Reply to the Reviewer

We thank the referee for the thoughtful and comprehensive suggestions and their overall positive attitude toward our manuscript. In the following we response on a point-by-point basis. Please note that the reviewer's comments are highlighted in red text color represents the revised text.

REVIEWER #1

Note that the red text color represents the modified text.

Reviewer Comments:

The manuscript investigates the transport of PM2.5 concentrations and the potential clustering of 336 cities in China over a 5-year period based on a complex network approach. Their results indicate that the probability density functions of the degrees, weighted degrees, and the lengths of links follow the power-law decay. In addition, the distributions of high-weighted degrees are aligned with high PM2.5 concentrations. This implies that high pollution in cities is not only caused by local emissions but also associated with transportation from other cities. They suggest that the Beijing-Tianjin-Hebei-Henan-Shandong (BTHHS) cluster is a key region to control pollution levels in China since this area exports most of the PM2.5 pollution to other cities. Thus, their discovery can help to identify the optimal collection of cities to take the same measures to control air pollution. Overall, the manuscript is interesting and useful for the implementation of air pollution control measures in China. The paper is well written and is publishable after they address the following comments:

Response: We thank the referee for the positive assessment of our paper.

Comments:

1. Section 2, Line 75. Please mention the time resolution of the PM2.5 concentrations. Are they hourly or daily?

Response:

We thank the referee for this kind and valuable comment. We have added the time resolution in the revised version.

Before

"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."

After:

"The daily $PM_{2.5}$ concentrations data for 336 cities over China from 1 January 2015 to 31 December 2019 are used in this study."

2. Is the word "for rea" in line 145 of the caption in Figure 2 a spelling mistake? Please check the text

carefully for similar errors?

Response:

We thank the referee for this kind and valuable comment. We have corrected the spelling language mistakes in the entire manuscript.

Before

"**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."

After

"**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 real (blue) and shuffled (red) data. (c), (d) Same as (a), (b) but for negative links."

3. "On-Page 7, line 170, "The average path length is 4.61 and 3.15 for the original and shuffled network, indicating that cities transport the PM2.5 concentrations to other cities crossed almost three other cities. PM2.5 cities have a higher clustering coefficient and lower average path length, compared with the shuffled network", but the average path length of the original network is 4.61, which is greater than 3.15 of the shuffled network, which is contrary to the description in the text. Please check carefully?

Response:

We thank the referee for raising this. We have corrected it in the revised version.

Before

"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."

After

"The average path length is 3.15 and 4.61 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."

*4. "*Page 7 Lines 175, Is the word "short distances (<1000 km) a mistake " Please check it*?*

Response:

We thank the referee for raising this. The word "short distances (<1000 km) " is a mistake and we have modified it in the revised version.

Before

"Moreover, these links are mainly short distances (⩽1000 km), whereas long distances (>1000 km) show few connections."

After

"Moreover, these links are mainly short distances (⩽20000 km), whereas long distances (>20000 km) show few connections."

5. The part before Section 3.2 on Page 8 refers to the pictures in the wrong order, e.g. Figure 6 in line 190 should actually be Figure 5. Figure 7 in line 200 should be Figure 6. Please check the text carefully.

Response:

We thank the referee for this kind and valuable comment. We have modified it in the revised version.

6. The title of Figure 6 is wrong, please check the typo in the full text carefully.

Response:

We thank the referee for this kind and valuable comment. We have modified it in the revised version.

Before

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

After

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

7. The Figures in the manuscript are not clear. Please provide a clear version.

Response:

We thank the referee for this kind and valuable comment. We have modified the Figures it in the revised version.

8. Please explain the connections and differences between others and this research in the Summary and discussion. And I also suggest elaborating on the limitations of the study in the summary and discussion.

Response:

We thank the referee for this kind and valuable comment. We have added them in the Summary and discussion in the revised version.

Before:

"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."

