https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions

Identifying the control cities of O_3 Pollution using Complex networks

Zhi-dan Zhao [1,2], Demei Xue [2,3], Haojun Sun [2,3], Weiping Wang [4], Na Ying[5]*

[1]China Complexity Computation Lab, Department of Computer Science, School of Engineering, Shantou University, Shantou 515063, China
[2]Key Laboratory of Intelligent Manufacturing Technology (Shantou University), Ministry of Education, Shantou 515063, China
[3]Complexity Computation Laboratory, Department of Computer Science, School of Engineering, Shantou University, Shantou 515063, China
[4]School of National Safety and Emergency Management, Beijing Normal University, Zhuhai 519087, China
[5]China State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing, 100012, China.

** Corresponding author: yingna@craes.org.cn*

Abstract. In recent years, ozone (O_3) pollution has been rapidly spreading, restricting further improvement of air quality in China. Investigating the interaction of O_3 concentration and identifying their driven cities are important for the prevention and control of O_3 pollution in China. However, the interaction between O_3 pollution between cities and their driven cities has not yet been revealed. In this study, we fill this gap based on the integration of complex network methods, the Louvain community partitioning algorithm and the maximum matching network control theory. O_3 network model exhibits a structured cluster framework, such as Northeast, North China, Sichuan and Chongqing, and Southeast coastal areas. And the driver nodes are mainly concentrated in the central region, while the non-driver nodes are mainly located in the coastal periphery. We also found that the proportion of driven nodes exhibits a positive relation with the threshold. In addition, the coincidence degree of the driven node is related to the choose of threshold. A closer threshold value corresponds to a higher coincidence ratio of the driven nodes. The correlation of driven nodes predicting non-driven nodes is stronger than non-driven nodes predicting driven nodes, suggesting that driven nodes have more influence in the O_3 network than non-driven nodes. The results provide scientific guidance for national O_3 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.

1 Introduction

With the rapid development of human society, environmental issues are receiving increasing attention (Hao et al., 2007; Wang and Hao, 2012). 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 (Wang et al., 2014). A large number of actions have been

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions Open Access EGU

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 (Xue et al., 2021; Geng et al., 2019). It's glad to see the $PM_{2.5}$ concentrations in China have significantly decreased, however, the concentration of ozone (O_3) is showing an upward and spreading trend, becoming a limiting factor for further improvement of our air quality (Wang et al., 2020; Guo et al., 2022). controlling O_3 pollution become urgent issues that need to be addressed in each city.

O_3 pollution is a secondary pollutant generated by photochemical reactions that have adverse effects on human health and ecosystems(Wei et al., 2022). Nitrogen oxides and VOCs are one of the main precursor pollutants for the formation of O_3(Lu et al., 2021; Wang et al., 2017; Shao et al., 2009; Mcdonald et al., 2018). The relationship between O_3, NO_x, and VOCs is not a simple linear dependence, but rather a non-linear variation(He et al., 2024). The photochemical reaction of O_3 is very complex and is influenced by both meteorological conditions and the control of precursor pollutants(Guan et al., 2023; Wang et al., 2023). Seasonal characteristics are observed in ozone pollution, as it is notably affected by the summer monsoon(Liu et al., 2018). Moreover, in China, there has been a positive correlation between surface O_3 concentration and the ENSO index during summers from 1990 to 2019(Yang et al., 2022). 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(Li et al., 2021). This suggests that net ozone production decreases as relative humidity increases(Kavassalis and Murphy, 2017).

In addition to industrial emissions and meteorological conditions, inter-regional atmospheric transport is an important factor in the formation of O_3 pollution (Qi et al., 2023; Qi et al., 2021; Yang et al., 2021; Shen et al., 2022). 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_3 (Fang et al., 2021; Kalsoom et al., 2021; Mao et al., 2022). 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_3 and managing joint prevention and control. In addition, research on identifying O_3 driven cities is even rarer.

In recent years, complex network methods have gradually been introduced into the field of atmospheric environment (Zhang et al., 2018). The ozone transport network in California has been constructed to examine the phenomenon of networked communities (Tian and Gunes, 2014). By applying optimized complex network approach and source resolution model, the transport patterns of O_3 in the Yangtze River Delta region has been explored (Wang et al., 2022a). 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_3 network is based on the cross-correlation function(Fan et al., 2017; Ying et al., 2019; Zhang et al., 2019). Subsequently, we employ the Louvain community partitioning algorithm to delineate communities within the ozone

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions Open Access EGU

network(Blondel et al., 2008). 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_3 pollution.

2 Materials and Methods

2.1 Data

The maximum daily average 8-hour (MDA8) O_3 concentration used in this study are acquired from the China National Environmental Monitoring Centre (CNEMC). The 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_3 network, we introduce the obtained O_3 data in this section, as well as the cross-correlation O_3 network construction method, the O_3 network community partition algorithm Louvain Community Classification and the driven node discovery method maximum matching algorithm(Blondel et al., 2008; Liu et al., 2011).

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_3 observation sites and use it as the link weight for constructing the O_3 network(Zhao et al., 2021; Ying et al., 2020a; Ying et al., 2020b; Wang et al., 2022b; Ying et al., 2022). The specific calculation process of the cross-correlation function algorithm is shown in the equation (1) and (2):

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions

$$X_{C_i,C_j}(\tau) = \frac{\sum_{t=1}^{L-\tau}\left(C_i(t)-\overline{C_i}\right)\left(C_j(t+\tau)-\overline{C_j}\right)}{\sqrt{\sum_{t=1}^{L-\tau}\left(C_i(t)-\overline{C_i}\right)^2}\cdot\sqrt{\sum_{t=1}^{L-\tau}\left(C_j(t+\tau)-\overline{C_j}\right)^2}} \quad (1)$$

$$X_{C_i,C_j}(-\tau) = \frac{\sum_{t=1}^{L-\tau}\left(C_i(t+\tau)-\overline{C_i}\right)\left(C_j(t)-\overline{C_j}\right)}{\sqrt{\sum_{t=1}^{L-\tau}\left(C_i(t+\tau)-\overline{C_i}\right)^2}\cdot\sqrt{\sum_{t=1}^{L-\tau}\left(C_j(t)-\overline{C_j}\right)^2}} \quad (2)$$

where τ 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, $\overline{C_i}$ and $\overline{C_j}$ 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.

$$W^{pos}_{C_i,C_j} = \frac{max\left(X_{C_i,C_j}\right) - mean\left(X_{C_i,C_j}\right)}{std\left(X_{C_i,C_j}\right)} \quad (3)$$

$$W^{neg}_{C_i,C_j} = \frac{min\left(X_{C_i,C_j}\right) - mean\left(X_{C_i,C_j}\right)}{std\left(X_{C_i,C_j}\right)} \quad (4)$$

$W^{pos}_{C_i,C_j}$ and $W^{neg}_{C_i,C_j}$ 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 $\tau^{pos}_{C_i,C_j} > 0$ and $W^{pos}_{C_i,C_j} \geq T_C$, there exists a link between city i and city j, with direction from city i to city j. The link weight is $W^{pos}_{C_i,C_j}$, and all eligible links are connected to form a positive O_3 network (PoN). When $\tau^{neg}_{C_i,C_j} < 0$ and $W^{neg}_{C_i,C_j} \leq -T_C$, there exist a link between city i and city j, and direction from city i to city j with the weight $W^{neg}_{C_i,C_j}$, 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

