Application of the Complex Network in the $PM_{2.5}$ Analysis of Air Pollution in Regions

XIAO Qin, Lu Yuting
(School of Sciences, Shanghai Institute of Technology, Shanghai 201418, China)

Abstract:The complex network was used to investigate $PM_{2.5}$ of the air pollution. The relevant data of $PM_{2.5}$ were analyzed with correlation and the complex network in regions in china was established. Through the study of the degree, community structure and motif, the results showed that the main polluted cities in China could be effectively analyzed by this method, and the air polluted cities having cluster phenomena needed to be treated as a whole, which was consistent with real conditions. Because of the fluidity of the air, this research provides guidance for analyzing the agglomeration of polluted cities.
Keywords: Complex network; Degree; Community structure; Motif

In recent years, with the rapid development ofchina's economy, the acceleration of industrialization and population growth, energy consumption is too fast, resulting in air pollution. China's air quality shows a downward trend, and smog weather occurs frequently. The main component of smog is atmospheric particulate matter, namely $PM_{2.5}$, whose concentration greatly exceeds the range specified by the state. The main harm of air pollution comes from $PM_{2.5}$, which causes serious damage to people's respiratory system, nervous system and skin tissue [1]. Therefore, $pm_{2.5}$ concentration index is an important detection index of ambient air quality. How to effectively control air pollution is an urgent problem to be solved.

It is worth noting that each heavy pollution is not an individual city, but covers multiple cities. The regional air pollution is not only related to the local pollution of the city, but also related to the air circulation habits of other cities near the region. It is urgent to control air pollution. It is necessary to carry out research from various aspects [2-3] such as blocks to alleviate and effectively reduce the national $PM_{2.5}$ index.

$PM_{2.5}$ is a very fine particle, which cannot be shielded by the respiratory tract. These fine

Fund Project: Shanghai Natural Science Foundation Project (16zr1447200); Supported by the key project of the 2018 graduation design thesis of Shanghai University of Applied Technology (3911lw180040)

About the author: Xiao Qin (1976-), female, lecturer, doctor, mainly engaged in complex network and financial sequence analysis
E-mail: sunshao3000@163.com

particles will enter the respiratory tract and lungs when people are unprepared. Induce rhinitis, pharyngitis, bronchitis, asthma and other respiratory diseases. In this environment for a long time, particles entering the lungs may also enter the blood through the alveoli, inducing a series of cardiovascular diseases such as lung cancer, myocardial ischemia, hypertension, etc.The research on the analysis and remediation of $PM_{2.5}$ has always been valued by people [4]. In foreign countries, the research on $PM_{2.5}$ was carried out earlier, and the research on the source analysis and characteristic spectrum of $PM_{2.5}$ was carried out more [5]. In 2014, elangasinghe et al. [6] used the method of artificial neural network for analysis. The research in China started relatively late. At first, the statistical data on $PM_{2.5}$ were few and incomplete. In recent years, the data has been gradually accumulated and unified, which is of great significance for effective and reasonable research on the very limited air data [7]. Yanghongbin et al. [8] analyzed the chemical composition and source of $PM_{2.5}$.Sun et al. [9] analyzed the concentration in winter and summer in Beijing and found that the concentration in winter was higher in Beijing. Yuxingna et al. [10] and Songyu et al. [11] proposed that there is a negative correlation between $PM_{2.5}$ and visibility of air quality in Beijing. Sufuqing et al. [12] proposed that Beijing has external transmission channels for air pollutants.

To reduce the $PM_{2.5}$ index, first of all, it is necessary to select the main pollution source cities and reduce several seriously polluted areas. In this way, pollution control can draw inferences from one instance. How to find out the main pollution source urban areas through the $PM_{2.5}$ index of each city? In this paper, the complex network method is used to analyze the air pollution areas. Most complex systems in nature can be described by complex networks. Generally, nodes represent different individuals in a real system, and edges represent some kind of connection between individuals. The research and analysis results of complex networks can deepen the understanding of many complex systems in the real world.

