mathcal{M}$ and $\Omega_1$ are bounded Borel sets. Let $\mathcal{M}$ be a manifold of the data set taken at time steps from 1 to $T+1$ for node $i$ as $(X_1^{(i)}, X_2^{(i)}, \ldots, X_T^{(i)}, X_{T+1}^{(i)})$ and let $\Omega_1$ be a set $\mathbf{X}^{(\mathbf{i})} = (X_1^{(i)}, X_2^{(i)}, \ldots, X_T^{(i)})$ taken at time steps from 1 to $T$ for node $i$. The geometric information flow $Geo(\cdot|\cdot)$ is defined in the sequel:*

$$Geo(X^{'(i)}|\mathbf{X}^{(\mathbf{i})}) = \mathcal{D}_2(\mathcal{M}) - \mathcal{D}_2(\Omega_1) \tag{4}$$

where $\mathcal{D}_2(\cdot)$ is the correlation dimension of the given dataset [33]. Then, the authors in [7] have defined correlation dimensional geometric information flow between two systems by the following:

**Definition 2** (Correlation Dimensional Geometric Information Flow [7]). *Consider $\mathbf{X}^{(i)} = (X_1^{(i)}, X_2^{(i)}, \ldots, X_T^{(i)})$ and $\mathbf{X}^{(j)} = (X_1^{(j)}, X_2^{(j)}, \ldots, X_T^{(j)})$ as time series measured at time steps from 1 to $T$ for nodes $i$ and $j$, respectively. The correlation dimensional geometric information flow from $j$ to $i$ is measured by using the conditional correlation dimension in (4) and is given by*

$$GeoC_{j \to i} := Geo(X^{'(i)}|\mathbf{X}^{(i)}) - Geo(X^{'(i)}|\mathbf{X}^{(i)}, \mathbf{X}^{(j)}). \tag{5}$$

It is clear that if $j$ influences $i$, then $GeoC_{j \to i} > 0$. However, if $j$ does not influence $i$, $GeoC_{j \to i} = 0$.

$GeoC$ is based on quantifying information flow between variables and how manifolds are mapped [7]. The study also investigates information flow between two systems, even if the observation set is not only a manifold but also a fractal. It lies on how the fractal dimension changes through the transformations [34].

*2.2. Geometric Causation of Information Flow in Networks*

We aim to extend the previous concept of correlation dimension geometric information flow to the networks. The idea is to make an analogy between the oCSE principle [25] and *GeoC* [7]. Hence, the extension of the proposed geometric measure $GeoC_{j \to i}$ leads us to solve this problem.

**Definition 3** (optimal Conditional Correlation Dimensional Geometric Information Flow (*oGeoC*)). *Let I, J, and K be subsets of nodes in a network. The correlation dimensional geometric information flow from J to I by conditioning on K is defined as*

$$GeoC_{J \to I|K} := Geo(X^{'(I)}|\mathbf{X}^{(K)}) - Geo(X^{'(I)}|\mathbf{X}^{(J)}, \mathbf{X}^{(K)}) \tag{6}$$

where $X^{'(I)}$ is the observation at $T+1$ for subset $I$ and $\mathbf{X}^{(J)} = (X_1^{(J)}, X_2^{(J)}, \ldots, X_T^{(J)})$ and $\mathbf{X}^{(K)} = (X_1^{(K)}, X_2^{(K)}, \ldots, X_T^{(K)})$ are time series measured at time steps from 1 to $T$ for nodes $J$ and $K$, respectively. It is clear that if $\mathbf{X}^{(I)} = \mathbf{X}^{(K)}$, then (6) becomes

$$GeoC_{J \to I|K} = GeoC_{J \to I|I} = GeoC_{J \to I} \tag{7}$$

as is denoted (5). Moreover, if $\mathbf{X}^{(K)} = \varnothing$, $GeoC_{\cdot \to \cdot \mid \cdot}$ simplifies to

$$GeoC_{J \to I \mid \varnothing} = Geo(X^{'(I)}) - Geo(X^{'(I)} \mid \mathbf{X}^{(J)}). \tag{8}$$

Consider the case where $J \subset K$. In this case, $Geo(X^{'(I)} \mid \mathbf{X}^{(J)}, \mathbf{X}^{(K)})$ reduces to $Geo(X^{'(I)} \mid \mathbf{X}^{(J)})$, which implies that $GeoC_{J \to I \mid K} = 0$. Furthermore, if $J \subset \mathcal{N}_I$ and $J \not\subset K$, then $GeoC_{J \to I \mid K} > 0$.

Using Definition 3 and its properties, we can quantify the information flow between variables geometrically. Moreover, we can identify direct influences and indirect influences by using similar algorithms that were earlier designed in [25]. "FORWARD GEOC" in Algorithm 1 computes $GeoC_{j \to i \mid \mathcal{K}}$ for each node $i \in \mathcal{V}$ and finds the maximum of $GeoC_{j \to i \mid \mathcal{K}}$ over $j$. Then, the algorithm discovers one of the causal parents of $i$ in each iteration and updates the $\mathcal{K}$ set iteratively until $GeoC_{j \to i \mid \mathcal{K}}$ reaches zero ($\varepsilon_{forward}$). In "BACKWARD GEOC", the candidate set of the causal parents of $i$ is used, and the algorithm calculates $GeoC_{j \to i \mid \mathcal{K}-\{j\}}$ over the $K$ set. If each $GeoC$ is zero ($\varepsilon_{backward}$), then the candidate set of the causal parent of $i$ is removed from the $K$ set. "BACKWARD GEOC" also returns the estimated adjacency matrix $\hat{\mathbf{A}}$ to present the graph.

$\varepsilon_{forward}$ and $\varepsilon_{backward}$ can be determined by setting a threshold or performing a significant test statistically [35]. We utilized a shuffle test to determine the zero ($\varepsilon_{forward}$) in Algorithm 1. The shuffle test procedure is presented in Appendix B. We set a threshold for $\varepsilon_{backward}$ in the backward algorithm. Note that the computation of the correlation dimension is required in Algorithm 1. Hence, the estimation of the correlation dimension is explained in Section Estimation of Correlation Dimension.

---

**Algorithm 1** *oGeoC* Algorithm

---

1: **procedure** FORWARD GEOC($\mathbf{x}_n^{(i)} \ \forall \ i = 1, 2, \ldots N$ and $n = 1, 2, \ldots, T$ ;
    $\mathcal{V}$: the vertex set; $\varepsilon_{forward}$: threshold for zero)
2:   *Initialize:* $\mathcal{N} \leftarrow \varnothing$
3:   **for** $i \in \mathcal{V}$ **do**
4:     *Initialize:* $\mathcal{K} \leftarrow \varnothing$, *index\_max* $\leftarrow \varnothing$
5:     **do**$\{\mathcal{K} \leftarrow \mathcal{K} \cup \{index\_max\} \}$
6:       **for** $j \in \mathcal{V} - \mathcal{K}$ **do**
7:         geoc\_j[j] = $GeoC_{j \to i \mid \mathcal{K}}$
8:       **end for**
9:       $max_{geoc} = \max_j\{geoc\_j\}$
10:       $index\_max = \arg\max\{geoc\_j\}$
11:     **while** $max_{geoc} > \varepsilon_{forward}$
12:     $\mathcal{N}[i] \leftarrow \mathcal{K}$
13:   **end for**
14:   **return** $\mathcal{I}, \mathcal{N}$
15: **end procedure**

