



Visibility Graph Network Analysis of Air Quality Data

Xinghua Fan¹, Qi Zhang¹, Li Wang², Jiuli Yin^{1,*}

¹Faculty of Science, Jiangsu University, Zhenjiang, China

²Helie Middle School, Wuxi, China

Email address:

fan131@ujs.edu.cn (Xinghua Fan), yjl@ujs.edu.cn (Jiuli Yin)

*Corresponding author

To cite this article:

Xinghua Fan, Qi Zhang, Li Wang, Jiuli Yin. Visibility Graph Network Analysis of Air Quality Data. *International Journal of Environmental Monitoring and Analysis*. Vol. 6, No. 3, 2018, pp. 110-115. doi: 10.11648/j.ijema.20180603.15

Received: August 19, 2018; **Accepted:** September 25, 2018; **Published:** October 18, 2018

Abstract: As air quality is closely related to human life and physical and mental health, the data of air quality has become a concern of the entire society. This study analyzes the characteristics of air quality data from a visibility graph networks point of view. The authors select eight monitoring stations in Beijing as samples. The time series of air quality data is mapped to a complex network based on the visibility graph algorithm. First, the authors study the topological structure of the networks for all the monitoring stations. Comparison results show that all constructed networks have similar structures in terms of the average path length, the network diameter, average clustering coefficient, density and the average degrees. Then the authors study the evolution of the visibility graph network for Huairou Town station for a long period of time. On the one hand, the value of the node degree indicates that the most important dates for air quality are the end of April, the beginning of May and the first three weeks of winter. On the other hand, the small-world properties of the networks reveals that the air quality data for the year 2014 is more stable without extreme fluctuations. This finding is consistent with the conclusion that air quality is largely affected by the weather while human activities play a more and more important role.

Keywords: Air Quality Index, Visibility Graph Algorithm, Complex Network, Topological, Measure, PM2.5

1. Introduction

Much of the developing world has suffered an air pollution problem due to the rapid economic growth and urbanization over the past decades [1, 2]. One specific type air pollution, commonly known as PM2.5 shorthand for particulate matter 2.5 micrometers in diameter or smaller is especially harmful. In China, air pollution is responsible for killing 1.2 to 2 million Chinese a year, about one sixth of all the premature deaths in the country [3]. Air pollution problem is particularly serious in big cities, such as Beijing, which was ranked the 13th most polluted city in the world in 2004 [4]. In the Chinese capital, the number of so-called haze days has increased dramatically since 2011 [5].

The Beijing Municipal Government has implemented numerous air pollution control measures since 2000 [6]. National Ambient Air Quality Standard of China (NAAQS-2012) was adopted by the government from 2013. Since then, data for air quality index (AQI) of all its monitoring sites in the city have been reported from the

website and the air quality is categorized according to its value. AQI is a standardized indicator that comprehensively considers the concentration of a diversity of airborne pollutants, including PM 2.5, PM 10, CO, NO₂ and O₃. AQI report published by China's Ministry of environmental protection is slightly different from the EPA (Environmental Protection Agency), US and is more suitable for the air quality analysis in China [7].

A large number of studies have been conducted on the sources, characteristics and seasonal variations of atmospheric pollutants. Elangasinghe et al. used artificial neural networks (ANN), combined with k-means clustering, to understand the complex time series of PM2.5 concentrations [8]. Li et al. collected samples of PM2.5 aerosol from a receptor site in the East China Sea to explore the seasonal variation and sources of polychlorinated biphenyl ethers [9]. Chen et al. used CMAQ model to simulate winter PM2.5 formation in the San Joaquin Valley [10]. The CMAQ model generally met the PM2.5 model performance criteria and was suitable for the State Implementation Plan (SIP) applications. Substantial efforts

have been made toward understanding PM2.5 pollution in China, especially in Beijing, such as source apportionment [11, 12], characteristics of PM2.5 [13], the spatial-temporal characterization [14], and monitoring and mitigation policies [15]. As a time series, air quality data was analyzed from the perspective of correlation complex networks [16]. These efforts have greatly enriched the knowledge of PM2.5 pollution and have done a remarkable job in helping inform pollution mitigation policies.

In the time series literature, there has been an increasing trend in investigating a time series through the mapping on networks or graphs [17]. Through this mapping, the dynamics of the time series is converted into the topology properties of the network, and vice versa, the characteristics of the time series can be inferred from the network. Time series can be mapped into networks with different methods [17] such as recurrence networks, correlation networks cycle networks, transition networks, and visibility graphs. Among these methods, visibility graphs [18] have advantages in maintaining time orders of the time series in the order of the nodes, the straightforward geometric interpretation of the original time series, a simple and fast speed in calculation without the need to create the state space which requires a large number of sampling points [19].

