https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

Earth System Dynamics
Discussions

Complex networks analysis of PM2.5: transport and clustering

Na Ying[1], Wansuo Duan[2], Zhidan Zhao[3], Jinfang Fan[4]

[1]China State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China.
[2]State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China.
[3]China Complexity Computation Lab, Department of Computer Science, School of Engineering, Shantou University, Shantou 515063, China
[4]School of Systems Science, Beijing Normal University, Beijing 100875, China

Correspondence to: Wansuo Duan (duanws@lasg.iap.ac.cn)

Abstract. Complex network theory has been applied to reveal the transport patterns and cooperative regions of fine (<2.5 μm) particulate matter ($PM_{2.5}$) in the whole of China over a long-term record. The results show the degrees, weighted degrees, and edge lengths of $PM_{2.5}$ cities follow power-law distributions. Cities in the Beijing-Tianjin-Hebei-Henan-Shandong (BTHHS) region have a strong ability to import $PM_{2.5}$ pollution to other cities. By analyzing the transport routes, we show that a mass of links extends southward from the BTHHS to the Yangtze River Delta (YRD) regions with one- or two-day time lags. Hence, we conclude that earlier emission reduction in BTHHS and early-warning measures in YRD will help to improve air quality in both regions. Moreover, significant links are concentrated in wintertime, suggesting the impact of the winter monsoon. In addition, cities have been divided into nine clusters according to their synchronicity characteristics. Cities in the same clusters should be regarded as a whole to control the level of air pollution. The results are derived by an economic approach of complex network theory, which avoids the time-consuming of traditional model simulation approach and suggests a highly efficient approach to the studies of transport and cluster of $PM_{2.5}$. This approach, beyond doubt, is certainly also applicable to the studies of other air pollutants such as ozone, NOx, and so on.

1 Introduction

The Earth behaves as a complex self-regulating system comprised of atmosphere, hydrosphere, cryosphere, lithosphere and biosphere, with highly nonlinear interactions and feedbacks between the component parts (Steffen et al 2015). With the increasing understanding of interactions between physical, chemical, biological and human processes, a new 'science of the Earth'–Earth System Science (ESS) has been initiated (Steffen et al 2020). Facilitated by its various tools and approaches, ESS has introduced some new concepts and theories, the most important of which is the concept of Anthropocene (Malm and Hornborg, 2015). In the Anthropocene era, haze events have occurred frequently in China, and the problem of air pollution has received wide attention from the government, scholars and the public (Huang *et al* 2014, Sheehan *et al* 2014).

Atmospheric fine particulate matter ($PM_{2.5}$) is the primary cause of haze pollution (Ding *et al* 2016, Cai *et al* 2017). It has adverse influences on human health, atmospheric visibility and global climate change (Liang *et al* 2016, Liao *et al* 2017). $PM_{2.5}$ pollution is generated from both anthropogenic and natural sources, including primary aerosols as well as secondary aerosols that are produced in the atmosphere through the chemistry of precursor gases (Squizzato *et al* 2012). In recent years, it has also been increasingly recognized that air pollution in a given area is influenced not only by the air pollutant emissions there but also by the transport of air pollutants from other regions. Based on trajectory clustering methods, Li *et al* (2015) concluded that regional $PM_{2.5}$ transmission has become the key factor driving severe haze in Beijing. By using the positive matrix factorization approach, Khuzestani *et al* (2017) revealed that remote transmission accounted for approximately 77% of the $PM_{2.5}$ concentration in the Ordos region. Furthermore, $PM_{2.5}$ transmissions are also examined using model simulations. For example, Wang *et al* (2014) quantified the regional contribution of $PM_{2.5}$ in southern Hebei by using Mesoscale Modeling System Generation 5(MM5) and the Models-3/Community Multiscale Air Quality (CMAQ) modeling system; Zhang *et al* (2017) investigated the effect of regional pollution transport based on the GEOS-Chem chemical transport model and its adjoint. These studies suggest that curbing air pollution has not been a local issue, and the regional coordinate could be an effective approach to improve the air quality of the regional atmospheric environment. In 2012, The 12th Five-Year Plan on Air Pollution Prevention and Control in Key Regions approved proposed to divide China into three key regions to jointly prevent air pollution, which is named as the Beijing-Tianjin-Hebei (BTH), Yangtze River Delta (YRD) and the Pearl River Delta (PRD), and major urban agglomerations such as Lanzhou-Xining, Wuhan and surrounding areas, Shaanxi and Guanzhong city (MEP, 2012). However, this kind of region division ignores the nonlinear transport characteristics of $PM_{2.5}$ concentrations; furthermore, considerable discrepancies exist in the above studies of $PM_{2.5}$ transmission in different cities/regions during different air pollution periods. Hence, the $PM_{2.5}$ transports in the whole of China over a long-time period have not been fully understood; furthermore, the traditional approaches adopted in the above studies dis not fully consider the nonlinear transport processes between cities.

During the last two decades, complex network theory has been applied to reveal the statistical and dynamic topological features in complex systems (Fountalis *et al* 2014, Feldhoff *et al* 2015). In complex networks, geographical locations are considered to be nodes. Links represent communications between time series of nodes, and their strength is measured by the cross-correlation between records (Castrejon-Pita and Read 2010). The network-theory based approach has been used to study teleconnection patterns (Zhou *et al* 2015, Boers *et al* 2019, Ying *et al* 2019), El Niño events (Yamasaki *et al* 2008, Ludescher *et al* 2013, 2014), North Atlantic Oscillation (Guez *et al* 2012), Atlantic Multidecadal Oscillation (Wyatt *et al* 2012) and Rossby waves (Wang *et al* 2013, Ying *et al* 2020). This approach is also useful in the studies of atmosphere environment systems, especially enabling us to investigate the nonlinear spatiotemporal dynamics between air pollution agents. Such nonlinear relationships are critical for assessing the intrinsic dynamics of atmospheric pollution systems, but traditional statistical or model simulation methods are difficult to reveal. The network-theory based approach has been used to uncover the correlation pattern of $PM_{2.5}$ concentrations (Zhang *et al* 2018), to analyze the $PM_{2.5}$ spillover routes in BTH cities (Li *et al* 2019), to discriminate between urban and rural tropospheric ozone (Rafael *et al* 2019), and to quantify the interaction between upper air conditions and surface $PM_{2.5}$ concentrations (Zhang *et al* 2019). It is obvious that complex network methods are valuable tools for depicting and quantifying air pollution transmission and cluster among cities.

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

Earth System Dynamics
Discussions

In the present study, we attempt to explore the transport and cluster of $PM_{2.5}$ based on complex networks, and in the next section, we introduce the data and methods. The patterns of $PM_{2.5}$ concentrations and their transport features and demarcation regions are presented in section 3. Finally, the summary and discussion are detailed in Section 4.

2 Data and methods

2.1 Data

The $PM_{2.5}$ concentrations data for 336 cities of China with a daily average from 1 January 2015 to 31 December 2019 are used in this study. These raw data were acquired from the China National Environmental Monitoring Centre (CNEMC). Then we pre-processed these data according to the needs of the Ambient Air Quality Standard on the validity of air pollutant concentration data. Specifically, the missing values in the $PM_{2.5}$ data are excluded; then the negative values and those larger than 900 mg/m^3 on a given day for a given year are removed and for these years we deleted the data corresponding to those days. As a result, we obtained data for 360 valid days per year (data on January 9, April 1, July 6, September 5, and November 29 are removed) and the total length is 5 × 360 (1800 days).

