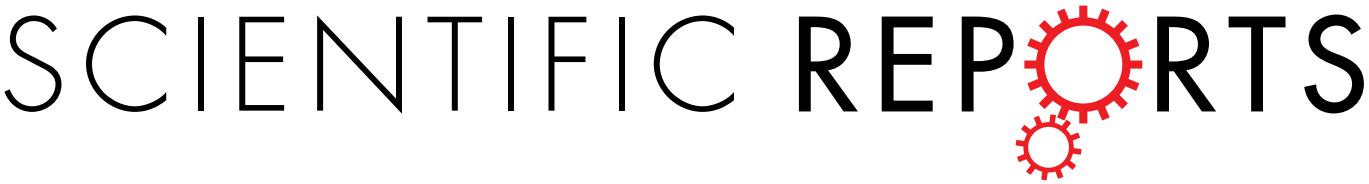

OPEN

Higher-order Network Analysis of Fine Particulate Matter ($PM_{2.5}$) Transport in China at City Level

Received: 26 July 2017
Accepted: 25 September 2017
Published online: 16 October 2017

Yufang Wang[1,3], Haiyan Wang[2], Shuhua Chang[3] & Maoxing Liu[4]

Specification of $PM_{2.5}$ transmission characteristics is important for pollution control and policymaking. We apply higher-order organization of complex networks to identify major potential $PM_{2.5}$ contributors and $PM_{2.5}$ transport pathways of a network of 189 cities in China. The network we create in this paper consists of major cities in China and contains information on meteorological conditions of wind speed and wind direction, data on geographic distance, mountains, and $PM_{2.5}$ concentrations. We aim to reveal $PM_{2.5}$ mobility between cities in China. Two major conclusions are revealed through motif analysis of complex networks. First, major potential $PM_{2.5}$ pollution contributors are identified for each cluster by one motif, which reflects movements from source to target. Second, transport pathways of $PM_{2.5}$ are revealed by another motif, which reflects transmission routes. To our knowledge, this is the first work to apply higher-order network analysis to study $PM_{2.5}$ transport.

Accompanying the world's fastest-growing industrialization and the consequent large amount of vehicle exhaust, China's increasing occurrences of haze, especially $PM_{2.5}$ (particulate matter smaller than 2.5 μm), have been linked to de\creased visibility, negative effects on human health, and influence on global climate. Air pollution has been one of the world's most important eco-environmental problems. In 2012, a new ambient air quality standard (GB 3095–2012) was set by the Chinese Environmental Protection Agency (EPA), which adds $PM_{2.5}$ into the existing list of regularly monitored species. $PM_{2.5}$ originates from many sources, such as road dust, vehicle exhaust, biomass burning, industrial emission and agriculture activities, as well as from regionally transported aerosols.

Regionally transported aerosols are an important factor for $PM_{2.5}$ pollution[1–6]. There are a number of studies on regional transport for $PM_{2.5}$. In[2], it was found that the air quality of Shanghai is largely influenced by the air masses from the north, east and west directions, accounting for 44.8%, 30.4%, and 24.8% of all the air masses respectively. In[3], the contribution of regional transport to $PM_{2.5}$ was estimated in Lingcheng on the North China Plain. The $PM_{2.5}$ from regional transport contributed 31.6% of the $PM_{2.5}$ concentrations, with only 15.4% from the local emissions.

It is debatable how far $PM_{2.5}$ can spread. A number of research works have studied $PM_{2.5}$ transport in local, regional, or long-range scale[1–8]. In the existing works on $PM_{2.5}$ transmission, it's unclear how "local," "regional," and "long-range" transport are defined and distinguished. Relative geographic distances are indispensable factors for determining the pollution level. In this paper, if cities are separated by more than a certain distance, we assume their $PM_{2.5}$ has no influence on each other.

In addition, the pollution level is highly influenced by meteorological conditions such as wind speed and wind direction[9–14], which dramatically influence the diffusion, accumulation, and transport of air pollutants[15,16]. Generally, greater wind speed leads to stronger turbulence, resulting in more favorable dispersion conditions for pollutants[17]. Wind direction significantly affects $PM_{2.5}$ transport because of the spatial distribution of pollution sources and air pollutants' transportation[18].

Mountains between cities are also a major factor influencing $PM_{2.5}$ concentration. Where mountains exist, air does not flow between the cities. As depicted in[19], Beijing is surrounded by mountains in three directions, and polluted air can not be easily expelled in that special geographical environment. Chongqing lies in a mountainous area of China. Influenced by the specific topographic condition, Chongqing is in the region of lowest wind speed

[1]Department of Statistics, Tianjin University of Finance and Economics, Tianjin, 300222, China. [2]School of Mathematical and Natural Sciences, Arizona State University, AZ, 85069, USA. [3]Coordinated Innovation Center for Computable Modeling in Management Science, Tianjin University of Finance and Economics, Tianjin, 300222, China. [4]Department of Mathematics, North University of China, Shanxi, 030051, China. Correspondence and requests for materials should be addressed to Y.W. (email: wangyufangminshan@163.com) or S.C. (email: shuhua55@126.com)