In recent decades, the research of complex networks has developed rapidly. There are many articles on complex networks, and the research objects of complex networks are diverse. Complex networks have gradually become a hot research field in the interdisciplinary fields of finance, sociology, computer science, information science and so on [13-20]. However, there are few studies on air quality using complex networks, especially on network models [21].

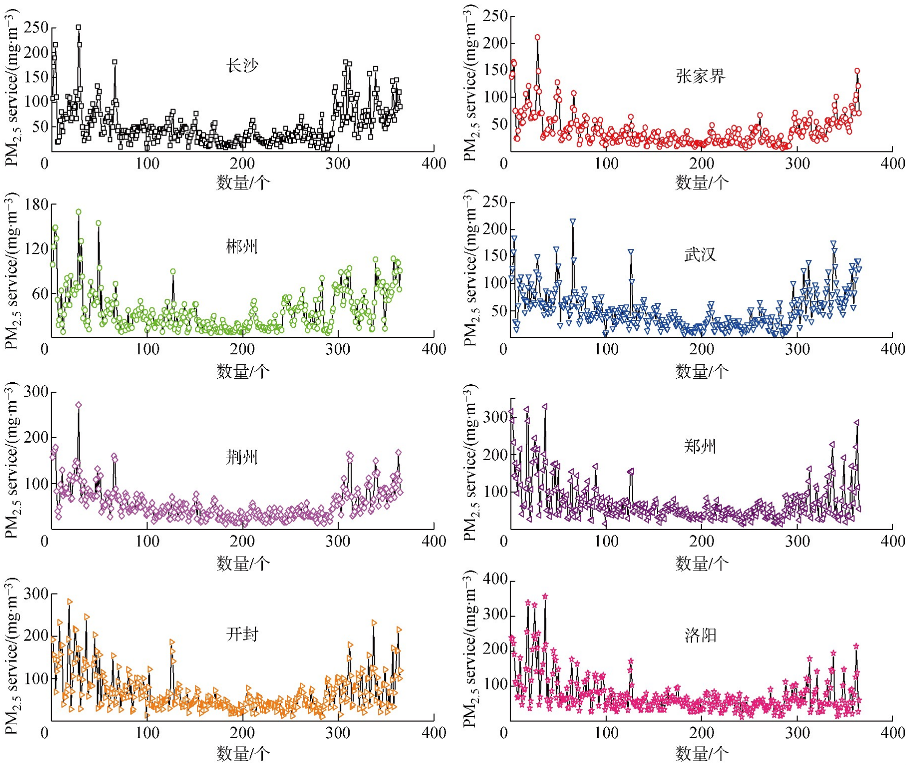
Wang Li [22] took the air quality data of 8 monitoring points in Beijing as the research object, reconstructed the phase space of the time series of air quality, and built a complex network. Using k-means clustering analysis method, three kinds of clustering results of air quality in Beijing are obtained. Zhangzhongyong et al. [23] analyzed the role of various factors in the urban environment, explored the physical process of $PM_{2.5}$ regional diffusion, constructed the capacity network model of urban $PM_{2.5}$ diffusion by using the shortest augmented chain algorithm based on the complex network theory, and revealed the diffusion path of $PM_{2.5}$ in Xi'an. Xue an et al. [24] took 161 cities as the points of the network, took the correlation of $PM_{2.5}$ mass concentration between cities and the ratio of distance as the weight of each side, built a weighted network, and used Girvan Newman algorithm to divide the network, so as to obtain the regional distribution of China's $PM_{2.5}$ pollution in different seasons and the difference of regional division of different seasons according to the regional degree of pollution in different seasons.