The anomalies records of $PM_{2.5}$ are adopted, where the anomalies are obtained by subtracting the daily averages and dividing them by the corresponding standard deviations and the function of the denominator is used to eliminate the effects of autocorrelations in the records.

2.2 Methods

The network construction includes three steps. First, we calculate the weight of the edges between nodes. Second, we apply a shuffled procedure to identify a certain threshold. Third, we calculate network typological metrics to determine the interaction strength between two nodes. Below, we detail each step.

Step 1. The calculation of the weight links between nodes

The anomalous $PM_{2.5}$ time series of each node *i* is represented as $\delta S_i(t)$, where *i* is the node index. Similar to earlier studies (Gozolchiani *et al* 2011, Ying *et al* 2020), we define $X_{i,j}(\tau)$ as the time-delayed cross-correlation function for $PM_{2.5}$ node (*i* and *j*), $\delta S_i(t)$ and $\delta S_j(t)$. For τ > 0,

$$X_{i,j}(\tau)=\frac{\langle \delta S_i(t-\tau)\delta S_j(t)\rangle-\langle \delta S_i(t-\tau)\rangle\langle \delta S_j(t)\rangle}{\sqrt{\left\langle\left(\delta S_j(t-\tau)-\langle \delta S_j(t-\tau)\rangle\right)^2\right\rangle}\bullet\sqrt{\left\langle\left(\delta S_j(t)-\langle \delta S_j(t)\rangle\right)^2\right\rangle}} \tag{1}$$

where τ denotes the time lag, which is in the range between −30 and +30 days. $X_{i,j}(\tau)=X_{j,i}(-\tau)$. The bracket is the average over the time period of our concerned. We quantify the strength of the correlations as follows (Gozolchiani *et al* 2011, Guez *et al* 2014):

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

Earth System Dynamics Discussions

$$W_{i,j}^{pos} = \frac{\max(X_{i,j}) - \mathrm{mean}(X_{i,j})}{\mathrm{std}(X_{i,j})} \tag{2}$$

$$W_{i,j}^{neg} = \frac{\min(X_{i,j}) - \mathrm{mean}(X_{i,j})}{\mathrm{std}(X_{i,j})} \tag{3}$$

In this approach, $\max()$, $\mathrm{mean}()$, $\min()$, and $std()$ denote the maximum, minimum, mean, and standard deviation of the cross-correlation function $X_{i,j}(\tau)$, respectively. The deviations in the link identification caused by persistence or autocorrelation in the records are reduced through dividing the $std(X_{i,j})$. We defined the maximum and minimum of $X_{i,j}$ as $P_{i,j}^{pos}$ and $P_{i,j}^{neg}$, respectively; $\tau_{i,j}^{pos}$ and $\tau_{i,j}^{neg}$ represent the maximum and minimum values of $X_{i,j}(\tau)$, respectively; and the sign of $\tau_{i,j}^{pos}$ (or $\tau_{i,j}^{neg}$) represent the direction of each positive (or negative) link. When $\tau_{i,j}^{pos}$ >0, the link is regarded as from node *i* pointing to node *j*. When $\tau_{i,j}^{pos}$ <0, the link is regarded as pointing away from node *j* to node *i*. The direction is undefined when $\tau_{i,j}^{pos}$ =0. The definitions are similar for the negative weighted links.

The adjacency matrix is defined as:

$$A_{i,j}^{pos} = (1-\delta_{i,j})H(W_{i,j}^{pos} - Q) \tag{4}$$

where $\delta_{i,j}$ is the Kronecker delta introduced to avoid self-loops in the network and $H(x)$ is the Heaviside step function ($H(x > 0) = 1$ and $H(x < 0) = 0$). Q denotes a certain threshold value. The definitions are similar for the negative weighted links. We constructed networks by pruning the links for which the statistical significance was below a certain threshold (Guez et al. 2014). The threshold is determined according to the shuffle method, which is explained in detail in the next section.

Step 2. The identification of the critical threshold

In the shuffled case, the order of years is permutated and the order of days within each year is maintained for each pair of nodes (Ying *et al* 2020). For each link, we selected one of two nodes randomly, then shuffled this time series by persisting the order of days in each year and changing the permutation of years several times. We then calculated the cross-correlation function and weight links for the shuffled datasets. The shuffling procedure represents the properties of statistical quantities and the autocorrelations of the original records, which may introduce unrealistic links. We only considered the link weights in the original network that are significantly higher than values in the shuffled case as a real link; otherwise, they are classed as spurious links. According to the principles mentioned above, figure 1 depicts a description of the research process and integration of analytical tools.

Step 3. The determination of network typological metrics

The degree is the most common application for measuring complex networks. A link that points toward a node is referred to as an in-degree link, and a link that points away from a node is considered as an out-degree link. The in- (or out-) weights

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

Earth System Dynamics
Discussions

degrees of node i is denoted as $\mathrm{In}(w)_i$ and $\mathrm{Out}(w)_i$, representing the total in-coming (or out-going) weighted links, respectively

$$\mathrm{In}(w)_i = \sum\nolimits_j A_{j,i} W_{j,i} \tag{5}$$

$$\mathrm{Out}(w)_i = \sum\nolimits_j A_{i,j} W_{i,j} \tag{6}$$

The In and Out fields represent a node's dependence on its surrounding nodes, and the influence of the node on the surroundings nodes, respectively. Nodes with higher values in the network indicate a larger amount of connection with other nodes, whereas lower values indicate that the node is isolated.

The Girvan Newman algorithm is used to explore regional division in the networks. In binary networks, the quality of community structure is typically measured by the modularity (Q) function (Newman, 2006). A high value of Q suggests a strong division of a network into clusters. Nodes in the same community may have the same properties. The Q in networks is defined as follows:

$$Q = \frac{1}{2M} \sum_{i,j} [(A_{i,j} - \frac{k_i k_j}{2M}) \delta(\sigma_i, \sigma_j)] \tag{7}$$

where k_i, k_j is the weight if node i and j, $A_{i,j}$ is the adjacency matrix, δ is the membership function and M is the number of edges.

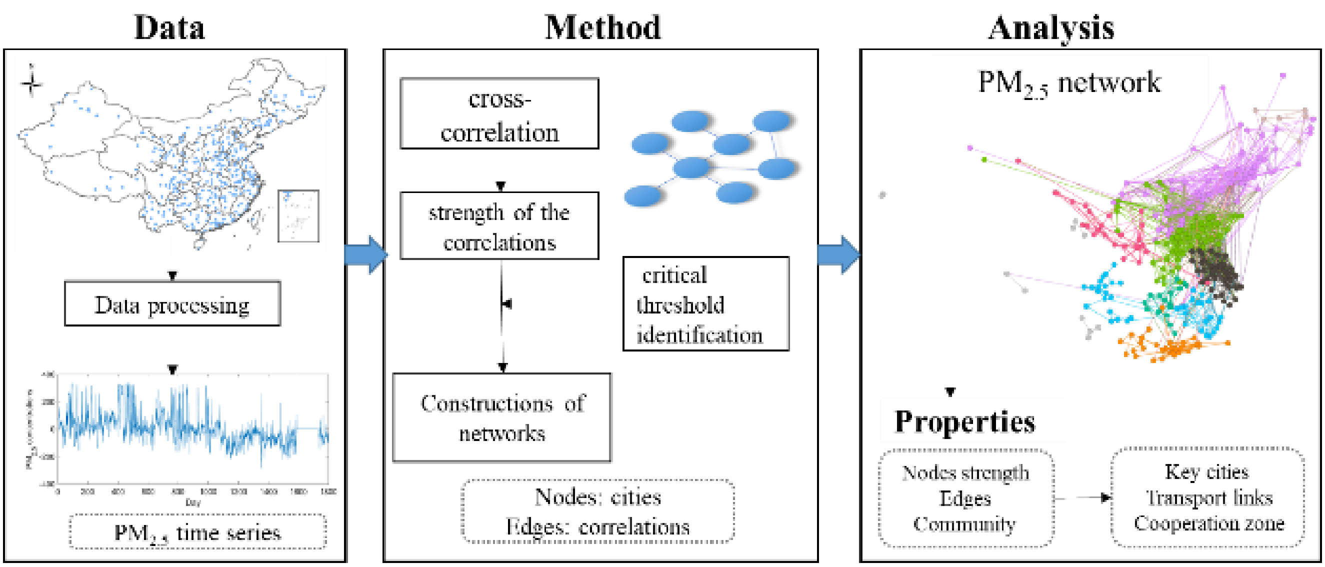

Figure 1. The flow chart of the method with complex network analysis.

