ISSN 1749-3889 (print), 1749-3897 (online)
International Journal of Nonlinear Science
Vol.30(2020) No.2/3,pp.129-135

Analysis of Evolution Law of PM2.5 Concentration Fluctuation Based on Complex Network

Jiang Xiao*
Energy Development and Environmental Protection Strategy Research Center, Jiangsu University, Zhenjiang, Jiangsu 212013, China

(Received 10 October 2020, accepted 10 December 2020)

Abstract: In this paper, PM2.5 concentration fluctuation network is constructed from the perspective of complex network, and its dynamic characteristics are studied. When constructing PM2.5 concentration fluctuation network, the sequence of concentration fluctuation rate is converted into a string composed of R, r, e, d, D characters. Previous studies on PM2.5 have not been done before. This conversion can better reflect the complexity of PM2.5 concentration fluctuation. According to the probability of occurrence of different characters in character sequence of PM2.5 concentration fluctuation, PM2.5 concentration fluctuation is divided into different periods. Concentration network is constructed for different periods and comparative study is carried out, which is not found in previous studies. Considering the time factor, the evolution rule of network nodes with time is given, the important nodes of PM2.5 concentration network are identified, and the relationship between the important nodes is analyzed.

Keywords: PM2.5 concentration; complex networks; dynamic characteristics

1 Introduction

With the rapid development of China's economy, the haze weather with PM2.5 as the main pollutant occurs frequently, so environmental protection is an urgent task.At present, the research on PM2.5 mainly focuses on the prediction model [1,2], source and characteristics [3,4], influencing factors, diffusion law, coping strategies [5,6], etc., while Chinese scholars have diversified theories and methods in PM2.5 research.In this paper, the daily average concentration data of PM2.5 is converted into a symbol sequence, the fluctuation period is divided, and the concentration fluctuation mode of PM2.5 is obtained.The complex network is used to construct the PM2.5 concentration fluctuation network in different periods, and the topological structure of PM2.5 concentration network in three periods is compared to seek the regularity of PM2.5 concentration network evolution and the main characteristics in different periods.

2 Data and Methods

2.1 Data Source and Processing of PM2.5 Concentration

The data are taken from the real data (original data) of Miyun Meteorological Station, Beijing, National Meteorological Administration from January 19, 2005 to December 31, 2015, which is the daily average of the PM2.5 concentration. The daily mean sequence of PM2.5 concentration in Miyun, Beijing are denoted as $T_{PM2.5}\,(t)$, $t = 1, 2, \cdots, N$, $N = 3706$. The variation sequence of PM2.5 in Miyun, Beijing are denoted as $\Delta T_{PM2.5}\,(t) = T_{PM2.5}\,(t) - T_{PM2.5}\,(t-1)$. Where $T_{PM2.5}\,(t)$ shows the current concentration of PM2.5 and $T_{PM2.5}\,(t-1)$ shows the previous concentration of PM2.5. For $E_{\Delta T_{PM2.5}} = \frac{\sum_{t=1}^{N-1} |\Delta T_{PM2.5}(t)|}{N-1}$, when $\Delta T_{PM2.5} > E_{\Delta T_{PM2.5}}$, it indicates a sharp rise in PM2.5 concentration; when $0 < \Delta T_{PM2.5} \leq E_{\Delta T_{PM2.5}}$,it indicates a rise in PM2.5 concentration; when $\Delta T_{PM2.5} = 0$,it indicates stable PM2.5

***Corresponding author**. *E-mail address*: xiaoyue_78@163.com

IJNS.2020.12.15

concentration;when $-E_{\Delta T_{PM2.5}} \leq \Delta T_{PM2.5} < 0$,it indicates a decrease in PM2.5 concentration;when $\Delta T_{PM2.5} < -E_{\Delta T_{PM2.5}}$, it indicates a sharp decrease in PM2.5 concentration.According to the change of $\Delta T_{PM2.5}$,it corresponds to a letter and gives the sequence of PM2.5 concentration change expressed by letters, so that $\Delta T_{PM2.5}$ corresponds to S_i.