This paper analyzes the 365 day $PM_{2.5}$ data of 68 regions in 2017. These 68 regions belong to the five major regions of China, namely, east China, west China, south China, central China and North China.By calculating the correlation between regions, the complex relationship network is obtained. The degree, community and module in the network are analyzed, and the main pollution source areas are obtained. These results can provide a reference for pollution control in major polluted urban areas.

1 Model establishment

1.1 Data

The $PM_{2.5}$ index selected in this paper takes 68 regions in China as the research object. The 68 regions are divided into five categories, which are from East China, north China, central China, south China and West China.There are 20 regions in East China, including Shanghai, hefei, huangshan, bengbu, nantong, nanjing, suzhou, yangzhou, hangzhou, ningbo, wenzhou, jiaxing, fuzhou, xiamen, quanzhou, jinan, qingdao, yantai, nanchang and Jingdezhen (the serial numbers are 1-20). There are [24] 18 regions in North China, including Beijing, tianjin, xilingol, baoding, hengshui, xingtai, handan, tangshan, cangzhou, zhangjiakou, chengde, qinhuangdao, baotou, hohhot, shijiazhuang, langfang, taiyuan and Yuncheng (serial numbers 21 – 38). There are 8 regions in Central China, namely Changsha, zhangjiajie, chenzhou, wuhan, jingzhou, zhengzhou, kaifeng and Luoyang (serial numbers 39 – 46 respectively). There are 8 regions in South China, namely Guangzhou, zhongshan, shenzhen, guilin, beihai, nanning, haikou and Sanya (serial numbers are 47 – 54 respectively). There are 14 regions in West China, including Xining, hainan, urumqi, aksu, turpan, lanzhou, tianshui, lhasa, chengdu, mianyang, chongqing, yinchuan, kunming and Lijiang (serial numbers are 55 – 68). The series length is $PM_{2.5}$ index of all dates from January 1, 2017 to December 31, 2017. The data comes from the weather report. The sequence diagram of central China is shown in Figure 1.



Changsha
Zhangjiajie
Binzhou
Wuhan
Jingzhou
Zhengzhou
Kaifeng
Quantity / piece

Fig. 1 Series of central China

1.2 PM2.5 Data collation and correlation calculation in various regions

Assume that the air pollution data of each region is:

$$Y = \{Y_{ij}, i = 1,2,\cdots,I;\ j = 1,2,\cdots,J\} \quad (1)$$

Where, PM2.5 Y_{ij} data of the ith region on the jth day, *i*=68, *j*=365. In order to be comparable, each data is divided by the root mean square of the data in other regions on that day as the de trend processing for the data:

$$X = \{X_{ij},\ i = 1,2,\cdots,I;\ j = 1,2,\cdots,J\}$$
$$X_{ij} = \frac{Y_{ij}}{\sqrt{\sum_{k=1}^{I} \frac{Y_{ij}^2}{I}}} \quad (2)$$

The Pearson correlation coefficient calculated for the data of each region is:

$$\rho_{ij} = \frac{\sum_{j=1}^{J} X_{rj} \cdot X_{ij} - \frac{1}{J}\left(\sum_{j=1}^{J} X_{rj} \cdot \sum_{j=1}^{J} X_{ij}\right)}{\sqrt{\left(\sum_{j=1}^{J} X_{rj}^2 - \left(\sum_{j=1}^{J} X_{rj}^2\right)\right)\left(\sum_{j=1}^{J} X_{ij}^2 - \left(\sum_{j=1}^{J} X_{ij}^2\right)\right)}} \quad (3)$$

Where, $\rho_{ij}(\Delta t) \in [-1,1]$ (from complete negative correlation -1 to complete positive correlation 1). If $\rho_{ij}(\Delta t) = 0$ air data is not relevant. The corresponding correlation coefficient matrix is symmetric.

Define distance matrix

$$d_{ij} = \sqrt{1 - |\rho_{ij}|} \quad (4)$$

Where, d_{ij} three definitions of distance are met $d_{ij} \in [0,1]$ and. d_{ij} Indicates the similarity between regions. The greater the similarity, the smaller the distance. This distance definition is used to build a complex network.