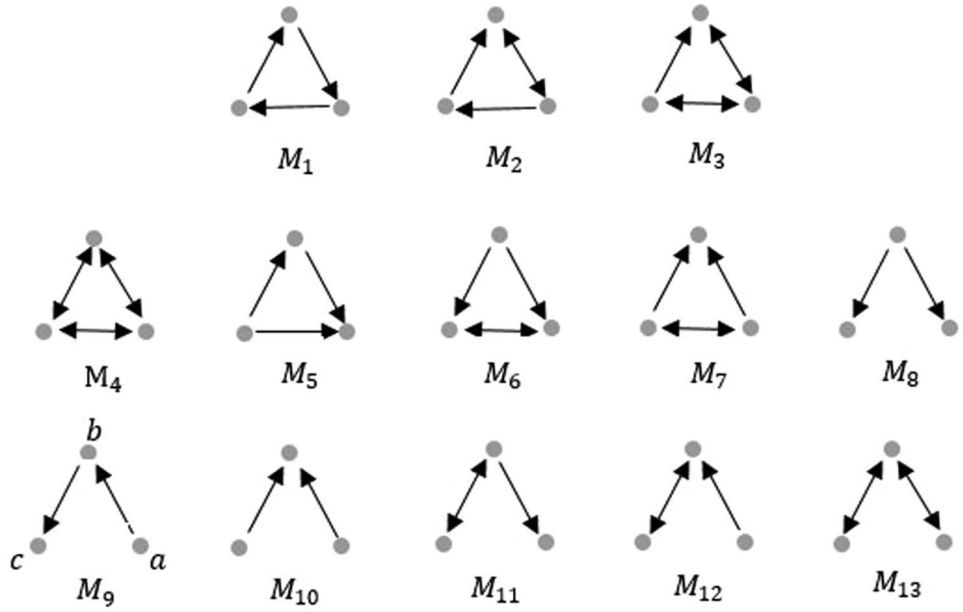

Figure 1. Triangular motifs.

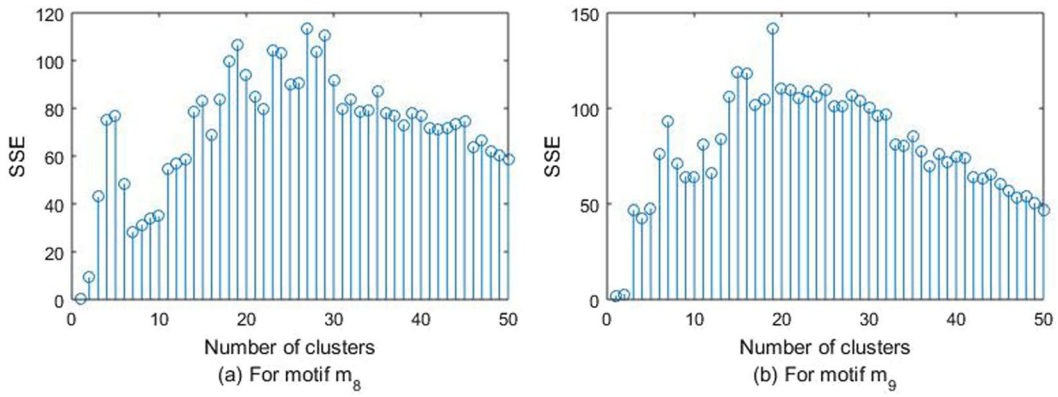

Figure 2. SSE varies with the number of clusters (K).

over China. In this paper, we considered thirteen major mountains in China to build a city-network, in which the $PM_{2.5}$ of any two cities has no reciprocal influence, if there is a mountain between them.

$PM_{2.5}$ has significant spatial and temporal characteristics in China[5,20]. Regionally, $PM_{2.5}$ concentrations are generally higher in northern regions than in southern regions and tend to be higher in inland regions than in the coastal regions. Seasonally, the level of $PM_{2.5}$ is highest in winter and lowest in summer. In wintertime, except for emissions from fossil fuel combustion and biomass burning, meteorological conditions largely contribute to the high concentrations of $PM_{2.5}$. More frequent occurrences of stagnant weather, less rainfall, and low temperature are not good for pollution dispersion. Therefore, we choose January of 2016 for this research.

Presently, most of the methods for studying $PM_{2.5}$ can be divided into two groups: deterministic and statistical approaches. Deterministic methods[21,22] mainly focus on the formation mechanism of $PM_{2.5}$ from the respective of meteorological-chemistry. In comparison, the statistical approaches, such as linear regression models[23,24], neural networks[25], and nonlinear regression models[26,27], aim to detect certain correlated patterns between air quality data and various selected predictors, thereby predicting the pollutant concentrations in future. Each approach addresses problems from different perspectives.

Network analysis is an important and global method to study relationships between objects[28,29], that can be organized into a graph. In graph theory, objects are presented as nodes and relationships between two nodes are presented as edges. Network analysis can group nodes into clusters whose members have certain common characteristics. In general, there are more connections between the nodes within a cluster than between the nodes in different clusters. Yang *et al.*[20] have applied the network tool in studying $PM_{2.5}$. In[20], the correlation between two $PM_{2.5}$ emission profiles are investigated, and then network analysis is applied to cluster cities in China. The network structure in their work is depicted at the level of individual nodes and edges, which are considered to be lower-order connectivity patterns of complex networks.