$$S_i = \begin{cases} R\ , & \Delta T_{PM2.5} > E_{\Delta T_{PM2.5}} \\ r\ , & 0 < \Delta T_{PM2.5} \leq E_{\Delta T_{PM2.5}} \\ e\ , & \Delta T_{PM2.5} = 0 \\ d\ , & -E_{\Delta T_{PM2.5}} \leq \Delta T_{PM2.5} < 0 \\ D\ , & \Delta T_{PM2.5} \leq -E_{\Delta T_{PM2.5}} \end{cases} \tag{1}$$

Where R, r, e, d, D respectively indicate that the PM2.5 concentration rises sharply, rises, stabilizes, decreases and decreases sharply. Based on the above ideas, the PM2.5 concentration sequence of Shangdian Zi in Miyun, Beijing can be converted into the corresponding symbol sequence.

$$FT_{PM2.5} = \{S_1, S_2, S_3, \cdots\}, S_i \in \{R, r, e, d, D\} \tag{2}$$

2.2 Classification of PM2.5 Concentration Fluctuation Period

On the basis of the data processing results in the previous section, taking 6 months as a time period, the probability of the occurrence of the characters e, r, R, d, D ,in each time period is counted respectively and denoted as Pe, Pr, PR, Pd, PD.The probability evolution image of occurrence of each character is shown in Figure 1 (a). As can be seen from Figure1 (a), in the process of fluctuation of PM2.5 concentration, the probability of occurrence of stable state e is the lowest to 0, while that of sharp rise state R , and that of sharp decrease state D and decrease state d alternate around July 2005. In order to further distinguish the time when PM2.5 concentration rises sharply,decreases sharply, rises and decreases,

$$\Pr d = \Pr + Pd, PRD = PR + PD \tag{3}$$

is given. Where $\Pr d$ shows the probability of occurrence of PM2.5 concentration rise and decrease, PRD shows the probability of a sharp rise and decrease of PM2.5 concentration. Figure 1 (b) shows the evolution trend of $\Pr d$ and PRD . It can be clearly seen from Figure 1(b) that the probability of sharp rise and decrease of PM2.5 concentration after July 26, 2005 is higher than that of rise and decrease of PM2.5 concentration,, indicating that the PM2.5 concentration fluctuation state has changed. The period from January 19, 2005 to July 26, 2005 is divided into the period of steady fluctuation of PM2.5 concentration, the period from July 27, 2005 to April 3, 2006 is divided into the period of sharp fluctuation of PM2.5 concentration, and the period from April 4, 2006 to December 31, 2015 is divided into the period of steady fluctuation of PM2.5 concentration. Continue to investigate the following probability during the period of dramatic change in PM2.5 concentration.

$$\Pr R = \Pr + PR, PdD = Pd + PD \tag{4}$$

Where $\Pr R$ denotes the probability of rising state of PM2.5 concentration in the period of violent fluctuation, PdD denotes the probability of falling state of PM2.5 concentration in the period of violent fluctuation. Evolution images of $\Pr R$ and PdD are obtained, as shown in Figure 1 (c). In summary, the fluctuation period of PM2.5 concentration can be divided into three periods.

3 Dynamic Characteristics of PM2.5 Concentration Fluctuation Network

3.1 Node Strength and Strength Distribution of PM2.5 Concentration Network On the basis of the daily mean value symbol sequence $FT_{PM2.5} = \{S_i\}\ i = 1, 2, 3, \cdots, 3705$ of PM2.5 concentration of Miyun , Beijing, the symbol sequence consists of five-day fluctuation symbols a week, and one day as the step of data sliding, 3701 PM2.5 concentration fluctuation modes (the same mode repeated statistics) are obtained respectively. PM2.5 data sliding forms a mode, the former mode is the basis of the latter, and transitivity and directivity exist between the modes. Each fluctuating mode is chosen as a node, and the directed transformation between modes is taken as an edge. The directed weighted network of PM2.5 concentration variation can be formed. Since the PM2.5 concentration fluctuation network is a directed weighted network, the node strength and strength distribution are analyzed below. The strength of nodes in weighted networks not only reflects the connection between nodes, but also reflects the weight of the connection edges between nodes. Node strength in directed network is divided into node's input strength and output strength [7,8],where the number of conversions from other nodes to node i is expressed by node i's input strength, and the number of conversions from node i to