1.3 Construction and heterogeneity of complex networks

Taking each region as a node and connecting the edges by the distance between regions, a complex network between regions can be established [25]. The methods for building complex networks include threshold method, visible graph, etc. [26]. In this paper, connecting the four regions with the smallest distance can explain the degree of correlation between regions. The smaller the correlation is, the less transmissibility of air pollution is

1.3.1 Node degree

In this paper, we first investigate the node degree in the properties of complex networks. The degree of a node refers to the number of other regions directly connected to the node, expressed in K. The greater K is, the closer it is to other regions, the greater the correlation is, and the more likely it is to be a seriously polluted region. It can be seen from table 1 that Beijing has the largest share, followed by Suzhou, langfang, changsha and Tianshui, followed by

Tianjin, shijiazhuang and other cities.

1.3.2 Community classification

Table 1 The first 13areas with larger degrees in complex

Region	Beijing	Suzhou	Langfang	Changsha	Tianshui	Tianjin	Tangshan
Nodal degree	11	9	8	8	8	7	7
Region	Zhangjiakou	Shijiazhuang	Zhangjiajie	Zhongshan	Xining	Lanzhou	
Nodal degree	7	7	7	7	7	7	

The heterogeneity of complex networks is also manifested in the existence of community structure. The network can be divided into multiple groups [27-28]. There are many edges in the group, and the internal structure is dense; however, there are few connecting edges between the group and its internal structure is loose. In general, a group is called a community. The structure of the community forms a hierarchical structure. The network is divided into seven blocks by using Newman Girvan's division method. At this time, the corresponding Q value is the largest, which is 0.655. Figure 2 shows the community distribution of the complex network, and Table 2 shows the community classification.

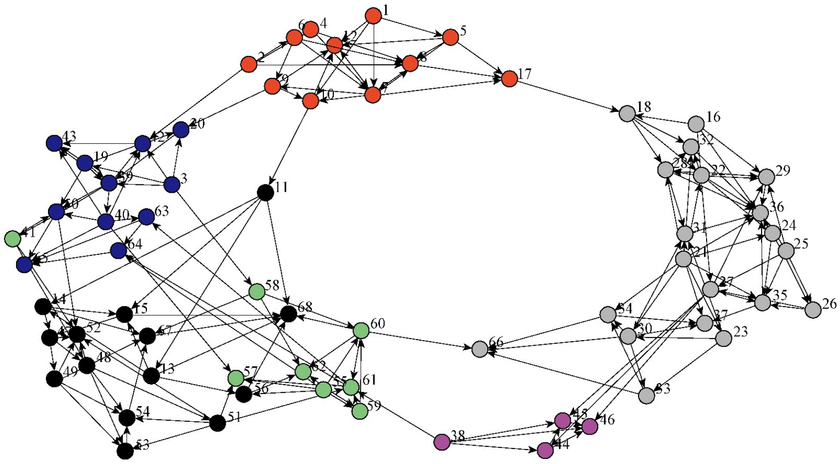


Fig. 2 Community distribution of complex network

Table 2 The classification of communit

Community	Region
1	Chenzhou (41)
2	Yuncheng (38), zhengzhou (44), kaifeng (45), luoyang (46)
3	Xining (55), urumqi (57), aksu (58), turpan (59), lanzhou (60), tianshui (61), lhasa (62)
4	Huangshan (3), nanchang (19), jingdezhen (20), changsha (39), zhangjiajie (40), wuhan (42), jingzhou (43), guilin (50), chengdu (63), mianyang (64), chongqing (65)
5	Shanghai (1), hefei (2), bengbu (4), nantong (5), nanjing (6), suzhou (7), yangzhou (8), hangzhou (9), ningbo (10), jiaxing (12), qingdao (17)
6	Wenzhou (11), fuzhou (13), xiamen (14), quanzhou (15), guangzhou (47), zhongshan (48), shenzhen (49), beihai (51), nanning (52), haikou (53), sanya (54), hainan (56), kunming (67), lijiang (68)

