

Article

Not peer-reviewed version

---

# Network Approach Identifies the Control Cities of O<sub>3</sub> Pollution

---

[Zhidan Zhao](#) , Demei Xue , Haojun Sun , [Weiping Wang](#) , [Na Ying](#) \*

Posted Date: 28 November 2023

doi: 10.20944/preprints202311.1827.v1

Keywords: ozone; complex network; maximum matching algorithm; community classification



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

# Network Approach Identifies the Control Cities of O<sub>3</sub> Pollution

Zhi-dan Zhao <sup>1,2</sup>, Demei Xue <sup>2,3</sup>, Haojun Sun <sup>2,3</sup>, Weiping Wang <sup>4</sup> and Na Ying <sup>5,\*</sup>

<sup>1</sup> China Complexity Computation Lab, Department of Computer Science, School of Engineering, Shantou University, Shantou 515063, China

<sup>2</sup> Key Laboratory of Intelligent Manufacturing Technology (Shantou University), Ministry of Education, Shantou 515063, China

<sup>3</sup> Complexity Computation Laboratory, Department of Computer Science, School of Engineering, Shantou University, Shantou 515063, China

<sup>4</sup> School of National Safety and Emergency Management, Beijing Normal University, Zhuhai 519087, China

<sup>5</sup> China State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing, 100012, China

\* Correspondence: yingna@craes.org.cn

**Abstract:** In recent years, ozone (O<sub>3</sub>) pollution has been rapidly spreading, restricting further improvement of air quality in China. Investigating the interaction of O<sub>3</sub> concentration and identifying their driven cities are important for the prevention and control of O<sub>3</sub> pollution in China. However, the complex interaction between O<sub>3</sub> pollution between the cities and their driven cities has not yet been revealed. In this study, we address this gap based on complex network methods. Specifically, an ozone relational network is constructed using an association calculation method. The driven nodes and spatial clusters were analyzed based on the maximum matching algorithm and the Louvain algorithm. The findings of the study reveal an aggregation phenomenon in the ozone network concerning distance. Furthermore, as the threshold  $T_c$  varies, the proportion of driven nodes exhibits a positive correlation. Moreover, a closer threshold value corresponds to a higher coincidence ratio of the driven nodes. The results provide scientific guidance for national O<sub>3</sub> pollution prevention and regional synergy formatting. Furthermore, the introduced network-based approaches offer a mythological framework for the study of air pollution in key cities and clusters.

**Keywords:** ozone; complex network; maximum matching algorithm; community classification

## 1. Introduction

With the rapid development of human society, environmental issues are receiving increasing attention [1,2]. Among these, air pollution is one of the world's most significant environmental problems, with a direct impact on the physical and mental health of its inhabitants [3]. A large number of actions have been taken to improve air quality. To improve air quality, the government released a series of policies, such as the State Council issued Air Pollution Prevention and Control Action Plan and the Blue Sky Defense War [4,5]. It's glad to see the PM<sub>2.5</sub> concentrations in China have significantly decreased, however, the concentration of ozone (O<sub>3</sub>) is showing an upward and spreading trend, becoming a limiting factor for further improvement of our air quality [6,7]. controlling O<sub>3</sub> pollution become urgent issues that need to be addressed in each city.

O<sub>3</sub> pollution is a secondary pollutant generated by photochemical reactions that have adverse effects on human health and ecosystems [8]. Nitrogen oxides and VOCs are one of the main precursor pollutants for the formation of O<sub>3</sub> [9–12]. The relationship between O<sub>3</sub>, NO<sub>x</sub>, and VOCs is not a simple linear dependence, but rather a non-linear variation [13]. The photochemical reaction of O<sub>3</sub> is very complex and is influenced by both meteorological conditions and the control of precursor pollutants [14,15]. Seasonal characteristics are observed in ozone pollution, as it is notably affected by the summer monsoon [16]. Moreover, in China, there has been a positive correlation between surface O<sub>3</sub> concentration and the ENSO index during summers from 1990 to 2019 [17]. A comprehensive study

on meteorological variables, including surface ozone, relative humidity, and temperature, was conducted in 74 major Chinese cities from 2017 to 2018. The findings revealed that ozone levels in each city decreased with increasing relative humidity, showing a negative overall correlation [18]. This suggests that net ozone production decreases as relative humidity increases [19].

In addition to industrial emissions and meteorological conditions, inter-regional atmospheric transport is an important factor in the formation of O<sub>3</sub> pollution [20–23]. Identifying driving cities is beneficial for managing atmospheric environmental quality. A large number of studies have applied numerical models, such as WRF-Chem, CMAQ, and CAMx, to analyze the transport characteristics of O<sub>3</sub> [24–26]. However, these studies mainly focus on limited cities and limited time. Research has pointed out that longer time scale analysis is of greater significance for studying the causes of O<sub>3</sub> and managing joint prevention and control. In addition, research on identifying O<sub>3</sub> driven cities is even rarer.

In recent years, complex network methods have gradually been introduced into the field of atmospheric environment [27]. A notable study conducted by Tian et al. delved into the ozone transport network in California, examining the phenomenon of networked communities. Similarly, Wang et al. applied an optimized complex network approach and source resolution model to analyze the transport patterns of O<sub>3</sub> in the Yangtze River Delta region [28,29]. Despite the abundance of valuable findings in ozone research, the application of complex network methods, especially control network theory, to comprehend certain ozone phenomena remains relatively rare.

In this study, the O<sub>3</sub> network is based on the cross-correlation function [30–32]. Subsequently, we employ the Louvain community partitioning algorithm to delineate communities within the ozone network [33]. We then employ the maximum matching control network theory to explore the driven nodes of the ozone network, among other aspects. The research findings reveal significant associations within the ozone network, primarily concentrated in regions such as Northeast China, North China, Sichuan-Chongqing, and Southeast coastal areas. Furthermore, network control theory demonstrates that the driven nodes within the ozone network are distinct from hub nodes. Additionally, we investigate the impact of the edge weight threshold on the maximum number of connected community nodes, the number of driven nodes, and the co-occurrence of driven nodes. Finally, we carefully examine the correlation between the predicted sequences and node distances for two distinct types of nodes. This comprehensive analysis uncovers a noteworthy and intricate relationship among these variables. Our research results provide reference for the development of strategies and countermeasures for O<sub>3</sub> pollution.

## 2. Materials and Methods

### 2.1. Data

The maximum daily average 8-hour (MDA8) O<sub>3</sub> concentration used in this study are acquired from the China National Environmental Monitoring Centre (CNEMC). The temporal span of the dataset encompasses the period from January 1, 2015, to December 31, 2019, encompassing a duration of 1,800 days. Our dataset comprises a grand total of 604,800 individual records.

### 2.2. Methods

In order to verify and understand the applicability of the complex network theory in the O<sub>3</sub> network, we introduce the obtained O<sub>3</sub> data in this section, as well as the cross-correlation O<sub>3</sub> network construction method, the O<sub>3</sub> network community partition algorithm Louvain Community Classification and the driven node discovery method maximum matching algorithm [33,34].

#### 2.2.1. Network construction

In recent years, more and more studies have focused on the use of complex network theories and methods to study climate phenomena. Among them, the cross-correlation function has achieved many meaningful results in the study of atmospheric gases, such as the influence of distance and so on. In this work, we employ the cross-correlation function to calculate the correlation between O<sub>3</sub>

observation sites and use it as the link weight for constructing the  $O_3$  network [35–39]. The specific calculation process of the cross-correlation function algorithm is shown in the equation (1) and (2):

$$X_{C_i,C_j}(\tau) = \frac{\sum_{t=1}^{L-\tau}(C_i(t)-\bar{C}_i)(C_j(t+\tau)-\bar{C}_j)}{\sqrt{\sum_{t=1}^{L-\tau}(C_i(t)-\bar{C}_i)^2} \cdot \sqrt{\sum_{t=1}^{L-\tau}(C_j(t+\tau)-\bar{C}_j)^2}} \quad (1)$$

$$X_{C_i,C_j}(-\tau) = \frac{\sum_{t=1}^{L-\tau}(C_i(t+\tau)-\bar{C}_i)(C_j(t)-\bar{C}_j)}{\sqrt{\sum_{t=1}^{L-\tau}(C_i(t+\tau)-\bar{C}_i)^2} \cdot \sqrt{\sum_{t=1}^{L-\tau}(C_j(t)-\bar{C}_j)^2}} \quad (2)$$

where  $\tau$  represents the delay time and  $\tau \geq 0$ , vectors  $C_i$  and  $C_j$  represent the values of  $O_3$  data for two cities  $i$  and  $j$ ,  $\bar{C}_i$  and  $\bar{C}_j$  are the averages of  $O_3$  for these two cities. Time delay  $\tau \in [-30,30]$ ,  $X_{C_i,C_j}(\tau)$  take the  $O_3$  data of city  $i$  as  $C[0:L_{max}-\tau]$ , the  $O_3$  data of city  $j$  as  $T[\tau:L_{max}]$ , while  $L_{max}$  is the total number of records of  $O_3$  data for a city.  $X_{C_i,C_j}(-\tau)$  take the  $O_3$  data of city  $i$  as  $C[\tau:L_{max}]$ , and the  $O_3$  data of city  $j$  as  $C[0:L_{max}-\tau]$ .