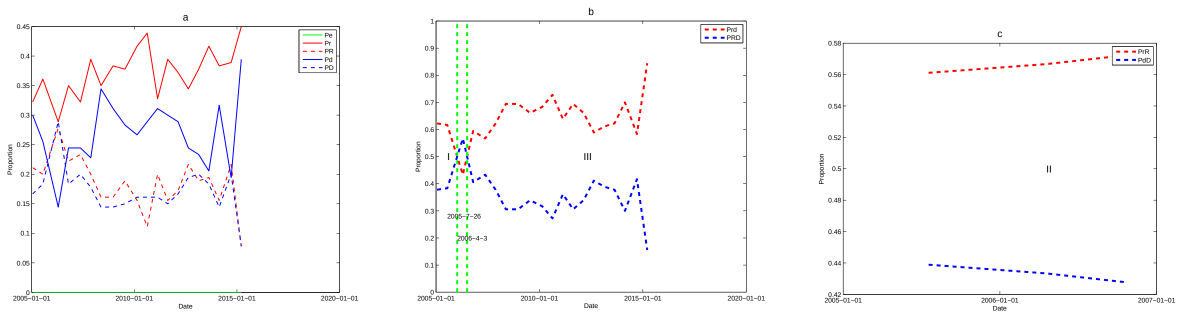

Figure 1: fluctuation state of PM2.5 concentration (a) The probability evolution image of the five states of the fluctuation , (b) The evolutional image of rising and falling and sharp rise and fall ,(c)The evolution image of fluctuation state during the period of violent fluctuation

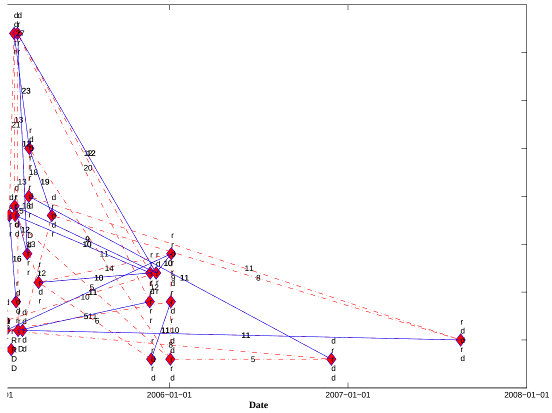

Figure 2: Important nodes and their connections in PM2.5 network

other nodes is expressed by node i's output strength, the formula is as follows.

$$s_i^{in} = \sum_{j=1}^{N} a_{ji}\omega_{ji}, s_i^{out} = \sum_{j=1}^{N} a_{ij}\omega_{ij} \tag{5}$$

Where a_{ji} is an adjacency matrix element, if v_j is connected to v_i , then $a_{ji} = 1$, otherwise $a_{ji} = 0$, ω_{ji} shows the weight of the edges of the connecting node v_j to v_i. According to the construction method of the concentration network mentioned above, the nodes are connected in time sequence, and the input and output strengths of the nodes are the same except for the first node and the last node. Therefore, this part only chooses the node's input strength for calculation and analysis. For the sake of convenience, the input strength is directly described as the strength in the following section.

In PM2.5 concentration fluctuation network, there are relatively few nodes with high strength, while most nodes have low strength. For PM2.5 concentration network, the total number of nodes in the network is 758, of which 25 nodes have strength over 17, accounting for 18.95 % of the total strength, that is, the strength of the 3.3 % node accounts for 18.95 % of the total strength.

The strength of nodes reflects the importance of nodes in the whole network at a certain level. The nodes with strength greater than 17 in the PM2.5 concentration network are calculated and counted, and the node name, the node appearance time, the node strength, and the relationship between nodes are obtained, as shown in Figure 2. Characters in the figue show the names of nodes, numbers on the connecting lines between nodes show the weights, red dots show the positive direction, and blue solid lines show the negative direction.