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions Open Access EGU

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 $0 \ldots k$ neighbor nodes, calculate the modular gain after putting node a into the community of any neighboring node ΔQ, 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(Liu et al., 2011). Here, we use the maximum matching control algorithm to study the O_3 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(Liu et al., 2011). Here we briefly describe the conversion steps. At the beginning, let $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, $\boldsymbol{V} = \{\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_n}\}$, $\boldsymbol{E} = \{(\boldsymbol{c_1}, \boldsymbol{c_2}), (\boldsymbol{c_1}, \boldsymbol{c_2}), (\boldsymbol{c_1}, \boldsymbol{c_4}), (\boldsymbol{c_5}, \boldsymbol{c_3}) \ldots\}$, $(\boldsymbol{c_1}, \boldsymbol{c_2})$ indicates there exist a link from $\boldsymbol{c_1}$ to $\boldsymbol{c_2}$. Constructing two disjoint subsets, denoted by $+$ as the source node, and denoted by $-$ as the target node, $\boldsymbol{V^+} = \{\boldsymbol{c_1^+}, \boldsymbol{c_2^+}, \ldots, \boldsymbol{c_n^+}\}, \boldsymbol{V^-} = \{\boldsymbol{c_1^-}, \boldsymbol{c_2^-}, \ldots, \boldsymbol{c_n^-}\}$. In a bipartite graph, any two edges in the $\boldsymbol{V^+}$ and $\boldsymbol{V^-}$ 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(Liu et al., 2019). The LSTM model learns the data that can be observed at the $\boldsymbol{t}$ moment and predicts the data at

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions Open Access EGU

the $\boldsymbol{t+1}$ moment. We used LSTM to perform univariate prediction of O_3 data, selected O_3 data for a total of 3 years from 2015 to 2017, and entered the O_3 data of the previous two and a half years to predict the O_3 data of the last half year. The steps are as follows,

Step 1: Pretreatment data, including loading O_3 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 MSELoss 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_3 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_3 network

In order to visually display the network connection between various monitoring sites in China. First, we constructed the O_3 link network according to the above link weight calculation method(Fan et al., 2017; Ying et al., 2019; Zhang et al., 2019; Zhao et al., 2021; Ying et al., 2020a; Ying et al., 2020b; Wang et al., 2022a; Ying et al., 2022). According to the method in Section II B we get the connection weights of 336 cities in China and build an O_3 network, which consists of 336×335=112560 edges. To better describe the network structure, we extract the edge whose weight $W^{pos}_{C_i,C_j} \geq T_C$ with T_C =3.8 and obtain 1942 edges, which satisfies the sparse network requirement $e < nlogn$(Girvan *et al* 2002). 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(Tian and Gunes, 2014; Liu et al., 2019). 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

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions
Open Access
EGU

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(Girvan and Newman, 2002; Newman, 2003). 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(Ding et al., 2013; Gao et al., 2016; Shu et al., 2019; Shu et al., 2020). 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.

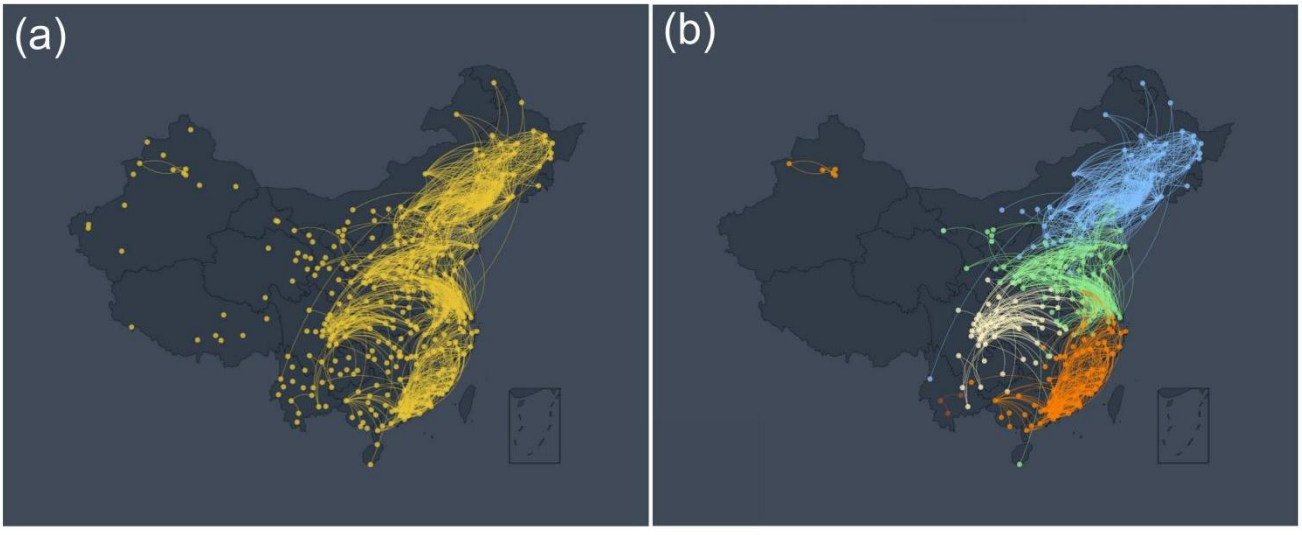


Figure 1. The O_3 Network and Community Structure. (a) A network of O_3 monitoring stations in China constructed according to the link weight calculation method. Yellow nodes represent O_3 observation stations, and links between stations indicate links between stations that meet the connectivity threshold TC = 3.8. (b) This figure is a network structure diagram

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions

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 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_3 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_3 link network. Future research should focus more on the explainable practical role of these network structures.

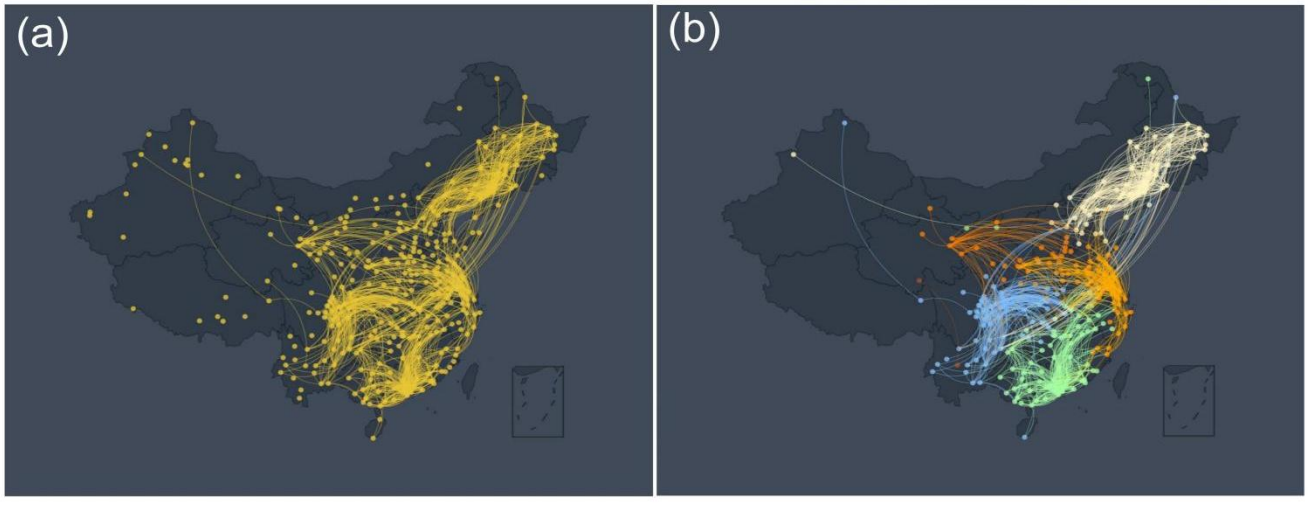


Figure 2. The O_3 Network and Community Structure from May to September. (a) A network of O_3 monitoring stations in China constructed according to the link weight calculation method. Yellow nodes represent O_3 observation stations, and links between stations indicate links between stations that meet the connectivity threshold TC = 3.3. (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.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions

To further verify the community structure presented by the O_3 network in different periods, we employ the Louvain community partitioning algorithm to divide the community of the O_3 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_3 network and also confirm the spatiotemporal characteristics of O3 distribution.