Equations (3) and (4) are used to calculate the positive and negative link weights of the  $O_3$  network.

$$W_{C_i,C_j}^{pos} = \frac{\max(X_{C_i,C_j}) - \text{mean}(X_{C_i,C_j})}{\text{std}(X_{C_i,C_j})} \quad (3)$$

$$W_{C_i,C_j}^{neg} = \frac{\min(X_{C_i,C_j}) - \text{mean}(X_{C_i,C_j})}{\text{std}(X_{C_i,C_j})} \quad (4)$$

$W_{C_i,C_j}^{pos}$  and  $W_{C_i,C_j}^{neg}$  are the positive and negative link weights respectively. In the process of building a city  $O_3$  network, the city is used as a node of the network, and the link weight threshold between nodes is set to  $T_C$ , when  $\tau_{C_i,C_j}^{pos} > 0$  and  $W_{C_i,C_j}^{pos} \geq T_C$ , there exists a link between city  $i$  and city  $j$ , with direction from city  $i$  to city  $j$ . The link weight is  $W_{C_i,C_j}^{pos}$ , and all eligible links are connected to form a positive  $O_3$  network (PoN). When  $\tau_{C_i,C_j}^{neg} < 0$  and  $W_{C_i,C_j}^{neg} \leq -T_C$ , there exist a link between city  $i$  and city  $j$ , and direction from city  $i$  to city  $j$  with the weight  $W_{C_i,C_j}^{neg}$ , which forms a negative  $O_3$  network (NoN). Interventionary studies involving animals or humans, and other studies that require ethical approval, must list the authority that provided approval and the corresponding ethical approval code.

### 2.2.2. Louvain Community Classification Algorithm

In this section, we introduce an algorithm to quickly divide the network community, which is Louvain algorithm. The algorithm is a method for extracting the structure of large-scale network communities with modularity as the optimization goal. It uses module gain to divide the community, which can quickly and effectively divide the community of the directed weighted network.

Louvain algorithm is a method of extracting large network community structure based on modular optimization, which uses module gain to divide the community, which can effectively divide the community and calculate fast. Applying the Louvain algorithm to the community division of the directed weighted  $O_3$  network can quickly obtain the community division results. Modularity is used to measure the quality of community division, the value range is  $[-1,1]$ , which is a measure of the density of links within the community and links between different communities. In highly modular communities, there are dense links between internal nodes, while sparse links between different communities nodes. Modularity calculation formula is defined as equation (5):

$$Q = \frac{1}{2m} \sum_{i,j} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (5)$$

The Louvain algorithm steps in dividing the community are as follows,

Step 1: Initialize each node in the network to a different community, the network with  $n$  nodes has  $n$  initial communities.

Step 2: Suppose the current node  $a$  has  $0 \dots k$  neighbor nodes, calculate the modular gain after putting node  $a$  into the community of any neighboring node  $\Delta Q$ , which is shown as equation (6). Then select the community with the largest gain value to place node  $a$ . If the modular gain is not positive, node  $a$  remains in the original community.

Step 3: Repeat step 2 until all nodes are in communities where they cannot get modular gain through movement.

$$\Delta Q = \left[ \frac{\sum_{in} + k_{i,in}}{2m} - \left( \frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[ \frac{\sum_{in}}{2m} - \left( \frac{\sum_{tot}}{2m} \right)^2 - \left( \frac{k_i}{2m} \right)^2 \right] \quad (6)$$

### 2.2.3. Maximum matching algorithm

With the introduction of the controllability of complex networks, more and more researchers focus on this field, and many interesting results have been achieved [34]. Here, we use the maximum matching control algorithm to study the O<sub>3</sub> network. The maximum matching algorithm is a structural controllability framework that can effectively identify the minimum set of driven nodes for a control network. The maximum matching in the network marks the maximum link set that does not share the starting node or the ending node, and the remaining nodes that are not matched are the minimum driven node set. These drive nodes enable control of the network. Generally, the maximum matching algorithm of a bipartite graph is used to find the minimum set of driven nodes of the network.

First, we need to convert the ordinary network into a bipartite graph network. The detailed conversion process can be found in reference [34]. Here we briefly describe the conversion steps. At the beginning, let  $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ ,  $\mathbf{V} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$ ,  $\mathbf{E} = \{(\mathbf{c}_1, \mathbf{c}_2), (\mathbf{c}_1, \mathbf{c}_2), (\mathbf{c}_1, \mathbf{c}_4), (\mathbf{c}_5, \mathbf{c}_3) \dots\}$ ,  $(\mathbf{c}_1, \mathbf{c}_2)$  indicates there exist a link from  $\mathbf{c}_1$  to  $\mathbf{c}_2$ . Constructing two disjoint subsets, denoted by + as the source node, and denoted by - as the target node,  $\mathbf{V}^+ = \{\mathbf{c}_1^+, \mathbf{c}_2^+, \dots, \mathbf{c}_n^+\}$ ,  $\mathbf{V}^- = \{\mathbf{c}_1^-, \mathbf{c}_2^-, \dots, \mathbf{c}_n^-\}$ . In a bipartite graph, any two edges in the  $\mathbf{V}^+$  and  $\mathbf{V}^-$  node sets are not attached to the same vertex, and finding the subset with the largest number of edges is the maximum matching of the graph.

### 2.2.4. LSTM time series forecasting

Long Short-Term Memory (LSTM) is a type of recurrent neural network, which is widely used in time series forecasting [40]. The LSTM model learns the data that can be observed at the  $t$  moment and predicts the data at the  $t + 1$  moment. We used LSTM to perform univariate prediction of O<sub>3</sub> data, selected O<sub>3</sub> data for a total of 3 years from 2015 to 2017, and entered the O<sub>3</sub> data of the previous two and a half years to predict the O<sub>3</sub> data of the last half year. The steps are as follows,

Step 1: Pretreatment data, including loading O<sub>3</sub> data sorted by date, differential conversion of data, conversion of data into supervised learning data, and data normalization.

Step 2: Construct LSTM model for predictive training. After the inverse scaling and inverse differential transformation of the predicted values, the MSE Loss between the real values and predicted values is calculated to obtain the prediction accuracy of the model, and the model parameters are adjusted through reverse gradient propagation to improve the prediction accuracy of the model.

Step 3: Predict O<sub>3</sub> data of the city with the above-trained model.

## 3. Results and discussion

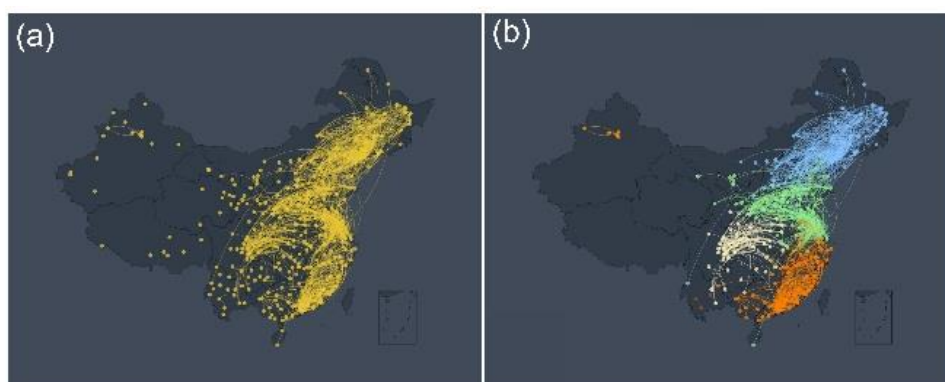
We present the main results of the correlated multi-layered networks composed of the TI and AQI as described above.

### 3.1. The characteristics of O<sub>3</sub> network

In order to visually display the network connection between various monitoring sites in China. First, we constructed the O<sub>3</sub> link network according to the above link weight calculation method [30–32,35–39]. According to the method in Section II B we get the connection weights of 336 cities in China and build an O<sub>3</sub> network, which consists of 336×335=112560 edges. To better describe the network structure, we extract the edge whose weight  $W_{c_i, c_j}^{pos} \geq T_c$  with  $T_c = 3.8$  and obtain 1942 edges, which satisfies the sparse network requirement  $e < n \log n$  [41]. Additionally, we discuss the effect of edge weights in Section III C. The extracted network in Figure 1 (a) indicates more

reasonable and understandable relationships. As can be seen from it, the network community structure can be observed through the above network construction method. For example, there are obvious community effects in Northeast China, North China, Sichuan-Chongqing, and Southeast coastal areas. These results are similar to the regional community effects seen in other previous meteorological studies [28,40]. Of course, we observe a sporadic and isolated distribution of nodes in the southwest and most of the northwest. One reason is that due to the limitation of data acquisition, currently there are only a small number of data collection sites in the southwest and northwest, resulting in a long distance between the sites and little mutual influence between the sites. As an illustration, consider the scenario where the distance between stations within a given community spans 500 kilometers, while the separation between stations in distinct communities extends to 1,500 kilometers. An additional factor contributing to this phenomenon could be attributed to the topographical features of the southwest and northwest regions. These areas are characterized by a multitude of mountains and basins, thereby impeding the free movement of air to a notable extent. It is therefore speculated that our  $O_3$  network construction method is effective and has practical value to a certain extent and can provide a new idea for the research and analysis of  $O_3$  to a certain extent.