After:

"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 a reference and basis for the development of effective regulatory policies for government to improve air quality. Previous researchers have demonstrated that the accumulated pollutants in the NCP can transport the pollution through the strong wind to the YRD based on traditional model simulation, which is similar to our study. We also observed links that transported from the BTHHS to the YRD regions show a 1- or 2-day time lag. The result is consistency with previous studies obtained from the WRF-Chem model. Hence, complex network methodologies are useful for the studies of the transport and cluster of air pollutants in faster and more economic ways. Furthermore, they are also potential in the studies of other air pollutants such as ozone, NOx, and so on.

In addition, the study have some limitations. The relations between $PM_{2.5}$ cities have been measured based on the lagged correlations, which have yielded useful results. However, the peak of cross-correlation in a correlogram may be spurious due to serial autocorrelation within each time series, which is another common feature in geophysical time series. Furthermore, the results cannot reveal causal relationships, which may suffer from problems related to interpretability."

9. *There are many grammatical mistakes in the article, so it is recommended to modify them carefully. Please see my more specific comments on the marked manuscript (attached with the comment file).*

Response:

We thank the referee for this kind and valuable comment. We have corrected the language mistakes in

the entire manuscript in the revised version.

Before:

"Complex networks analysis of $PM_{2.5}$: transport and clustering"

"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."

"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)."

"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."

"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."

"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. "

"$W_{i,j}^{pos}$ 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."

"It is found that the degrees, weighted degrees, and edge lengths conform to power-law distributions."

After:

"Complex network analysis of fine particulate matter ($PM_{2.5}$): transport and clustering"

"Here complex network theory has been applied to reveal the transport patterns and cooperative regions of fine particulate matter ($PM_{2.5}$) over China over from 2015 to 2019. The results show that the degrees, weighted degrees, and edge lengths of $PM_{2.5}$ cities follow power-law distributions. We find that the cities in the Beijing-Tianjin-Hebei-Henan-Shandong (BTHHS) region have a strong ability to export $PM_{2.5}$ 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, all cities have been divided into nine clusters according to their spatial correlations. We suggest that the cities in the same clusters should be regarded as a whole to control the level of air pollution. This approach, beyond doubt, is certainly also applicable to the studies of other air pollutants such as ozone, NOx, and so on."

"The Earth system 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 more 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)."

"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 as well as the transport distances have not been fully understood; furthermore, the traditional approaches adopted in the above studies dis not fully consider the nonlinear transport processes between cities."

"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 to remove the seasonal cycle."

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

"$W_{i,j}^{pos}$ 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 and the similarity of weather patterns (Liu et al., 2022)."

"It is found that the degrees, weighted degrees, and edge lengths conform to power-law distributions which is associated with some climate and weather phenomena such as the tropical circulations and cyclones (Pierrehumbert, 1986)"

■ **End of response to Reviewer #1**

REVIEWER #2

Using the complex network theory, the authors studied the particulate matter (PM2.5) transport pattern and routes around China in a more efficient way. They showed evidently that PM2.5 can transport from Beijing-Tianjin-Hebei-Henan-Shandong (BTHHS) region to Yangtze River Delta (YRD) region with one-or two-day time lags, and then they divide 284 cities in China into 9 clusters according to their synchronicity characteristics. This work can give us some advice on inter-city cooperation governance to solve the haze pollution problem, especially in winter when the pollution transport is the most severe. After reading this paper, I have some questions and advice as follows:

Response: We thank the referee for the positive assessment of our paper.

Comments:

1. *In the introduction section, line 50, the authors mentioned that there are considerable discrepancies in the current studies of PM2.5 transmission in different cities/regions during different air pollution periods. I would suggest the authors add one or two examples here to better introduce the "discrepancies".*

Response:

We thank the referee for this kind and valuable comment. We have added them in the revised version.

Before

"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."

After:

"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. For example, the transport from BTH region to the YRD is significant during the hazing periods (Huang et al., 2020). High $PM_{2.5}$ in the southwest and south of Beijing is related to the $PM_{2.5}$ transmission in Baoding and Hengshui in Hebei Province, and Dezhou, Liaocheng, Heze, Jining, and Zaozhuang in Shandong Province (Li et al., 2015). 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."