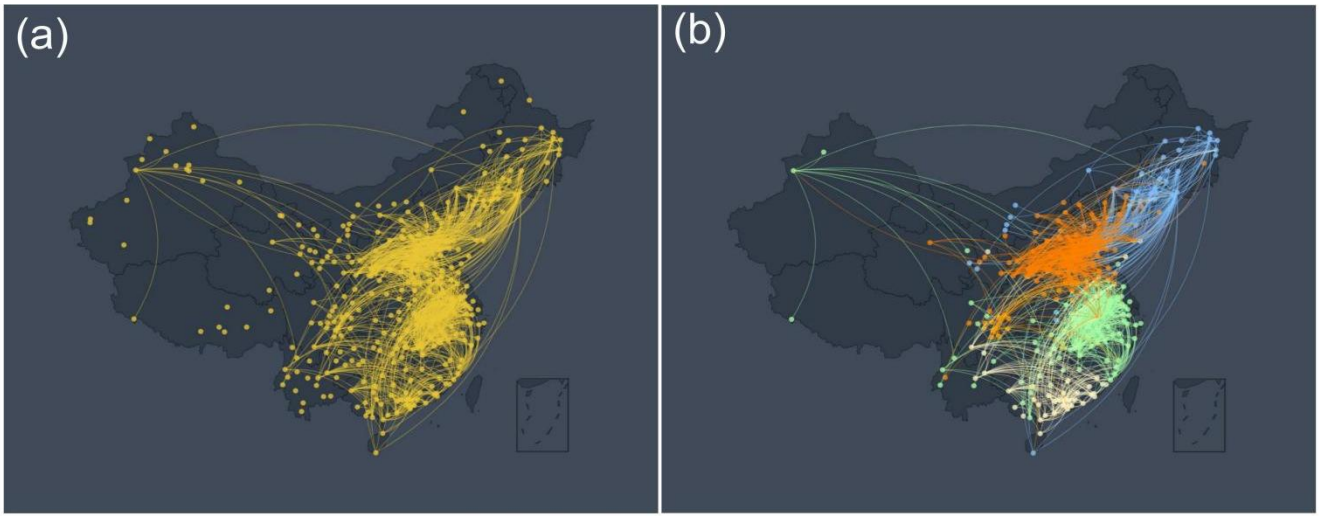


Figure 3. The O_3 Network and Community Structure from October to February. (a) A network of O_3 monitoring stations in China constructed according to the link weight calculation method. Yellow nodes represent O_3 observation stations, and links between stations indicate links between stations that meet 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 Drivens 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(Liu et al., 2011). 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 Tc =3.8.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions
Open Access
EGU

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_3 control network and different periods, we compare the O_3 control network in two different periods from May to September and from October to February. The O_3 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_3 control network during the two

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions
Open Access
EGU

different periods. Although the methods and processing procedures used in these two networks are the same as before, their results are significantly different.

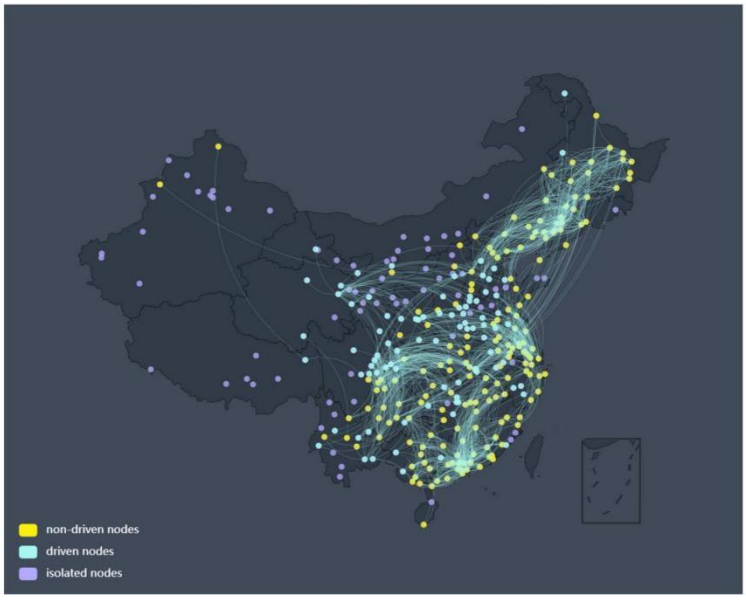


Figure 5. O_3 Network from May to September from the Perspective of Network Control Theory. The figure shows the isolated, driven, and non-driven nodes of the O_3 network with a threshold of TC = 3.3. Yellow represents non-driven nodes, cyan for driven nodes, and purple for isolated nodes.

Figure 5 is a schematic diagram of the control network for O_3 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(Shu et al., 2019; Shu et al., 2020). These results indicate that it is meaningful to use control theory to understand the O_3 network. Future research should focus more on understanding the practical implications of seasonal changes in O_3 network nodes over time.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics
Discussions
Open Access
EGU

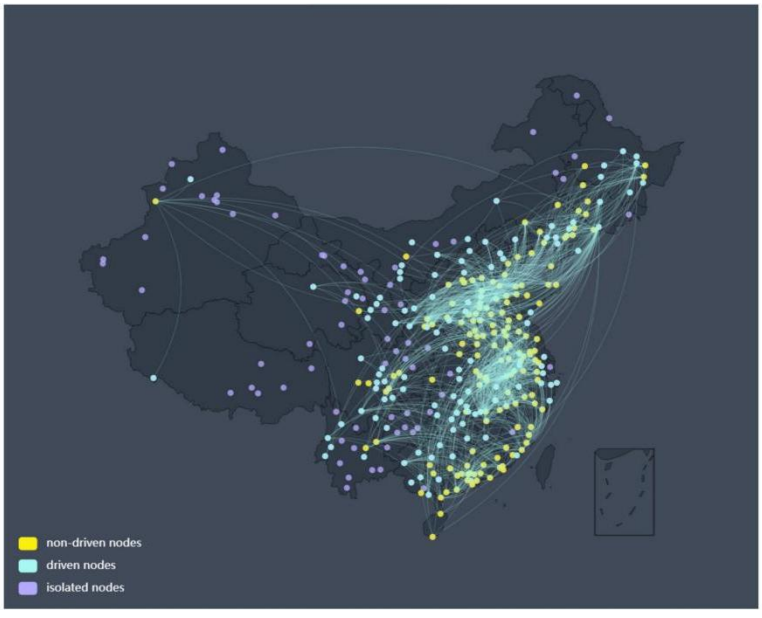
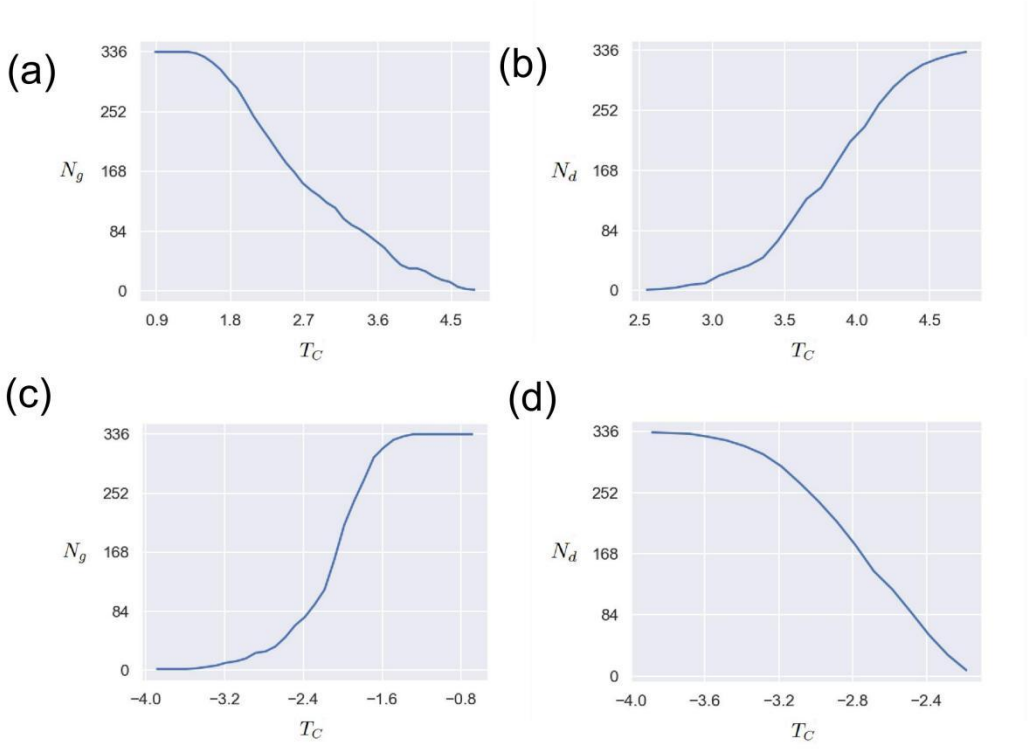