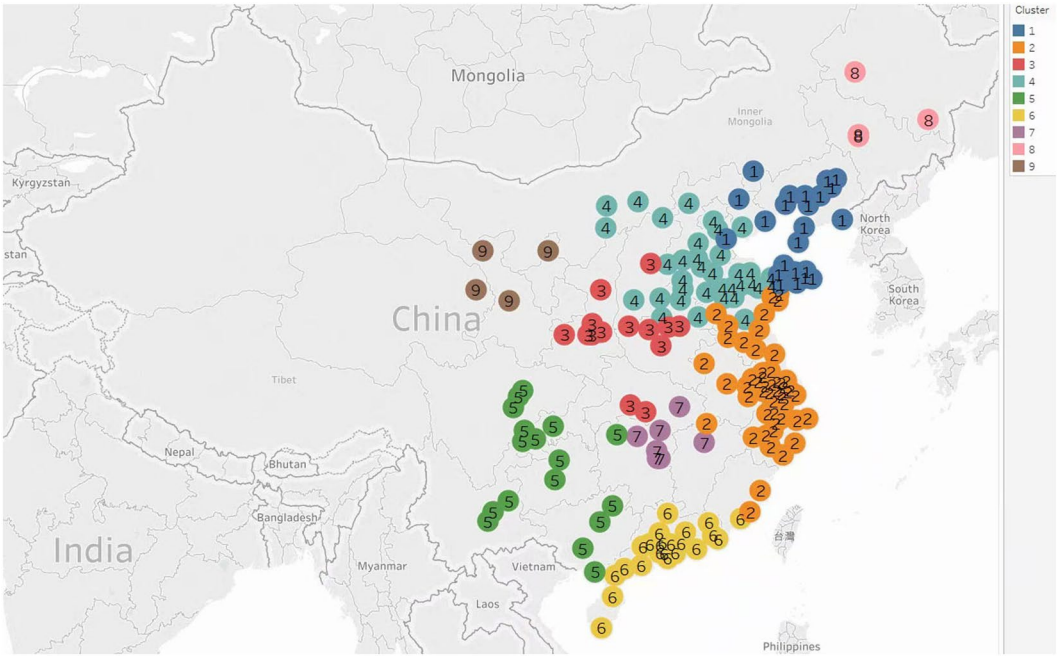

Figure 3. Nine clusters obtained by m_8-motif spectral clustering algorithm. Tableau Public 10.3 (https://public.tableau.com/) was used to create the map.

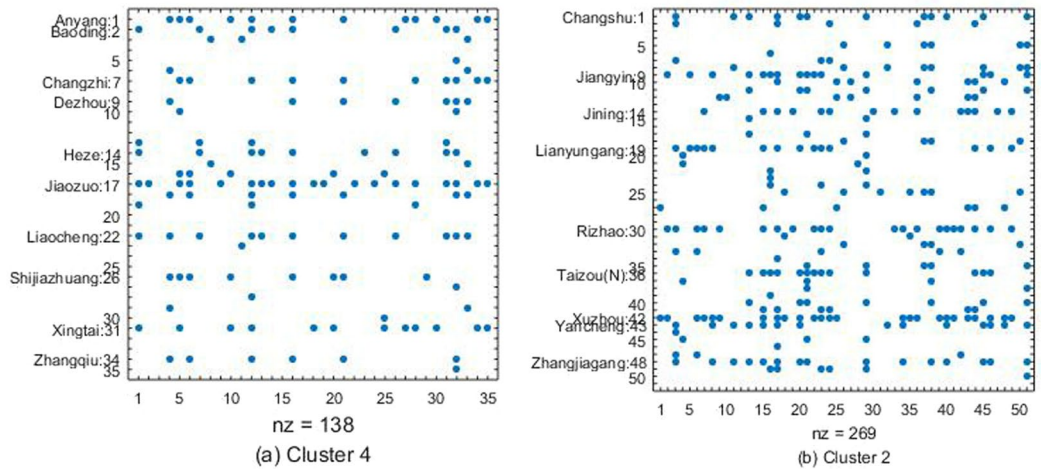

Figure 4. Spy plot of two representative clusters of Fig. 3 in January of 2016. The major potential $PM_{2.5}$ contributors in each cluster are marked in the plot. The number order in the spy plot is the ID in Supplementary Table S1.

Using higher-order organizations of complex networks as the basic building blocks of complex network can help us understand the fundamental structures of complex systems. The most common higher-order organization of complex networks is network motifs[30,31]. In particular, three-node motifs (Fig. 1) appear frequently in networks. In air traffic patterns, M_8–M_{13} are fundamental units of network. M_7–M_7 are structural hubs in the brain. A generalized framework[32] is developed for clustering networks based on higher-order connectivity patterns. In[32], different network motifs can result in different higher-order clusters. Motifs (M_5, M_6, or M_8) depict differing hierarchical flow between species in the Florida Bay ecosystem food web.

In this paper, we apply motif-based higher-order organization of complex networks to study $PM_{2.5}$ transmission and analyze structures in each city-cluster by motif analysis. Specifically, this paper aims to cluster 189 cities in China and identify major potential $PM_{2.5}$ contributors and regional transport pathways in each cluster. We first build an adjacency matrix of the complex network, combining geographic distance, wind speed, wind direction, mountains, and $PM_{2.5}$ concentration. Then the cities are clustered by using the motif-based higher-order organization of complex networks. Then, we apply motif analysis to identify the structure in each cluster.

Figure 5. 20 clusters obtained by m_9-motif spectral clustering algorithm. Tableau Public 10.3 (https://public.tableau.com/) was used to create the map.

To our best knowledge, this is the first work to apply higher-order organization of complex networks to $PM_{2.5}$ transmission. Network analysis not only gives a global view to examine $PM_{2.5}$ transmission, but also reveals an internal structure of pollution between cities in China. This research can provide valuable information for the Chinese government to implement air pollution control.

Results

In Figs 3 and 5, a circle with its cluster number represents a city. The cities in a cluster are more densely connected with each other but sparsely connected with the cities in other clusters. In accordance with specific characteristics of $PM_{2.5}$ emissions in China, a cluster will usually consist of cities in the same province or close geographical proximity.

Clustering 189 cities into groups and identifying major potential pollution contributors in each cluster by motif m_8. Motif m_8 is chosen to identify major potential pollution contributors in each cluster. After we perform a higher-order spectral clustering algorithm, three connected components and some isolated points are included (see the Supplementary Table S1) in the $m8$-motif adjacency matrix of 189 cities. The largest connected component contains 170 cities, which form seven clusters. The number of total clusters $K = 7$ makes SSE relatively smaller, which can be seen from Fig. 2(a). Here SSE is defined at the end of this paper. The other two connected components compose cluster 8 (Changchun, Daqing, Jilin, and Mudanjiang) and cluster 9 (Jinchang, Lanzhou, Xining, and Yinchuan) respectively. Thus, nine clusters (see Fig. 3 and Supplementary Table S1) are obtained by a motif m_8-based spectral clustering algorithm. The remaining 11 isolated cities can be explained by their geographic characteristics. Kelamayi, Wulumuqi, and Kuerler are located in the Mongolia Autonomous Region, and Jiayuguan is near the Mongolia Autonomous Region. Lhasa is a plateau area. Qiqihaer and Haerbin are in the most northerly province of China. Two representative clusters, cluster 2 (including Shanghai) and cluster 4 (including Beijing), are illustrated below.