To further verify the community structure presented by the  $O_3$  network, we use the network community partition algorithm to divide the community of the  $O_3$  network. Network community division is a commonly used network analysis technique, which has a large number of applications in social, biological, and other networks [41,42]. In this work, we employ the Louvain community partitioning algorithm, which is especially suitable for the current  $O_3$  network due to that it can quickly and effectively divide the community of directed weighted networks. Figure 1(b) shows several obvious community structures after segmentation according to the Louvain community partitioning algorithm with  $T_C = 3.8$ . These community structures corroborate our intuitive observations in Figure 1(a). For similar reasons as in Figure 1(a), community network structures rarely appear in northwest and southwest China. Nevertheless, the structure of the community partitioning algorithm can still confirm our intuitive conjecture that the community structure exists in the  $O_3$  network. This result implies that there are spatial and temporal characteristics of regional aggregation among the  $O_3$  networks, e.g., the Yangtze River Delta [43–46]. In general, the  $O_3$  detection data can be correlated according to network analysis techniques, and obvious community structure can be observed, suggesting the spatiotemporal characteristics of  $O_3$  distribution.



**Figure 1.** The  $O_3$  Network and Community Structure. (a) A network of  $O_3$  monitoring stations in China constructed according to the link weight calculation method. Yellow nodes represent  $O_3$  observation stations, and links between stations indicate links between stations that meet the connectivity threshold  $T_C = 3.8$ . (b) This figure is a network structure diagram after the community division of Figure 1(a) according to the Louvain community partitioning algorithm. The nodes and edges of different colors represent different network communities. For example, blue nodes represent the Northeast China community, green nodes represent the North China community, orange nodes represent the Sichuan-Chongqing community, and yellow nodes represent the Southeast coastal areas community.

represent the Southeast coastal community, and yellow nodes represent the Sichuan-Chongqing community.

The aim of this study is to observe the network connectivity between various monitoring points in China during different periods. Initially, we constructed the O<sub>3</sub> link network according to the above link weight calculation method. As can be seen from Figure 2(a) and Figure 3(a), the network community structure can be clearly observed, and these results are similar to Figure 1(a). To ensure comparability between different networks, we take the same number of edges in different networks. The edge thresholds of the corresponding networks for these two time periods are  $T_c=3.3$  and  $T_c=3.1$ , respectively. However, an interesting phenomenon is that although there are obvious community structures in the network at different periods, there are significant differences in the size of these community structures. These results suggest that different times have important effects on the construction of the O<sub>3</sub> link network. Future research should focus more on the explainable practical role of these network structures.

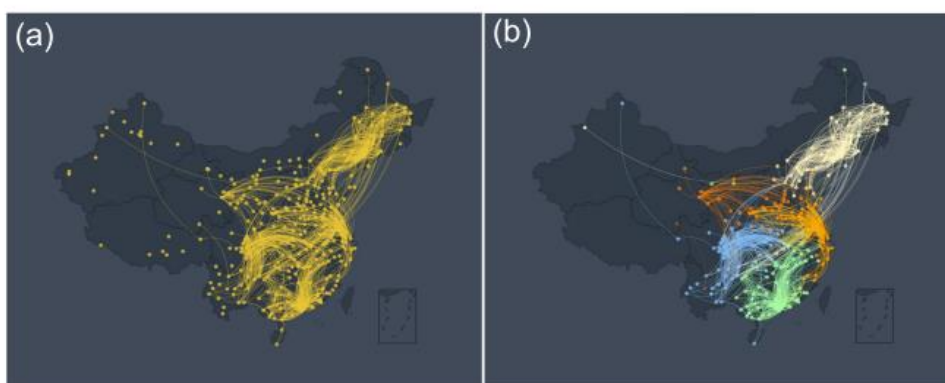
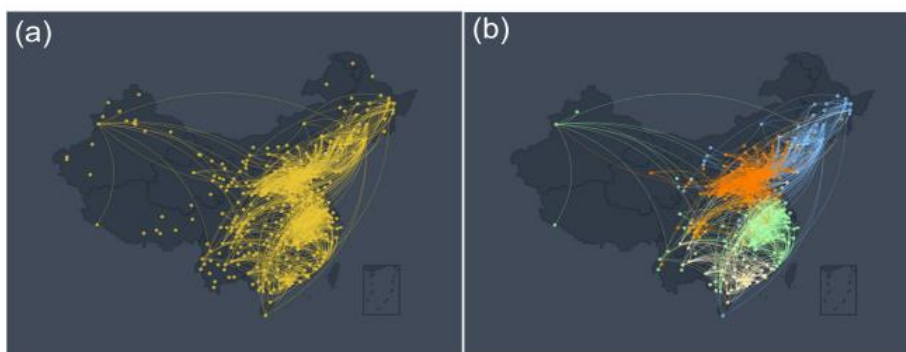


Figure 2. The O<sub>3</sub> Network and Community Structure from May to September. (a) A network of O<sub>3</sub> monitoring stations in China constructed according to the link weight calculation method. Yellow nodes represent O<sub>3</sub> observation stations, and links between stations indicate links between stations that meet the connectivity threshold  $T_c = 3.3$ . (b) This figure is a network structure diagram after the community division of Figure 2 (a) according to the Louvain community partitioning algorithm. The nodes and edges of different colors represent different network communities. For example, yellow nodes represent the Northeast China community, orange nodes represent the North China community, green nodes represent the Southeast coastal community, and blue nodes represent the Sichuan-Chongqing community.

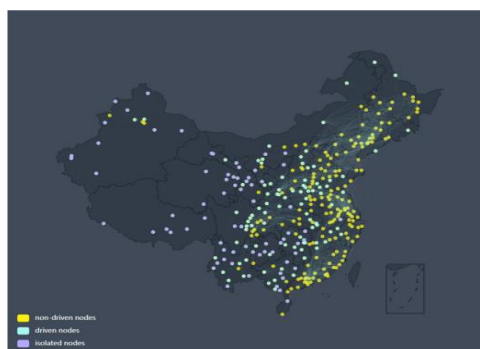
To further verify the community structure presented by the O<sub>3</sub> network in different periods, we employ the Louvain community partitioning algorithm to divide the community of the O<sub>3</sub> network. Obviously, Figure 2(b) and Figure 3(b) show several obvious community structures divided according to the Louvain community division algorithm into two different periods respectively. Their thresholds are  $T_c=3.3$  and  $T_c=3.1$ , respectively. These community structures corroborate our intuitive observations in Figure 2(a) and Figure 3(a). At the same time, these results indicate that the community structure formed has changed slightly over time. For example, there are certain differences in the community structure between North China and the Yangtze River Delta. In general, these results verify the existence of a distinct community structure in the O<sub>3</sub> network and also confirm the spatiotemporal characteristics of O<sub>3</sub> distribution.



**Figure 3.** The  $O_3$  Network and Community Structure from October to February. (a) A network of  $O_3$  monitoring stations in China constructed according to the link weight calculation method. Yellow nodes represent  $O_3$  observation stations, and links between stations indicate links between stations that meet the connectivity threshold  $TC = 3.1$ . (b) This figure is a network structure diagram after the community division of Figure 3 (a) according to the Louvain community partitioning algorithm. The nodes and edges of different colors represent different network communities. For example, blue represents the Northeast China community, green represents the Yangtze River Delta community, orange represents the North China community, and yellow represents the Southeast coastal community.

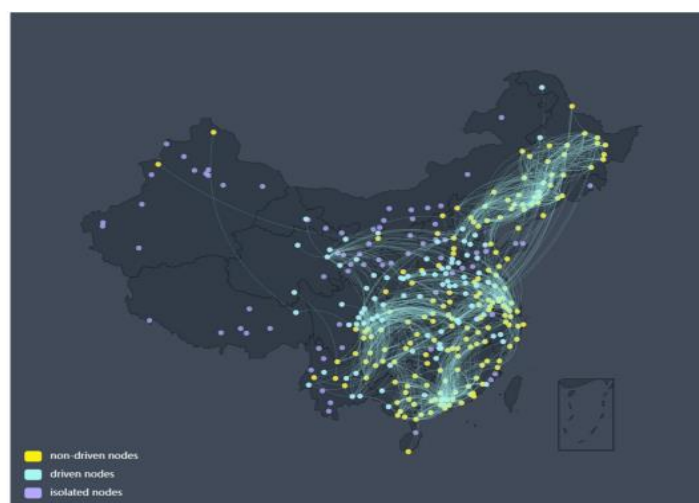
### 3.2. Control Algorithms and Driven Nodes

This study aims to explore the controllability of the  $O_3$  network, we select the maximum matching algorithm to distinguish the driven and non-driven nodes of the network [47]. This is due to the steps of the maximum matching algorithm and the results are easy to understand. Figure 4 shows the schematic diagram of the control network obtained by the maximum matching method. The edge threshold in this network is the same as the previous one and still takes  $T_c = 3.8$ . Here, the yellow nodes represent the non-driven nodes, and the cyan nodes indicate driven nodes. The outcomes depicted in Figure 4 incontrovertibly establish the presence of driven nodes within the network. For instance, driven nodes predominantly inhabit the central region, while non-driven nodes are primarily situated in coastal areas. The distribution of these driving nodes in space concurs with the region-dependent phenomena observed in conventional  $O_3$  studies. These results fill the gap in our understanding of the  $O_3$  network-driven nodes and have certain enlightening significance. This method of using control theory to understand the  $O_3$  network, especially its drivers, is novel and interesting. Future work should therefore focus on understanding the practical significance and use of driven nodes in  $O_3$  networks. This suggests that the network control theory and technology can be tried to be applied to studying  $O_3$  networks.