7	Jinan (16), yantai (18), beijing (21), tianjin (22), xilingol (23), baoding (24), hengshui (25), xingtai (26), handan (27), tangshan (28), cangzhou (29), zhangjiakou (30), chengde (31), qinhuangdao (32), baotou (33), hohhot (34), shijiazhuang (35), langfang (36), taiyuan (37), yinchuan (66)

It can be seen from Figure 2 and Table 2 that the distribution of these seven communities, of which the bold ones are areas with high degree. Community 1 is central China, community 2 is North China and central China, community 3 is all West China, community 4 is East China, central China, south China and West China, community 5 is all East China, community 6 is East China, south China and West China, and community 7 is East China, north China and West China.It can be seen that air pollution is somewhat related to the region.

The frequency of recurrence in is much higher than that in the random network. In fact, a motif is a large number of small-scale subgraphs with the same structure in the network. This seed graph depicts the specific patterns of interconnection within the network from the local level. Shenorr et al. [33] found various models in biological networks, neural networks, food chains and technical networks, and carried out research on the structural design principle of complex networks. Figure 3 shows 15 common subgraphs. Taking 3 nodes as an example, there are 13 common motifs.

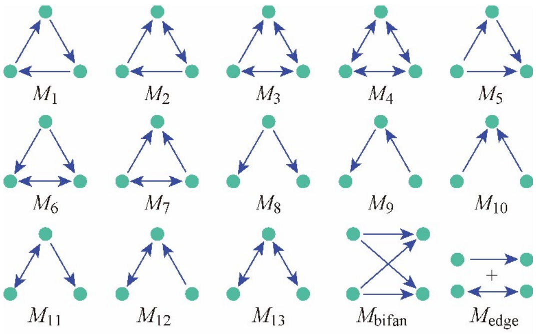
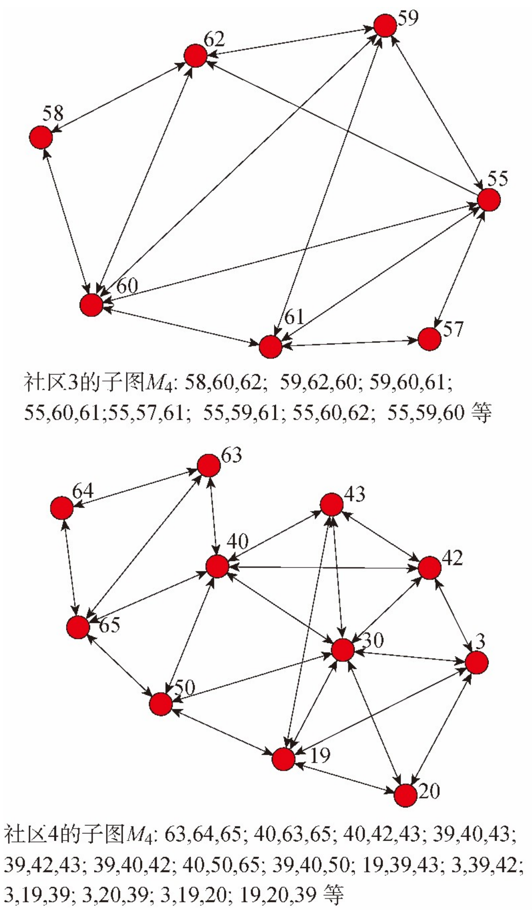