From the emergence time of important nodes, it can be seen from Figure 2 that in the PM2.5 concentration network, the nodes with high strength often appear earlier, but the nodes with earlier time will definitely become the nodes with high

Table 1: The fitting parameters of the strength distribution of the network nodes of PM2.5 concentration fluctuation at different periods

Period	γ	R^2
I	3.0660	0.9803
II	4.1341	0.9379
III	1.7653	0.9581

strength. The 25 nodes with node strength exceeding 17 are derived from the first 468 nodes in the fluctuation network of PM2.5 concentration. The nodes with the greatest strength are $DDRRR$ and $drrrr$, which appeared on February 18, 2005 and February 25, 2005. These two nodes are the 29th and 34th nodes in the network. The weight in Figure 2 reflects the connection between nodes with strong strength. The connection between nodes with strong strength is closely related. The average contribution of the connection between important nodes to the strength of nodes is 90.28%. The nodes with strong node strength tend to connect with each other, which reflects the obvious positive correlation characteristics of the network. Although the conversion between PM2.5 concentration fluctuation states is frequent and complex, its core fluctuation states are reflected in the first 3% nodes. The essential characteristics of PM2.5 concentration fluctuation can be approximated by using the fluctuation states reflected by the first 3% nodes and their conversion relations.

There are fewer nodes with higher node strength, and most of them have lower node strength. By using the least squares method, the double logarithmic curve of the node strength of the PM2.5 concentration fluctuation network is regressed, and the result is $y = -1.6564x - 0.4468$, and the trend line correlation coefficient R^2 is 0.9366. It can be seen that the reliability of the results is high, indicating that the PM2.5 network as a whole obeys the power law distribution, and the power exponent value γ of the PM2.5 concentration fluctuation network is 1.6564.

Using the least squares, the double logarithmic curves of the node strengths of the PM2.5 concentration fluctuation network in different periods are regressed, and the node strengths of the PM2.5 concentration fluctuation network in different periods are fitted. The parameters are shown in Table 1.It can be seen from Table 1 that the PM2.5 concentration fluctuation network in different periods obeys the power law distribution, and the power exponents of the network are different in different periods. Generally speaking, the power exponent of the stationary period fluctuating network is less than the violent fluctuation period. It shows that the power law distribution of node strength is higher during the period of sharp fluctuation of PM2.5. In summary, the power law distribution of PM2.5 concentration fluctuation network is the highest in stage II violent fluctuation period, followed by stage I stationary fluctuation period, and the lowest in stage III stationary fluctuation period, which reflects the complex dynamic essential characteristics of PM2.5 fluctuation.

3.2 Agglomeration Effect Analysis of Fluctuation Modes in PM2.5 Concentration Network

Local structural properties of networks are usually reflected by clustering coefficients, and the degree of connection between neighboring nodes is reflected by clustering coefficients. The established PM2.5 concentration fluctuation network is a weighted network, and the aggregation coefficient of a node in the network can be calculated by the following formula[7,8].

$$C_i = \frac{1}{s_i\left(k_i - 1\right)} \sum_{j,k} \frac{\omega_{ij} + \omega_{ik}}{2} a_{ij} a_{jk} a_{ik} \tag{6}$$

Where ω_{ij} is the edge weight of two nodes (i, j), s_i is the strength of nodes,$s_i = \sum_j \omega_{ij}$, $k_i$is the degree of nodes, $\sum_{k>j} a_{ij} a_{jk} a_{ik}$ is the total number of triangles containing nodes i . Obviously $0 \leq C \leq 1$, if $C = 0$ denotes that all nodes are isolated, and $C = 1$ denotes that the network is globally coupled, that is, any two nodes in the network are connected.