In cluster 4, there are 31 cities, covering most of northern China. They are shown in Fig. 4(a) and Supplementary Table S1. From the spy plots, we can observe that some cities of y-axis direction correspond to more dots in the horizontal line, which indicates that they have more out-direction arrow lines than other cities in the network subgraph of the cluster, such as Anyang, Baoding, Jiaozuo, Xingtai, and some cities, which are labeled in the spy plot. They are major potential $PM_{2.5}$ contributors of Cluster 4. This is in agreement with the results of[5,33]. They concluded that the above cities are the heavily haze-affected cities in Beijing-Tianjin-Hebei, and that pollution from Shandong and Henan provinces by regional transport is also an important factor for the $PM_{2.5}$ of North China.

In cluster 2, there are 51 cities, including all the cities from the Yangtze River delta and some of Shandong's coastal cities, as shown in Fig. 4(b) and Supplementary Table S1. Jining, Xuzhou, Yancheng, Zhangjiagang, and some cities that are labeled in the spy plot are the potential $PM_{2.5}$ contributors of cluster 2. Most of the potential $PM_{2.5}$ contributors are in the north part of the cluster; this agrees with the spatial characteristics of $PM_{2.5}$[5,20]. Note

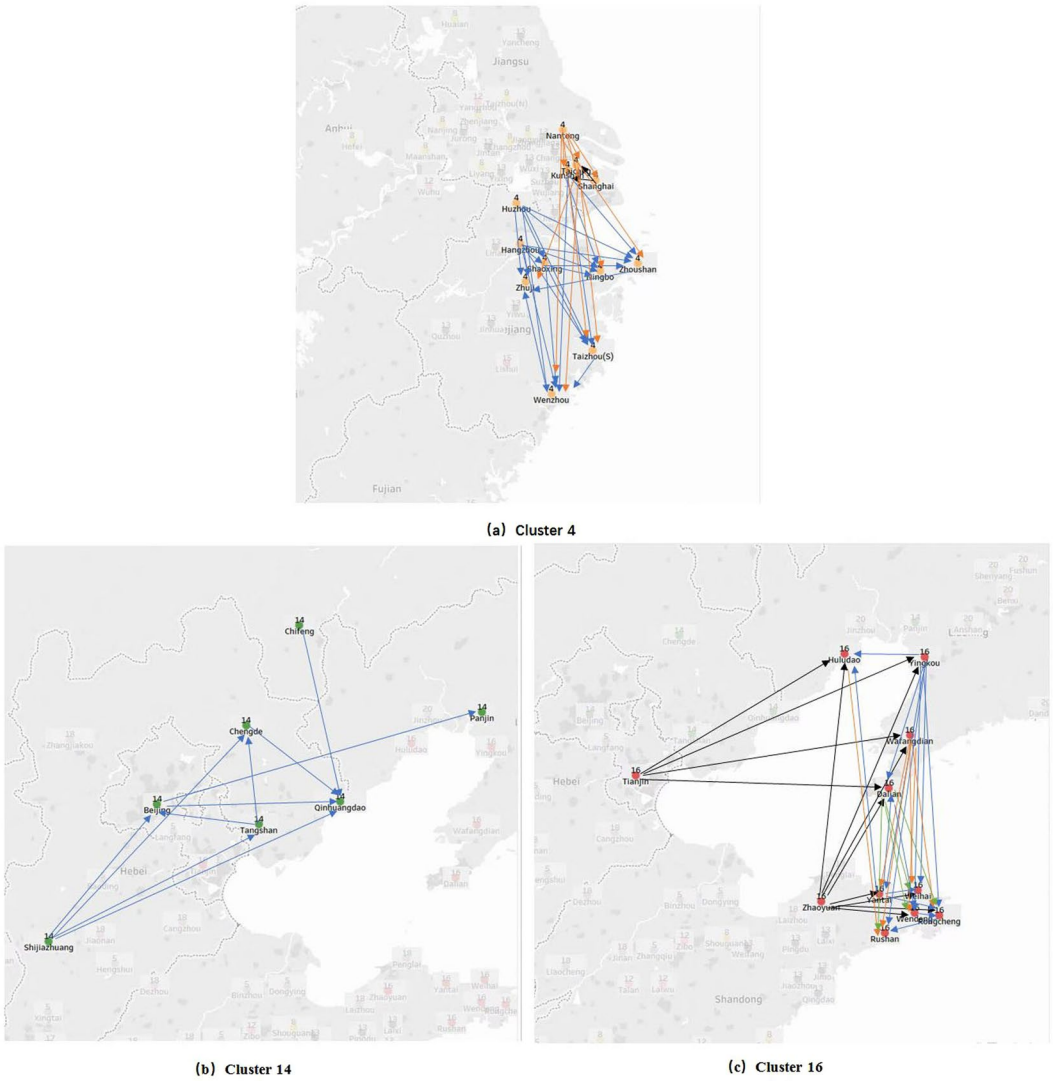

Figure 6. m_9-motif analysis for three representative clusters obtained by motif spectral clustering algorithm based on January of 2016. Tableau Public 10.3 (https://public.tableau.com/) was used to create these maps.

that although Shanghai is a metropolis, it is not a major source of pollution in the cluster. Our conclusion accords with the results of[2].

Clustering 189 cities into groups and identifying transport pathways in each cluster by motif m_9. We choose motif m_9 to identify transport pathways in each cluster. We can see, from Fig. 5, that only 166 cities are shown and they are clustered into 20 groups. The remaining 23 cities are isolated and these isolated cities can also be explained by their geographic characteristics. Some of them are from the Mongolia Autonomous Region, the plateau area, the most northerly part of China, or from Gansu and Ningxia. All the clustering results are listed in Supplementary Table S2.

In Fig. 2, SSE is smaller when the number of total clusters $K = 10$. However, clustering with $K = 10$ leads to more cities in each cluster. It's difficult to see the transport pathways clearly when more cities appear in each cluster. Therefore, we choose $K = 20$ to cluster cities and identify transport pathways in each cluster through motif analysis.