2. *In the introduction section, I would suggest the authors add a few more sentences to show why complex network analyses are important and should be used in the analysis. Compared to the traditional approaches, what are the advantages of the complex network analysis?*

Response:

We thank the referee for this kind and valuable comment. We have added them in the revised version.

Before

"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).

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."

After

"Methods are required that help to unveil the transport processes at the national scale. Also, it is important to quantify their spatial and temporary interactions 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)."

"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. In addition, 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."

3. In line 82, what do you mean by "eliminate the effects of autocorrelations in the records"? Do you mean Eqs. (2) and (3)?

Response:

We thank the referee for this kind and valuable comment. The eliminate the effects of autocorrelations in the records In line 82 is not mean *Eqs. (2) and (3)*. We have modified it in the revised version.

Before

"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."

After

"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 to remove the seasonal cycle."

4. In line 128, I guess only zero values indicate that the node is isolated, right?

Response:

We thank the referee for raising this. We have modified according to the comments.

Before

"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."

After

"Nodes with higher values in the network indicate a larger amount of connection with other nodes, whereas zero values indicate that the node is isolated."

5. In line 159, what is the definition of the "clustering coefficient"?

Response:

We thank the referee for this kind and valuable comment. We have added the definition in the revised version.

Before

"The clustering coefficient, which indicates the degree of connection of the network, is 0.46."

After

"The clustering coefficient measures the probability that the adjacent nodes of a node are connected. If one city has a high clustering coefficient, there are close connections between its neighbours. In this paper, the clustering coefficient is 0.46."

6. In line 191, it should be figure 5.

Response:

We thank the referee for this kind and valuable comment. We have corrected the mistakes in the revised version.

7. In line 200, it should be figure 6.

Response:

We thank the referee for this kind and valuable comment. We have corrected the mistakes in the revised version.

8. The figure caption for Fig. 6 is not correct.

Response:

We thank the referee for this kind and valuable comment. We have corrected the mistakes in the revised version.

Before

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

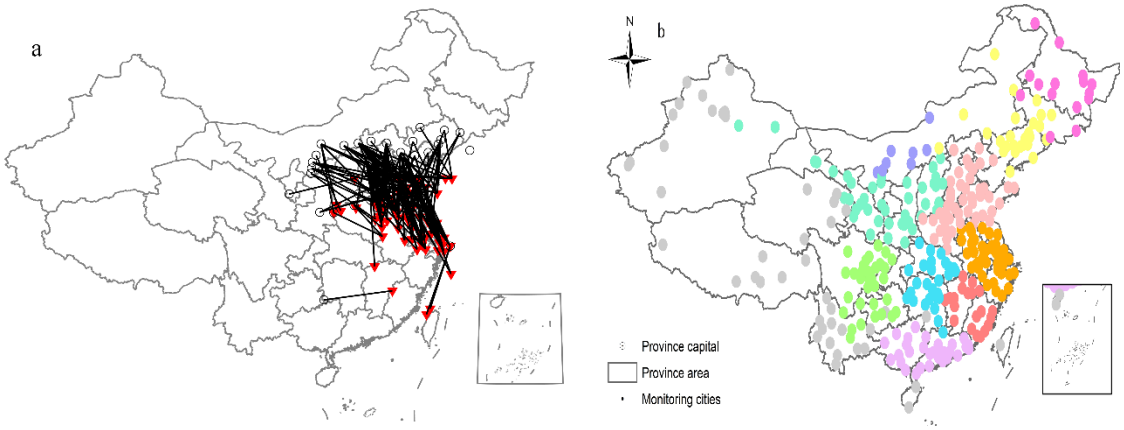
After

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

9. Figure 7(a) is not very clear, I would suggest the authors improve the resolution of the figure?

Response:

We thank the referee for this kind and valuable comment. We have modified the Figures it in the revised version.



10. In line 134, "the weight if node i and j" should be "the weight of node i and j"?.

Response:

We thank the referee for this kind and valuable comment. We have corrected the mistakes in the revised version.

Before:

"where k_i, k_j *is the weight if node i and j,"*

After

"where k_i, k_j *is the weight of node i and j,"*

■ End of response to Reviewer #2