The average value of the clustering coefficient is 0.0004165, and there are only three nodes with the clustering coefficient not 0, which are in the order of appearance: $Drrrr, rrrrr$. The corresponding clustering coefficients are 0.0865, 0.0833, 0.1458, and the node strengths are 52, 12, and 24, respectively. However, the clustering coefficient of PM2.5 concentration fluctuation network increases with the change of node strength, which indicates that PM2.5 concentration fluctuation network is not completely random, and clustering characteristics of PM2.5 network may occur in small or large groups. The actual concentration of PM2.5 varies with time, sometimes there are clusters of fluctuations of PM2.5 concentration on a small time scale, and sometimes on a large time scale. By studying the clustering coefficient of the PM2.5 concentration fluctuation network, the grouping of PM2.5 concentration fluctuations can be referenced and helped.

The average value of the clustering coefficient is 0.00042154 in stable fluctuation period and there are only three nodes with the clustering coefficient not 0, which are in the order of appearance: $drrrr, rrrrr, Drrrr$. The corresponding clustering coefficients are 0.08650.15220.0758, and the node strengths are 522311, respectively. The average value of the

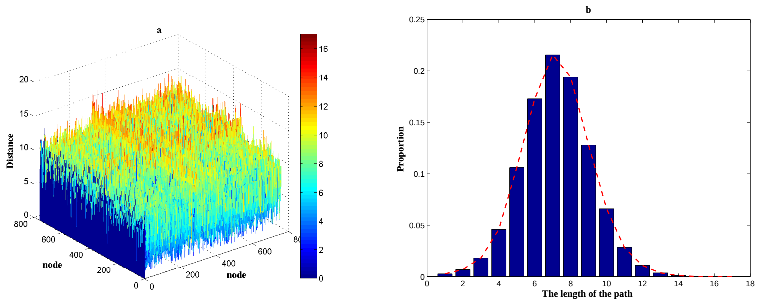

Figure 3: Node distance of PM2.5 concentration network (a) Distance between any two nodes, (b) Distribution of node distance

clustering coefficient is 0 in violent fluctuation period and the clustering coefficient of all nodes is 0. The number of nodes whose clustering coefficient is not 0 in the period of violent fluctuation is less than that in the period of stable fluctuation, and the average clustering coefficient in the period of violent fluctuation is less than that in the period of stalbe fluctuation, which indicates that the network is closer in the period of stable fluctuation and the PM2.5 concentration in the period of violent fluctuation shows higher complexity.

3.3 Intermodal Transition Period Analysis of PM2.5 Concentration Network Fluctuation

The distance between two nodes i and j is denoted as d_{ij}. The diameter D of a network is defined as the maximum distance between all nodes, i.e.

$$D = \max d_{ij} \tag{7}$$

The average distance between any two nodes is called the average path length L of the network. The formula for calculating L in directed networks is as follows[7,8]:

$$L = \frac{1}{N\left(N-1\right)} \sum_{i \neq j} d_{ij} \tag{8}$$

Where N is the number of network nodes.

For PM2.5 concentration network, Floyd algorithm is used to calculate the distance between any two nodes in the network, as shown in Figure3 (a), and different path length distributions are shown in Figure3 (b).

From Figure 3 (a, b), the network diameter of PM2.5 concentration is 17, the distance between nodes is 5, 6, 7, 8, 9, accounting for 81.63% of the total, and the average path length of the network is 7.1899. The results show that there are short-range correlations among the fluctuation modes in PM2.5 concentration network. The mode conversion of PM2.5 concentration fluctuation is frequent, and it can be done once in 7-8 days. These properties make it possible to predict the fluctuation of PM2.5 concentration in the future. The path length distribution between PM2.5 network nodes is calculated in the period of stable fluctuation and violent fluctuation, respectively.

In the period of violent fluctuation, the diameter D of PM2.5 concentration network is 17, the distance L between nodes is 5, 6, 7, 8, 9, accounting for 80.31% of the total, and the average path length of the network is 7.293. This shows that in the period of violent fluctuation, the mode of PM2.5 concentration fluctuation is converted once through 7-8 days. In the period of stable fluctuation, the diameter D of PM2.5 concentration network is 44, the distance L between nodes is 6, 7, 8, 9, accounting for 67.09% of the total, and the average path length of the network is 15.151. This shows that in the period of stable fluctuation, the mode of PM2.5 concentration fluctuation is converted once through 15-16 days. It can be seen that the diameter of PM2.5 concentration network is shorter and the fluctuation period is shorter during the period of violent fluctuation.