16: **procedure** BACKWARD GEOC($\mathbf{x}_n^{(i)} \ \forall \ i = 1, 2, \ldots N$ and $n = 1, 2, \ldots, T$ ;
    Set of nodes $\mathcal{I} \subset \mathcal{V}$; and set of nodes $\mathcal{N} \subset \mathcal{V}$: the candidate set of causal parents of $\mathcal{I}$;
    $\varepsilon_{backward}$: threshold for zero)
17: *Initialize:* $\hat{\mathbf{A}} = \mathbf{0}_{N \times N}$ whose elements $[\hat{a}_{ij}]$
18:   **for** $i \in \mathcal{I}$ **do**
19:     $\mathcal{K} \leftarrow \mathcal{N}[i]$
20:     **for** $j \in \mathcal{K}$ **do**
21:       **if** $GeoC_{j \to i \mid \mathcal{K}-\{j\}} > \varepsilon_{backward}$ **then**
22:         $\hat{a}_{ij} = 1$
23:       **end if**
24:     **end for**
25:   **end for**
26:   **return** $\hat{\mathbf{A}}$: estimated adjacency matrix
27: **end procedure**

Estimation of Correlation Dimension

$\mathcal{D}_2(\cdot)$ is used to estimate $GeoC_{J\to I|K}$. Hence, the correlation dimension and its implementation details are discussed in this section.

Consider the probability of a trajectory being within a ball $\mathcal{B}_\epsilon(\mathbf{x})$ of radius $\epsilon$ around $\mathbf{x}$, which is defined as $p_\epsilon(\mathbf{x}) = \int_{\mathcal{P}_{\epsilon(\mathbf{x})}} d\mu(\mathbf{x})$. Then, the generalized correlation integral becomes [33]

$$C_q(\epsilon) = \int_{\mathbf{x}} p_\epsilon(\mathbf{x})^{q-1} d\mu(\mathbf{x}). \tag{9}$$

The correlation integral in Equation (9) can be rewritten using a Heaviside step function as follows

$$C_q(\epsilon) = \int_{\mathbf{x}} \left[ \int_{\mathbf{y}} \Theta(\epsilon - ||\mathbf{x} - \mathbf{y}||) \, d\mu(\mathbf{y}) \right]^{q-1} d\mu(\mathbf{x}). \tag{10}$$

Here, $\Theta$ is the Heaviside step function, which is defined as $\Theta(x) = 1$ if $x > 0$ and $\Theta(x) = 0$ otherwise.

Grassberger and Proccacia have discussed the correlation integral for the case of $q = 2$ [36,37]. From a finite set of observations of $\mathbf{x}_i$, the estimation of the correlation integral is given as

$$\hat{C}(\epsilon) = \frac{1}{[T(T-1)]^{q-1}} \sum_{i=1}^{T} \left[ \sum_{i\neq j} \Theta(\epsilon - ||\mathbf{x}_i - \mathbf{x}_j||) \right]^{q-1} \tag{11}$$

where $T$ is the number of samples. The modified version of the correlation sum when $q = 2$ is described in [38]

$$\hat{C}(\epsilon) = \frac{2}{T(T-1)} \sum_{i=1}^{T} \sum_{j\neq i+1}^{T} \Theta(\epsilon - ||\mathbf{x}_i - \mathbf{x}_j||). \tag{12}$$

The summation terms count the pairs $(\mathbf{x}_i, \mathbf{x}_j)$ for which distance $||\mathbf{x}_i - \mathbf{x}_j|| < \epsilon$. It is expected that $\hat{C}(\epsilon)$ scales a power law, $\hat{C} \propto \epsilon^{\mathcal{D}_2}$ when $N \to \infty$ and $\epsilon \to 0$. The correlation dimension $\mathcal{D}_2$ is defined as

$$d(T, \epsilon) = \frac{\partial \ln C(\epsilon, T)}{\partial \ln \epsilon}$$
$$\mathcal{D}_2 = \lim_{\epsilon \to 0} \lim_{N \to \infty} d(T, \epsilon). \tag{13}$$

A well-known technique for estimating $\mathcal{D}_2$ involves obtaining the slope of the $\ln C(\epsilon, T) / \ln \epsilon$ curve in linear regions for $T \gg 0$ [39]. First, $\ln C(\epsilon, T)$ is plotted against $\ln \epsilon$ by increasing $\epsilon$ until $\ln C(\epsilon, T)$ no longer changes with increasing $\ln \epsilon$. Then, the slope of the $\ln C(\epsilon, T) / \ln \epsilon$ in the linear region is determined using a numerical estimation method, particularly least squares estimation. Clearly, the estimation of $\mathcal{D}_2$ depends on the number of samples $T$, minimum radius $\epsilon_{min}$, maximum radius $\epsilon_{max}$ and the number of radius steps $\#_{rs}$ between $\epsilon_{min}$ and $\epsilon_{max}$. The details are demonstrated in Appendix A.

## 3. Results

In this section, we present some examples to demonstrate the performance of the proposed method.

**Example 1.** *In the first example, we choose the logistic map as a dynamical system in Equation (2), and its state equation is defined as*

$$x_{n+1} = f(x_n) = a x_n (1 - x_n) \quad x_0 \in [0, 1]. \tag{14}$$

*Here, for $a = 4$, the system is chaotic. The number of nodes in the network is $N = 7$ in (2), and we consider directed and bidirectional coupled networks as shown in Figure 1a,b. The coupling*

strength is $\sigma = 0.1$, and the inner coupling matrix is $\kappa = 1$. We choose the coupling function as $g(x_n^{(i)}, x_n^{(j)}) = f(x_n^{(j)}) - f(x_n^{(i)})$. The number of the permutation is selected as $N_p = 100$ to determine the zero $\varepsilon_{forward}$, and the significance threshold is $\theta = 0.01$. The performance of the proposed algorithm is defined in terms of the true positive rate (TPR), the false positive rate (FPR) and the receiver operating characteristic (ROC) curve [40].
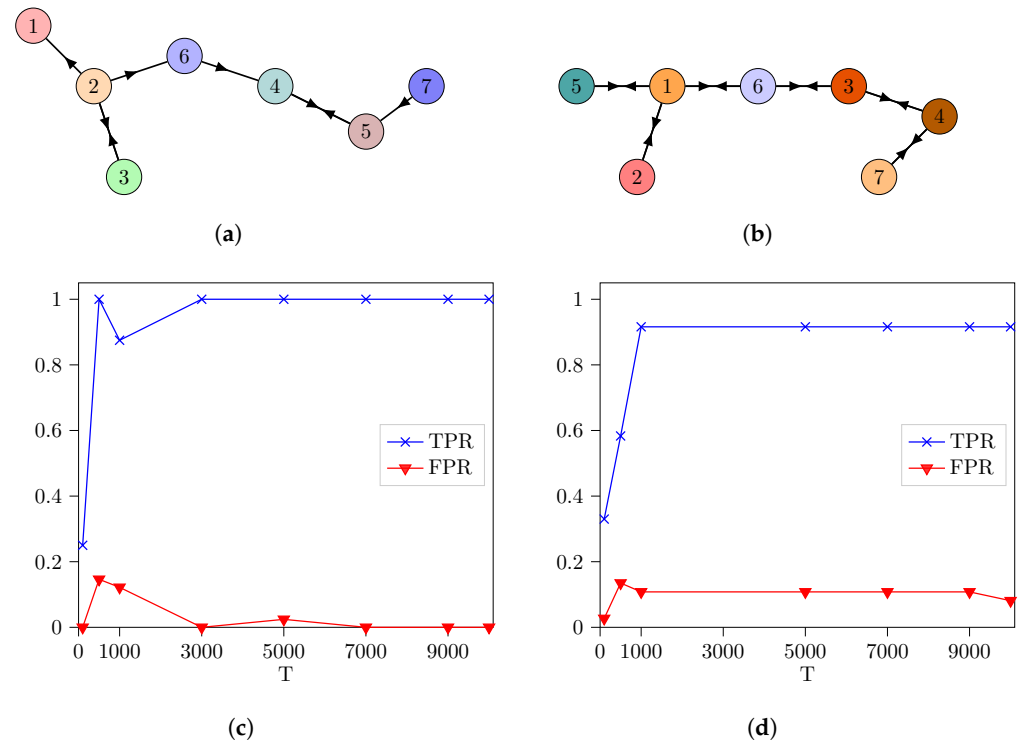