Figure 6. O_3 Network from October to February from the Perspective of Network Control Theory. The figure shows the isolated, driven, and non-driven nodes of the O_3 network with a threshold of TC = 3.1. Yellow represents non-driven nodes, cyan for driven nodes, and purple for isolated nodes.

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.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.



285 **Figure 7.** The effect of threshold TC versus the number of GCC Ng and the number of driven nodes Nd. (a) The relationship between the threshold TC and the giant connected community nodes Ng. The horizontal axis is the value of threshold TC, and the vertical axis is the value of Ng. The network edge weights are positive. (b) The relationship between the threshold TC and the number of driven nodes Nd. The horizontal axis is the value of threshold TC, and the vertical axis is the value of Nd. The network edge weights are positive. (c) The relationship between the threshold TC and the giant connected
290 community nodes Ng. The horizontal axis is the value of threshold TC, and the vertical axis is the value of Ng. The network edge weights are negative. (d) The relationship between the threshold TC and the number of driven nodes Nd. The horizontal axis is the value of threshold TC, and the vertical axis is the value of Nd. The network edge weights are negative.

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
295 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_3 network.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions

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.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

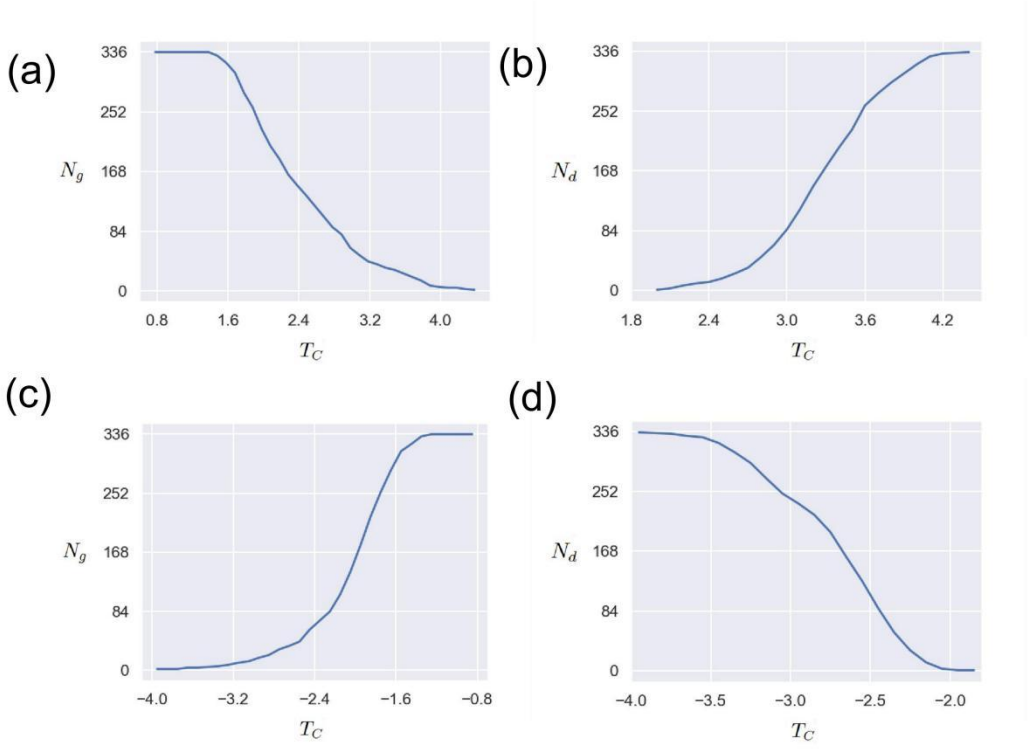


Figure 8. The effect of threshold TC versus the number of GCC Ng and the number of driven nodes Nd from May to September. (a) The relationship between the threshold TC and the giant connected community nodes Ng. The horizontal axis is the value of threshold TC, and the vertical axis is the value of Ng.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

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

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions

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 Figure7(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. 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.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