Varying air quality data can be thought as the output of a complex system of which the governing rules are hard to be understood. This paper applies the visibility graph networks (VGN) [18] to disclose the characteristics of air quality data. When the air quality data are mapped onto the visibility graph networks, dynamics underlying the complex system are reflected by the topological measures of the VGNs. A global view of correlation of the data and key dates is displayed by the visualization of the VGN. In this case, researchers can better understand the dynamics of air quality data, while policymakers can make more comprehensive and objective assessments of existing environmental policies to improve local air quality.

2. Methodology

2.1. Visibility Graph

In general, a complex network can be represented by an undirected and unweighted graph $G = (V, E)$, which consists of a set of nodes or vertices $V = \{v_1, v_2, \dots, v_N\}$ and a set of edges or links $E = \{e_1, e_2, \dots, e_E\}$.

The Visibility Graph algorithm [18] maps a time series $\{y_t\}$ into a complex network according to the visibility criteria. A node in the visibility graph is a point or location in Euclidean plane determined by the observations of the time series. Two points (t_a, y_a) and (t_b, y_b) (assume time $t_a < t_b$ and y_a and y_b are the values of the two data points, respectively) in the time series are mutually visible if and only if for any $t_c (t_a < t_c < t_b)$, the following stands

$$y_c < y_a + (y_b - y_a) \frac{t_c - t_a}{t_b - t_a}. \tag{1}$$

Two mutually visible nodes are then connected by an edge.

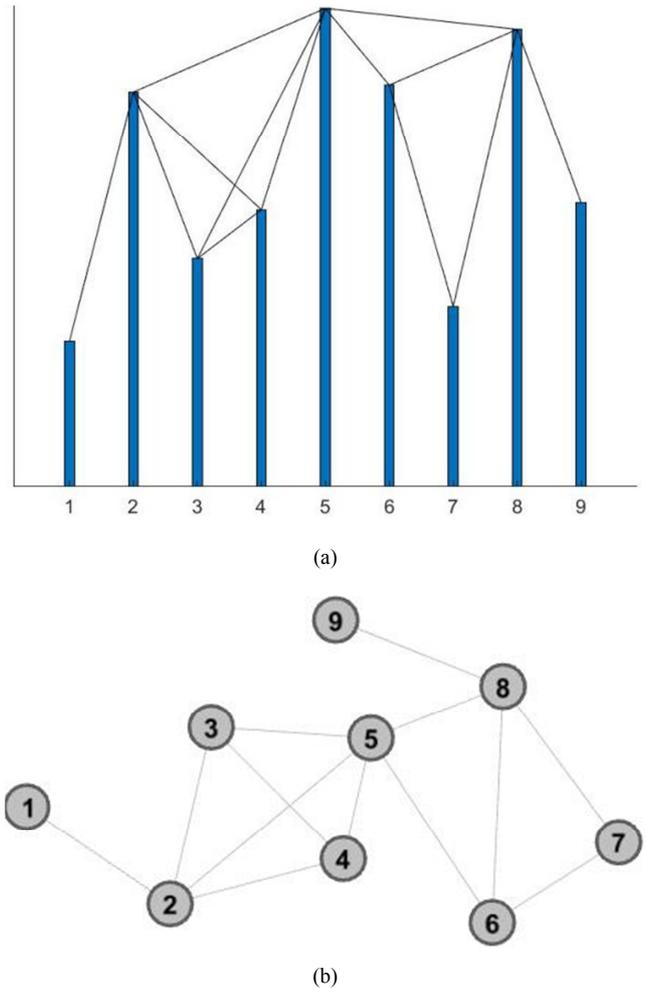


Figure 1. Illustration of the visibility graph network constructed from a time series. (a) A simple time series represented by vertical bars. The lines indicate the network connection established according to the visibility criterion of Eq.1. (b) The network emerging from the time series in (a).

Figure 1 illustrates the procedure of converting the time series y (Figure 1(a)) to its VG (Figure 1(b)). The gray line between y_i and y_j in Figure 1(a) indicates the two observations can see each other. The VG is always connected because each view is certainly connected to its previous and next observations except for the first and last observations. The visibility criteria ensure that VGNs are invariant under vertical (linear) rescaling, translation and superposition of a linear trend of a time series [18].

2.2. Measurements of Network Structure

The degree of a node $i(k_i)$ is the number of connections of node i . As undirected network, the degree is defined as:

$$k_i = \sum_j a_{ij} = \sum_{i=1} a_{ij}. \tag{2}$$

The average degree of a network is the average of k_i for all nodes in the network, as:

Table 1. Topological properties of eight visible graph networks.