Fig. 3 The fifteen major directed sub-graphs

1.4 Phantom

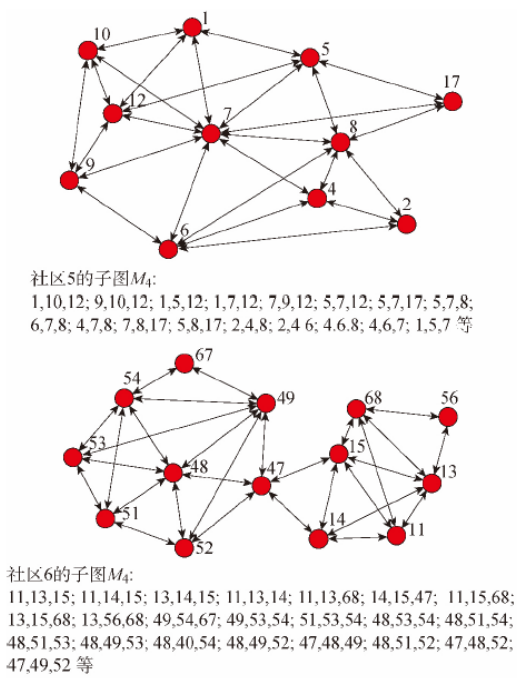
In order to reveal the structural principle of complex networks, the concept of "motif" is introduced [29-32]. Motif is the basic mode of network construction. The frequency of its recurrence in the network is much higher than that in the random network. In fact, a motif is a large number of small-scale subgraphs with the same structure in the network. This seed graph depicts the specific patterns of interconnection within the network from the local level. Shenorr et al. [3] found various models in biological networks, neural networks, food chains and technical networks, and carried out research on the structural design principle of complex networks. Figure 3 shows 15 common subgraphs. Taking 3 nodes as an example, there are 13 common motifs.

Use the method in reference [34] to find 13 seed graphs of 3 nodes in 7 communities of the complex network in five regions. Especially, the subgraphs of regions with large degree are more concerned. Among the 13 seed maps, Figure 4 and Figure 3 are more distributed. Take Figure 4 as an example. The subgraph distribution corresponding to Table 1 is found.



Subgraph of community 3
Subgraph of community 4

Fig.4 The distribution of sub-graphs M4 Of community 3 And 4



Subgraph of community 5
Subgraph of community 6

Fig. 5 The distribution of sub-graphs M4 Of community 5, 6 And 7

2 Results and discussion

68 cities in five regions of China were selected to construct the complex network of air pollution with the correlation size, and the heterogeneity of the corresponding network was studied. Firstly, the size of node degree is analyzed, and 13 regions with large degree are found, including Beijing, suzhou, langfang, changsha, tianshui, tianjin, tangshan, zhangjiakou, shijiazhuang, zhangjiajie, zhongshan, xining and Lanzhou. These are areas with relatively serious air pollution.

Then the established network is divided into communities. When n=7, the Q value is the largest. The 68 districts were divided into 7 communities. The areas in these seven communities are not completely divided according to East China, north China and other regions.For example, community 1 is in Central China, community 3 is all in West China, and community 5 is all in East China. Some communities belong to several regions. For example, community 2 refers to North China and central China, community 4 refers to East China, central China, south China and West China, community 6 refers to East China, south China and West China, and community 7 refers to East China, north China and West China.

Finally, different communities are analyzed by using the method of motif classification in literature. It is found that M4 and M13 are common. Taking M4 sub motif as an example, communities 3~7, which are more moderate, are searched. In the process of analysis, it was found that the sub motifs in these heavily polluted areas also appeared clustering.

Table 3 The clustering phenomenon of sub-graphs in different communities

Community	Region
3	Xining (55), turpan (59), lanzhou (60), tianshui (61), lhasa (62)
4	Huangshan (3), nanchang (19), changsha (39), zhangjiajie (40), wuhan (42), jingzhou (43), guilin (50), chengdu (63), chongqing (65)
5	Shanghai (1), bengbu (4), nantong (5), nanjing (6), suzhou (7), yangzhou (8), ningbo (10), jiaxing (12)
6	Two clusters: Fuzhou (13), xiamen (14), quanzhou (15), lijiang (68); zhongshan (48), shenzhen (49), beihai (51), nanning (52), haikou (53), sanya (54)
7	Three clusters: Beijing (21), tianjin (22), tangshan (28), qinhuangdao (32) and Langfang (36); hengshui (25), xingtai (26), handan (27), shijiazhuang (35); beijing (21), xilingol (23), cangzhou (29), zhangjiakou (30), chengde (31), baotou (33), hohhot (34)