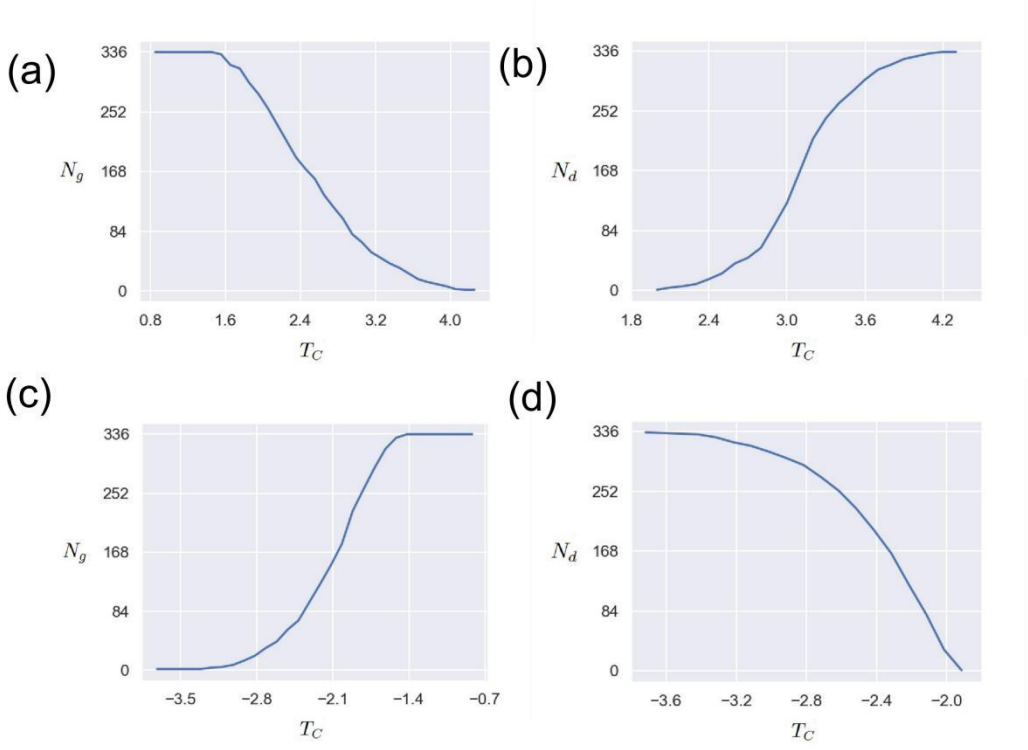


Figure 9. The effect of threshold TC versus the number of GCC Ng and the number of driven nodes Nd from October to February. (a) The relationship between the threshold TC and the giant connected community nodes Ng. The horizontal axis is the value of threshold TC, and the vertical axis is the value of Ng. The network edge weights are positive. (b) The relationship between the threshold TC and the number of driven nodes Nd. The horizontal axis is the value of threshold TC, and the vertical axis is the value of Nd. The network edge weights are positive. (c) The relationship between the threshold TC and the giant connected community nodes Ng. The horizontal axis is the value of threshold TC, and the vertical axis is the value of Ng. The network edge weights are negative. (d) The relationship between the threshold TC and the number of driven nodes Nd. The horizontal axis is the value of threshold TC, and the vertical axis is the value of Nd. 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(Jaccard, 1901). 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 $N_d^{T_C^m}$ represent the set of driven nodes when the thresholds are T_C^m and T_C^n, respectively.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions

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(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).

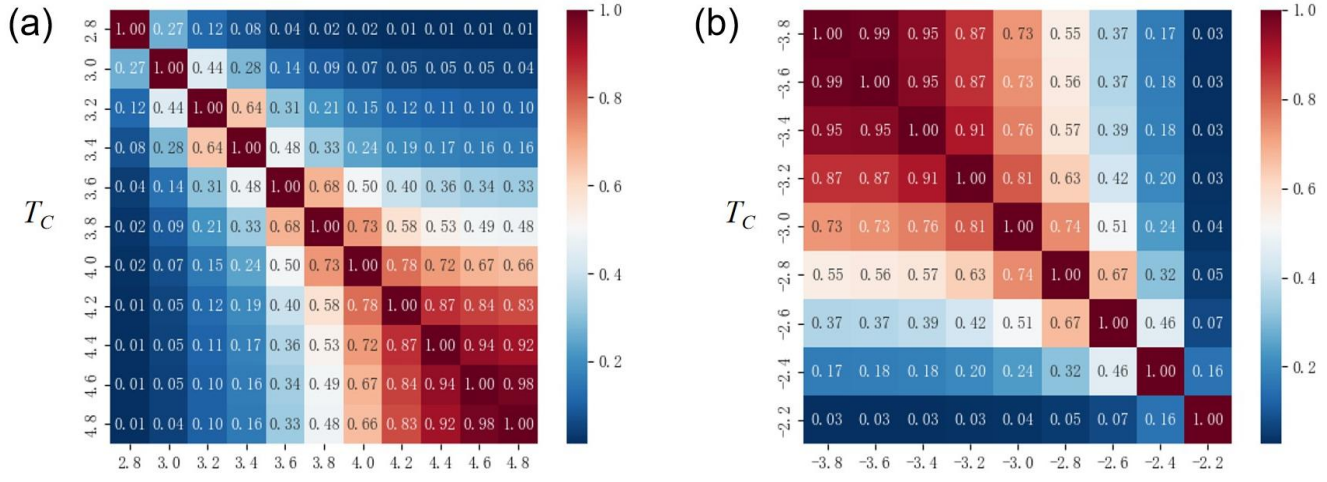


Figure 10. Heat map of co-occurrence rate of driven nodes under different thresholds TC. (a) The horizontal and vertical axes represent different thresholds TC. The higher the Jaccard coefficient of the co-occurrence of the driven nodes, the darker red the pattern color is, and vice versa, the darker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

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics
Discussions
Open Access
EGU

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(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(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.

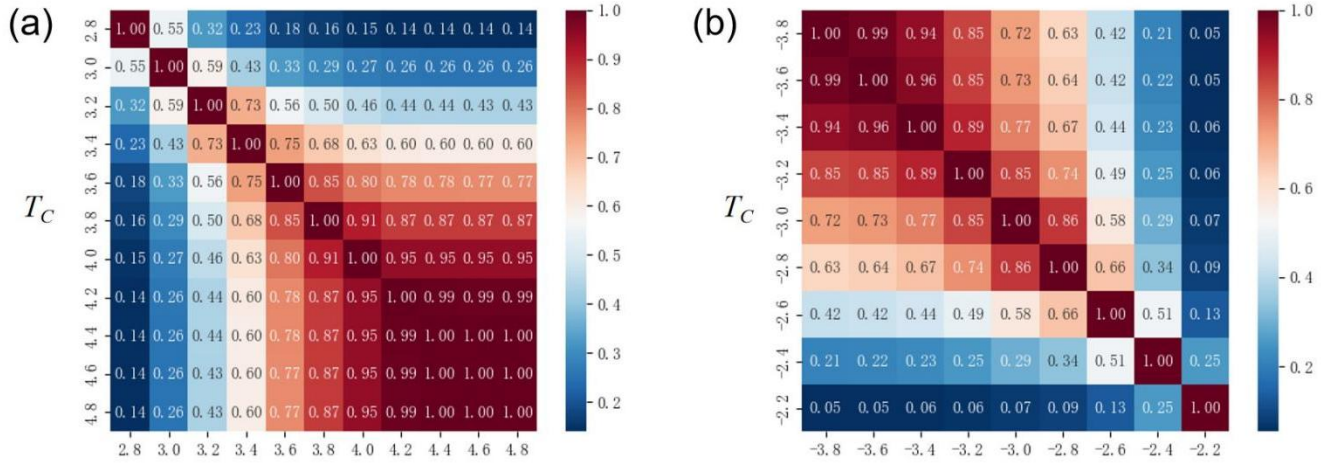


Figure 11. Heat map of co-occurrence rate of driven nodes under different thresholds TC from May to September. (a) The horizontal and vertical axes represent different thresholds TC. The higher the Jaccard coefficient of the co-occurrence of the driven nodes, the darker red the pattern color is, and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

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions
Open Access
EGU

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.