**Figure 1.** The performance of the proposed algorithm for the networks in (**a**,**b**). (**a**) A network with directed coupling consisting of $N = 7$ nodes and 8 links. (**b**) A network with bidirectional coupling consisting of $N = 7$ nodes and 12 links. In both networks, each circle represents a logistic map. (**c**) TPRs and FPRs are plotted against various sample sizes ($T$) for the network in Figure (**a**). (**d**) We also illustrated TPRs and FPRs with respect to ($T$) for the network in Figure (**b**). In both simulations, the number of permutations is $N_p = 100$, and the significance threshold is $\theta = 0.01$.

Figure 1c,d demonstrate the TPRs and FPRs with respect to various sample sizes (T) when the network is selected in Figure 1a,b, respectively. As the number of samples increases, the TPR reaches one, and the FPR becomes zero as depicted in Figure 1c. We could find all links correctly for the network in Figure 1a using the *oGeoC* algorithm. The TPR is almost one when $T$ is increased (i.e., the algorithm misses only one link as a false negative in Figure 1b), and FPR drops when $T > 9000$, as shown in Figure 1d.

**Example 2.** *In this example, we investigate the case of randomly coupled networks. We use the Erdős–Rényi (ER) model to generate random graphs [41], creating random couplings with a probability of p = 0.1 for the networks. The number of nodes is selected as N = 20. In particular, we choose the same dynamical system as the logistic map and the network parameters in Example 1. However, the number of independent trials is 10.*

We show only one of the realizations in Figure 2a. We illustrate TPRs and FPRs for different sample sizes in Figure 2b. The points represent the average values of TPRs and FPRs over 10 trials. The maximum point of the error bar indicates the highest value of TPRs and FPRs, while the minimum point represents their lowest values. Additionally, the ROC curve is demonstrated in Figure 2c with $T = 500$ and $T = 1000$ for the network in

Figure 2a. The number of permutations is chosen as $N_p = 100$. The significance threshold $\theta$ varies from 0.01 to 0.99 in Figure 2c to extract the ROC curve.

TPR approaches one, and the interval between the maximum and minimum values of the error bars is decreased when $T$ is increased, as shown in Figure 2b. FPR is slightly reduced but not equal to zero for $T = 10,000$. In summary, the algorithm achieves to find true links in the networks; it detects a small number of links as false positives.

The ROC curves in Figure 2c indicate that the performance of the proposed algorithms is improved when the number of samples is increased as expected. However, when $\theta$ is increased, the performance of algorithms decreases even if $N_p$ is large.



(**a**)



(**b**)                    (**c**)
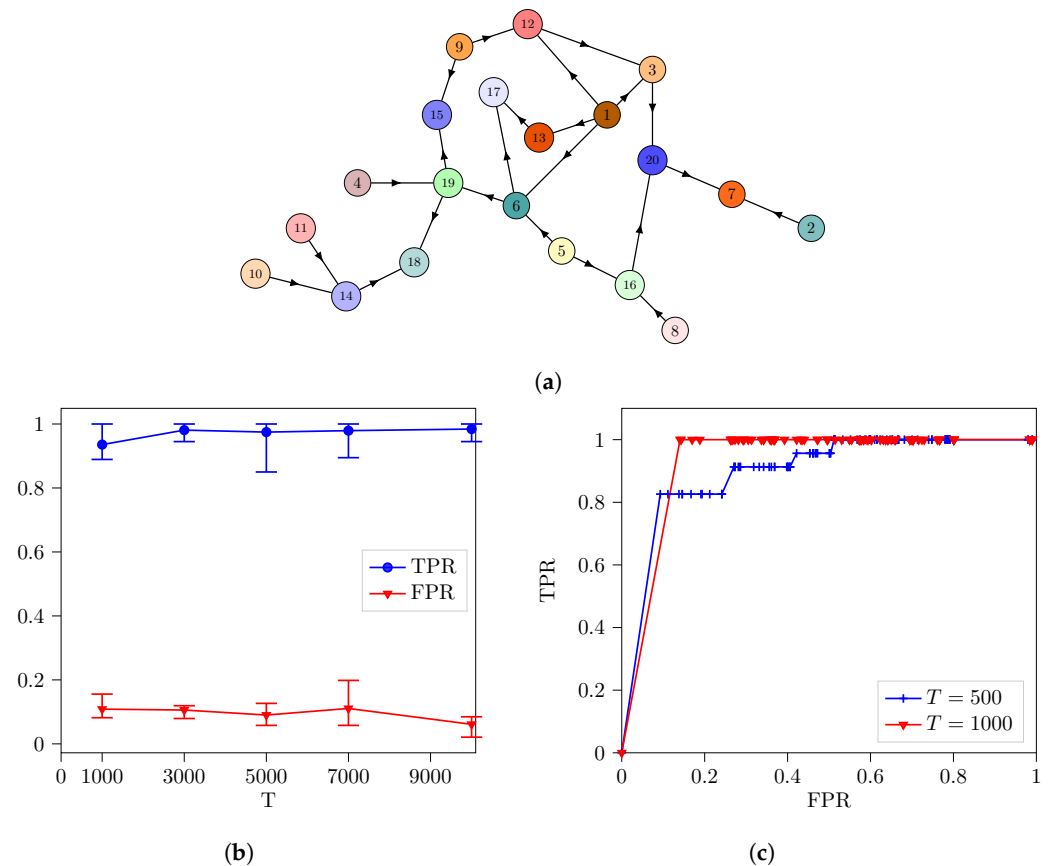
**Figure 2.** The performance of the proposed algorithm for the networks randomly coupled according to the Erdős–Rényi (ER) model with the probability of $p = 0.1$. The simulations are repeated 10 times for different networks. (**a**) An illustration of one of the networks. (**b**) Error bar points show the mean of TPRs and FPRs with respect to different sample sizes. The maximum and minimum points of the error bar indicate the highest and lowest values of TPRs and FPRs. The number of permutations is again $N_p = 100$ with $\theta = 0.01$. (**c**) ROC curve for $T = 500$ and $T = 1000$. Here, $N_p = 100$, but $\theta$ is varied from to 0.01 to 0.99 to plot the ROC curve.

## 4. Discussion

In this study, we investigated the causal inference of networks using only geometric interpretations. We utilized the conditional correlation dimensional geometric information flow measure based on the correlation dimension to accomplish this. We proposed the *oGeoC* principle, which allows us to find causal and noncausal parents of a node, thereby identifying direct and indirect links through geometric sense. We tested our proposed method on coupled logistic maps. Our findings revealed that we could find the links when the number of samples was large enough. False positives decreased when the observations were sufficient and the significance level $\theta$ was selected appropriately.

It is also important to note that the number of observations is a vital parameter when estimating *GeoC*. Although *GeoC* detects causal relationships between systems, the estimation requires a large number of observations to estimate $\mathcal{D}_2$ accurately [42].