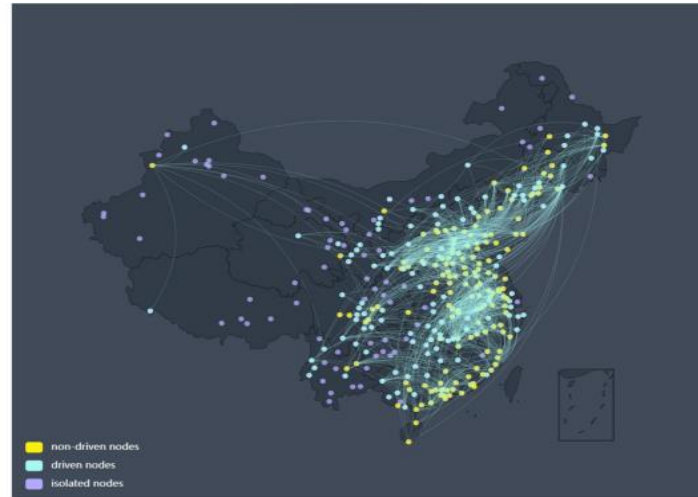
**Figure 4.**  $O_3$  Network from the Perspective of Network Control Theory. The figure shows the isolated, driven, and non-driven nodes of the  $O_3$  network with a threshold of  $TC = 3.8$ . Yellow represents non-driven nodes, cyan for driven nodes, and purple for isolated nodes.

To gain a deeper understanding of the relationship between the O<sub>3</sub> control network and different periods, we compare the O<sub>3</sub> control network in two different periods from May to September and from October to February. The O<sub>3</sub> control networks for two different periods are shown in Figure 5 and Figure 6, respectively. Similar to the previous results in Figure 2 and Figure 3, the edge thresholds of the corresponding networks in these two time periods are  $T_c = 3.3$  and  $T_c = 3.1$ , respectively. From these figures, it can be seen that there are significant differences in the O<sub>3</sub> control network during the two different periods. Although the methods and processing procedures used in these two networks are the same as before, their results are significantly different.



**Figure 5.** O<sub>3</sub> Network from May to September from the Perspective of Network Control Theory. The figure shows the isolated, driven, and non-driven nodes of the O<sub>3</sub> network with a threshold of  $T_c = 3.3$ . Yellow represents non-driven nodes, cyan for driven nodes, and purple for isolated nodes.

Figure 5 is a schematic diagram of the control network for O<sub>3</sub> data from May to September obtained by the maximum matching method. To ensure comparability between different networks, we take the same number of edges in different networks. The edge threshold in this network is  $T_c = 3.3$ . Similar to Figure 4, yellow nodes represent the non-driven nodes, and cyan nodes indicate driven nodes. Comparing Figure 4, Figure 5 and Figure 6, it can be seen that in different periods, the control network presents significant differences, and the nodes change from driven nodes to non-driven nodes in different periods, and vice versa. For example, some nodes in the Yangtze River Delta are non-driven nodes from May to September and change into driven nodes from October to February. These results are similar to the seasonal variation of climate characteristics in the Yangtze River Delta region in previous studies [45,46]. These results indicate that it is meaningful to use control theory to understand the O<sub>3</sub> network. Future research should focus more on understanding the practical implications of seasonal changes in O<sub>3</sub> network nodes over time.

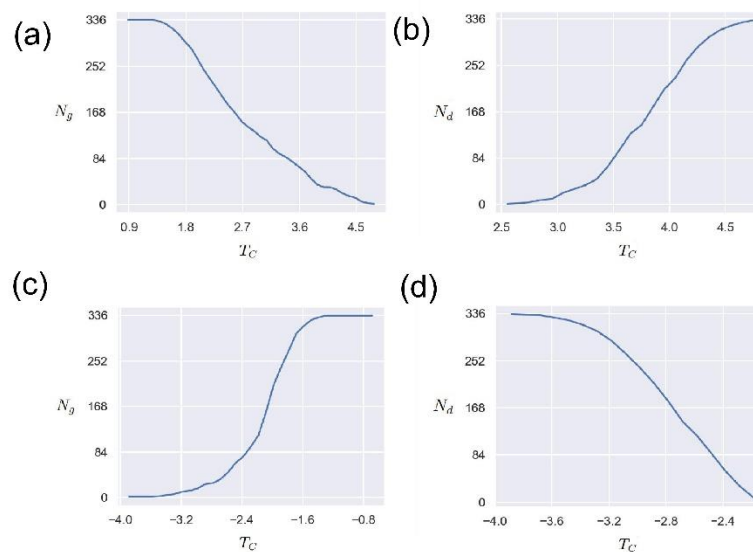


**Figure 6.** O<sub>3</sub> Network from October to February from the Perspective of Network Control Theory. The figure shows the isolated, driven, and non-driven nodes of the O<sub>3</sub> network with a threshold of TC = 3.1. Yellow represents non-driven nodes, cyan for driven nodes, and purple for isolated nodes.

### 3.3. Weights effect

In order to study the influence of the network edge weight threshold  $T_c$ , we study the changes in the number of nodes ( $N_g$ ) in the network's giant connected community (GCC) and the number of driven nodes ( $N_d$ ) when the threshold  $T_c$  changes in this section. The maximum connected community of the network is one of the important indicators reflecting the connectivity of the network, that is, the number of nodes contained in a connected maximum sub-network.

Figure 7 (a) shows the variation of the number of nodes  $N_g$  in the giant connected community with the threshold  $T_c$ . It can be seen that for the positive edge network, the threshold  $T_c$  has an obvious anti-correlation with the number of nodes in the giant connected community. This means that the larger the threshold, the smaller the number of giant connected community nodes. This is because the larger the threshold  $T_c$  is, the less the number of eligible edges is, and the corresponding network connectivity is reduced. Such results imply that the threshold  $T_c$  can be used as an indicator to measure and control the connectivity of the O<sub>3</sub> network.



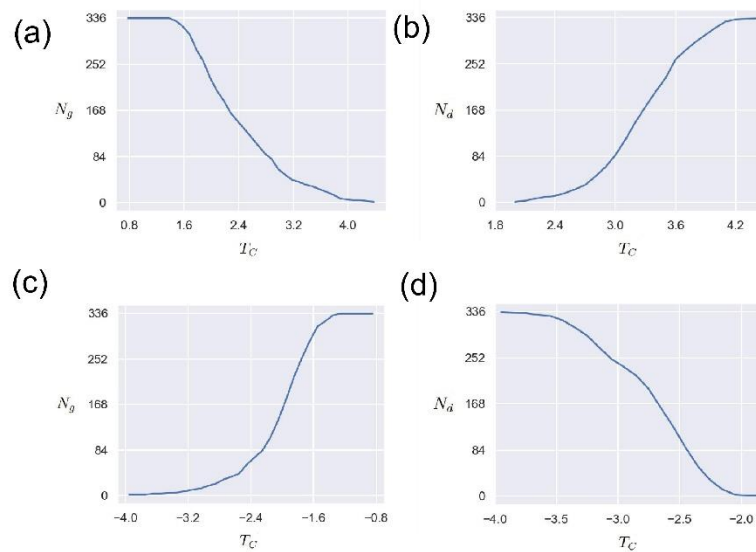
**Figure 7.** The effect of threshold TC versus the number of GCC  $N_g$  and the number of driven nodes  $N_d$ . (a) The relationship between the threshold TC and the giant connected community nodes  $N_g$ .

The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_g$ . The network edge weights are positive. (b) The relationship between the threshold  $T_C$  and the number of driven nodes  $N_d$ . The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_d$ . The network edge weights are positive. (c) The relationship between the threshold  $T_C$  and the giant connected community nodes  $N_g$ . The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_g$ . The network edge weights are negative. (d) The relationship between the threshold  $T_C$  and the number of driven nodes  $N_d$ . The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_d$ . The network edge weights are negative.