Cluster 4 (including Shanghai), cluster 14 (including Beijing), and cluster 16 (including Tianjin) are shown in Fig. 6. For cluster 4, the $PM_{2.5}$ transport pathway originates from Nantong and Huzhou in northwest to Wenzhou, Taizhou(s), Ningbo and Zhoushan in southeast. Shanghai is also generally downwind of the most developed and polluted YRD region in special meteorological conditions, which accords with[2]. For cluster 14, the main $PM_{2.5}$ transport pathway is from Shijiazhuang to the northeast of the cluster and the detailed $PM_{2.5}$ transport pathway is shown in Fig. 6(b). Shijiazhuang is a key controlling point because of its relative high $PM_{2.5}$ concentration and its location upwind of other cities in the cluster. Beijing's pollution is partly from Shijiazhuang, as described in[5]. Cluster 16 includes Tianjin and cities of Liaotung peninsula, Shangdong peninsula. All of the cities are around Bohai; therefore wind affecting these cities varies frequently and wind directions are not all the same in different cities at the same time. One possible $PM_{2.5}$ transport pathway originates from Tianjin, halfway between

Huludao, Yingkou,Wafangdian and Dalian in Liaotung peninsula, and arrives at Yantai, Weihai and some cities in Shangdong peninsula. Another possible $PM_{2.5}$ transport pathway is from Zhaoyuan to Yantai, Weihai and some cities in Shangdong peninsula, or to Huludao, Yingkou and some cities in Liaotung peninsula.

Discussion

In this paper, higher-order organization of complex network and spectral clustering methods are used to group cities in China. We obtain two major conclusions: specifically, major potential $PM_{2.5}$ contributors and $PM_{2.5}$ transport pathways. Clustering of complex networks often provides a global view of the underlying networks. Through the new clustering method, we presents a new framework to investigate the transmission of PM 2.5 among major cities in China.

In general, statistical methods tend to apply data over a long period. The complex network we use in this paper intends to analyze the relationship of nodes in a certain state and reveals the essential structure of a complex system. We intend to use the clustering method to identify the city-network, to aggregate cities, and to identify major potential $PM_{2.5}$ contributors and transport pathways. As a result, in this study we collect data only for a short period. Specifically, only January data are used; this is justified for several reasons. $PM_{2.5}$'s concentration shows an apparent seasonal pattern. High–frequency and high–concentration $PM_{2.5}$ days usually occur in winter[9,34]. This is mainly due to meteorological conditions. In a short period, some meteorological conditions that affects the $PM_{2.5}$ can be thought to be relatively stable, and this is helpful for simplifying models. This is the main reason we choose data from one month for our study. As a result, less important factors such as temperatures and atmospheric pressure can be ignored; thus, more important factors can be considered in a relatively simple model. As in[1,9], we can ignore atmospheric pressure and temperature and consider major meteorological factors such as wind speed and wind direction that influence $PM_{2.5}$ concentration in this paper. Wind speed and wind direction vary in each city constantly and they drive air pollution transport between cities. In constructing the adjacency matrix for the network, we choose the monthly prevailing wind direction and monthly average wind speed. This approach allows us to better describe the fact of the frequent change of wind speed and direction in the present study. We believe the data from January suffice for identifying major potential pollution contributors and pollution transport pathways.

We assume that $PM_{2.5}$ in city i has no influence on city j, if the straight-line geographical distance of the two cities is more than 500 kilometers. $PM_{2.5}$ flow will dissipate during the propagation. In addition, when the straight-line geographical distance of the two cities is more than 200 kilometers, we assume $PM_{2.5}$ in city i has influence on city j, only if city i's $PM_{2.5}$ concentration is higher than city j's at certain extent. We use "500 kilometers" and "200 kilometers" as the dividing values, mainly inspired by[9], which concluded that aerosol nucleation and growth processes occur on the regional (several hundred kilometers) to urban (less than 100 kilometers) scales. Although there are many research works on regional transport of $PM_{2.5}$, it is an open question as to how far $PM_{2.5}$ can travel. In addition, we believe that many physical, biological and social models, for example[35–44], could be used for estimating/predicting the long range transport of $PM_{2.5}$.

Because meteorological conditions are complex, additional factors affecting $PM_{2.5}$ should be considered in future studies. In addition, $PM_{2.5}$ in city i has influence on other cities and the incidence should be inversely proportional to the geographic distance. Therefore, it is more important that weighted complex networks should be considered in future.

In this paper, we consider some meteorological conditions and geographical data to cluster cities in China and identify the inner structure of each cluster. However, $PM_{2.5}$ transmission between cities is a very complex issue. Economy, population, in-vehicle commuting, and many others are also indispensable factors that influence $PM_{2.5}$ transmission. More economic factors and social factors will be considered in our future work.

Data and Methods

Data. In this paper, we focus on the top 189 pollution–monitoring cities in China's mainland, which cover all 34 provincial-level regions of China. The most polluted and the major cities are all included, such as Beijing, Shanghai and Guangzhou.

Data from January 2016 are used in this work to identify major $PM_{2.5}$ pollution contributors and transport pathways in each cluster. The data that we collect in this paper are as follows: (1) $PM_{2.5}$ monthly average concentration is calculated based on ground air quality monitoring data from China's National Environmental Monitoring. (2) The geo-location information in the forms of latitude and longitude of 189 cities are from Google Earth. (3) Thirteen major mountains with high altitudes in China (see Supplementary Table S3) are included in this paper. (4) Wind speed and wind direction data is from the China Meteorological Administration. Wind directions are classified into eight directions (e.g., N, E, S, W, W-S, E-S, W-N, E-N), We use the monthly prevailing wind direction of each city in January. The scaling of wind speed is based on the Jenks Natural Breaks Classification method[10]. Wind speed(ws) is divided into eight levels: $ws \leq 0.7m/s$ (Level-1), $0.7<ws \leq 1.1m/s$ (Level-2), $1.1 < ws \leq 1.6m/s$ (Level-3), $1.6<ws \leq 2.1m/s$ (Level-4), $2.1<ws \leq 2.7m/s$ (Level-5), $2.7<ws \leq 3.4m/s$ (Level-6), $3.4<ws \leq 4.4m/s$ (Level-7) and $ws>4.4m/s$ (Level-8). We use the monthly average wind speed. Here "monthly average" means the arithmetic average of the mean concentration levels or mean wind speed of each day in a calendar month.