Table 3 shows the cluster phenomenon of motifs, in which the blackbody part is the area with more connections, indicating that the air pollution in these cities is related. The air pollution index changes according to the clustering phenomenon. This clustering phenomenon should be considered when controlling air pollution. It can prevent the phenomenon that the air pollution in one area is transferred from other relevant areas, and the effect of air pollution control is not obvious. For example, in the process of air pollution control and treatment in Beijing, if these clustered areas can be treated together, the effect of treatment should be better than that of treating Beijing alone. If we want to control air pollution together, we need to work together.

3 Conclusion

In this paper, the complex network method is used to analyze 68 different regions, and the correlation network between air pollution cities is obtained. The properties of the corresponding network are analyzed, including node degree, community structure and motif.In real life, the phenomenon of air pollution is not a single phenomenon in a city, but a phenomenon occurring together in most areas. At the same time, due to the fluidity of the air, it is complex to analyze the problem of air pollution. Through the analysis of the network, the phenomenon of urban agglomeration is found, and it is concluded that the treatment of air pollution needs to treat the relevant cities together, so as to achieve an immediate effect faster.

References:

[1] KAMPA M, CASTANAS E.Human health effectsof air pollution[J]. Environmental Pollution, 2008, 151(2):362-367.

[2] BOVE M C, BROTTO P, CASSOLA F, et al.Anintegrated PM2.5 source apportionment study:Positive matrix factorisation vs the chemical transportmodels camx[J]. Atmospheric Environment, 2014, 94:274-286.

[3] WANG F, CHEN D S, CHENG S Y, et al.Identification of regional atmospheric PM10, transport pathways using HYSPLIT, MM5-CMAQ and synopticpressure pattern analysis[J]. Environmental Modelling &Software, 2010.25(8):927-934.

[4] Yangxinxing, fenglihua, wei Peng Atmospheric particulate matter PM2.5 and its hazards [J]. Preface science, 2012, 6 (21): 22-30

[5] RODRIGUEZ S, QUEROL X, ALASTUEY A, etal. Comparative PM10-PM2.5 source contributionstudy at rural, urban and industrial sites during pmepisodes in Eastern Spain[J]. Science of the Total Environment, 2004, 328(1-3:1-26.