Figure 7 (b) presents the relationship between the number of driven nodes  $N_d$  and the threshold  $T_C$ . As can be seen from Figure 7 (b), as the threshold increases, the number of driven nodes also increases, demonstrating a positive correlation. This phenomenon is understandable. As the threshold increases, the connectivity of the  $O_3$  network decreases, and the number of isolated nodes increases, thereby driving the increase in the number of nodes. It should be noted that this positive correlation is not linear, but a nonlinear relationship. Future research should pay more attention to the significance of these nonlinear relationships in practical  $O_3$  application scenarios. In conclusion, these results provide an attempt to understand the  $O_3$  network through network control theory, and the network weight threshold  $T_C$  has an important influence. According to the cross-correlation function network construction method, the case where the edge weight takes a negative value is also a direction that needs to be explored. Therefore, here we explore the effect of the threshold  $T_C$  on the maximum number of connected community nodes  $N_g$  and the number of driven nodes  $N_d$  when the edge weight is negative. Figure 7(c) and Figure 7(d) display the changing trends of  $N_g$  and  $N_d$  when the edge connection weight threshold is negative, respectively. It is worth noting that these trends are the opposite of those in Figure 7(c) and Figure 7(d). That is, with the increase of the threshold  $T_C$ ,  $N_g$  increases with the increase, but  $N_d$  shows a downward trend. These results are reasonable. One reasonable explanation is that as the threshold  $T_C$  increases, more and more qualified edges are connected, so  $N_g$  increases accordingly. For  $N_d$ , the stronger the network connectivity is, the stronger its controllability is, and the fewer driven nodes to be controlled. It is difficult to find a meaningful threshold, or what are the criteria and operating steps for taking a suitable threshold. These issues are worthy of in-depth research and analysis. Overall, these curves indicate that the complex network method can be applied to the analysis of  $O_3$  networks, and how selecting the threshold is a key point.

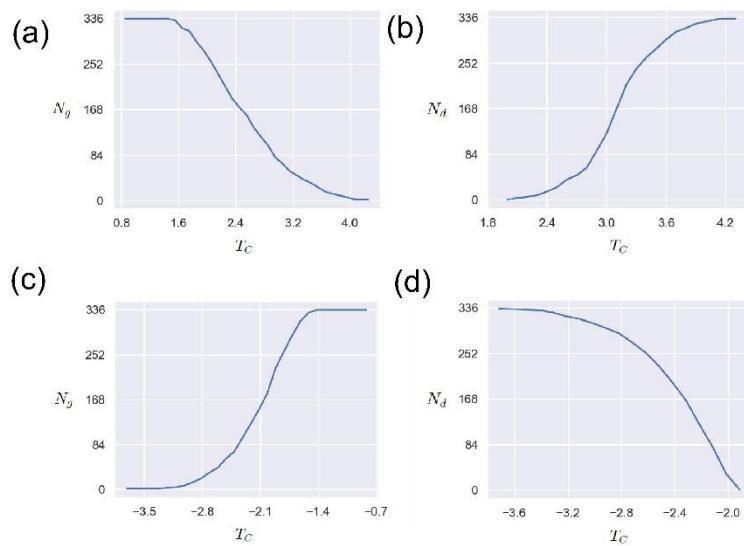
Figure 8(a) and Figure 9(a) show the change in the number of nodes  $N_g$  in the giant connected community of the  $O_3$  network for two different periods when the thresholds are  $T_C = 3.3$  and  $T_C = 3.1$ . It can be seen that for the positive edge network, the threshold  $T_C$  has an obvious inverse correlation with the number of nodes in the giant connected community. Such results are similar to those of Figure 7(a). This means that the larger the threshold, the smaller the number of connected mega-community nodes. This is because the larger the threshold  $T_C$  is, the less the number of eligible edges is, and the corresponding network connectivity is reduced.

Figures 8(b) and Figure 9(b) show the relationship between the number of driving nodes  $N_d$  and the threshold  $T_C$  in the  $O_3$  network in two different periods, respectively. A similar trend can be observed from these two figures, that is, as the threshold  $T_C$  increases, the number of driving nodes  $N_d$  increases too, showing a positive correlation. This is due to the fact that as the threshold increases, the connectivity of the  $O_3$  network decreases and the number of isolated nodes increases, thereby driving the increase in the number of nodes. It should be noted that this positive correlation is not linear, but a nonlinear relationship. The above results are similar to those of Figure 7(b), however, it can be observed that their thresholds  $T_C$  have different ranges. These results further verify that the network weight threshold  $T_C$  has a consistent and stable important influence on the  $O_3$  network.



**Figure 8.** The effect of threshold  $T_C$  versus the number of GCC  $N_g$  and the number of driven nodes  $N_d$  from May to September. (a) The relationship between the threshold  $T_C$  and the giant connected community nodes  $N_g$ . The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_g$ . The network edge weights are positive. (b) The relationship between the threshold  $T_C$  and the number of driven nodes  $N_d$ . The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_d$ . The network edge weights are positive. (c) The relationship between the threshold  $T_C$  and the giant connected community nodes  $N_g$ . The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_g$ . The network edge weights are negative. (d) The relationship between the threshold  $T_C$  and the number of driven nodes  $N_d$ . The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_d$ . The network edge weights are negative.

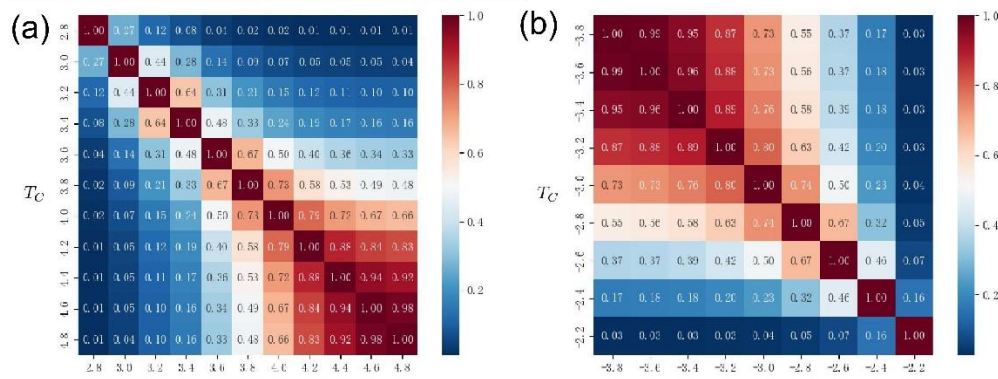
Similar to Figure 7(c) and Figure 7(d), Figure 8(c), Figure 9(c), Figure 8(d) and Figure 9(d) present the relationship between the threshold  $T_C$  on the maximum number of connected community nodes  $N_g$  and the number of driven nodes  $N_d$  when the edge weights are negative at two different periods. It can be seen from the above figures that the trend of the negative edge weights in the two different periods is basically similar. That is, as the threshold  $T_C$  increases,  $N_g$  increases with the increase, but  $N_d$  tends to decrease. These results demonstrate that although the range of the threshold  $T_C$  varies in different periods, the impact of the threshold value on the maximum number of connected community nodes  $N_g$  and the number of driven nodes  $N_d$  in the network is consistent. These results indicate that it is a feasible scheme to study the  $O_3$  networks using  $T_C$ ,  $N_g$ , and  $N_d$ . Future research should focus more on exploring the practical implications of these results.



**Figure 9.** The effect of threshold  $T_C$  versus the number of GCC  $N_g$  and the number of driven nodes  $N_d$  from October to February. (a) The relationship between the threshold  $T_C$  and the giant connected community nodes  $N_g$ . The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_g$ . The network edge weights are positive. (b) The relationship between the threshold  $T_C$  and the number of driven nodes  $N_d$ . The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_d$ . The network edge weights are positive. (c) The relationship between the threshold  $T_C$  and the giant connected community nodes  $N_g$ . The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_g$ . The network edge weights are negative. (d) The relationship between the threshold  $T_C$  and the number of driven nodes  $N_d$ . The horizontal axis is the value of threshold  $T_C$ , and the vertical axis is the value of  $N_d$ . The network edge weights are negative.

The coincidence degree of the driven node is a problem that needs attention, which can reflect the reliability of the driven node to a certain extent. Here, we propose to use the Jaccard coefficient to represent the coincidence of the driven nodes at two different thresholds [48]. That is the ratio of intersection over union  $\rho = \frac{N_d^{T_C^m} \cap N_d^{T_C^n}}{N_d^{T_C^m} \cup N_d^{T_C^n}}$ , where  $N_d^{T_C^m}$  and  $N_d^{T_C^n}$  represent the set of driven nodes when the thresholds are  $T_C^m$  and  $T_C^n$ , respectively.

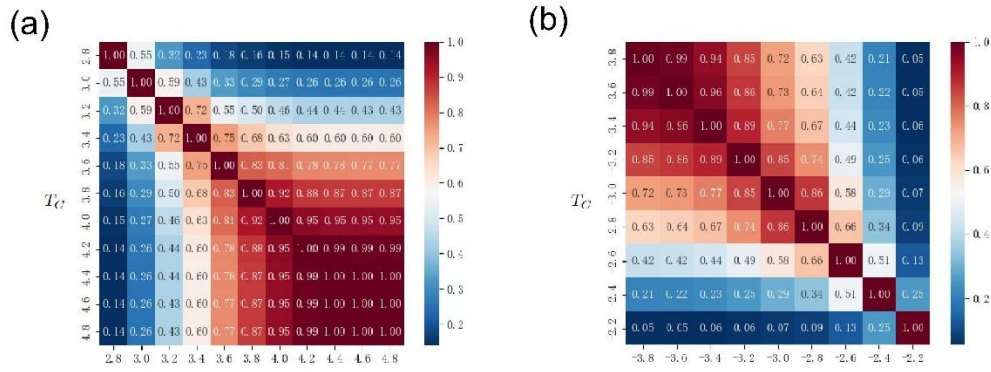
Figure 10(a) presents the Jaccard coefficients of the driven nodes for different thresholds  $T_C$ . In general, the closer the threshold, the higher the Jaccard coefficient. This result indicates that the closer the two thresholds are, the more identical nodes are obtained for the two sets of driven nodes. Meanwhile, an interesting phenomenon is that when both thresholds are large, the Jaccard coefficient is large, and vice versa. Such results on the one hand confirm that the size of the edge connection threshold  $T_C$  controls the number of driven nodes by controlling the network connectivity. On the other hand, it demonstrates that there are some deep mechanisms in the  $O_3$  network, that is, some nodes could be selected repeatedly instead of just randomly selected. Future research should pay more attention to the mechanism behind the above phenomenon. The above results indicate that the  $O_3$  network can be understood and analyzed using controllability methods, and is a research area worthy of further understanding.