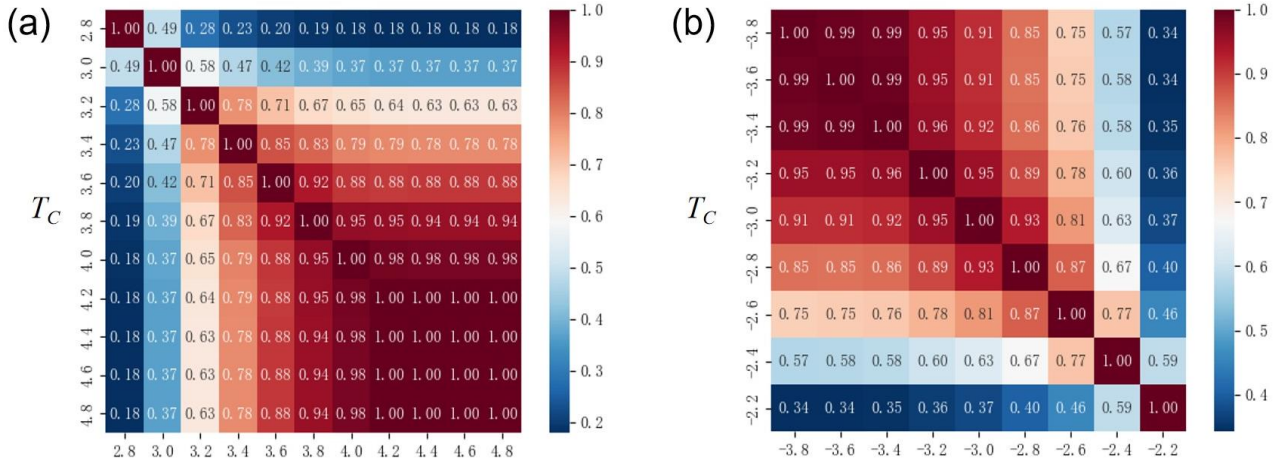


Figure 12. Heat map of co-occurrence rate of driven nodes under different thresholds TC from October to February. (a) The horizontal and vertical axes represent different thresholds TC. The higher the Jaccard coefficient of the co-occurrence of the driven nodes, the darker red the pattern color is, and 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 TC. 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.

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(Ludescher et al., 2021), for example, defining the influence of different variables by the predictability of one variable on another variable(Sugihara et al., 2012). 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 $\widetilde{C}_{nd}$ of the non-driven node; then calculate the Spearman correlation ρ between the predicted non-driven O_3 data sequence $\widetilde{C}_{nd}$ and the original non-driven sequence C_{nd}; finally, Figures 13(a-d) displays the relationship of Spearman correlation ρ 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

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics
Discussions
Open Access
EGU

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 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.

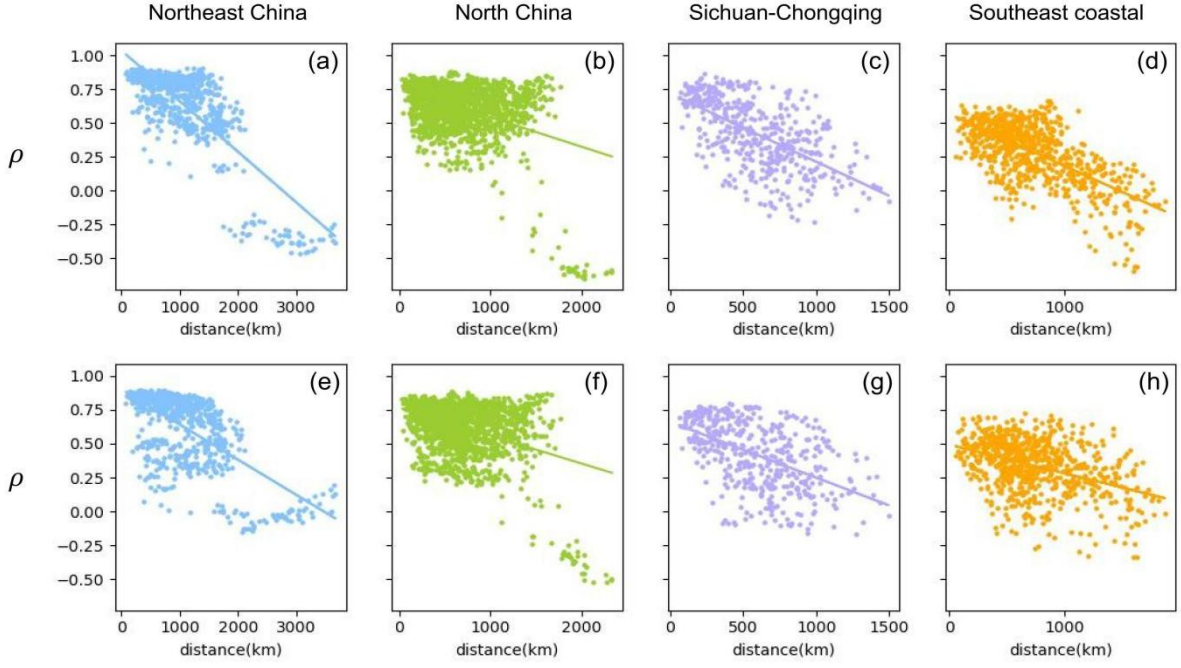


Figure 13. Scatter diagram of the relationship between the distance of city Ci and city Cj and the Spearman coefficient of their O_3 data in the Northeast China, North China, Sichuan-Chongqing and Southeast coastal areas. Figures (a-d) are the data of the direction of the driven to non-driven nodes, the vertical axis represents the Spearman coefficient between the O_3 data of the city Ci (Ci ∈ driven nodes) and the city Cj (Cj ∈ non-driven nodes), the horizontal axis represents the geographical distance between the city Ci and the city Cj. 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 Ci (Ci ∈ non-driven nodes) and city Cj (Cj ∈ driven nodes). The fitted slopes of scattered data are as follows, -3.74×10^{-4}, -2.17×10^{-4}, -5.04×10^{-4}, -3.82×10^{-4},

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions

-2.60×10^{-4}, -2.06×10^{-4}, -4.12×10^{-4}, -2.11×10^{-4}. Overall, we observe the distance between nodes is negatively correlated with their Spearman coefficient.

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 spread of the O_3 pollution into winter and spring due to the climate change, the study of O_3 has attracted more and more attention. Traditional research on O_3 generally focuses on the influence of different factors, etc. These results have greatly enriched our understanding and cognition of O_3. However, the problem of O_3 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_3 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_3 in China.

Our findings illustrate that the O_3 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_3 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_3 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_3 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.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions

Data availability

The study is based on publicly available data sets as described in the Methods section. Model and analysis scripts and outputs are available on request from the corresponding author.

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.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

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).

References

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E.: Fast unfolding of communities in large networks, J. Stat. Mech: Theory Exp., 2008, P10008, https://doi.org/10.1088/1742-5468/2008/10/P10008, 2008.

Ding, A. J., Fu, C. B., Yang, X. Q., Sun, J. N., Zheng, L. F., Xie, Y. N., Herrmann, E., Nie, W., Petäjä, T., Kerminen, V. M., and Kulmala, M.: Ozone and fine particle in the western Yangtze River Delta: an overview of 1 yr data at the SORPES station, Atmos. Chem. Phys., 13, 5813-5830, https://doi.org/10.5194/acp-13-5813-2013, 2013.

Fan, J. F., Meng, J., Ashkenazy, Y., Havlin, S., and Schellnhuber, H. J.: Network analysis reveals strongly localized impacts of El Niño, Proc. Natl. Acad. Sci. U.S.A., 114, 7543-7548, https://doi.org/10.1073/pnas.1701214114, 2017.

Fang, T. T., Zhu, Y., Wang, S. X., Xing, J., Zhao, B., Fan, S. J., Li, M. H., Yang, W. W., Chen, Y., and Huang, R. L.: Source impact and contribution analysis of ambient ozone using multi-modeling approaches over the Pearl River Delta region, China, Environ. Pollut., 289, 117860, https://doi.org/10.1016/j.envpol.2021.117860, 2021.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions Open Access EGU

Gao, J. H., Zhu, B., Xiao, H., Kang, H. Q., Hou, X. W., and Shao, P.: A case study of surface ozone source apportionment during a high concentration episode, under frequent shifting wind conditions over the Yangtze River Delta, China, Sci. Total Environ., 544, 853-863, https://doi.org/10.1016/j.scitotenv.2015.12.039, 2016.