[6] ELANGASINGHE M A, SINGHAL N, DIRKS KN, et al.Complex time series

analysis of PM10andPM2.5for a coastal site using artificial neural networkmodelling and k-means clustering[J]. Atmosphericenvironment, 2014(94):106-116.
[7] Zhu Tan, wu Lin, bixiaohui, et al Study on Optimization Technology of atmospheric particulate source apportionment receptor model [J]. China Environmental Science 2010, 30 (7): 865-870
[8] Yanghongbin, zouxudong, wanghongyu, et al Research progress and Prospect of PM2.5 in atmospheric environment [J]. Journal of Meteorology and environment, 2012, 28 (3): 77-82
[9] SUN Y, WANG Y, ZHUANG G.The airborne particulate pollution in Beijing concentration, composition, distribution and sources[J]. Atmosphere Environment, 2004, 38(35):5991-6004.
[10] Yu Xingna, lixinshu, deng zenran, et al Aerosol optical characteristics during haze in Beijing [J]. Environmental Science, 2012, 33 (4): 1057-1062
[11] Songyu, tangxiaoyan, fang Chen, et al Relationship between visibility decline and particulate pollution in Beijing [J]. Journal of environmental science, 2003, 23 (4): 468-471
[12] Sufuqing, gaoqingxian, zhangzhigang, et al Transport channel of foreign pollutants in Beijing boundary layer [J]. Environmental science research, 2004, 17 (1): 26-29
[13] FAN X H, WANG L, XU H H, et al.Characterizing air quality data from complex network perspective[J]. Environmental Science and Pollution Research International, 2016, 23(4):3621-3631.
[14] GU C, COOMANS C P, HU K, et al.Lack of exercise leads to significant and reversible loss of scale invariance in both aged and young mice[J]. PNAS, 2015, 112(8), 2320-2324.
[15] YANG T G, GU C G, YANG H J, et al.Longrange correlations in sentence series from a story ofthe stone[J]. Plos One, 2016, 11(9):e0162423.
[16] QIU L, YANG T G, YIN Y H, et al.Multifractalsembedded in short time series:An unbiased estimation of probability moment[J]. Phy. Rev. E, 2016, 94(6):062201.
[17] YANG Y, GU C, XIAO Q, et al.Evolution of scaling behaviors embedded in sentence series from a story of the stone [J]. Plos One, 2017, 12(2):e0171776.
[18] Xiaoqin Application of network in correlation analysis of stock market [J]. China management science, 2016, 24 (SI): 470-474
[19] Xiaoqin Application of network in correlation analysis of stock market [c].// Proceedings of the 18th China Annual Conference on management science [s.l.]:[s.n.], 2016, 476-480.
[20] Wangyanjuan, qijingchao, yan Jiebing, etc The complex network of University Relations in China [J]. Science Bulletin, 2012, 10:875-884
[21] Liu Shan, peng Lin, wenyanping, et al Pollution characteristics of organic carbon and elemental carbon in PM2.5 in Taiyuan [J]. Environmental Science, 2015, 36 (2): 396-401
[22] Wang Li Air quality data analysis based on complex network theory [d]. Master Thesis of Jiangsu University, 2017
[23] Zhangzhongyong, wangzhongjun Research on urban PM2.5 diffusion network model [J]. China environmental monitoring, 2014 (6): 129-132
[24] Xue an, geng Enze China PM2.5 regional division based on complex network [J].

Journal of applied basic and Engineering Sciences, 2015 (SI): 68-78
[25] Mayubo, gao Guangguang Research on Beijing Tianjin Hebei haze pollution network based on node importance evaluation [J]. Journal of environmental science, 2018, 38 (6): 2287-2296
[26] KRUSKAL J B.On the shortest spanning sub-tree ofagraph and traveling salesman problem[J]. Proc ammath Soc, 1956, 7(1):48-50.
[27] YANG Y, YANG H.Complex network-based timeseries analysis[J]. Physica A:Statistical Mechanics&Its Applications, 2008, 387(5):1381-1386.
[28] WANG J, YANG H.Complex network-based analysis of air temperature data in china[J]. Mod physlett B, 2009, 23(14):1781-1789.
[29] XU X, ZHANG J, SMALL M.Superfamily phenomena and motifs of networks induced from time series[J]. Proceedings of the National Academy of Sciences, 2008, 105(50):19601-19605.
[30] NEWMAN M E J.The structure and function ofcomplex networks[J]. SIAM Rev, 2003, 45:167256.
[31] FORTUNATO S.Community detection in graph[J]. Phys Rep, 2010, 486:75-174.
[32] RAVASZ E, BARAB□SI A L.Hierarchical organization in complex networks[J]. Phys Rev E Statistical Nonlinear &Soft Matter Physics, 2003, 67(2):026112.
[33] SHENORR S, MILO R, MANGAN S, et al.Network motifs in the transcriptional regulatory networkof Escherichia coli[J]., Nature Genetics, 2002, (31):64-68.
[34] AUSTIN R B, DAVID F G, JURE L.Supplementary materials for higher-order organization of complexnetwork[J]. Science, 2016, 353(6295):163-166.

(Editor: Chen Hong)