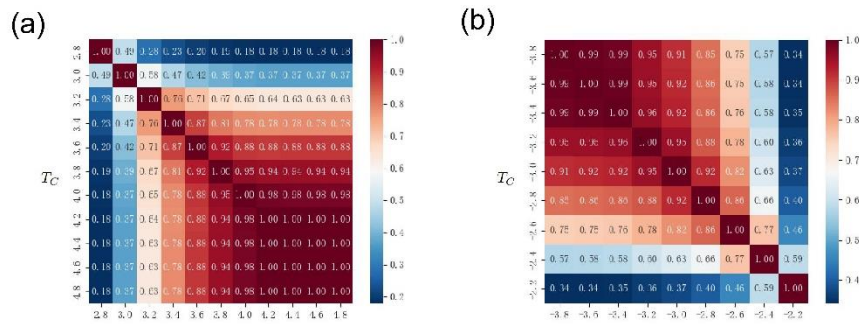
**Figure 10.** Heat map of co-occurrence rate of driven nodes under different thresholds  $T_c$ . (a) The horizontal and vertical axes represent different thresholds  $T_c$ . The higher the Jaccard coefficient of the co-occurrence of the driven nodes, the darker red the pattern color is, and vice versa, the darker blue. This figure displays the case where the edge weights take a positive value. (b) The horizontal and vertical axes represent different thresholds  $T_c$ . The higher the Jaccard coefficient of the co-occurrence of the driven nodes, the darker red the pattern color is, and vice versa, the darker blue. This figure displays the case where the edge weights take a negative value.

Figure 10(b) presents the Jaccard coefficients of the co-occurrence of drive nodes with different thresholds  $T_c$  when the edge weights take negative values. In general, the closer the threshold, the higher the Jaccard coefficient. This result is similar to that in Figure 10(a) when the edge weights are taken as positive values. However, an interesting phenomenon is that when both thresholds are small, the Jaccard coefficient of the co-occurrence of the driven node is large, and vice versa. This result is opposite to the result when Figure 10(a) is connected with a positive value. A possible explanation is that in the case of negative edge weights, the smaller the threshold  $T_c$ , the more isolated nodes in the network, so the higher the co-occurrence ratio of driven nodes. This result is basically consistent with the previous findings in Figures 7(c) and (d).

Figure 11(a) and Figure 12(a) show the Jaccard coefficients of the driving nodes with different thresholds  $T_c$  in two time periods, respectively. Similar to the results in Figure 10(a), the closer the threshold, the higher the Jaccard coefficient. That is, the closer the two thresholds are, the more identical nodes are obtained by the two groups of driven nodes. Meanwhile, an interesting phenomenon is that when both thresholds are large, the Jaccard coefficient is large, and vice versa. Moreover, when the threshold is large, the coincidence degree of the Jaccard coefficients of the two different periods is 1. On the one hand, this result is similar to the phenomenon in Figure 10(a), that is, the size of the edge connection threshold  $T_c$  is controlled by Network connectivity to control the number of driven nodes. On the other hand, if divided by different periods, a stronger repeated selection mechanism can be observed. That is to say, there is a stable strong link relationship between some nodes. Future research should focus on the underlying mechanisms leading to these phenomena.



**Figure 11.** Heat map of co-occurrence rate of driven nodes under different thresholds  $T_C$  from May to September. (a) The horizontal and vertical axes represent different thresholds  $T_C$ . The higher the Jaccard coefficient of the co-occurrence of the driven nodes, the darker red the pattern color is, and vice versa, the darker blue. This figure displays the case where the edge weights take a positive value. (b) The horizontal and vertical axes represent different thresholds  $T_C$ . The higher the Jaccard coefficient of the co-occurrence of the driven nodes, the darker red the pattern color is, and vice versa, the darker blue. This figure displays the case where the edge weights take a negative value.



**Figure 12.** Heat map of co-occurrence rate of driven nodes under different thresholds  $T_C$  from October to February. (a) The horizontal and vertical axes represent different thresholds  $T_C$ . The higher the Jaccard coefficient of the co-occurrence of the driven nodes, the darker red the pattern color is, and vice versa, the darker blue. This figure displays the case where the edge weights take a positive value. (b) The horizontal and vertical axes represent different thresholds  $T_C$ . The higher the Jaccard coefficient of the co-occurrence of the driven nodes, the darker red the pattern color is, and vice versa, the darker blue. This figure displays the case where the edge weights take a negative value.

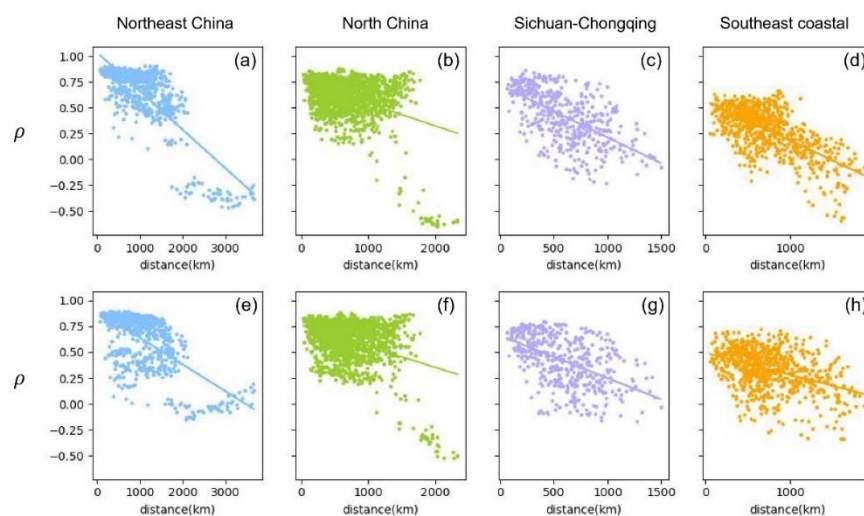
Figure 11(b) and Figure 12(b) display the Jaccard coefficients for the co-occurrence of driving nodes with different thresholds  $T_C$  when edge weights take negative values. These results are similar to those in Figure 10(b), the closer the two thresholds are, the higher the Jaccard coefficient. When both thresholds are small, the co-occurrence Jaccard coefficient of the driving node is large, and vice versa. This result is in contrast to the positive-valued edge results shown in Figure 11(a) and Figure 12(a). The possible reason for this phenomenon is that in the case of negative edge weights, the smaller the threshold  $T_C$ , the more isolated nodes in the network, and therefore the higher the co-occurrence rate of driving nodes.

### 3.4. Geographic distance and Spearman coefficient

In order to further distinguish driven nodes from non-driven nodes, we use prediction methods to measure the difference between the two types of nodes. An increasing number of studies use predictive methods to understand the correlation between different quantities [49], for example,

defining the influence of different variables by the predictability of one variable on another variable [50]. Figure 13 shows the degree of correlation between the predicted results and the true sequence for the driven and non-driven nodes. On the one hand, we employ the LSTM algorithm and use the  $O_3$  data sequence  $C_d$  of the driven node to predict the  $O_3$  data sequence  $\tilde{C}_{nd}$  of the non-driven node; then calculate the Spearman correlation  $\rho$  between the predicted non-driven  $O_3$  data sequence  $\tilde{C}_{nd}$  and the original non-driven sequence  $C_{nd}$ ; finally, Figures 13(a-d) displays the relationship of Spearman correlation  $\rho$  and distance from driven nodes to non-driven nodes in four different regions. Here we use the Spearman correlation coefficient calculation because the method has a better correlation calculation effect on nonlinear data. It can be seen from these figures that there is a significant negative correlation between the Spearman coefficient and the distance. These results are consistent with the trends of other pollutants changing with distance obtained in the past. It should be noted that in the Northeast and North China regions, there are different node clusters, which may be due to the relatively wide and uneven geographical distribution of cities in these regions.

On the other hand, we apply the same method and obtain the Spearman correlation of non-driven nodes predicting driven node sequences, and the relationship between  $\rho$  and distance is shown in Figures 13(e-h). A similar trend was observed in all cases, with no sudden changes. This trend further validates the role of distance in  $O_3$  variation between different sites. Comparing Figures 13(a-d) with Figures 13(e-h), it can be seen that the negative correlation of driven nodes predicting non-driven nodes is stronger than that of non-driven nodes predicting driven nodes. These results suggest that driven nodes have more influence in the  $O_3$  network than non-driven nodes.