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

Earth System Dynamics
Discussions

3. Results

3.1 Characteristics of the PM2.5 network

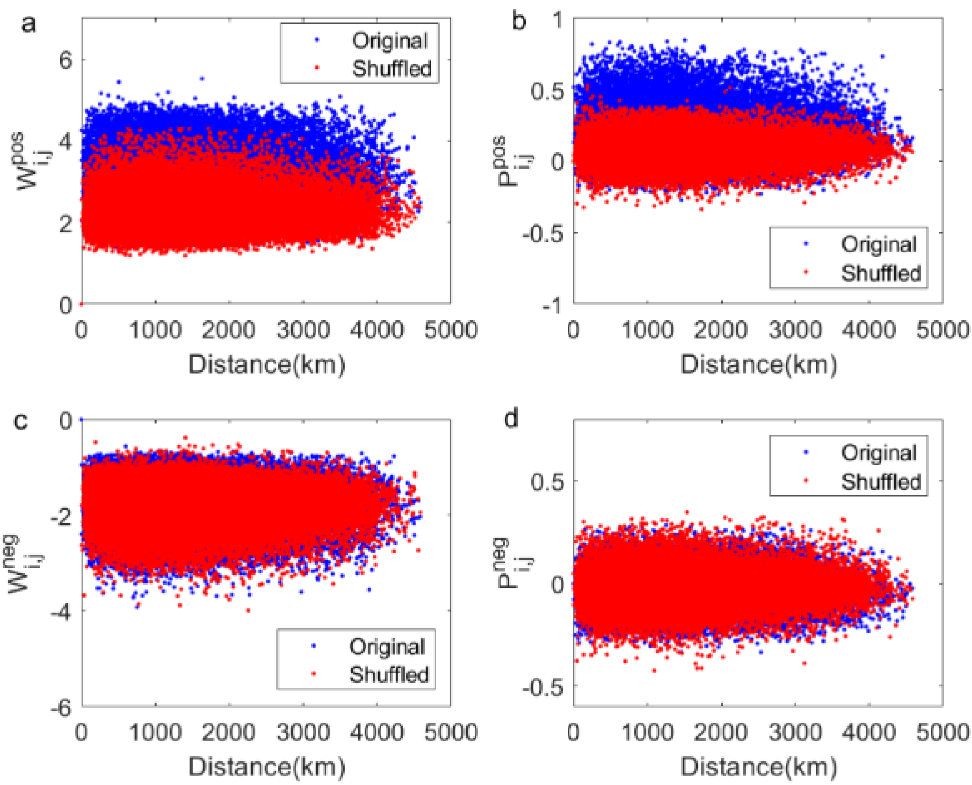

Figure 2. Positive link weights as a function of geographical distances $D_{i,j}$ for (a) $W_{i,j}^{pos}$ and (b) $P_{i,j}^{pos}$ for rea (blue)and shuffled (red) data. (c), (d) Same as (a), (b) but for negative links.

The function of positive link weights$W_{i,j}^{pos}$ and geographical distances $D_{i,j}$ for the original and the shuffled networks are shown in figures 2(a). $W_{i,j}^{pos}$ values in the original network are greater than those in the shuffled network, indicating that the stronger positive links are the result of information transport of $PM_{2.5}$ concentrations. For the relation between the largest cross-correlation $P_{i,j}^{pos}$versus $D_{i,j}$, we observe that the values in the shuffled case are smaller than those in the original case (figures 2(b)), which is in agreement with the pattern of $W_{i,j}^{pos}$. In the negative case (figures 2 (c) and (d)), there is no distinct difference between the original network and the shuffled network.

Figure 3 shows the probability density function (PDF) of links in the original network and the shuffled network. The PDF of positive links weights has a long tail in the original data, which is not presented in the link weights of the shuffled networks.

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

The PDF of negative link weights is a signature of random behavior, which continues to indicate that the many significant positive links are not likely to occur by chance. As a result, we consider links that are separated from the shuffled links. Both $W_{i,j}^{pos}$ and $P_{i,j}^{pos}$ can be used as a measure of the strength of links. In our analysis, positive link weights of 4.2 are the threshold, and accordingly, gain the adjacency matrix of the network.

In the network, 284 cities are connected by $PM_{2.5}$ concentrations with 3930 links among cities. The clustering coefficient, which indicates the degree of connection of the network, is 0.46. We also analyze the shuffled network with the same number of edges. 337 cities are connected and the value in the shuffled network is 0.07, suggesting $PM_{2.5}$ cities are more connected to each other. The density of networks is 0.05 in the original network, while the value is 0.03 in the shuffled network. It reflects the degree of completeness of the network, and high values mean strong connections between cities. The average path length is 4.61 and 3.15 for the original and shuffled network, indicating that cities transport the $PM_{2.5}$ concentrations to other cities crossed almost three other cities. $PM_{2.5}$ cities have a higher clustering coefficient and lower average path length, compared with the shuffled network, demonstrating cities with higher $PM_{2.5}$ concentrations can quickly affect their surrounding cities.

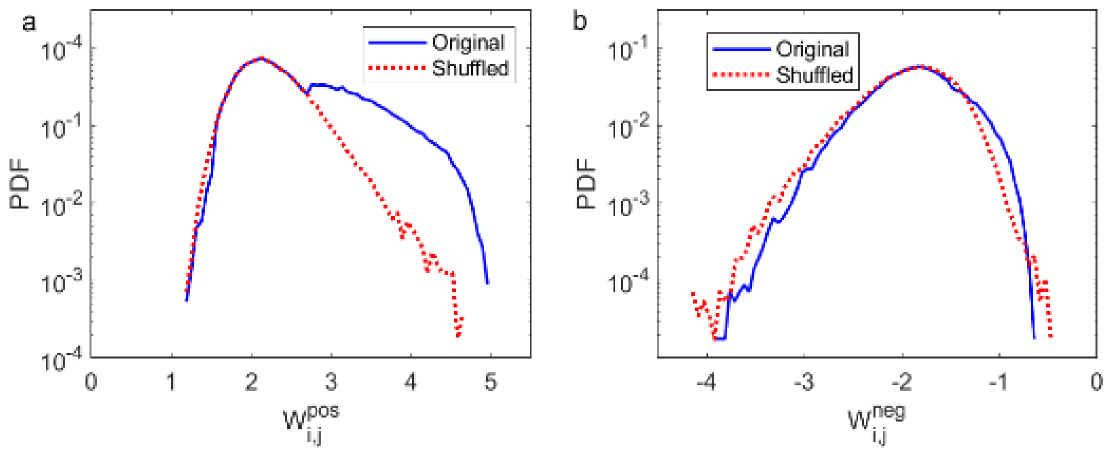

Figure 3. PDF of positive (a) and negative (b) link weights for original data and shuffled data. The blue lines represent the original data and the red dash lines denote the shuffled cases.