We obtained $\mathcal{D}_2$ by finding the slope of the $\ln C(\epsilon, T) / \ln \epsilon$ curve over its linear regions. In this technique, the selection of the minimum radius $\epsilon_{min}$ and the maximum radius $\epsilon_{max}$ plays a key role. If the dimensionality of the system is high, finding a linear region in the curve can be problematic even with a large number of observations. Hence, in case of large networks, it is important to increase the number of observations and to select the radius $\epsilon_{min}$ and $\epsilon_{max}$ by considering the linear region of the $\ln C(\epsilon, T) / \ln \epsilon$ curve.

Furthermore, we plotted ROC curves according to the significance level $\theta$ when the number of shuffles was fixed. If the number of shuffles is large enough, the *oGeoC* algorithm can determine the causal links while removing the noncausal links. When the significance level $\theta$ is increased, the false positive rate reaches one even if the number of shuffles is large, as the ROC curve depicts.

As a direction for future work, it would be interesting to apply the *oGeoC* principle to real data. In real data analysis, it may not be possible to observe all the states of the network nodes. In this case, it is necessary to reconstruct the phase space from a single observation of a node. The embedding dimension parameters (e.g., the delay of embedding and the embedding dimension) are significant factors when reconstructing the state space using Takens' embedding theorem [39]. When the length of the data is sufficient, and there is no noise in the data, there exists a diffeomorphism between the reconstructed state space and the original space. It ensures the invariants of the system, such as $\mathcal{D}_2$, are preserved in the reconstructed phase space. However, the observations are generally too short or noisy in real data. Therefore, the choice of the embedding dimension parameters is essential for an accurate estimation of $\mathcal{D}_2$.

Determining $\mathcal{D}_2$ can also become challenging in noisy conditions. The studies in [33,43] have shown that we may not find the significant scaling region in the correlation sum and the linear interval of the correlation dimension when the dataset has a 2% noise level. Therefore, we may not obtain a reliable estimation of *GeoC* in the presence of noise.

In addition, it is known that $\mathcal{D}_2$ estimation can be made with several techniques [44–46]. The *oGeoC* principle with different $\mathcal{D}_2$ estimation techniques or noisy data can be investigated for further exploration.

Although existing studies and our proposed method deal with detecting causal inference in deterministic systems, the question of which method to use in stochastic systems, especially in weak and moderate coupling, remains unresolved. Another open question is whether different dynamic properties of interacting systems might bias the estimation of the causal direction. In a recent study, the ability of a state–space correspondence method to identify causal direction in nonlinear bivariate stochastic processes was investigated to solve these problems [47].

In our case, it is known that the *oGeoC* principle utilizes the estimation of the correlation dimension. When the variance noise level of nonlinear bivariate stochastic processes increases, the estimation of the correlation dimensions for the systems becomes less reliable. As a result, it becomes difficult to identify causal and noncausal parents of a node, and the performance of the proposed algorithms will be reduced accordingly. A potential solution to these challenges can be to examine the estimation of the correlation dimension using the newly proposed techniques in [44–46] for nonlinear bivariate stochastic processes.

To conclude, we showed that the *oGeoC* principle can detect the causal parents of a node in a network when the observations are long enough and entirely noise-free. It will be interesting to apply the *oGeoC* principle to real data where the causal relations are unknown, including time series of air quality, temperature, and humidity. Future studies should also explore applications involving stochastic interactions, such as determining causality in physiological control mechanisms, brain activity interactions, and coupled ocean–atmosphere chaotic systems. We aim to test this principle in a more detailed study and explore their causal relationships.

## Appendix A. Estimation of Correlation Dimension

Given a time series $\{\mathbf{x}_n\}_{t=1}^T$, our goal is to estimate $\mathcal{D}_2$. Although $\mathcal{D}_2$ is invariant, the correlation sum Equation (12) is not invariant for a given $\epsilon$ [33]. Therefore, the correlation sum in Equation (12) is calculated for several radii at first. Then, the curve of $\ln \hat{C}(\epsilon) / \ln(\epsilon)$ is plotted. The slope of the curve is computed over a linear region. In this paper, we use the *Linear Regression* model to determine the slope. The pseudo-code is given in Algorithm A1.

In our simulations, we choose the minimum radius as $\epsilon_{max} = 0.0562$, the maximum radius as $\epsilon_{max} = 0.630$, and the number of radius steps as $\#_{rs} = 50$ for the coupled logistic networks. We utilize *kDTree* to count the pairs inside the ball of each $\epsilon$ and use *multiprocessing* tools when estimating $\mathcal{D}_2$.

---

**Algorithm A1** $\mathcal{D}_2$ Estimation

---

1: **procedure** $\mathcal{D}_2$ ESTIMATION($\mathbf{x}, \epsilon_{min}, \epsilon_{max}, \#_{rs}$)
2:     *Initialize:* step_$\epsilon = (\epsilon_{max} - \epsilon_{min})/\#_{rs}$,
3:     *slope_array* $\leftarrow [\,]$
4:     *radius_array* $\leftarrow [\,]$
5:     **for** $r = \epsilon_{min}$ **to** $\epsilon_{max}$ **step** step_$\epsilon$ **do**
6:         Calculate $\hat{C}(r)$ using $\mathbf{x}$
7:         *slope_array* $\leftarrow$ *slope_array* $+ [\ln \hat{C}(r)]$
8:         *radius_array* $\leftarrow$ *radius_array* $+ [\ln(r)]$
9:     **end for**
10:     $(\mathcal{D}_2, residuals) \leftarrow$ Linear_Regression(*radius_array*, *slope_array*)
11:     **return** $\mathcal{D}_2$
12: **end procedure**

---

## Appendix B. Shuffle Test to Determine the Zero

There are two thresholds, such as $\varepsilon_{forward}$ and $\varepsilon_{backward}$, to determine the zero in the *GeoC* algorithm. We perform a statistical significance test to check whether *GeoC* is greater than zero. The idea is to obtain an empirical cumulative distribution from the shuffled *GeoC* and use it to determine the significance level.

To achieve this, we generate a random permutation array that shuffles the time and surrogate $\mathbf{x}_n^{(j)}$ according to the permutation as $\mathbf{x}_n^{(j^\star)}$. Then, $GeoC_{j^\star \to i \mid \mathcal{K}}$ is calculated. The procedure is repeated $N_p$ times, and the shuffled *GeoC*s are sorted in ascending order. The significance threshold $\theta$ and $N_p$ are used to set the predefined index, which determines the significance level. $\varepsilon$ is the *predefined index*$^{th}$ element of the ascending list of shuffled *GeoC* values. The pseudo-code is given in Algorithm A2.

The determination of $\varepsilon$ depends significantly on the number of shuffles, $N_p$ and $\theta$. As the $N_p$ increases and $\theta$ decreases, we can determine zero significantly.