Geng, G. N., Xiao, Q. Y., Zheng, Y. X., Tong, D., Zhang, Y. X., Zhang, X. Y., Zhang, Q., He, K. B., and Liu, Y.: Impact of China's Air Pollution Prevention and Control Action Plan on PM2.5 chemical composition over eastern China, Sci. China Earth Sci., 62, 1872-1884, 10.1007/s11430-018-9353-x, 2019.

Girvan, M. and Newman, M. E. J.: Community structure in social and biological networks, Proc. Natl. Acad. Sci., 99, 7821-7826, https://doi.org/10.1073/pnas.122653799, 2002.

Guan, Y. N., Liu, X. J., Zheng, Z. Y., Dai, Y. W., Du, G. M., Han, J., Hou, L. a., and Duan, E.: Summer O3 pollution cycle characteristics and VOCs sources in a central city of Beijing-Tianjin-Hebei area, China, Environ. Pollut., 323, 121293, https://doi.org/10.1016/j.envpol.2023.121293, 2023.

Guo, Y., Li, K., Zhao, B., Shen, J., Bloss, W. J., Azzi, M., and Zhang, Y. P.: Evaluating the real changes of air quality due to clean air actions using a machine learning technique: Results from 12 Chinese mega-cities during 2013–2020, Chemosphere., 300, 134608, https://doi.org/10.1016/j.chemosphere.2022.134608, 2022.

Hao, J. M., He, K. B., Duan, L., Li, J. H., and Wang, L. T.: Air pollution and its control in China, Front. Environ. Sci. Eng. China., 1, 129-142, https://doi.org/10.1007/s11783-007-0024-2, 2007.

He, L., Duan, Y. S., Zhang, Y., Yu, Q., Huo, J. T., Chen, J., Cui, H. X., Li, Y. W., and Ma, W. C.: Effects of VOC emissions from chemical industrial parks on regional O3-PM2.5 compound pollution in the Yangtze River Delta, Sci. Total Environ., 906, 167503, https://doi.org/10.1016/j.scitotenv.2023.167503, 2024.

Jaccard, P.: Distribution de la flore alpine dans le Bassin des Dranses et dans quelques régions voisines, Bulletin de la Société Vaudoise des Sciences Naturelles., 37, 547-579, https://doi.org/10.5169/seals-266440, 1901.

Kalsoom, U., Wang, T. J., Ma, C. Q., Shu, L., Huang, C. W., and Gao, L. B.: Quadrennial variability and trends of surface ozone across China during 2015–2018: A regional approach, Atmos. Environ., 245, 117989, https://doi.org/10.1016/j.atmosenv.2020.117989, 2021.

Kavassalis, S. C. and Murphy, J. G.: Understanding ozone-meteorology correlations: A role for dry deposition, Geophys. Res. Lett., 44, 2922-2931, https://doi.org/10.1002/2016GL071791, 2017.

Li, M. Y., Yu, S. C., Chen, X., Li, Z., Zhang, Y. B., Wang, L. Q., Liu, W. P., Li, P. F., Lichtfouse, E., Rosenfeld, D., and Seinfeld, J. H.: Large scale control of surface ozone by relative humidity observed during warm seasons in China, Environ. Chem. Lett., 19, 3981-3989, https://doi.org/10.1007/s10311-021-01265-0, 2021.

Liu, C. Q., Geng, H., Shen, P., Wang, Q. G., and Shi, K.: Coupling detrended fluctuation analysis of the relationship between O3 and its precursors –a case study in Taiwan, Atmos. Environ., 188, 18-24, https://doi.org/10.1016/j.atmosenv.2018.06.022, 2018.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Liu, J. D., Wang, L. L., Mingge, L., Liao, Z. H., Sun, Y., Tao, S., Gao, W. K., Wang, Y. H., Li, Y., Ji, D. S., Hu, B., Kerminen, V.-M., Wang, Y. S., and Kulmala, M.: Quantifying the impact of synoptic circulation patterns on ozone variability in northern China from April to October 2013–2017, Atmos. Chem. Phys., 19, 14477-14492, https://doi.org/10.5194/acp-19-14477-2019, 2019.

Liu, Y.-Y., Slotine, J.-J., and Barabási, A.-L.: Controllability of complex networks, Nature., 473, 167-173, https://doi.org/10.1038/nature10011, 2011.

Lu, X., Ye, X. P., Zhou, M., Zhao, Y. H., Weng, H. J., Kong, H., Li, K., Gao, M., Zheng, B., Lin, J. T., Zhou, F., Zhang, Q., Wu, D. M., Zhang, L., and Zhang, Y. H.: The underappreciated role of agricultural soil nitrogen oxide emissions in ozone pollution regulation in North China, Nat. Commun., 12, 5021, https://doi.org/10.1038/s41467-021-25147-9, 2021.

Ludescher, J., Martin, M., Boers, N., Bunde, A., Ciemer, C., Fan, J., Havlin, S., Kretschmer, M., Kurths, J., Runge, J., Stolbova, V., Surovyatkina, E., and Schellnhuber, H. J.: Network-based forecasting of climate phenomena, Proc. Natl. Acad. Sci., 118, e1922872118, https://doi.org/10.1073/pnas.1922872118, 2021.

Mao, J. Y., Yan, F. H., Zheng, L. M., You, Y. C., Wang, W. W., Jia, S. G., Liao, W. H., Wang, X. M., and Chen, W. H.: Ozone control strategies for local formation- and regional transport-dominant scenarios in a manufacturing city in southern China, Sci. Total Environ., 813, 151883, https://doi.org/10.1016/j.scitotenv.2021.151883, 2022.

McDonald, B. C., de Gouw, J. A., Gilman, J. B., Jathar, S. H., Akherati, A., Cappa, C. D., Jimenez, J. L., Lee-Taylor, J., Hayes, P. L., McKeen, S. A., Cui, Y. Y., Kim, S.-W., Gentner, D. R., Isaacman-VanWertz, G., Goldstein, A. H., Harley, R. A., Frost, G. J., Roberts, J. M., Ryerson, T. B., and Trainer, M.: Volatile chemical products emerging as largest petrochemical source of urban organic emissions, Science., 359, 760-764, https://doi.org/10.1126/science.aaq0524, 2018.

Newman, M. E. J.: The Structure and Function of Complex Networks, SIAM Rev., 45, 167-256, https://doi.org/10.1137/S003614450342480, 2003.

Qi, H. Y., Duan, W. J., Cheng, S. Y., Huang, Z. J., and Hou, X. S.: Spatial clustering and spillover pathways analysis of O3, NO2, and CO in eastern China during 2017–2021, Sci. Total Environ., 904, 166814, https://doi.org/10.1016/j.scitotenv.2023.166814, 2023.

Qi, M. Y., Wang, L. T., Ma, S. M., Zhao, L., Lu, X. H., Liu, Y. Y., Zhang, Y., Tan, J. Y., Liu, Z. T., Zhao, S. T., Wang, Q., and Xu, R. G.: Evaluation of PM2.5 fluxes in the "2+26" cities: Transport pathways and intercity contributions, Atmos. Pollut. Res., 12, 101048, https://doi.org/10.1016/j.apr.2021.03.011, 2021.

Shao, M., Zhang, Y. H., Zeng, L. M., Tang, X. Y., Zhang, J., Zhong, L. J., and Wang, B. G.: Ground-level ozone in the Pearl River Delta and the roles of VOC and NOx in its production, J. Environ. Manage., 90, 512-518, https://doi.org/10.1016/j.jenvman.2007.12.008, 2009.