The degree of a node is one of the most important statistical properties in networks. The weighted degree characterizes the total strength of correlation of the node with surrounding cities. The PDF of degrees, weighted degrees, and edge lengths of the nodes are shown in figure 4. It is found that the degrees, weighted degrees, and edge lengths conform to power-law distributions. The power-law exponents are 1.3, 1.2, and 1.5, with R-squared values 0.71, 0.70, and 0.63, respectively. These links are heterogeneous, with few nodes possessing the majority of links in the network. Most of the $PM_{2.5}$ concentration links remain confined to a handful of cities. Moreover, these links are mainly short distances (≤1000 km), whereas long distances (>1000 km) show few connections.

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

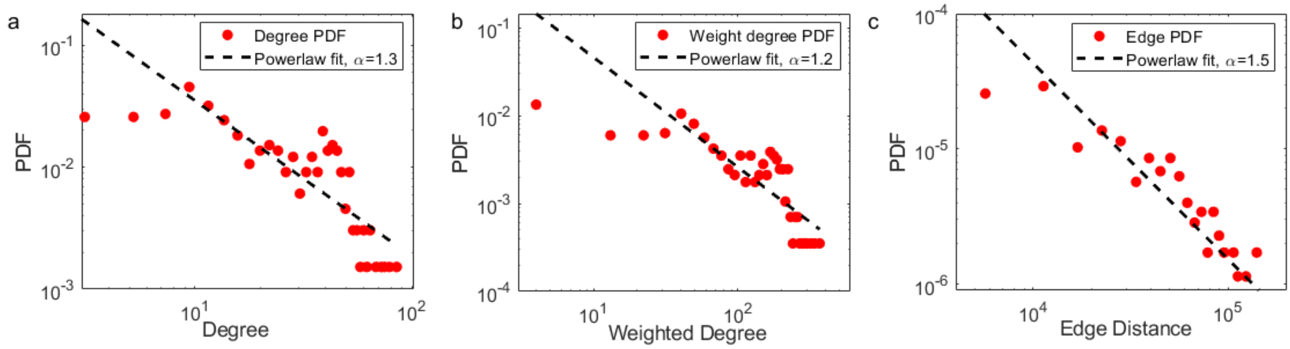

Figure 4. (a) the PDF of degree (red dots) and the power law fit curve (black line); (b) PDF of weighted degrees (red dots) and the power law fit curve (black line). (c) PDF of edge lengths (km) (red dots) and the power law fit curve (black line).

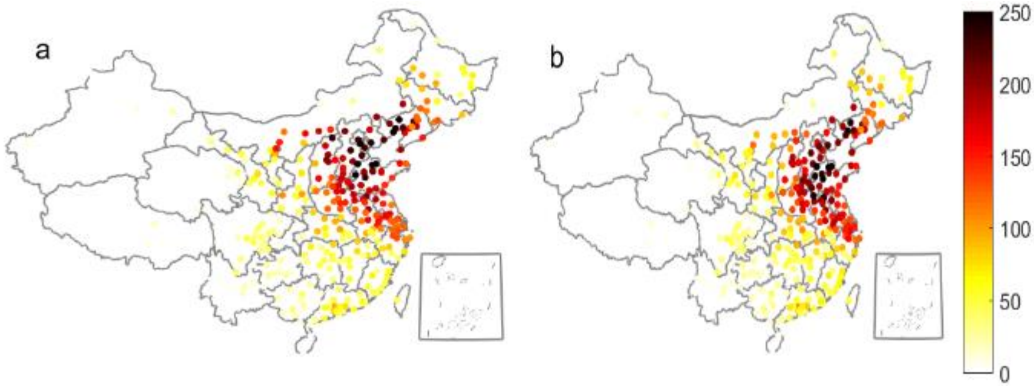

Figure 5. Distribution of in- weighted degree (a) and out- weighted degree (b) in the network of each node for positive cases.

To examine a node's dependence or influencing role on the other nodes, we analyze the patterns of in- and out-weighted degrees. The direction of links is determined by the sign of the time delay, which quantifies the incoming or outgoing nodes. Links with zero-time delay represent bidirectional links. The in-weighted degree of each node measures incoming links towards the target city and high values indicate a stronger export effect from source cities to the target city. Out-weighted degrees denote the strength of outgoing links to other cities, and higher values suggest that more cities transfer their $PM_{2.5}$ concentrations to the target city. Figure 6 presents the spatial distribution of in- and out-weighted degrees for the whole years. Different colors represent the ability to transmission. Regions in BTHHS, YRD, and northwest China show significant synchronicity with the rest of the provinces in terms of $PM_{2.5}$ mass concentrations. These regions correspond to regions with high mean $PM_{2.5}$ concentrations. Furthermore, we observe that the distribution of the in-weighted degree is similar to that of the out-weighted degree, which indicates these cities are both recipients and senders in the networks. This suggests that their pollution is not only due to the local emissions but also imported from other cities. Therefore, solving air pollution should not only rely on reducing emissions in a single city, but rather on developing inter-city cooperation. Compared with the out-

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

Earth System Dynamics
Discussions

weighted degrees, in-weighted degrees are stronger over the BTHHS region. These cities (sending cities) can also export $PM_{2.5}$ concentrations to other cities (recipient cities). In addition, the values of in-/out- weighted degrees display remarkable differences in different seasons, as shown in figure 7. The weighted degrees in summer and autumn are small (figure 7(b) and (c)). In winter and spring, especially in wintertime the values of in-/out- weighted degrees are significant, and their patterns are similar to that of the whole year.