---

**Algorithm A2** Shuffle Test

---

1: **procedure** SHUFFLE TEST($\mathbf{x}_n^{(i)}$, $\mathbf{x}_n^{(j)}$ $\mathbf{x}_n^{(\mathcal{K})}$ for $n = 1, 2, \ldots, T$ ; , $N_p$ : number of shuffles, $\theta$: significance threshold)
2:     *Initialize: array_index $\leftarrow int(N_p \times (1 - \theta))$,*
3:     *shuffle_array $\leftarrow [\,]$*
4:     **for** *trial $\in N_p$* **do**
5:         Generate a random permutation array that shuffles n
6:         Shuffle $\mathbf{x}_n^{(j)}$ according the permutation and obtain shuffled times series $\mathbf{x}_n^{(j^\star)}$
7:         Calculate $GeoC_{j^\star \to i \mid \mathcal{K}}$
8:         *shuffle_array $\leftarrow$ shuffle_array $+ [GeoC_{j^\star \to i \mid \mathcal{K}}]$*
9:     **end for**
10:    *ascending_Geo_array $\leftarrow$* sort(*shuffle_array*)
11:    *$\varepsilon \leftarrow$ ascending_Geo_array[array_index]*
12:    **return** $\varepsilon$
13: **end procedure**

---

**Appendix C. Illustration of $\mathcal{D}_2$ Estimations for the Networks**

To demonstrate $\mathcal{D}_2$ estimations, we show one step of the proposed forward algorithm for the networks in Figures 1a and 2a, respectively.

First, we start with the network in Figure 1a. Let us randomly choose the third node ($i = 3$) from the network. The forward algorithm calculates $GeoC_{j \to 3 \mid \varnothing}$ for all $j$ and returns the maximum value of $GeoC_{j \to 3 \mid \varnothing}$ for all $j$ and its index of this maximum value in the first iteration. In our case, we found that the maximum $GeoC$ as $max_{geoc} = GeoC_{2 \to 3 \mid \varnothing} = 0.163$ for the third node. When Equation (8) is written for $i = 3$ and $j = 2$, it becomes

$$
\begin{aligned}
GeoC_{2 \to 3 \mid \varnothing} &= Geo(X'^{(3)}) - Geo(X'^{(3)} | X^{(2)}) \\
&= \mathcal{D}_2(\mathcal{M}_{X'^{(3)}}) - \mathcal{D}_2(\mathcal{M}_{(X'^{(3)}, X^{(2)})}) + \mathcal{D}_2(\mathcal{M}_{X^{(2)}}).
\end{aligned} \tag{A1}
$$

In the estimation of $GeoC_{2 \to 3 \mid \varnothing}$, we require to estimate the correlation dimension of $\mathcal{D}_2(\mathcal{M}_{X'^{(3)}})$, $\mathcal{D}_2(\mathcal{M}_{(X'^{(3)}, X^{(2)})})$ and $\mathcal{D}_2(\mathcal{M}_{X^{(2)}})$ for given datasets $X'^{(3)}$, $(X'^{(3)}, X^{(2)})$ and $X^{(2)}$, respectively. We use Algorithm A1 to estimate $\mathcal{D}_2(\cdot)$. We plot the curves of $\ln \hat{C}(\epsilon) / \ln(\epsilon)$ of the datasets $X'^{(3)}$, $(X'^{(3)}, X^{(2)})$ and $X^{(2)}$ in Figure A1. The blue circle points demonstrate the natural logarithm of correlation sum versus the natural logarithm of radius. The red line determines the slope of the blue points by using least squares estimation. In our case, the slope of the red line gives us the estimated correlation dimension.

In the algorithm, we determine whether the maximum value of $GeoC_{\cdot \to \cdot \mid \cdot}$ is statistically significant or not. We found that $GeoC_{2 \to 3 \mid \varnothing} > 0$. Hence, the index of the maximum $GeoC_{j \to 3 \mid \varnothing}$ becomes $\mathcal{K} = index\_max = 2$.

In the second iteration, we need to compute $GeoC_{j \to 3 \mid 2}$ for all $j$ except for $\mathcal{K} = 2$. Our results reveal that the maximum value of $GeoC_{j \to 3 \mid 2}$ becomes $GeoC_{7 \to 3 \mid 2} = 0.001$ when $j = 7$. If $j = 7$, $i = 3$, and $\mathcal{K} = 2$, Equation (6) is expressed by

$$
\begin{aligned}
GeoC_{7 \to 3 \mid 2} &= Geo(X'^{(3)} | X^{(2)}) - Geo(X'^{(3)} | X^{(2)}, X^{(7)}) \\
&= \mathcal{D}_2(\mathcal{M}_{(X'^{(3)}, X^{(2)})}) - \mathcal{D}_2(\mathcal{M}_{X^{(2)}}) - \mathcal{D}_2(\mathcal{M}_{(X'^{(3)}, X^{(2)}, X^{(7)})}) + \mathcal{D}_2(\mathcal{M}_{(X^{(2)}, X^{(7)})}).
\end{aligned} \tag{A2}
$$

Again, we illustrate the curves of $\ln \hat{C}(\epsilon) / \ln(\epsilon)$ and corresponding estimated correlation dimensions ($\mathcal{D}_2(\mathcal{M}_{(X'^{(3)}, X^{(2)})})$, $\mathcal{D}_2(\mathcal{M}_{X^{(2)}})$, $\mathcal{D}_2(\mathcal{M}_{(X'^{(3)}, X^{(2)}, X^{(7)})})$, and $\mathcal{D}_2(\mathcal{M}_{(X^{(2)}, X^{(7)})})$) in Figure A2.
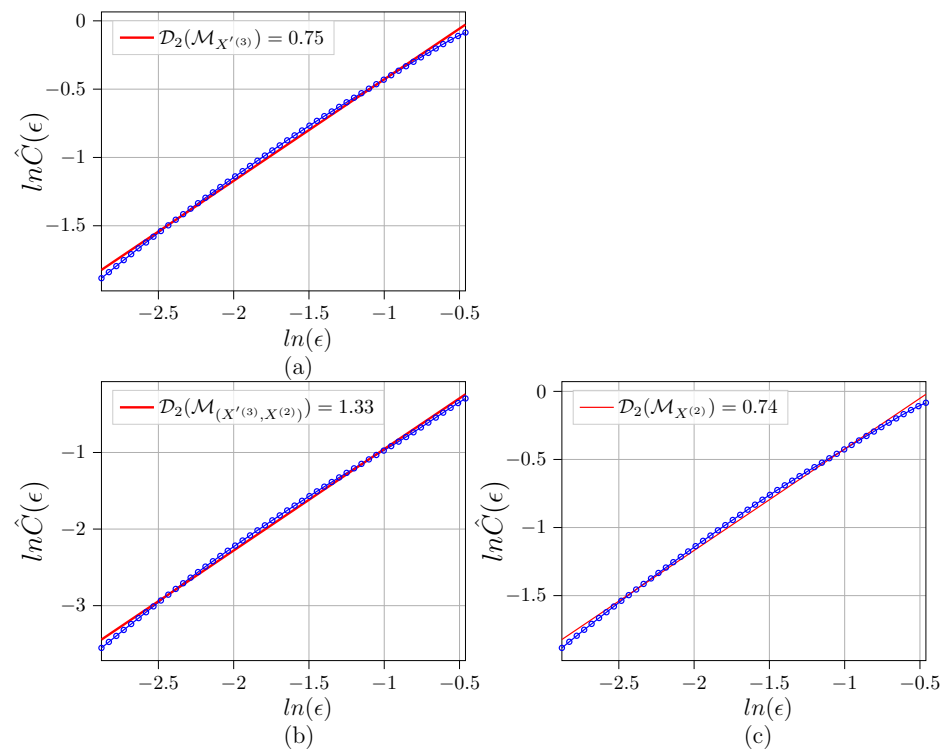
**Figure A1.** The curve of $\ln \hat{C}(\epsilon) / \ln(\epsilon)$ for a given dataset of (**a**) $X^{'(3)}$, (**b**) $(X^{'(3)}, X^{(2)})$, (**c**) $X^{(2)}$ for the network in Figure 1a. The estimated correlation dimension is shown in the legend. The number of observations is $T = 10,000$.
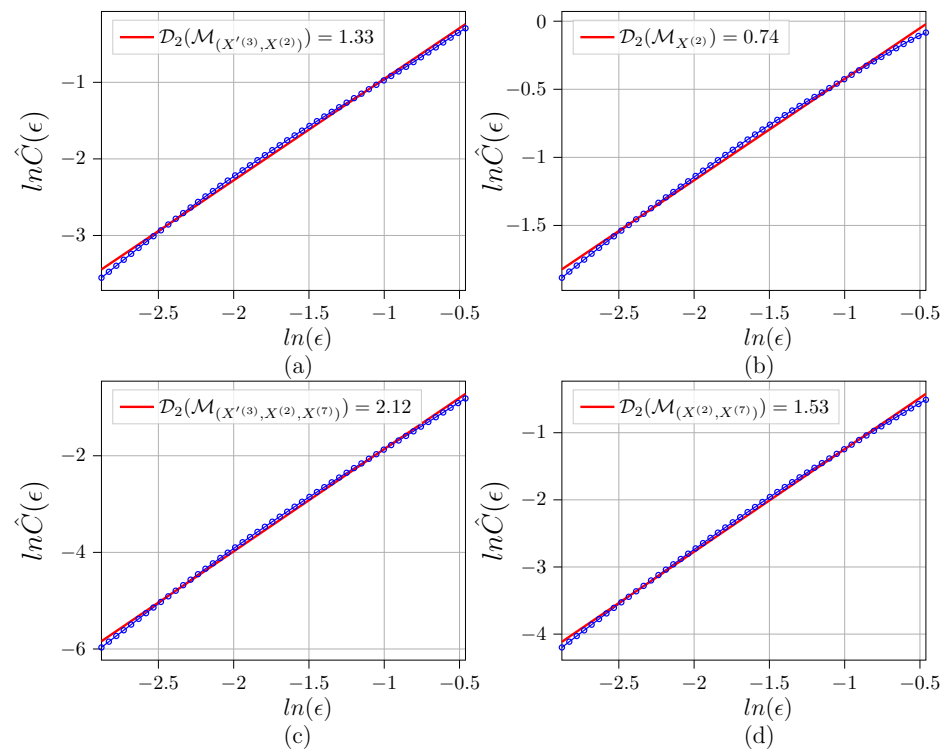