**Figure 13.** Scatter diagram of the relationship between the distance of city  $C_i$  and city  $C_j$  and the Spearman coefficient of their  $O_3$  data in the Northeast China, North China, Sichuan-Chongqing and Southeast coastal areas. Figures (a-d) are the data of the direction of the driven to non-driven nodes, the vertical axis represents the Spearman coefficient between the  $O_3$  data of the city  $C_i$  ( $C_i \in$  driven nodes) and the city  $C_j$  ( $C_j \in$  non-driven nodes), the horizontal axis represents the geographical distance between the city  $C_i$  and the city  $C_j$ . Figures (e-h) are the data of the direction of the non-driven to driven nodes, and the vertical axis represents the Spearman coefficient between the  $O_3$  data of city  $C_i$  ( $C_i \in$  non-driven nodes) and city  $C_j$  ( $C_j \in$  driven nodes). The fitted slopes of scattered data are as follows,  $-3.74 \times 10^{-4}$ ,  $-2.17 \times 10^{-4}$ ,  $-5.04 \times 10^{-4}$ ,  $-3.82 \times 10^{-4}$ ,  $-2.60 \times 10^{-4}$ ,  $-2.06 \times 10^{-4}$ ,  $-4.12 \times 10^{-4}$ ,  $-2.11 \times 10^{-4}$ . Overall, we observe the distance between nodes is negatively correlated with their Spearman coefficient.

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

#### 4. Conclusions

The rapid development of human society has made life more and more convenient, but it has also caused more and more environmental problems. In particular, with the destruction of the O<sub>3</sub> layer by modern human pollution, the study of O<sub>3</sub> has attracted more and more attention. Traditional research on O<sub>3</sub> generally focuses on the influence of different factors, etc. These results have greatly enriched our understanding and cognition of O<sub>3</sub>. However, the problem of O<sub>3</sub> is a multifaceted and complex system problem. Therefore, using the theory and method of the complex network to analyze and study such a complex system as O<sub>3</sub> has great theoretical and practical application value. In this research, we integrated the complex network theories and technologies from the network construction method cross-correlation function, the Louvain community partitioning algorithm and the maximum matching network control theory, and systematically studied the characteristics of O<sub>3</sub> in China.

Our findings illustrate that the O<sub>3</sub> network prominently exhibits a structured community framework, delineating regions such as Northeast China, North China, Sichuan-Chongqing, and the Southeast coastal areas. Notably, driving nodes are predominantly concentrated within central regions, in contrast to non-driven nodes, which are primarily situated along coastal perimeters. Furthermore, the threshold denoted as  $T_c$  exhibits a conspicuous inverse relationship with the number of nodes constituting the giant connected community. Moreover, as the threshold value increases, there is a corresponding augmentation in the count of driving nodes, indicating a positive correlation. It is noteworthy that these trends manifest in an opposing manner within the Negative O<sub>3</sub> network.

Additionally, an intriguing observation emerges wherein significant interplay occurs between the Jaccard coefficient and both thresholds, whereby a larger value is observed when thresholds are substantial, and vice versa. In the context of the negative O<sub>3</sub> network, a parallel trend emerges, where a diminutive Jaccard coefficient aligns with diminutive thresholds, and vice versa. Furthermore, a discernible pattern of substantial negative correlation surfaces between the Spearman coefficient and the distance, as evident from the depicted figures.

Our study presents a valuable endeavor to probe the O<sub>3</sub> network through the lens of complex network theory and methodologies, yielding substantive and insightful outcomes. These encouraging results warrant broader validation across larger geographical expanses. Subsequent research should underscore considerations concerning threshold selection and its subsequent interpretation, among other pertinent aspects.

**Author Contributions:** Z.Z. and D.X. conducted the experiments; H.S. and W.W. guided the experiments; Z.Z. processed the data results and wrote the manuscript; D.X. and H.S. modified the grammar and format of the whole article; W.W. and N.Y. guided the writing of manuscripts. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the National Natural Science Foundation of China (Grant No. 62176148; No. 42205062); the Basic and Applied Basic Research of Colleges and Universities in Guangdong Province (Special Projects in Digital Economic: 2021ZDZX3025); the Scientific Research Foundation of Shantou University (Grant No. NTF19015); the 2020 Li Ka Shing Foundation Cross-Disciplinary Research (Grant No. 2020LKSFG09D).

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** The authors would like to thank the science teams of CNEMC for providing data used in this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. J. M. Hao, K. B. He, L. Duan, J. H. Li, L. T. Wang. Air pollution and its control in China. *Front. Environ. Sci. Eng. China* **2007**, 1, 129-142.
2. S. X. Wang, J. M. Hao. Air quality management in China: Issues, challenges, and options. *J Environ Sci (China)* **2012**, 24, 2-13.
3. S. X. Wang, B. Zhao, S. Y. Cai, Z. Klimont, C. P. Nielsen, T. Morikawa, J. H. Woo, Y. Kim, X. Fu, J. Y. Xu, J. M. Hao, K. B. He. Emission trends and mitigation options for air pollutants in East Asia. *Atmospheric Chem. Phys.* **2014**, 14, 6571-6603.
4. The State Council of the People's Republic of China. Air Pollution Prevention and Control Action Plan. [http://www.gov.cn/zwggk/2013-09/12/content\\_2486773.html](http://www.gov.cn/zwggk/2013-09/12/content_2486773.html) (2013) (access:10 Sep 2013).
5. Three-year Action Plan to Win the Blue Sky Defense War. [http://www.gov.cn/zhengce/content/2018/7/03/content\\_5303158.html](http://www.gov.cn/zhengce/content/2018/7/03/content_5303158.html).
6. Wang, Y., Gao, W., Wang, S., Song, T., Gong, Z., Ji, D., Wang, L., Liu, Z., Tang, G., Huo, Y., Tian, S., Li, J., Li, M., Yang, Y., Chu, B., Petaja, T., Kerminen, V., He, H., Hao, J., Kulmala, M., Wang, Y., Zhang, Y. Contrasting trends of PM<sub>2.5</sub> and surface-O<sub>3</sub> concentrations in China from 2013 to 2017. *Natl. Sci. Rev.* **2020**, 7, 1331-1339.
7. Y. Guo, K. Li, B. Zhao, J. Shen, W. J. Bloss, M. Azzi, Y. Zhang. Evaluating the real changes of air quality due to clean air actions using a machine learning technique: results from 12 Chinese megacities during 2013-2020. *Chemosphere* **2020**, 300, Article 134608.
8. J. Wei, Z. Li, K. Li, R. R. Dickerson, R. T. Pinker, J. Wang, X. Liu, L. Sun, W. Xue, M. Cribb. Full-coverage mapping and spatiotemporal variations of ground-level O<sub>3</sub> (O<sub>3</sub>) pollution from 2013 to 2020 across China. *Remote Sens. Environ.* **2022**, 270, Article 112775.
9. X. Lu, X. P. Ye, M. Zhou, Y. Zhao, H. Weng, H. Kong, K. Li, M. Gao, B. Zheng, J. Lin, F. Zhou, Q. Zhang, D. Wu, L. Zhang & Y. Zhang. The underappreciated role of agricultural soil nitrogen oxide emissions in O<sub>3</sub> pollution regulation in North China. *Nat. Commun.* **2021**, 12, 5021-5029.
10. T. Wang, L. K. Xue, P. Brimblecombe, Y. F. Lam, L. Li, & L. Zhang. O<sub>3</sub> pollution in China: a review of concentrations, meteorological influences, chemical precursors, and effects. *Sci. Total Environ.* **2017**, 575, 1582-1596.
11. M. Shao, Y. Zhang, L. Zeng, X. Tang, J. Zhang, L. Zhong, B. Wang. Ground-level O<sub>3</sub> in the Pearl River Delta and the roles of VOC and NO<sub>x</sub> in its production. *J. Environ. Manag.* **2009**, 90, 512-518.
12. B. C. McDonald, J. A. de Gouw, J. B. Gilman, S. H. Jathar, A. Akherati, C. D. Cappa, J. L. Jimenez, J. Lee-Taylor, P. L. Hayes, S. A. McKeen, Y. Y. Cui, S. W. Kim, D. R. Gentner, G. Isaacman-VanWertz, A. H. Goldstein, R. A. Harley, G. J. Frost, J. M. Roberts, T. B. Ryerson, M. Trainer. Volatile chemical products emerging as largest petrochemical source of urban organic emissions. *Science* **2018**, 359, 760-764.
13. L. He, Y. S. Duan, Y. Zhang, Q. Yu, J. T. Huo, J. Chen, H. X. Cui, Y. W. Li, W. C. Ma. Effects of VOC emissions from chemical industrial parks on regional O<sub>3</sub>-PM<sub>2.5</sub> compound pollution in the Yangtze River Delta. *Sci. Total Environ* **2023**, 906, 167503.
14. Y. A. Guan, X. J. Liu, Z. Y. Zheng, Y. W. Dai, G. M. Du, J. Han, L. A. Hou, E. R. Duan. Summer O<sub>3</sub> pollution cycle characteristics and VOCs sources in a central city of Beijing-Tianjin-Hebei area, China. *Environ. Pollut.* **2023**, 323, Article 121293.
15. R. P. Wang, W. J. Duan, S. Y. Cheng, X. Q. Wang. Nonlinear and lagged effects of VOCs on SOA and O<sub>3</sub> and multi-model validated control strategy for VOC sources. *Sci. Total Environ.* **2023**, 887, Article 164113.