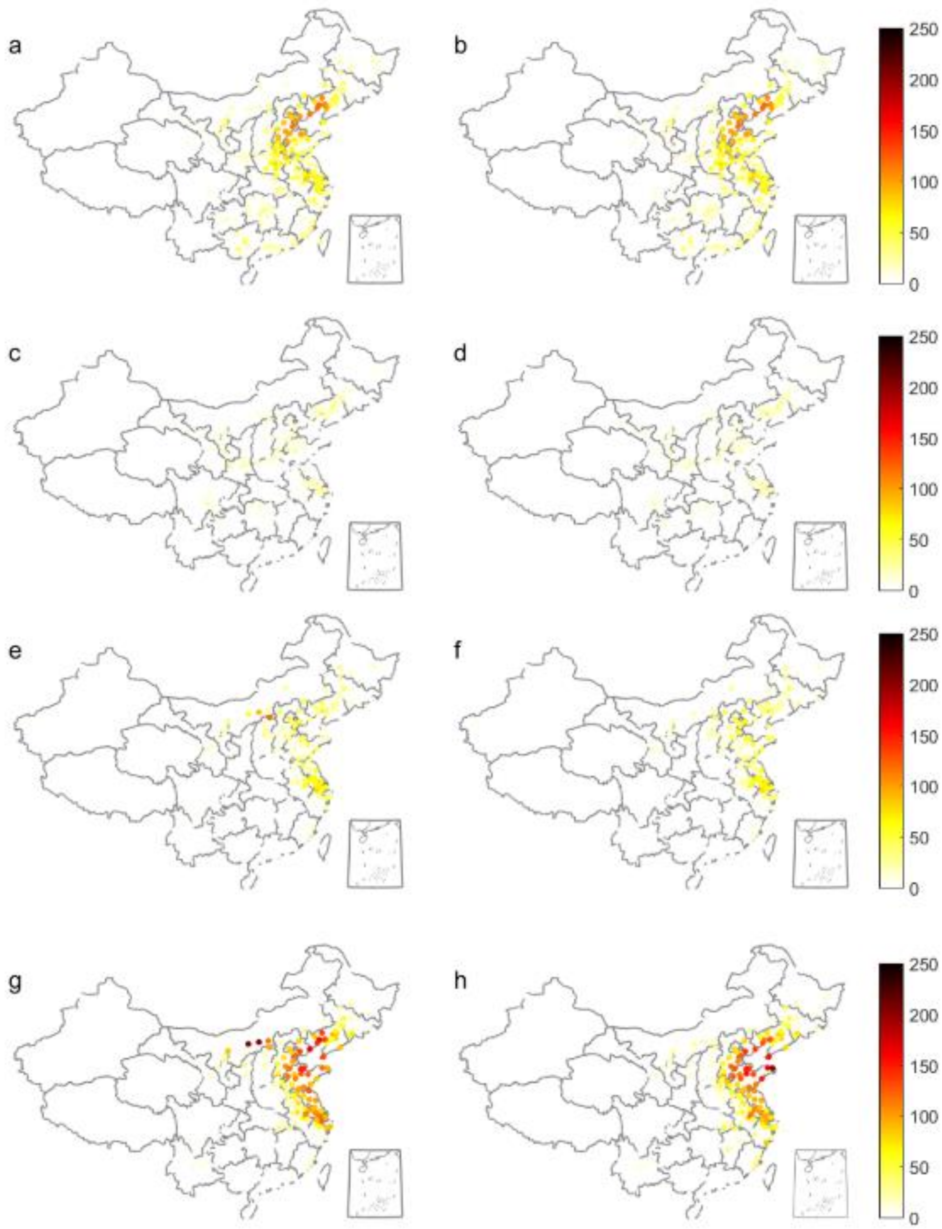

Figure 6. Distribution of in- weighted degree (a) and out- weighted degree (b) in the network of each node for positive cases.

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

3.2 Routes and clustering of the PM2.5

Both in- and out-weighted degrees offer information in terms of nodes (cities). It is reported that urban air quality can be substantially influenced by atmospheric transport of $PM_{2.5}$ pollution from distant cities. An analysis of the edges can contribute to revealing the transport routes of $PM_{2.5}$ among cities. A recent study found that $PM_{2.5}$ concentrations over a distance of 1000 km were related to a typical cyclonic scale within the Rossby waves (Zhang *et al* 2019). Here we discuss the transport path within 1000 km and only focus on positive time lags. This is since they are typical links that are related to different climate processes, and they enable detailed comparisons with the previous literature. The transport routes show that southward propagation is predominant in the sub-network (figure 7(a) (Zhang and Cao 2015)). We focus on two groups of connections that belong to different regions. The first one is links traveled from the Gobi Desert over southwestern parts of Mongolia and the Badain Jaran Desert to the BTH regions. The second one is links transported from the BTHHS to the YRD regions and these links show a 1- or 2-day time lag. This is consistent with previous studies obtained from the WRF-Chem model (Huang *et al* 2020). The outbreak of YRD pollution usually peaks with a time lag of 1–2 days after that in the BTHHS. The government in YRD should implement early warning measures to prevent the negative influence from BTHHS, while the government in BTHHS should take steps to improve air quality by re-adjusting and optimizing the industrial structure, reducing the ratio of heavy industry and developing clean energy.

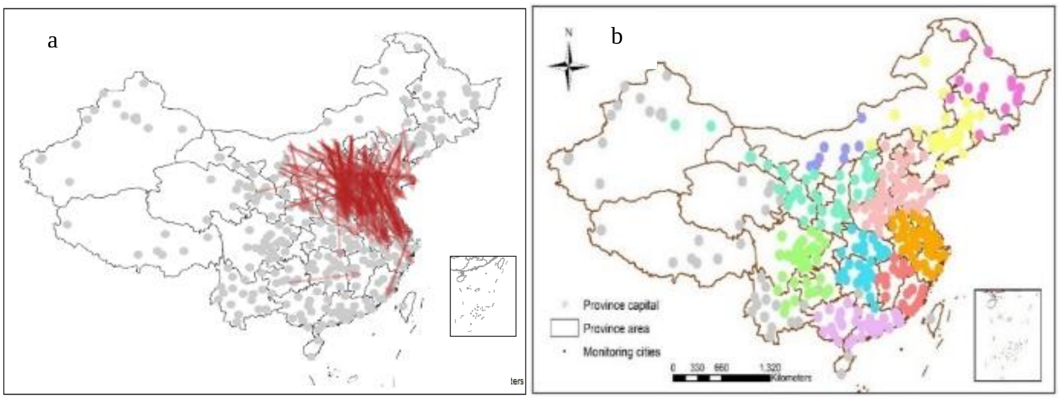

Figure 7. (a) Map of $PM_{2.5}$ transport links among the monitoring cities in China. (b)The cluster regions of $PM_{2.5}$ concentrations. Different colors represent different communities.

In addition, we also analyzed the transport routes in different seasons (figure 8). The transport routes are significant in autumn and winter, especially in wintertime. It means the routes features in winter are dominant over the whole year. Hence, the southwestern links are related to the East Asia winter monsoon.

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

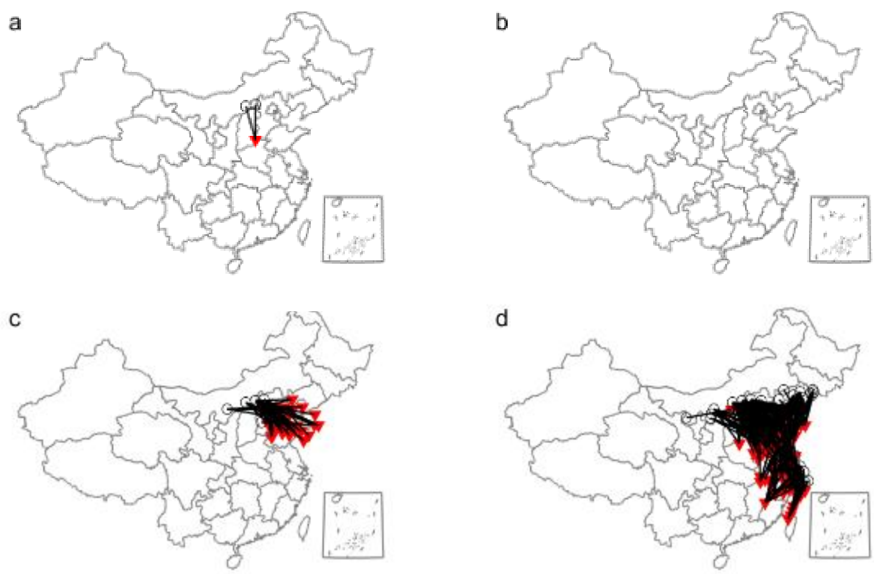

Figure 8. Distribution of transport paths in the network for spring (a), summer (b), autumn (c) and winter (d).

In complex networks, nodes that are closely related to each other are more likely to be grouped in the same cluster. Hence, cities are tightly bound to cities in the same cluster and uncorrelated to cities in other clusters. The pollution transport routes presented above indicate that curbing air pollution is more than just a local issue. In the following, we investigate the cluster features of our networks by utilizing the modularity algorithm described above. Considering a larger Q value means a more accurate community structure for network segmentation, we calculate the Q value at each division to obtain a better result. Here, 284 cities are divided into 9 clusters, where the Q value obtains the maximum value (0.56). The results present a strong regional character regional division, shown in figures 7(b). Cities having the same color represent the same cluster, which could be considered for collaborative governance. These nine regions include the above-mentioned three key regions: BTH regions, YRD region (containing Shanghai, Jiangsu, Anhui and Zhejiang province), and the PRD area (including Guangdong and Guangxi). The other interconnected areas are Heilongjiang and Jilin provinces, Jilin and Liaoning province (northeast China), Hunan and Hubei province (central China), and Jiangxi-Fujian, Guizhou-Chongqing-Sichuan, and Shanxi-Shaanxi-Ningxia-Gansu.