**Figure A2.** The curve of $\ln \hat{C}(\epsilon) / \ln(\epsilon)$ for a given dataset of (**a**) $(X^{'(3)}, X^{(2)})$, (**b**) $X^{(2)}$, (**c**) $(X^{'(3)}, X^{(2)}, X^{(7)})$, (**d**) $(X^{(2)}, X^{(7)})$ for the network in Figure 1a. The estimated correlation dimension is shown in the legend. The number of observations is $T = 10,000$.

In the second iteration, the maximum *GeoC* is not statistically significant ($GeoC_{7 \to 3 \mid 2} < 0$). Thus, the algorithm stops in the second iteration. We determine the causal parents of the third node as $\mathcal{K} = 2$ for the network in Figure 1a as expected.

Second, we continue with the network in Figure 2a. We randomly choose the seventh node ($i = 7$) from the network. The algorithm finds the maximum *GeoC* as $GeoC_{2 \to 7 \mid \varnothing}$ in the first iteration, and it is statistically significant ($GeoC_{2 \to 7 \mid \varnothing} > 0$). When estimating $GeoC_{2 \to 7 \mid \varnothing}$, the algorithm computes the correlation dimension of the datasets $X'^{(7)}$, $(X'^{(7)}, X^{(2)})$, and $X^{(2)}$ using the same technique in Algorithm A1. The curve of estimation of $\ln \hat{C}(\epsilon) / \ln(\epsilon)$ and the corresponding estimated $\mathcal{D}_2(\cdot)$ is shown in Figure A3.
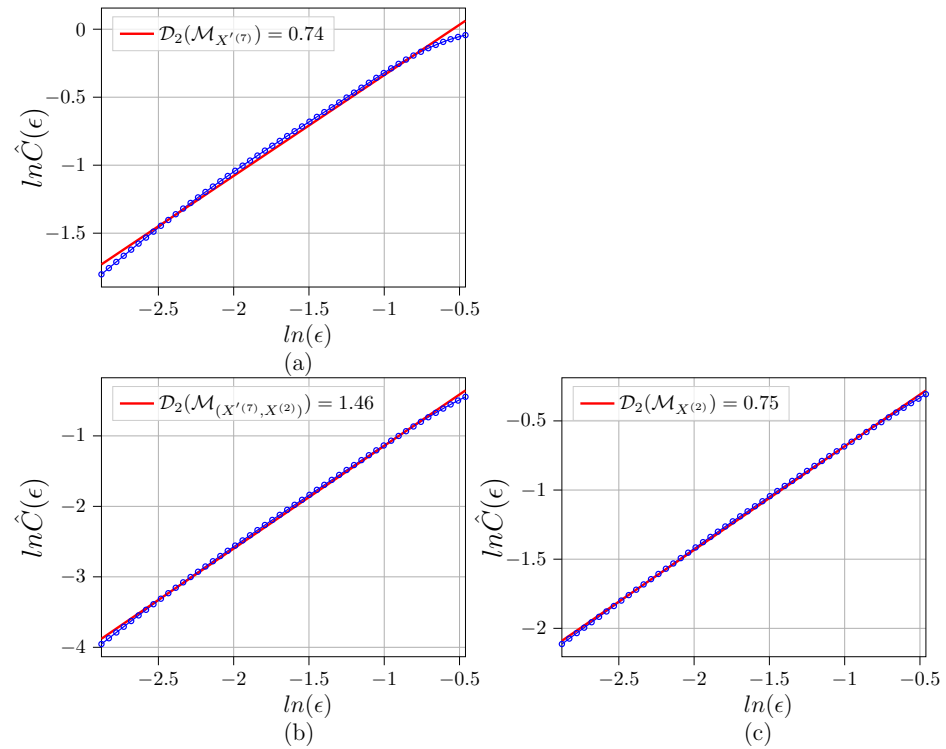


**Figure A3.** The curve of $\ln \hat{C}(\epsilon) / \ln(\epsilon)$ for a given dataset of (**a**) $X'^{(7)}$, (**b**) $(X'^{(7)}, X^{(2)})$, (**c**) $X^{(2)}$ for the network in Figure 2a. The estimated correlation dimension is shown in the legend. The number of observations is $T = 10,000$.

In the second iteration, the maximum *GeoC* is achieved $GeoC_{20 \to 7 \mid 2}$. We demonstrate the curve of $\ln \hat{C}(\epsilon) / \ln(\epsilon)$ and the estimations of $\mathcal{D}_2(\cdot)$ of the dataset $(X'^{(7)}, X^{(2)})$, $X^{(2)}$, $(X'^{(7)}, X^{(2)}, X^{(20)})$, and $(X^{(2)}, X^{(20)})$ in Figure A4. The algorithm determines $GeoC_{20 \to 7 \mid 2} > 0$ and updates $\mathcal{K} = (2, 20)$ as a candidate set of causal parents.