3.4 Transitional Modal Analysis of PM2.5 Concentration Network Fluctuation

The node betweenness reflects the role and influence of nodes in the network[7,8]. For directed networks, the node betweenness B_i is defined as

$$B_i = \sum_{j \neq l \neq i} \left[N_{jl}\left(i\right)/N_{jl}\right] \tag{9}$$

Where N_{jl} is the number of shortest paths from node v_j to node v_l ,and N_{jl} is the number of shortest paths from n-ode v_j to node v_l passing through node v_i . In the PM2.5 concentration fluctuation network, the top three nodes are

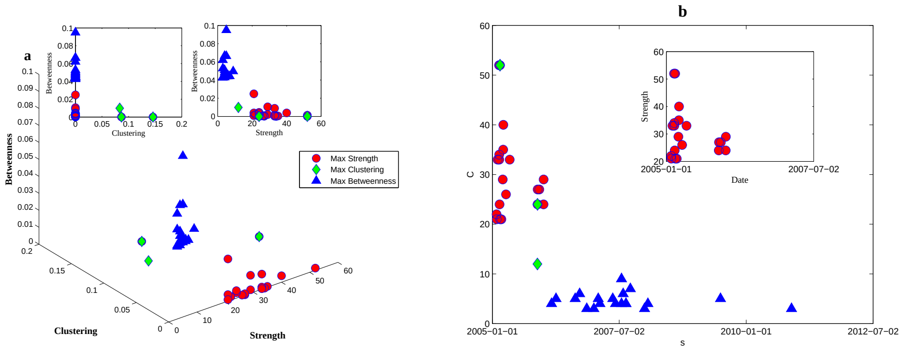

Figure 4: Node strength, clustering coefficient and node betweenness in PM2.5 concentration network (a) Nodes with larger node strength, clustering coefficient and node betweenness,(b)Time of first appearance of nodes with larger node strength, clustering coefficient and node betweenness

$rRDRD, rrDrr, andrDrrR$. Their node betweenness are 0.0949, 0.0665 and 0.0664 respectively, while their node strength are 5, 4, 5 respectively. Their first occurrence time is February 2, 2007, November 3, 2006 and November 4, 2006, respectively. In summary, the nodes with smaller strength have larger betweenness.

During the period of stable fluctuation, the top three nodes in PM2.5 concentration network are $rRDRD$, $rDrrR$ and $RDRDd$. Their node betweenness are 0.1064, 0.069, 0.0664, and their node strengths are 5, 5 and 3, respectively. During the period of violent fluctuation, the top three nodes in PM2.5 concentration network are rrdrrrdrrRRDdrr. Their node betweenness are 0.22250.2080.2046, and their node strengths are 2, 2 and 3, respectively. The node betweenness of PM2.5 concentration network node in the period of violent fluctuation is larger than that in the period of stable fluctuation. It can be seen that in the period of violent fluctuation, the median capability of network node is on the rise. In PM2.5 concentration network, nodes with low node strength act as the main intermediary function. Nodes with high node strength mainly transit through nodes with low node strength.

Based on the above analysis of node strength, clustering coefficient and node betweeness, the relationship among them is investigated. The top ten important nodes with node strength, clustering coefficient and node betweeness in PM2.5 concentration network are selected, and the relationship among them is shown in Figure 4(a) and the time distribution of their first occurrence is obtained. As shown in Figure 4 (b).