4. Summary and discussion

In the Anthropocene era, the atmospheric environment issue is increasingly prominent, which brings challenges to the realization of sustainable development. Despite great efforts has been taken by the Chinese government, some cities in China are still plagued by haze pollution. Air pollution was partly related to the transmission from other regions, controlling air

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

pollution needs to consider regional transmission and cooperation. The emergence and application of complex networks could enhance our understanding of the dynamics process of $PM_{2.5}$. This paper analyses the transport routes and joint clusters over China based on a network theory-based approach.

By constructing $PM_{2.5}$ networks based on complex network approaches, it is found that the PDF of the degrees, weighted degrees, and edge lengths of $PM_{2.5}$ cities follow a power-law distribution, which indicates the variability of $PM_{2.5}$ concentrations in China is not random. Hence, it is reasonable to analyze the transmission and cooperation regions of $PM_{2.5}$ from the perspective of whole national evolution over a long period of time. To quantify the relations of $PM_{2.5}$ among cities, the patterns of weighted degree are investigated. Higher weighted degrees are overserved in the BTH regions, which is consistent with the patterns of high levels of $PM_{2.5}$ concentrations. Cites in the BTH region have stronger strength to export their $PM_{2.5}$ pollution to other cities. The distributions of weighted degrees exhibit significant differences in seasons, with the largest in winter and the least in summer.

Based on the $PM_{2.5}$ networks, the transport links and collaborative regions are analyzed. It showed that a dense of links traveled from the Gobi Desert over southwestern parts of Mongolia and the Badain Jaran Desert to the BTH regions. The other group extends southward from BTH to the YRD regions and then south to Fujian province with a one- or two-day time lag. This is consistent with previous studies obtained from the WRF-Chem model (Huang *et al* 2014). In winter, although we get a similar transmission pattern, it possesses a strong intensity. We demonstrate that the possible reason is resulted from the influence of cold fronts, which, exactly, disperses the $PM_{2.5}$ accumulated in the North China Plain to the Yangtze River Delta region and thus, leads to the propagation of $PM_{2.5}$ from the BTH region to the YRD region. Hence, links BTH to the YRD region obtained from the whole year are related to the cold front occurring in wintertime.

Besides, we also performed the communities detection based on the synchronicity of $PM_{2.5}$ concentrations, and obtained 9 clusters. Cities in the same regions should join together to control air pollution. This result provides theoretical support for the JPCAP proposed by the national government. Regional cooperation should be promoted in these regions to implement regional policies to improve air quality.

A central implication of this study is that the transmission and collaborative regions can be explored via the complex network approach. For traditional model simulation, numerous parameters are needed in the simulation process. In contrast, complex network theory is performed based on time series of field observations, so the estimation process is faster and more economic. As our analysis is based on long-time $PM_{2.5}$ records in China, rather than a particular region or period of air pollution, it may provide reference and basis for the development of effective regulatory policies for government to improve air quality. In this paper, we demonstrate the applicability of complex network methodology for the studies of the transport and cluster of air pollutants in faster and more economic ways. It is expected that complex network methods are also potential in the studies of other air pollutants such as ozone, NOx, and so on.

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

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

NY developed the research idea, NY developed the model and performed the analysis. All authors discussed the results and contributed to the writing of the paper.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This study is supported by Budget Surplus of Central Financial Science and Technology Plan (Grant No. 2021-JY-15) and National Key Research and Development Program of China (Grant No. 2019YFC0214201).

References

Boers, N., Goswami, B., Rheinwalt, A., Bookhagen, B., Hoskins, B., and Kurths J.: Complex networks reveal global pattern of extreme-rainfall teleconnections, Nature, 566, 373-377, https://doi.org/10.1038/s41586-018-0872-x, 2019.

Cai, S., Wang, Y., Zhao, B., Wang, S., Chang, X., and Hao, J.: The impact of the "Air Pollution Prevention and Control Action Plan" on $PM_{2.5}$ concentrations in Jing-Jin-Ji region during 2012-2020, Sci. Total. Environ., 580, 197-209, https://doi.org/10.1016/j.scitotenv.2016.11.188, 2017.

Castrejon-Pita, A. A. and Read, P. L.: Synchronization in a Pair of Thermally Coupled Rotating Baroclinic Annuli: Understanding Atmospheric Teleconnections in the Laboratory, Phys. Rev. Lett., 104, https://doi.org/10.1103/PhysRevLett.104.204501, 2010.

Ding, A. J., Huang, X., Nie, W., Sun, J. N., Kerminen, V. -M., Petaja, T., Su, H., Cheng, Y. F., Yang, X. -Q., Wang, M. H., Chi, X. G., Wang, J. P., Virkkula, A., Guo, W. D., Yuan, J. Wang, S. Y., Zhang, R. J., Wu, Y. F., Song, Y. Zhu, T., Zilitinkevich, S. Kulmala, M., and Fu, C. B.: Enhanced haze pollution by black carbon in megacities in China, Geophys. Res. Lett., 43, 2873-79, https://doi.org/10.1002/2016GL067745, 2016.

Feldhoff, J. H., Lange, S., Volkholz, J., Donges, J. F., Kurths, J., and Gerstengarbe, F-W.: Complex networks for climate model evaluation with application to statistical versus dynamical modeling of South American climate, Clim. Dynam., 44, 1567-81, https://doi.org/10.1007/s00382-014-2182-9, 2015.

Fountalis, I., Bracco, A., and Dovrolis, C.: Spatio-temporal network analysis for studying climate patterns, Clim. Dynam., 42, 879-99, https://doi.org/10.1007/s00382-013-1729-5, 2014.

Gozolchiani, A., Havlin, S., and Yamasaki, K.: Emergence of El Nino as an autonomous component in the climate network, Phys. Rev. Lett., 107, 148501, https://doi.org/10.1103/PhysRevLett.107.148501, 2011.

Guez, O., Gozolchiani, A., Berezin, Y., Brenner, S., and Havlin, S.: Climate network structure evolves with North Atlantic Oscillation phases, Epl, 98, 38006, https://doi.org/10.1209/0295-5075/98/38006, 2012.

Guez, O. C., Gozolchiani, A., and Havlin, S.: Influence of autocorrelation on the topology of the climate network, Phys. Rev. E, 90, https://doi.org/10.1103/PhysRevE.90.062814, 2014.

Guo, S., Hu, M., Zamora, M. L., Peng, J., Shang, D., Zheng, J., Du, Z., Wu, Z., Shao, M., Zeng, L., Molina, M. J., and Zhang, R.: Elucidating severe urban haze formation in China, Proc. Natl Acad. Sci. USA, 111, 17373-78, https://doi.org/10.1073/pnas.1419604111, 2014.

Huang, R-J., Zhang, Y., Bozzetti, C., Ho, K-F., Cao, J-J., Han, Y., Daellenbach, K. R., Slowik, J. G., Platt, S. M., Canonaco, F., Zotter, P., Wolf, R., Pieber, S. M., Bruns, E. A., Crippa, M., Ciarelli, G., Piazzalunga, A., Schwikowski, M., Abbaszade, G., Schnelle-Kreis, J., Zimmermann, R., An, Z., Szidat, S., Baltensperger, U., Haddad, I. E., and Prévôt, A. S. H.: High secondary aerosol contribution to particulate pollution during haze events in China, Nature, 514, 218-22, https://doi.org/10.1038/nature13774, 2014.