In the third iteration, $GeoC_{10 \to 7 \mid (2,20)}$ takes the maximum value by calculating the correlation dimension of the datasets $(X'^{(7)}, X^{(2)}, X^{(20)})$, $(X^{(2)}, X^{(20)})$, $(X'^{(7)}, X^{(2)}, X^{(7)}, X^{(10)})$, and $(X^{(2)}, X^{(7)}, X^{(10)})$. The curves of $\ln \hat{C}(\epsilon) / \ln(\epsilon)$ of these datasets and their estimated $\mathcal{D}_2(\cdot)$ are represented in Figure A5. In this case, the algorithm decides $GeoC_{10 \to 7 \mid (2,20)} < 0$ and returns the causal parents of the seventh node as $\mathcal{K} = \{2, 20\}$.
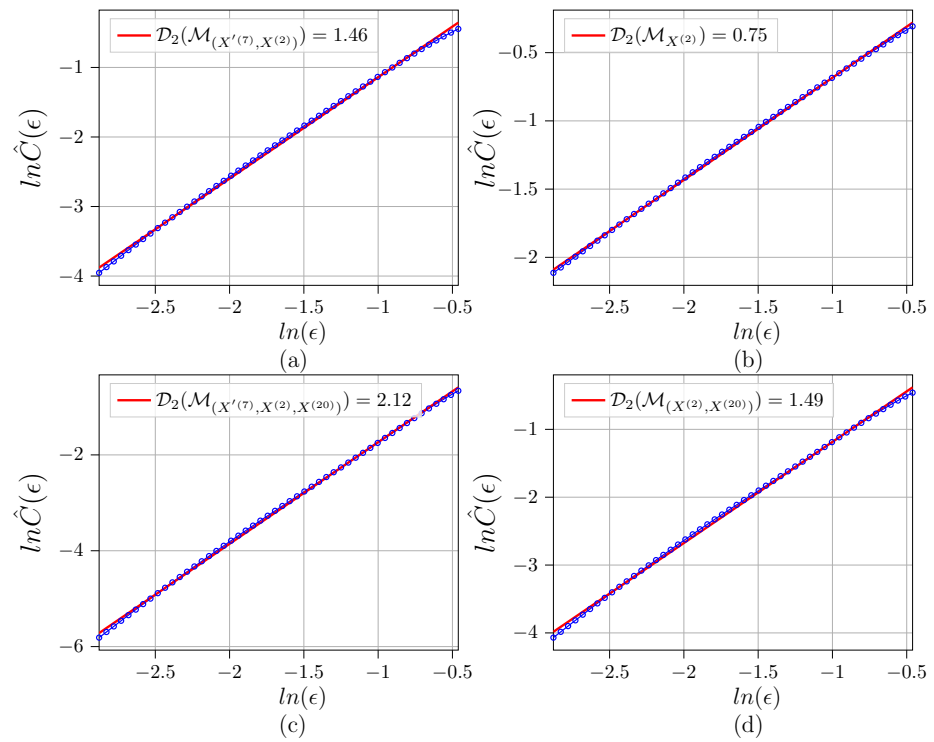
**Figure A4.** The curve of $\ln \hat{C}(\epsilon) / \ln(\epsilon)$ for a given dataset of (**a**) $(X'^{(7)}, X^{(2)})$, (**b**) $X^{(2)}$, (**c**) $(X'^{(7)}, X^{(2)}, X^{(20)})$, (**d**) $(X^{(2)}, X^{(20)})$ for the network in Figure 2a. The estimated correlation dimension is shown in the legend. The number of observations is $T = 10,000$.
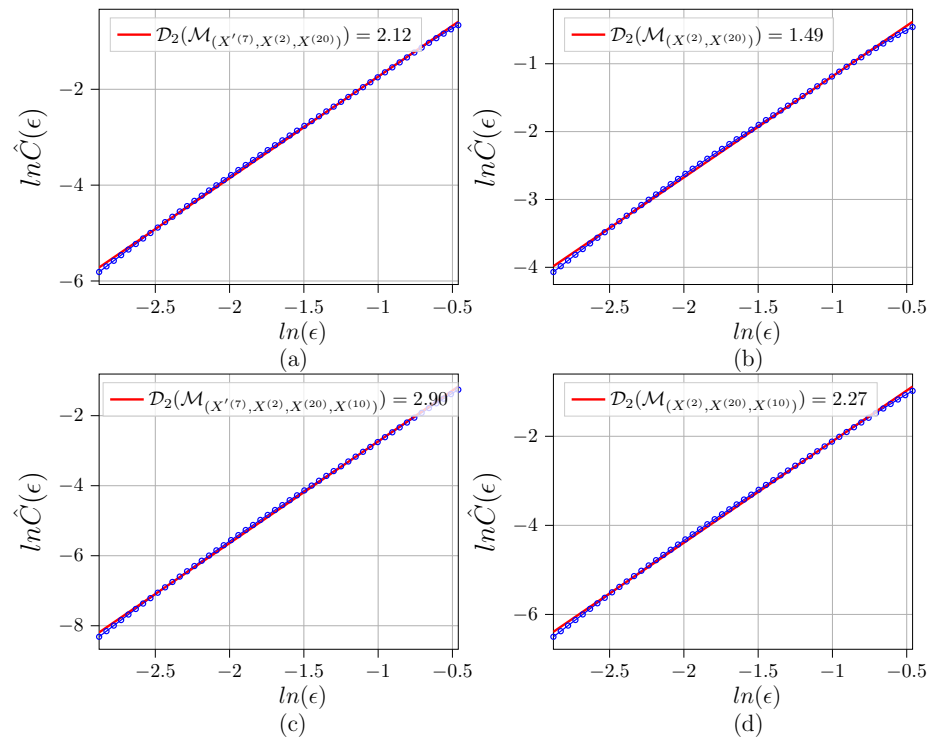


**Figure A5.** The curve of $\ln \hat{C}(\epsilon) / \ln(\epsilon)$ for a given dataset of (**a**) $(X'^{(7)}, X^{(2)}, X^{(20)})$, (**b**) $(X^{(2)}, X^{(20)})$, (**c**) $(X'^{(7)}, X^{(2)}, X^{(7)}, X^{(10)})$, (**d**) $(X^{(2)}, X^{(7)}, X^{(10)})$ for the network in Figure 2a. The estimated correlation dimension is shown in the legend. The number of observations is $T = 10,000$.