Motif-based higher-order spectral clustering algorithm. The motif-based higher-order spectral clustering algorithm in the supplementary materials of[32] unifies motif analysis[30] and k-means spectral clustering[45] to reveal new organizational patterns and modules in complex systems. We use the method to cluster 189 cities in China and identify major potential $PM_{2.5}$ contributors and $PM_{2.5}$ transport pathways in clusters. The major steps are listed below.

(1) Building the adjacency matrix A of the network and choosing motif M of interest. Specifically, in this

paper, matrix A is built as follows: m_8 and m_9 are chosen as the building modules to reveal the essential structures of the complex network; m_8 reflects reveal the relationship between source and victims; m_9 reflects the transmission route.

(2) Computing the motif adjacency matrix W_M, whose entry $W_M(i, j)$ equals the number of the motif instances of motif M with node i and node j.

(3) Clustering 189 cities by spectral clustering algorithm through the motif adjacency matrix W_M.

a) Computing the normalized motif Laplacian $L_M = I - D_M^{-1/2} W_M D_M^{-1/2}$, where D_M is diagonal matrix with $(D_M)_{ii} = \sum_j (W_M)_{ij}$.

b) Forming matrix X, st.$X = [x_1, x_2, \cdots, x_k]$, where $x_1, x_2, \cdots, x_k$ are the k largest eigenvectors of L_M.

c) Calculating matrix Y, whose entry is $Y_{ij} = X_{ij}/(\sum_j X_{ij}^2)^{1/2}$.

d) Taking each row of Y as a point in R^k and cluster the points into k clusters via k-means method[45]. In this paper, the optimal number of cluster(K) is chosen as follows, which is inspired by[46].

e) City j is assigned to cluster j if and only if row j of matrix Y is assigned to cluster j.

(4) Analyzing every cluster using the motifs of (1). This paper applies motif m_8 to analyze major potential contributors. In the social network graph, a node with the largest numbers of edges is commonly considered as source, from which information begins to disperse[47]. After using motif m_8 to cluster 189 cities in the complex network, a city with more out-direction arrow lines shows that it has relatively high $PM_{2.5}$ concentration and it has a high influence ratio on the other cities of the cluster. The city can be regarded as one major $PM_{2.5}$ pollution contributor in the cluster. This can be seen through the spy plot, which illustrates the network structure of the cluster. Motif m_9 helps find $PM_{2.5}$ transport pathway in every cluster. In Fig. 1, m_9 corresponds to the $PM_{2.5}$ flow from city a to city b, then from city b to city c.

Building an Adjacency Matrix. A network can be represented as a matrix, which is called the sociomatrix[44] or adjacency matrix. Suppose the number of nodes is n. Let V and E be the sets of nodes and edges in the network, respectively. Then the adjacency matrix of the network can be expressed by matrix $A \in \{0, 1\}_{nn}$. An entry $A_{ij} \in \{0, 1\}$ denotes whether there is a link between node v_i and node v_j. If node v_i and node v_j are adjacent, then $A_{ij} = 1$. Otherwise, $A_{ij} = 0$. If the network is undirected, the adjacency matrix A is symmetric. However, in some situations, interactions between two different individuals are directional. In Twitter, for example, one user x follows another user y, but user y does not necessarily follow user x. In this case, the follower- followee network is directed and asymmetrical.

Based on $PM_{2.5}$ monthly-average concentration, geographic distance between cities, monthly prevailing wind direction, monthly average wind speed, and mountains between cities (189 cities in China in January 2016), the detailed procedure for building the adjacency matrix is as follows:

(1) Adjacency matrix based on distance (A_1): Based on the latitudes and longitudes of 189 cities, the relative geographic distances are calculated. The entry $A_1(i, j) = 0$, if the relative geographic distance is more than 500 kilometers. Otherwise, $A_1(i, j) = 1$. The assumption is plausible because $PM_{2.5}$ of each city has no effect on another city, if they are distant from each other.

We choose 500 kilometers, because it is an empirical value through numerical simulation. This is in agreement with[9], which found that aerosol nucleation and growth processes occur on the regional (several hundred kilometers) to urban (less than 100 kilometers) scales.

(2) Adjacency matrix based on mountain (A_2): In the planimetric map, a major mountain can be expressed by a line segment through its latitudes and longitudes, which is called a mountain-segment. Therefore the 13 major mountains we considered are depicted by 13 different line segments. Meanwhile, there is a line segment between any two cities, which is called a cities-segment. If the cities-segment between city i and city j has a cross point with any of the 13 mountain-segments, the entry $A_2(i, j) = 0$. Otherwise, $A_2(i, j) = 1$.

(3) Adjacency matrix based on wind (A_3): Wind speed and wind direction jointly affect the propagation of $PM_{2.5}$ flow. Due to the wind direction, the effect from wind on $PM_{2.5}$ transmission is directional. Specifically, $PM_{2.5}$ of city i may flow to city j. But it is possible that $PM_{2.5}$ of city j may not be blown to city i.

We assume city i's $PM_{2.5}$ has no effect on any other cities, if wind level is less than 2. When the speed level is more than 2 (more than 1.1 m/s), the wind direction is a key point, determining whether $PM_{2.5}$ flowing from city i affects city j. Specifically, in the planimetric map, there is a directional line segment from city i to city j. If the angle θ_{ij}, between city i's wind direction and the directional line segment from city i to city j, is less than 90 degree, we think city i's wind can flow to city j.

The overall effect of wind from city i to city j is calculated by $A_3(i, j) = w_i cos(\theta_{ij})$, where w_i is the wind speed of city i.