Li, H., Qi, Y., Li, C., and Liu, X.: Routes and clustering features of PM2.5 spillover within the Jing-Jin-Ji region at multiple timescales identified using complex network-based methods, J. Clean. Prod., 209, 1195-205, https://doi.org/10.1016/j.jclepro.2018.10.284, 2019.

Liang, C. S., Duan, F. K., He, K. B., and Ma, Y. L.: Review on recent progress in observations, source identifications and countermeasures of $PM_{2.5}$, Environ. Int., 86, 150-70, https://doi.org/10.1016/j.envint.2015.10.016, 2016.

Liao, T., Wang, S., Ai, J., Gui, K., Duan, B., Zhao, Q., Zhang, X., Jiang, W., and Sun, Y.: Heavy pollution episodes, transport pathways and potential sources of PM2.5 during the winter of 2013 in Chengdu (China), Sci. Total Environ., 584, 1056-65, https://doi.org/10.1016/j.scitotenv.2017.01.160, 2017.

Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S., and Schellnhuber, H. J.: Improved El Nino forecasting by cooperativity detection, Proc. Natl Acad. Sci. USA, 110, 11742-45, https://doi.org/10.1073/pnas.1309353110, 2013.

Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S., and Schellnhuber, H. J.: Very early warning of next El Nino, Proc. Natl Acad. Sci. USA, 111, 2064-66, https://doi.org/10.1073/pnas.1323058111, 2014.

Rafael, C. C., Javier, G. G., Ariza-Villaverde, A. B., Gutierrez, de. Rave. E., and Jimenez-Hornero, F. J.: Can complex networks describe the urban and rural tropospheric O_3 dynamics?, Chemosphere, 230, 59-66, https://doi.org/10.1016/j.chemosphere.2019.05.057, 2019.

https://doi.org/10.5194/esd-2022-9
Preprint. Discussion started: 9 March 2022

Earth System Dynamics
Discussions

Sheehan, P., Cheng, E., English, A., and Sun, F.: China's response to the air pollution shock, Nat Clim Chang, 4, 306-09, https://doi.org/10.1038/nclimate2197, 2014.

Squizzato, S., Masiol, M., Innocente, E., Pecorari, E., Rampazzo, G., and Pavoni, B.: A procedure to assess local and long-range transport contributions to $PM_{2.5}$ and secondary inorganic aerosol, J. Aerosol Sci., 46, 64-76, https://doi.org/10.1016/j.jaerosci.2011.12.001, 2012.

Wang, S., Zhou, C., Wang, Z., Feng, K., and Hubacek, K.: The characteristics and drivers of fine particulate matter (PM2.5) distribution in China, J. Clean. Prod., 142, 1800-09, https://doi.org/10.1016/j.jclepro.2016.11.104, 2017.

Wang, Y., Gozolchiani, A., Ashkenazy, Y., Berezin, Y., Guez, O., and Havlin, S.: Dominant Imprint of Rossby Waves in the Climate Network, Phys. Rev. Lett., 111, https://doi.org/10.1103/PhysRevLett.111.138501, 2013.

Boers, N., Goswami, B., Rheinwalt, A., Bookhagen, B., Hoskins, B., and Kurths, J.: Complex networks reveal global pattern of extreme-rainfall teleconnections, Nature, 566, 373-77, https://doi.org/10.1038/s41586-018-0872-x, 2019.

Cai, S., Wang, Y., Zhao, B., Wang, S., Chang, X., and Hao, J.: The impact of the "Air Pollution Prevention and Control Action Plan" on PM2.5 concentrations in Jing-Jin-Ji region during 2012-2020, Sci. Total Environ., 580, 197-209, https://doi.org/10.1016/j.scitotenv.2016.11.188, 2017.

Castrejon-Pita, A. A., and Read, P. L.: Synchronization in a Pair of Thermally Coupled Rotating Baroclinic Annuli: Understanding Atmospheric Teleconnections in the Laboratory, Phys. Rev. Lett., 104, https://doi.org/10.1103/PhysRevLett.104.204501, 2010.

Ding, A. J., Huang, X., Nie, W., Sun, J. N., Kerminen, V. -M., Petaja, T., Su, H., Cheng, Y. F., Yang, X. -Q., Wang, M. H., Chi, X. G., Wang, J. P., Virkkula, A., Guo, W. D., Yuan, J., Wang, S. Y., Zhang, R. J., Wu, Y. F., Song, Y., Zhu, T., Zilitinkevich, S., Kulmala, M., and Fu, C. B.: Enhanced haze pollution by black carbon in megacities in China, Geophys. Res. Lett., 43, 2873-79, https://doi.org/10.1002/2016GL067745, 2016.

Feldhoff, J. H., Lange, S., Volkholz, J., Donges, J. F., Kurths, J., and Gerstengarbe F-W.: Complex networks for climate model evaluation with application to statistical versus dynamical modeling of South American climate, Clim. Dynam., 44, 1567-81, https://doi.org/10.1007/s00382-014-2182-9, 2015.

Fountalis, I., Bracco, A., and Dovrolis, C.: Spatio-temporal network analysis for studying climate patterns, Clim. Dynam., 42, 879-99, https://doi.org/10.1007/s00382-013-1729-5, 2014.

Gozolchiani, A., Havlin, S., and Yamasaki, K.: Emergence of El Nino as an autonomous component in the climate network, Phys. Rev. Lett., 107, 148501, https://doi.org/10.1103/PhysRevLett.107.148501, 2011.

Guez, O., Gozolchiani, A., Berezin, Y., Brenner, S., and Havlin, S.: Climate network structure evolves with North Atlantic Oscillation phases, Epl, 98, 38006, https://doi.org/10.1209/0295-5075/98/38006, 2012.

Guez, O. C., Gozolchiani, A,, and Havlin, S.: Influence of autocorrelation on the topology of the climate network, Phys. Rev. E, 90, 062814, https://doi.org/10.1103/PhysRevE.90.062814, 2014.

Guo, S., Hu, M., Zamora, M. L., Peng, J., Shang, D., Zheng, J., Du, Z., Wu, Z., Shao, M., Zeng, L., Molina, M. J., and Zhang, R.: Elucidating severe urban haze formation in China, Proc. Natl Acad. Sci. USA, 111, 17373-78, https://doi.org/10.1073/pnas.1419604111, 2014.

Hu, W., Downward, G. S., Reiss, B., Xu, J., Bassig, B. A., Hosgood 3rd, H. D., Zhang, L., Seow, W. J., Wu, G., Chapman, R. S., Tian, L., Wei, F., Vermeulen, R., and Lan, Q.: Personal and Indoor $PM_{2.5}$ Exposure from Burning Solid Fuels in Vented and Unvented Stoves in a Rural Region of China with a High Incidence of Lung Cancer, Environ. Sci. Technol., 48, 8456-64, https://doi.org/10.1021/es502201s, 2014.

Huang, R. J., Zhang, Y., Bozzetti, C., Ho, K. F., Cao, J. J., Han, Y., Daellenbach, K. R., Slowik, J. G., Platt, S. M., Canonaco, F., Zotter, P., Wolf, R., Pieber, S. M., Bruns, E. A., Crippa, M., Ciarelli, G., Piazzalunga, A., Schwikowski, M., Abbaszade, G., Schnelle-Kreis, J., Zimmermann, R., An, Z., Szidat, S., Baltensperger, U., Haddad, I. E., and Prévôt, A. S. H.: High secondary aerosol contribution to particulate pollution during haze events in China, Nature, 514, 218-22, https://doi.org/10.1038/nature13774, 2014.