## References

1. Sudu Ambegedara, A.; Sun, J.; Janoyan, K.; Bollt, E. Information-theoretical noninvasive damage detection in bridge structures. *arXiv* **2016**, arXiv:1612.09340. [CrossRef] [PubMed]
2. Runge, J.; Bathiany, S.; Bollt, E.; Camps-Valls, G.; Coumou, D.; Deyle, E.; Glymour, C.; Kretschmer, M.; Mahecha, M.D.; Muñoz-Marí, J.; et al. Inferring causation from time series in Earth system sciences. *Nat. Commun.* **2019**, *10*, 2553. [CrossRef] [PubMed]
3. Runge, J.; Gerhardus, A.; Varando, G.; Eyring, V.; Camps-Valls, G. Causal inference for time series. *Nat. Rev. Earth Environ.* **2023**, *4*, 487–505. [CrossRef]
4. Seth, A.K.; Chorley, P.; Barnett, L.C. Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. *Neuroimage* **2013**, *65*, 540–555. [CrossRef] [PubMed]
5. Sugihara, G.; May, R.; Ye, H.; Hsieh, C.h.; Deyle, E.; Fogarty, M.; Munch, S. Detecting causality in complex ecosystems. *Science* **2012**, *338*, 496–500. [CrossRef]
6. Bollt, E.M. Open or closed? Information flow decided by transfer operators and forecastability quality metric. *arXiv* **2018**, arXiv:1804.03687. [CrossRef]
7. Surasinghe, S.; Bollt, E.M. On geometry of information flow for causal inference. *Entropy* **2020**, *22*, 396. [CrossRef]
8. Hendry, D.F. The nobel memorial prize for clive wj granger. *Scand. J. Econ.* **2004**, *106*, 187–213. [CrossRef]
9. Wiener, N. The theory of prediction. In *Modern Mathematics for Engineers*; McGraw Hill: New York, NY, USA, 1956.
10. Granger, C.W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, 424–438. [CrossRef]
11. Marinazzo, D.; Pellicoro, M.; Stramaglia, S. Kernel method for nonlinear Granger causality. *Phys. Rev. Lett.* **2008**, *100*, 144103. [CrossRef] [PubMed]
12. Barrett, A.B.; Barnett, L.; Seth, A.K. Multivariate Granger causality and generalized variance. *Phys. Rev. E* **2010**, *81*, 041907. [CrossRef] [PubMed]
13. Marinazzo, D.; Liao, W.; Chen, H.; Stramaglia, S. Nonlinear connectivity by Granger causality. *Neuroimage* **2011**, *58*, 330–338. [CrossRef] [PubMed]
14. Rulkov, N.F.; Sushchik, M.M.; Tsimring, L.S.; Abarbanel, H.D. Generalized synchronization of chaos in directionally coupled chaotic systems. *Phys. Rev. E* **1995**, *51*, 980. [CrossRef] [PubMed]
15. Schiff, S.J.; So, P.; Chang, T.; Burke, R.E.; Sauer, T. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Phys. Rev. E* **1996**, *54*, 6708. [CrossRef] [PubMed]
16. Arnhold, J.; Grassberger, P.; Lehnertz, K.; Elger, C.E. A robust method for detecting interdependences: application to intracranially recorded EEG. *Phys. D Nonlinear Phenom.* **1999**, *134*, 419–430. [CrossRef]
17. Quiroga, R.Q.; Kraskov, A.; Kreuz, T.; Grassberger, P. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Phys. Rev. E* **2002**, *65*, 041903. [CrossRef]
18. Andrzejak, R.G.; Kraskov, A.; Stögbauer, H.; Mormann, F.; Kreuz, T. Bivariate surrogate techniques: Necessity, strengths, and caveats. *Phys. Rev. E* **2003**, *68*, 066202. [CrossRef]
19. Chicharro, D.; Andrzejak, R.G. Reliable detection of directional couplings using rank statistics. *Phys. Rev. E* **2009**, *80*, 026217. [CrossRef]
20. Ye, H.; Deyle, E.R.; Gilarranz, L.J.; Sugihara, G. Distinguishing time-delayed causal interactions using convergent cross mapping. *Sci. Rep.* **2015**, *5*, 14750. [CrossRef]
21. Mønster, D.; Fusaroli, R.; Tylén, K.; Roepstorff, A.; Sherson, J.F. Causal inference from noisy time-series data—Testing the Convergent Cross-Mapping algorithm in the presence of noise and external influence. *Future Gener. Comput. Syst.* **2017**, *73*, 52–62. [CrossRef]
22. Breston, L.; Leonardis, E.J.; Quinn, L.K.; Tolston, M.; Wiles, J.; Chiba, A.A. Convergent cross sorting for estimating dynamic coupling. *Sci. Rep.* **2021**, *11*, 20374. [CrossRef] [PubMed]
23. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461. [CrossRef] [PubMed]
24. Sun, J.; Bollt, E.M. Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings. *Phys. D Nonlinear Phenom.* **2014**, *267*, 49–57. [CrossRef]
25. Sun, J.; Taylor, D.; Bollt, E.M. Causal network inference by optimal causation entropy. *SIAM J. Appl. Dyn. Syst.* **2015**, *14*, 73–106. [CrossRef]
26. Lord, W.M.; Sun, J.; Ouellette, N.T.; Bollt, E.M. Inference of causal information flow in collective animal behavior. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2016**, *2*, 107–116. [CrossRef]
27. Janjarasjitt, S.; Loparo, K. An approach for characterizing coupling in dynamical systems. *Phys. D Nonlinear Phenom.* **2008**, *237*, 2482–2486. [CrossRef]
28. Krakovská, A.; Budáčová, H.; Interdependence measure based on correlation dimension. In Proceedings of the 9th International Conference on Measurement, Bratislava, Slovakia, 27–30 May 2013; pp. 31–34.
29. Krakovská, A.; Jakubík, J.; Budáčová, H.; Holecyová, M. Causality studied in reconstructed state space. Examples of uni-directionally connected chaotic systems. *arXiv* **2015**, arXiv:1511.00505.
30. Krakovská, A. Correlation dimension detects causal links in coupled dynamical systems. *Entropy* **2019**, *21*, 818. [CrossRef]

31. Takens, F. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence*; Springer: New York, NY, USA, 1981; pp. 366–381.
32. Cummins, B.; Gedeon, T.; Spendlove, K. On the efficacy of state space reconstruction methods in determining causality. *SIAM J. Appl. Dyn. Syst.* **2015**, *14*, 335–381. [CrossRef]
33. Kantz, H.; Schreiber, T. *Nonlinear Time Series Analysis*; Cambridge University Press: Cambridge, UK, 2003.
34. Sauer, T.; Yorke, J.A.; Casdagli, M. Embedology. *J. Stat. Phys.* **1991**, *65*, 579–616. [CrossRef]
35. Good, P. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*; Springer: New York, NY, USA, 2005.
36. Grassberger, P.; Procaccia, I. Characterization of strange attractors. *Phys. Rev. Lett.* **1983**, *50*, 346. [CrossRef]
37. Grassberger, P.; Procaccia, I. Measuring the strangeness of strange attractors. *Phys. D Nonlinear Phenom.* **1983**, *9*, 189–208. [CrossRef]
38. Theiler, J. Efficient algorithm for estimating the correlation dimension from a set of discrete points. *Phys. Rev. A* **1987**, *36*, 4456. [CrossRef] [PubMed]
39. Abarbanel, H. *Analysis of Observed Chaotic Data*; Springer Science & Business Media: New York, NY, USA, 2012.
40. Sammut, C.; Webb, G.I. *Encyclopedia of Machine Learning*; Springer Science & Business Media: New York, NY, USA, 2011.
41. Erdős, P.; Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **1960**, *5*, 17–60.
42. Eckmann, J.P.; Ruelle, D. Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems. *Phys. D Nonlinear Phenom.* **1992**, *56*, 185–187. [CrossRef]
43. Schreiber, T. Determination of the noise level of chaotic time series. *Phys. Rev. E* **1993**, *48*, R13. [CrossRef]
44. Ji, C.; Zhu, H.; Jiang, W. A novel method to identify the scaling region for chaotic time series correlation dimension calculation. *Chi. Sci. Bull.* **2011**, *56*, 925–932. [CrossRef]
45. Krakovská, A.; Chvosteková, M. Simple correlation dimension estimator and its use to detect causality. *Chaos Solitons Fractals* **2023**, *175*, 113975. [CrossRef]
46. Makarov, V.A.; Muñoz, R.; Herreras, O.; Makarova, J. Correlation dimension of high-dimensional and high-definition experimental time series. *Chaos Interdisc. Nonlinear Sci.* **2023**, *33*, 123114. [CrossRef]
47. Porta, A.; de Abreu, R.M.; Bari, V.; Gelpi, F.; De Maria, B.; Catai, A.M.; Cairo, B. On the validity of the state space correspondence strategy based on k-nearest neighbor cross-predictability in assessing directionality in stochastic systems: Application to cardiorespiratory coupling estimation. *Chaos An Interdisc. J. Nonlinear Sci.* **2024**, *34*, 053115. [CrossRef]