(4) Adjacency matrix based on $PM_{2.5}$ (A_4): The paper aims to study the major pollution contributors and the pollution transport pathways. And we are particularly interested in that how a city with high $PM_{2.5}$ concentration affects a city with low concentration. Specifically, two situations are considered below.

Situation One: When the geographic distance is less than 200 kilometers, the $PM_{2.5}$ of city i has effect on city j, as long as the $PM_{2.5}$ concentration of it is higher than city j's, then $A_4(i, j) = 1$. Otherwise, $A_4(i, j) = 0$.

Situation Two: $PM_{2.5}$ flow will dissipate during the propagation. Therefore, when the geographic distance is more than 200 kilometers, if and only if city *i*'s $PM_{2.5}$ concentration is $\alpha \times d_{ij}$ higher than city *j*'s, then $A_4(i, j) = 1$. Otherwise, $A_4(i, j) = 0$. Here d_{ij} is the geographic distance between city *i* and city *j* and $\alpha = 0.01$ is an empirical threshold value through numerical simulation. Better α should be considered in future work according to the meteorological condition. $\alpha \times d_{ij}$ is a degree of $PM_{2.5}$ concentration, increasing with the geographic distance d_{ij}.

Clustering and motif analysis are based on the adjacency matrix $A = A_1 \circ A_2 \circ A_3 \circ A_4$, where "$\circ$" is the Hadamard (entry-wise) product. Namely, $PM_{2.5}$'s propagation is the combined effects of geographic distance, mountain, wind and $PM_{2.5}$ concentration.

Selecting K. As in[46], the sum of the squared distance between each member of a cluster and its cluster centroid (SSE) is defined as

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} dist(c_i, x)^2,$$

where x is a city; c_i is the centroid of cluster C_i; C_i is the i th cluster (cluster *i*); dist is the the standard Euclidean distance between two cities of Euclidean space. K is the number of clusters, and the optimal value is chosen from 2 to 50, which makes dist smallest. The K larger than 50 has not much meaning for clustering 189 cities, which leads to too-detailed clustering.

References

1. Zheng, G. *et al*. Exploring the severe winter haze in Beijing: the impact of synoptic weather, regional transport and heterogeneous reactions. *Atmospheric Chemistry and Physics* **15**, 2969–2983 (2015).
2. Wang, H. *et al*. Chemical composition of pm 2.5 and meteorological impact among three years in urban Shanghai, China. *Journal of Cleaner Production* **112**, 1302–1311 (2016).
3. Chen, D. *et al*. Estimating the contribution of regional transport to pm 2.5 air pollution in a rural area on the North China plain. *Science of The Total Environment* **583**, 280–291 (2017).
4. Xiong, Y., Zhou, J., Schauer, J. J., Yu, W. & Hu, Y. Seasonal and spatial differences in source contributions to pm 2.5 in Wuhan, China. *Science of the Total Environment* **577**, 155–165 (2017).
5. Zhang, Y. L. & Cao, F. Fine particulate matter (pm2.5) in China at a city level. *Scientific Reports* **5**, 14884 (2015).
6. Liu, J. *et al*. Source apportionment using radiocarbon and organic tracers forpm2. 5 carbonaceous aerosols in Guangzhou, south China: Contrasting local-and regional-scale haze events. *Environmental science & technology* **48**, 12002–12011 (2014).
7. Baker, J. A cluster analysis of long range air transport pathways and associated pollutant concentrations within the UK. *Atmospheric Environment* **44**, 563–571 (2010).
8. Saliba, N. A., Kouyoumdjian, H. & Roumié, M. Effect of local and long-range transport emissions on the elemental composition of pm 10–2.5 and pm 2.5 in Beirut. *Atmospheric Environment* **41**, 6497–6509 (2007).
9. Guo, S. *et al*. Elucidating severe urban haze formation in China. *Proceedings of the National Academy of Sciences* **111**, 17373–17378 (2014).
10. Zhang, B. *et al*. Influences of wind and precipitation on different-sized particulate matter concentrations (pm2.5, pm10, pm2.5–10). Meteorology and Atmospheric Physics 1–10 (2017).
11. Adams, H., Nieuwenhuijsen, M. & Colvile, R. Determinants of fine particle (pm 2.5) personal exposure levels in transport microenvironments, London, UK. *Atmospheric Environment* **35**, 4557–4566 (2001).
12. Guerra, S. *et al*. Effects of wind direction on pm10 and pm2. 5 concentrations in southeast Kansas. Proceedings of the Air & Waste Management Association (2004).
13. Nguyen, M.-V., Park, G.-H. & Lee, B.-K. Correlation analysis of size-resolved airborne particulate matter with classified meteorological conditions. *Meteorology and Atmospheric Physics* **129**, 35–46 (2017).
14. Westervelt, D. *et al*. Quantifying pm 2.5-meteorology sensitivities in a global climate model. *Atmospheric Environment* **142**, 43–56 (2016).
15. Jacob, D. J. & Winner, D. A. Effect of climate change on air quality. *Atmospheric Environment* **43**, 51–63 (2009).
16. Pearce, J. L., Beringer, J., Nicholls, N., Hyndman, R. J. & Tapper, N. J. Quantifying the influence of local meteorology on air quality using generalized additive models. *Atmospheric Environment* **45**, 1328–1336 (2011).
17. Tian, G., Qiao, Z. & Xu, X. Characteristics of particulate matter (pm 10) and its relationship with meteorological factors during 2001–2012 in Beijing. *Environmental Pollution* **192**, 266–274 (2014).
18. Zhou, W., Tie, X., Zhou, G. & Liang, P. Possible effects of climate change of wind on aerosol variation during winter in Shanghai, China. *Particuology* **20**, 80–88 (2015).
19. Yang, F. *et al*. Characteristics of pm 2.5 speciation in representative megacities and across China. *Atmospheric Chemistry and Physics* **11**, 5207–5219 (2011).
20. Yan, S. & Wu, G. Network analysis of fine particulate matter (pm2. 5) emissions in China. *Scientific Reports* **6**, 33227 (2016).
21. Chuang, M.-T., Zhang, Y. & Kang, D. Application of wrf/chem-madrid for real-time air quality forecasting over the southeastern United States. *Atmospheric Environment* **45**, 6241–6250 (2011).
22. Yahya, K., Zhang, Y. & Vukovich, J. M. Real-time air quality forecasting over the southeastern united states using wrf/chem-madrid: Multiple-year assessment and sensitivity studies. *Atmospheric Environment* **92**, 318–338 (2014).
23. Li, C., Hsu, N. C. & Tsay, S.-C. A study on the potential applications of satellite data in air quality monitoring and forecasting. *Atmospheric Environment* **45**, 3663–3675 (2011).
24. Benas, N., Beloconi, A. & Chrysoulakis, N. Estimation of urban pm10 concentration, based on modis and meris/aatsr synergistic observations. *Atmospheric environment* **79**, 448–454 (2013).
25. Mao, X., Shen, T. & Feng, X. Prediction of hourly ground-level pm 2.5 concentrations 3 days in advance using neural networks with satellite data in eastern China. Atmospheric Pollution Research https://doi.org/10.1016/j.apr.2017.04.002 (2017).
26. Emili, E. *et al*. Pm 10 remote sensing from geostationary seviri and polar-orbiting modis sensors over the complex terrain of the european alpine region. *Remote sensing of environment* **114**, 2485–2499 (2010).
27. Tian, J. & Chen, D. A semi-empirical model for predicting hourly ground-level fine particulate matter (pm 2.5) concentration in southern ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sensing of Environment* **114**, 221–229 (2010).
28. Rosvall, M., Esquivel, A. V., Lancichinetti, A., West, J. D. & Lambiotte, R. Memory in network flows and its effects on spreading dynamics and community detection. *Nature communications* **5**, 4630 (2014).
29. Leskovec, J., Lang, K. J., Dasgupta, A. & Mahoney, M. W. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* **6**, 29–123 (2009).