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions Open Access EGU

Shen, L. J., Liu, J., Zhao, T. L., Xu, X. D., Han, H., Wang, H. L., and Shu, Z. Z.: Atmospheric transport drives regional interactions of ozone pollution in China, Sci. Total Environ., 830, 154634, https://doi.org/10.1016/j.scitotenv.2022.154634, 2022.

Shu, L., Wang, T. J., Han, H., Xie, M., Chen, P. L., Li, M. M., and Wu, H.: Summertime ozone pollution in the Yangtze River Delta of eastern China during 2013–2017: Synoptic impacts and source apportionment, Environ. Pollut., 257, 113631, https://doi.org/10.1016/j.envpol.2019.113631, 2020.

Shu, L., Wang, T. J., Xie, M., Li, M. M., Zhao, M., Zhang, M., and Zhao, X. Y.: Episode study of fine particle and ozone during the CAPUM-YRD over Yangtze River Delta of China: Characteristics and source attribution, Atmos. Environ., 203, https://doi.org/10.1016/j.atmosenv.2019.01.044, 2019.

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., and Munch, S.: Detecting Causality in Complex Ecosystems, Science., 338, 496-500, https://doi.org/10.1126/science.1227079, 2012.

Tian, G. X. and Gunes, M. H.: Complex Network Analysis of Ozone Transport, Complex Networks V, Cham, 87-96, https://doi.org/10.1007/978-3-319-05401-8_9, 2014.

Wang, Q., Wang, X. H., Huang, R. Z., Wu, J. B., Xiao, Y., Hu, M., Fu, Q. Y., Duan, Y. S., and Chen, J. M.: Regional Transport of PM2.5 and O3 Based on Complex Network Method and Chemical Transport Model in the Yangtze River Delta, China, J. Geophys. Res.: Atmos., 127, e2021JD034807, https://doi.org/10.1029/2021JD034807, 2022. a.

Wang, R. P., Duan, W. J., Cheng, S. Y., and Wang, X. Q.: Nonlinear and lagged effects of VOCs on SOA and O3 and multi-model validated control strategy for VOC sources, Sci. Total Environ., 887, 164113, https://doi.org/10.1016/j.scitotenv.2023.164113, 2023.

Wang, S. X. and Hao, J. M.: Air quality management in China: Issues, challenges, and options, J. Environ. Sci. China., 24, 2-13, https://doi.org/10.1016/S1001-0742(11)60724-9, 2012.

Wang, S. X., Zhao, B., Cai, S. Y., Klimont, Z., Nielsen, C. P., Morikawa, T., Woo, J. H., Kim, Y., Fu, X., Xu, J. Y., Hao, J. M., and He, K. B.: Emission trends and mitigation options for air pollutants in East Asia, Atmos. Chem. Phys., 14, 6571-6603, https://doi.org/10.5194/acp-14-6571-2014, 2014.

Wang, T., Xue, L. K., Brimblecombe, P., Lam, Y. F., Li, L., and Zhang, L.: Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects, Sci. Total Environ., 575, 1582-1596, https://doi.org/10.1016/j.scitotenv.2016.10.081, 2017.

Wang, W. P., Yang, S. N., Yin, K., Zhao, Z. D., Ying, N., and Fan, J. F.: Network approach reveals the spatiotemporal influence of traffic on air pollution under COVID-19, Chaos., 32, 041106, https://doi.org/10.1063/5.0087844, 2022. b.

Wang, Y. H., Gao, W. K., Wang, S., Song, T., Gong, Z. Y., Ji, D. S., Wang, L. L., Liu, Z. R., Tang, G. Q., Huo, Y. F., Tian, S. L., Li, J. Y., Li, M. G., Yang, Y., Chu, B. W., Petäjä, T., Kerminen, V.-M., He, H., Hao, J. M., Kulmala, M., Wang, Y. S.,

https://doi.org/10.5194/esd-2024-4
Preprint. Discussion started: 25 April 2024
© Author(s) 2024. CC BY 4.0 License.

Earth System Dynamics Discussions

595 and Zhang, Y. H.: Contrasting trends of PM2.5 and surface-ozone concentrations in China from 2013 to 2017, Natl. Sci. Rev., 7, 1331-1339, https://doi.org/10.1093/nsr/nwaa032, 2020.

Wei, J., Li, Z. Q., Li, K., Dickerson, R. R., Pinker, R. T., Wang, J., Liu, X., Sun, L., Xue, W. H., and Cribb, M.: Full-coverage mapping and spatiotemporal variations of ground-level ozone (O3) pollution from 2013 to 2020 across China, Remote Sens. Environ., 270, 112775, https://doi.org/10.1016/j.rse.2021.112775, 2022.

600 Xue, T., Zhu, T., Peng, W., Guan, T. J., Zhang, S. Q., Zheng, Y. X., Geng, G. N., and Zhang, Q.: Clean air actions in China, PM2.5 exposure, and household medical expenditures: A quasi-experimental study, PLoS Med., 18, e1003480, 10.1371/journal.pmed.1003480, 2021.

Yang, W. Y., Du, H. Y., Wang, Z. F., Zhu, L. L., Wang, Z., Chen, X. S., Chen, H. S., Wang, W., Zhang, R. Q., Li, J., and Dao, X.: Characteristics of regional transport during two-year wintertime haze episodes in North China megacities, Atmos.

605 Res., 257, 105582, https://doi.org/10.1016/j.atmosres.2021.105582, 2021.

Yang, Y., Li, M. Y., Wang, H. L., Li, H. M., Wang, P. Y., Li, K., Gao, M., and Liao, H.: ENSO modulation of summertime tropospheric ozone over China, Environ. Res. Lett., 17, 034020, 10.1088/1748-9326/ac54cd, 2022.

Ying, N., Duan, W., Zhao, Z., and Fan, J.: Complex network analysis of fine particulate matter (PM2.5): transport and clustering, Earth Syst. Dynam., 13, 1029-1039, https://doi.org/10.5194/esd-13-1029-2022, 2022.

610 Ying, N., Zhou, D., Chen, Q. H., Ye, Q., and Han, Z. G.: Long-term link detection in the CO2 concentration climate network, J. Cleaner Prod., 208, 1403-1408, https://doi.org/10.1016/j.jclepro.2018.10.093, 2019.

Ying, N., Zhou, D., Han, Z. G., Chen, Q. H., Ye, Q., and Xue, Z. G.: Rossby Waves Detection in the CO2 and Temperature Multilayer Climate Network, Geophys. Res. Lett., 47, e2019GL086507, https://doi.org/10.1029/2019GL086507, 2020. a.

Ying, N., Zhou, D., Han, Z. G., Chen, Q. H., ye, Q., Xue, Z. G., and Wang, W. P.: Climate networks suggest Rossby-waves-

615 related CO_2 concentrations to surface air temperature, Europhys. Lett., 132, https://doi.org/10.1209/0295-5075/132/19001, 2020. b.

Zhang, Y. W., Chen, D. A., Fan, J. F., Havlin, S., and Chen, X. S.: Correlation and scaling behaviors of fine particulate matter (PM2.5) concentration in China, Europhys. Lett., 122, 58003, https://doi.org/10.1209/0295-5075/122/58003, 2018.

Zhang, Y. W., Fan, J. F., Chen, X. S., Ashkenazy, Y., and Havlin, S.: Significant Impact of Rossby Waves on Air Pollution

620 Detected by Network Analysis, Geophys. Res. Lett., 46, 12476-12485, https://doi.org/10.1029/2019GL084649, 2019.

Zhao, Z.-D., Zhao, N., and Ying, N.: Association, Correlation, and Causation Among Transport Variables of PM2.5, Front. Phys., 9, 684104, https://doi.org/10.3389/fphy.2021.684104, 2021.