16. C. Q. Liu, H. Geng, P. Shen, Q. G. Wang, K. Shi. Coupling detrended fluctuation analysis of the relationship between O<sub>3</sub> and its precursors—a case study in Taiwan. *Atmos. Environ.* **2018**, 188, 18–24.
17. Y. Yang, M. Li, H. Wang, H. Li, P. Wang, K. Li, M. Gao, and H. Liao. Enso modulation of summertime tropospheric O<sub>3</sub> over china. *Environ. Res. Lett.* **2022**, 17, 034020.
18. M. Li, S. Yu, X. Chen, Z. Li, Y. Zhang, L. Wang, W. Liu, P. Li, E. Lichtfouse, D. Rosenfeld, and J. H. Seinfeld. Large scale control of surface. O<sub>3</sub> by relative humidity observed during warm seasons in china. *Atmos. Environ.* **2021**, 19, 3981-3989.
19. S. C. Kavassalis and J. G. Murphy. Understanding O<sub>3</sub>-meteorology correlations: A role for dry deposition. *Geophys. Res. Lett.* **2017**, 44, 2922–2931.
20. H. Qi, W. Duan, S. Cheng, Z. Huang, X. Hou. Spatial clustering and spillover pathways analysis of O<sub>3</sub>, NO<sub>2</sub>, and CO in eastern China during 2017-2021. *Sci. Total Environ.* **2023**, 904, 166814.
21. M. Y. Qi, L. T. Wang, S. M. Ma, L. Zhao, X. H. Lu, Y. Y. Liu, Y. Zhang, J. Y. Tan, Z. T. Liu, S. T. Zhao, Q. Wang, R. G. Xu. Evaluation of PM<sub>2.5</sub> fluxes in the "2+26" cities: transport pathways and intercity contributions. *Atmos. Pollut. Res.* **2021**, 12(5), Article 101048.
22. W. Y. Yang, H. Y. Du, Z. F. Wang, L. L. Zhu, Z. Wang, X. S. Chen, H. S. Chen, W. Wang, R. Q. Zhang, J. Li, X. Dao. Characteristics of regional transport during two-year wintertime haze episodes in North China megacities. *Atmos. Res.* **2021**, 257, Article 105582.
23. L. Shen, J. Liu, T. Zhao, X. Xu, H. Han, H. Wang, Z. Shu. Atmospheric transport drives regional interactions of O<sub>3</sub> pollution in China. *Sci. Total Environ.* **2022**, 830, 154634.
24. T. T. Fang, Y. Zhu, S. X. Wang, J. Xing, B. Zhao, S. J. Fan, M. H. Li, W. W. Yang, Y. Chen, R. L. Huang. Source impact and contribution analysis of ambient O<sub>3</sub> using multi-modeling approaches over the Pearl River Delta region, China. *Environ. Pollut.* **2021**, 289, Article 117860.
25. U. Kalsoom, T. Wang, C. Ma, L. Shu, C. Huang, L. Gao. Quadrennial variability and trends of surface O<sub>3</sub> across China during 2015–2018: a regional approach. *Atmos. Environ.* **2021**, 245, Article 117989.
26. Mao, J., Yan, F., Zheng, L., You, Y., Wang, W., Jia, S., Liao, W., Wang, X., Chen, W. O<sub>3</sub> control strategies for local formation and regional transport dominant scenarios in a manufacturing city in southern China. *Sci. Total Environ.* **2022**, 813, 151883.
27. Y. Zhang, D. Chen, J. Fan, S. Havlin, and X. Chen. Correlation and scaling behaviors of fine particulate matter (PM<sub>2.5</sub>) concentration in china. *EPL* **2018**, 122, 58003.
28. G. Tian, M. H. Gunes. Complex network analysis of O<sub>3</sub> transport in Complex Networks V (Springer) **2014**, 87–96.
29. Q. Wang, X. Wang, R. Huang, J. Wu, Y. Xiao, M. Hu, Q. Fu, Y. Duan, and J. Chen. Regional transport of PM<sub>2.5</sub> and O<sub>3</sub> based on complex network method and chemical transport model in the yangtze river delta, China. *J. Geophys. Res-Atmos.* **2022**, 127, e2021JD034807.
30. J. Fan, J. Meng, Y. Ashkenazy, S. Havlin, and H. J. Schellnhuber. Network analysis reveals strongly localized impacts of el niño. *Proc. Nat. Acad. Sci. U.S.A.* **2017**, 114, 7543-7548.
31. N. Ying, D. Zhou, Q. Chen, Q. Ye, and Z. Han. Long-term link detection in the CO<sub>2</sub> concentration climate network. *J. Clean. Prod.* **2019**, 208, 1403–1408.
32. Y. Zhang, J. Fan, X. Chen, Y. Ashkenazy, and S. Havlin. Significant impact of rossby waves on air pollution detected by network analysis. *Geophys. Res. Lett.* **2019**, 46, 12476–12485.
33. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, Article P10008.
34. Y. Y. Liu, J. J. Slotine, and A. L. Barabási. Controllability of complex networks. *Nature* **2011**, 473, 167–173.

35. Z. D. Zhao, N. Zhao, and N. Ying. Association, correlation, and causation among transport variables of PM<sub>2.5</sub>. *Front. Phys.* **2021**, 9.
36. N. Ying, D. Zhou, Z. Han, Q. Chen, Q. Ye, and Z. Xue. Rossby waves detection in the CO<sub>2</sub> and temperature multilayer climate network. *Geophys. Res. Lett.* **2020**, 47, e2019GL086507.
37. N. Ying, D. Zhou, Z. Han, Q. Chen, Q. Ye, Z. Xue, and W. Wang. Climate networks suggest rossby-waves-related CO<sub>2</sub> concentrations to surface air temperature. *EPL* **2020**, 132, 19001.
38. W. Wang, S. Yang, K. Yin, Z. Zhao, N. Ying, and J. Fan. Network approach reveals the spatiotemporal influence of traffic on air pollution under covid-19. *Chaos* **2022**, 32, 041106.
39. N. Ying, W. Duan, Z. Zhao, and J. Fan. Complex networks analysis of PM<sub>2.5</sub>: transport and clustering. *Earth Syst. Dynam. Discuss.* **2022**, 1-18.
40. J. Liu, L. Wang, M. Li, Z. Liao, Y. Sun, T. Song, W. Gao, Y. Wang, Y. Li, D. Ji, B. Hu, V. M. Kerminen, Y. Wang, and M. Kulmala. Quantifying the impact of synoptic circulation patterns on O<sub>3</sub> variability in northern china from april to october 2013–2017. *Atmos. Chem. Phys.* **2019**, 19, 14477–14492.
41. M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 7821-7826.
42. M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.* **2003**, 45, 167-256.
43. A. J. Ding, C. B. Fu, X. Q. Yang, J. N. Sun, L. F. Zheng, Y. N. Xie, E. Herrmann, W. Nie, T. Petäjä, V. M. Kerminen, and M. Kulmala. O<sub>3</sub> and fine particle in the western yangtze river delta: an overview of 1 yr data at the sorpes station. *Atmos. Chem. Phys.* **2013**, 13, 581-5830.
44. J. Gao, B. Zhu, H. Xiao, H. Kang, X. Hou, and P. Shao. A case study of surface O<sub>3</sub> source apportionment during a high concentration episode, under frequent shifting wind conditions over the yangtze river delta, China. *Sci. Total Environ.* **2016**, 544, 853-863.
45. L. Shu, T. Wang, M. Xie, M. Li, M. Zhao, M. Zhang, and X. Zhao. Episode study of fine particle and O<sub>3</sub> during the capum-yrd over yangtze river delta of china: Characteristics and source attribution. *Atmos. Environ.* **2019**, 203, 87-101.
46. L. Shu, T. Wang, H. Han, M. Xie, P. Chen, M. Li, and H. Wu. Summertime O<sub>3</sub> pollution in the yangtze river delta of eastern china during 2013-2017: Synoptic impacts and source apportionment. *Environ. Pollut.* **2020**, 257, 113631 (2020).
47. Y.-Y. Liu, J.-J. E. Slotine, and A. L. Barabasi. Controllability of complex networks. *Nature* **2011**, 473, 167-173.
48. P. Jaccard. Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines. *Bull Soc Vaudoise Sci Nat* **1901**, 37, 241-272.
49. J. Ludescher, M. Martin, N. Boers, A. Bunde, C. Ciemer, J. Fan, S. Havlin, M. Kretschmer, J. Kurths, J. Runge, V. Stolbova, E. Surovyatkina, and H. J. Schellnhuber. Network-based forecasting of climate phenomena. *Proc. Nat. Acad. Sci. U.S.A.* **2021**, 118, e1922872118.
50. G. Sugihara, R. May, H. Ye, C.-H. Hsieh, E. Deyle, M. Fogarty, and S. Munch. Detecting causality in complex ecosystems. *Science* **2012**, 338, 496-500.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.