As can be seen from Figure4(a), in PM2.5 concentration networks nodes with smaller clustering coefficient and node betweenness have larger node strength, nodes with smaller node strength and clustering coefficient have larger node betweenness, and nodes with smaller node betweenness and smaller node strength have larger clustering coefficient. On the whole, in PM2.5 concentration network, the characteristics of smaller clustering coefficient, smaller average path length, smaller average node betweenness and larger average node strength are showed ,which are different from those of random network and chaotic network[9]. From the first appearance time of nodes with larger node strength, clustering coefficient and node betweenness in PM2.5 concentration network, it can be seen from Figure 4 (b) that the first appearance time of nodes with larger node strength (shown in the figure$\bigcirc$) are earlier. It can be seen that in PM2.5 concentration network, the nodes with larger node strength tend to appear in the early stage, while the nodes with larger clustering coefficient (shown in the figure$\diamondsuit$) appear in relatively dispersed time, but still tend to be in the early stage. Nodes with larger node betweenness (shown in the figure Δ) appear at the most dispersed time, which may occur at any time. Once they occur, it means that PM2.5 concentration is in a transitional stage of fluctuation, which is helpful to grasp the regularity of PM2.5 concentration change.

4 Conclusions

In this paper, PM2.5 concentration fluctuation network is constructed from the perspective of complex network, and its dynamic characteristics are studied. When constructing PM2.5 concentration fluctuation network, the sequence of concentration fluctuation rate is converted into a string composed of R, r, e, d, D characters. Previous studies on PM2.5 have not been done before. This conversion can better reflect the complexity of PM2.5 concentration fluctuation. According to the probability of occurrence of different characters in character sequence of PM2.5 concentration fluctuation, PM2.5

concentration fluctuation is divided into different periods. Concentration network is constructed for different periods and comparative study is carried out, which is not found in previous studies. Considering the time factor, the evolution rule of network nodes with time is given, the important nodes of PM2.5 concentration network are identified, and the relationship between the important nodes is analyzed. In this paper, the real data of PM2.5 concentration are analyzed, and the results are obtained on the basis of data samples. Due to the limitation of collected data, the results of the study inevitably have some errors. However, it is very meaningful to obtain the characteristics of PM2.5 concentration fluctuation network in different fluctuation periods by means of complex network method. These conclusions can help us to understand the essential characteristics of PM2.5 concentration fluctuation scientifically, and provide a theoretical basis for further studying the stability of other related data fluctuation with the help of complex network theory and constructing a new PM2.5 concentration prediction method in the future.

Acknowledgments

This paper was supported by the following foundations: the Major Program of the National Natural Science Foundation of China (Grants No.71690242), the Program of National Natural Science of China(Grants Nos. 51976085 & 71673116), the Program of Social Science Foundation of Jiangsu Province (Grants No. 18EYB020), the Major Research Project of the National Natural Science Foundation of China (Grants No.91546118).

References

[1] P. Perez, E. Gramsch. Forecasting hourly PM2.5 in Santiago de Chile with emphasis on night episodes. *Atmospheric Environment*, 2016, 124: 22-27.

[2] J. Chen, B. Zou, M. Wang, et al. Bayesian network in PM2.5 concentration prediction considering data characteristics. *Environmental Science and Technology*, 2015, 38(11): 49-56.

[3] C. Milando, L. Huang, S. Batterman. Trends in PM2.5 emissions, concentrations and apportionments in Detroit and Chicago. *Atmospheric Environment*, 2016, 129: 197-209.

[4] S. Yan, G. Wu. Network analysis of fine particulate matter (PM2.5) emissions in China. *Scientific Report*, 2016,6:33227.

[5] Atkinson R. W., Kang S., Anderson H. R., et al. Epidemiological time series studies of PM2.5 and daily mortality and hospital admissions: a systematic review and meta-analysis. *Thorax*, 2014,69(7):660-665.

[6] Z. Shen, J. Cao, L. Zhang et al. Retrieving historical ambient PM2.5 concentrations using existing visibility measurements in Xi'an, Northwest China. *Atmospheric Environment*, 2016, 126: 15-20.

[7] H. An, X. Gao, W. Fang, et al. The role of fluctuating modes of autocorrelation in crude oil prices. *Physica A*, 2014, 393: 382390.

[8] H. An, X. Gao, W. Fang, et al. Research on patterns in the fluctuation of the co-movement between crude oil futures and spot prices: A complex network approach. *Applied Energy*, 2014,136: 10671075.

[9] Z. Lei, Z. Gong, Z. Rong , et al. An approach to research the topology of Chinese temperature sequence based on complex network. *Acta Physica Sinica*, 2008, 57(11): 73807389.