Huang, X., Ding, A., Wang, Z., Ding, K., Gao, J., Chai, F., and Fu, C.: Amplified transboundary transport of haze by aerosol-boundary layer interaction in China, Nat. Geosci., 13, 428-34, https://doi.org/10.1038/s41561-020-0583-4, 2020.

Li, H., Qi, Y., Li, C., and Liu, X.: Routes and clustering features of PM2.5 spillover within the Jing-Jin-Ji region at multiple timescales identified using complex network-based methods, J. Clean Prod., 209, 1195-205, https://doi.org/10.1016/j.jclepro.2018.10.284, 2019

Liang, C., Duan, F. K., He, K., and Ma, Y.: Review on recent progress in observations, source identifications and countermeasures of $PM_{2.5}$, Environ Int., 86, 150-70, https://doi.org/10.1016/j.envint.2015.10.016, 2016.

Liao, T., Wang, S., Ai, J., Gui, K., Duan, B., Zhao, Q., Zhang, X., Jiang, W., and Sun, Y.: Heavy pollution episodes, transport pathways and potential sources of PM2.5 during the winter of 2013 in Chengdu (China), Sci. Total Environ., 584, 1056-65, https://doi.org/10.1016/j.scitotenv.2017.01.160, 2017.

Liu, H., Fang, C., Zhang, X., Wang, Z., Bao, C., and Li, F.: The effect of natural and anthropogenic factors on haze pollution in Chinese cities: A spatial econometrics approach, J. Clean Prod., 165, 323-33, https://doi.org/10.1016/j.jclepro.2017.07.127, 2017.

Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S., Schellnhuber, H. J.: Improved El Nino forecasting by cooperativity detection, Proc. Natl Acad. Sci. USA, 110, 11742-45, https://doi.org/10.1073/pnas.1309353110, 2013

Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S., and Schellnhuber, H. J.: Very early warning of next El Nino, Proc. Natl Acad. Sci. USA, 111, 2064-66, https://doi.org/10.1073/pnas.1323058111, 2014.

Rafael, C. C., Javier, G. G., Ariza-Villaverde, A. B., Gutierrez, de. Rave. E., Jimenez-Hornero, F. J.: Can complex networks describe the urban and rural tropospheric O_3 dynamics?, Chemosphere, 230, 59-66, https://doi.org/10.1016/j.chemosphere.2019.05.057, 2019.

Sheehan, P., Cheng, E., English, A., and Sun, F.: China's response to the air pollution shock, Nat. Clim. Change, 4, 306-09, https://doi.org/10.1038/nclimate2197, 2014.

Squizzato, S., Masiol, M., Innocente, E., Pecorari, E., Rampazzo, G., and Pavoni, B.: A procedure to assess local and long-range transport contributions to PM2.5 and secondary inorganic aerosol, J. Aerosol Sci., 46, 64-76, https://doi.org/10.1016/j.jaerosci.2011.12.001, 2012.

Wang, S., Zhou, C., Wang, Z., Feng, K., and Hubacek, K.: The characteristics and drivers of fine particulate matter (PM2.5) distribution in China, J. Clean Prod., 142, 1800-09, https://doi.org/10.1016/j.jclepro.2016.11.104, 2017.

Wang, Y., Gozolchiani, A., Ashkenazy, Y., Berezin, Y., Guez, O., and Havlin, S.: Dominant Imprint of Rossby Waves in the Climate Network, Phys. Rev. Lett, 111, https://doi.org/10.1103/PhysRevLett.111.138501, 2013.

Wyatt, M. G., Kravtsov, S., and Tsonis, A. A.: Atlantic Multidecadal Oscillation and Northern Hemisphere's climate variability, Clim. Dyn., 38, 929-49, https://doi.org/10.1007/s00382-011-1071-8, 2012.

Yamasaki, K., Gozolchiani, A., and Havlin, S.: Climate networks around the globe are significantly affected by El Nino, Phys. Rev. Lett., 100, https://doi.org/10.1103/PhysRevLett.100.228501, 2008.

Ying, N., Zhou, D., Chen, Q., Ye, Q., and Han, Z.: Long-term link detection in the CO2 concentration climate network, J. Clean Prod., 208, 1403-08, https://doi.org/10.1016/j.jclepro.2018.10.093, 2019.

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

Zhang, Y. L., and Cao, F.: Fine particulate matter ($PM_{2.5}$) in China at a city level, Sci. Rep., 5, https://doi.org/10.1038/srep14884, 2015

Zhang, Y., Chen, D., Fan, J., Havlin, S., Chen, X.: Correlation and scaling behaviors of fine particulate matter ($PM_{2.5}$) concentration in China, Epl, 122, https://doi.org/10.1209/0295-5075/122/58003, 2018.

Zhang, Y., Fan, J., Chen, X., Ashkenazy, Y., and Havlin, S.: Significant Impact of Rossby Waves on Air Pollution Detected by Network Analysis, Geophys. Res. Lett., 46, 12476-85, https://doi.org/10.1029/2019GL084649, 2019.

Zhou, D., Gozolchiani, A., Ashkenazy, Y., and Havlin, S.: Teleconnection Paths via Climate Network Direct Link Detection, Phys. Rev. Lett., 115, 268501, https://doi.org/10.1103/PhysRevLett.115.268501, 2015.

Wyatt, M. G., Kravtsov, S., and Tsonis, A. A.: Atlantic Multidecadal Oscillation and Northern Hemisphere's climate variability, Clim. Dynam., 38, 929-49, https://doi.org/10.1007/s00382-011-1071-8, 2012.

Xu, B., and Lin, B.: Regional differences of pollution emissions in China: contributing factors and mitigation strategies, J. Clean. Prod., 112, 1454-63, https://doi.org/10.1016/j.jclepro.2015.03.067, 2016.

Yamasaki, K., Gozolchiani, A., and Havlin, S.: Climate networks around the globe are significantly affected by El Nino, Phys. Rev. Lett., 100, https://doi.org/10.1103/PhysRevLett.100.228501, 2008.

Ying, N., Zhou, D., Chen, Q., Ye, Q., and Han, Z.: Long-term link detection in the CO2 concentration climate network, J. Clean. Prod., 208, 1403-08, 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 CO_2 and Temperature Multilayer Climate Network, Geophys. Res. Lett., 47, https://doi.org/10.1029/2019GL086507, 2020.

Zhang, Y. L., and Cao, F.: Fine particulate matter (PM2.5) in China at a city level, Sci. Rep., 5, https://doi.org/10.1038/srep14884, 2015.

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

Zhang, L., Liu, L., Zhao, Y., Gong, S., Zhang, X., Henze, D. K., Capps, S. L., Fu, T. M., Zhang, Q., and Wang, Y.: Source attribution of particulate matter pollution over North China with the adjoint method, Environ. Res. Lett., 10, https://doi.org/10.1088/1748-9326/10/8/084011, 2015

Zhang, Y., Fan, J., Chen, X., Ashkenazy, Y., and Havlin, S.: Significant Impact of Rossby Waves on Air Pollution Detected by Network Analysis, Geophys. Res. Lett., 46, 12476-85, https://doi.org/10.1029/2019GL084649, 2019.

Zhou, D., Gozolchiani, A., Ashkenazy, Y., and Havlin, S.: Teleconnection Paths via Climate Network Direct Link Detection, Phys. Rev. Lett., 115, 268501, https://doi.org/10.1103/PhysRevLett.115.268501, 2015.