30. Milo, R. *et al*. Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
31. Yavero ğlu, Ö. N. *et al*. Revealing the hidden language of complex networks. *Scientific Reports* **4**, 4547 (2014).
32. Benson, A. R., Gleich, D. F. & Leskovec, J. Higher-order organization of complex networks. *Science* **353**, 163–166 (2016).
33. Huang, R.-J. *et al*. High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* **514**, 218–222 (2014).
34. Yang, J. *et al*. Concentrations and seasonal variation of ambientpm2.5 and associated metals at a typical residential area in Beijing, China. *Bulletin of environmental contamination and toxicology* **94**, 232–239 (2015).
35. Chen, M. H., Wang, L., Sun, S. W., Wang, J. & Xia, C. Y. Evolution of cooperation in the spatial public goods game with adaptive reputation assortment. *Physics Letters A* **380**, 40–47 (2016).
36. Chen, M. H., Wang, L., Wang, J., Sun, S. W. & Xia, C. Y. Impact of individual response strategy on the spatial public goods game within mobile agents. *Applied Mathematics and Computation* **251**, 192–202 (2015).
37. Xia, C. Y., Miao, Q., Wang, J. & Ding, S. Evolution of cooperation in the traveler's dilemma game on two coupled lattices. *Applied Mathematics and Computation* **246**, 389–398 (2014).
38. Sun, G. Q., Wang, C. H. & Wu, Z. Y. Pattern dynamics of a Gierer–Meinhardt model with spatial effects. *Nonlinear Dynamics* **88**, 1385–1396 (2017).
39. Sun, G. Q., Wang, S. L., Ren, Q., Jin, Z. & Wu, Y. P. Effects of time delay and space on herbivore dynamics: linking inducible defenses of plants to herbivore outbreak. *Scientific Reports* **5**, 11246 (2015).
40. Li, L. Patch invasion in a spatial epidemic model. *Applied Mathematics and Computation* **258**, 342–349 (2015).
41. Sun, G. Q. *et al*. Transmission dynamics of cholera: Mathematical modeling and control strategies. *Communications in Nonlinear Science and Numerical Simulation* **45**, 235–244 (2017).
42. Li, L. Monthly Periodic Outbreak Of Hemorrhagic Fever With Renal Syndrome In China. *Journal of Biological Systems* **24**, 519–533 (2016).
43. Sun, G. Q., Jusup, M., Jin, Z., Wang, Y. & Wang, Z. Pattern transitions in spatial epidemics: Mechanisms and emergent properties. *Physics of Life Reviews* **19**, 43–73 (2016).
44. Wasserman, S. & Faust, K. Social network analysis: Methods and applications, vol. 8 (Cambridge university press, 1994).
45. Ng, A. Y., Jordan, M. I. & Weiss, Y. On spectral clustering: Analysis and an algorithm. In Advances in neural information processing systems, 849–856 (2002).
46. Baxter, L. K. & Sacks, J. D. Clustering cities with similar fine particulate matter exposure characteristics based on residential infiltration and in-vehicle commuting factors. *Science of the Total Environment* **470**, 631–638 (2014).
47. Valente, T. W. Network interventions. *Science* **337**, 49–53 (2012).

Acknowledgements

The authors would like to acknowledge the help of Lipeng Song, Ying Yan, John Priniski and Adrian Avram for data collections. This project was supported in part by the Major Research Plan of the National Natural Science Foundation of China (91430108), the National Basic Research Program (2012CB955804), the National Natural Science Foundation of China (11171251, 11771322, 11571324), the National Science Foundation (DMS-1737861), and the Major Program of Tianjin University of Finance and Economics (ZD1302).

Author Contributions

Y.W. and H.W. initiated the idea and built the model. Y.W. and S.C. performed the analysis. M.L. provided partial data. Y.W. prepared the figures, wrote the manuscript. All authors contributed to the scientific discussion and revision of the article.Additional information.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-13614-7.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017