	A1	A2	A3	A4	A5	A6	A7	A8
Average Path length	4.265	4.74	4.45	4.532	4.592	4.49	4.426	4.526
Diameter	8	10	8	9	9	9	10	9
Average Clustering coefficient	0.767	0.767	0.764	0.757	0.766	0.766	0.771	0.763
Density	0.01	0.01	0.01	0.01	0.01	0.01	0.011	0.01
Average Degree	7.499	7.244	7.488	7.42	7.255	7.425	7.841	7.214

$$\langle k \rangle = \frac{1}{N} \sum_i k_i. \quad (3)$$

The degree and the degree distribution are the common measurements to analyze complex networks. The degree of a node is the number of edges incident with it, and degree distribution $p(k)$ is defined as the probability of a node with degree k .

The diameter D of a network is the largest of the shortest path distances between any pair of nodes, and the average path length $\langle d_{ij} \rangle$ is the average distance of a network connecting any pair of nodes i and j .

$$D = \max_{ij} d_{ij},$$

$$\langle d_{ij} \rangle = \frac{1}{N(N-1)} \sum_{i,j} d_{ij}. \quad (4)$$

For the overall network structure, the average path length and distance is an important measured characteristic [20].

The clustering coefficient of a node i is a measure of network transitivity, expressing the extent to which neighbors of a node are neighbors of each other, and is defined:

$$c_i = \frac{2e_i}{k_i(k_i-1)}. \quad (5)$$

where k_i is the number of neighbors of node i and e_i is the number of connected pairs between all neighbors of node i . The average clustering coefficient $\langle c_i \rangle$ of a network is the average c_i as follows

$$\langle c_i \rangle = \frac{1}{N} \sum_{i=1} c_i. \quad (6)$$

The density is defined as the number of edges divided by the largest number of edges possible.

3. Data and Experimental Results

3.1. Data

The authors collected AQI of eight monitoring stations in Beijing: Huairou Town (A1), Dingling Changping (A2), Changping Town (A3), East 4th north ring road (A4), North Xizhimen Street (A5), Fengtai Gardon (A6), Liangxiang Fangshan (A7) and Yizhuang Developmetn Zone (A8). The authors refer these stations as A1-A8 for simplicity. The sample stations are illustrated in Figure 2. Those eight monitoring sites located in different areas in Beijing. A2 is used as a background site, A6 is an urban site, A4 and A5 are traffic monitoring sites while A1, A3, A7 and A8 are suburban sites. Time period of the data covers from 1 January 2013 to 31 December 2014. The AQI data was acquired from the website (www.bjmemc.com.cn/xgzs_sjcx.action). There

are totally 730 observations. Additional time period from 2011 to 2016 is adopted for a further study of Huairou Town station (A1).

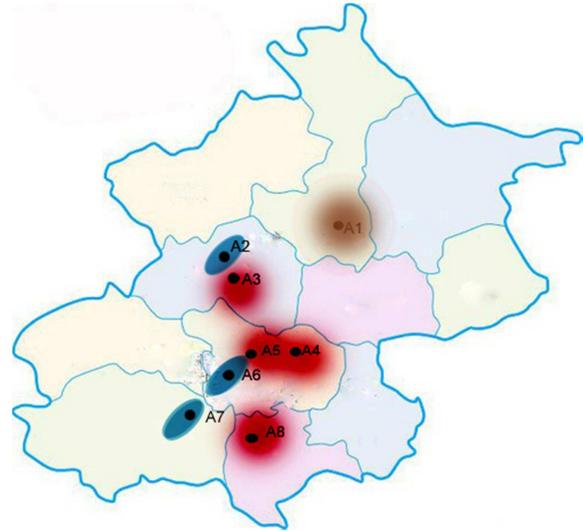


Figure 2. Illustration of eight sample monitoring stations. Dots represent sample sites. The star represents the Forbidden City. Station in the same colors belongs to the same group [16].

Based on the visibility graph, the authors mapped the air quality of eight monitoring stations into complex networks. The authors use daily air quality data as a node in a complex network, using the visible algorithm, the connection between the nodes is determined to form the edge of the complex network. Then, the air quality time sequence of the eight monitoring stations is mapped to complex networks.

3.2. Topological Properties

The authors calculate the five important topological characteristics (the average path length, the network diameter, average clustering coefficient, density and the average degrees) of the complex networks. The results are shown in Tab 1. Average path length ranges from 4.265 to 4.740 with the minimum value in A1 and the maximum in A2. The averaged value is 4.503, meaning that each of node can be arrived through 4.503 days. Diameter varies in the set of 8, 9 and 10 with A1 has the smallest value while A2 and A7 have the largest. This indicates that the longest time between any two nodes is about 8-10 days. Average clustering coefficient is about 0.7, which is a relative great. This means that the VGNs has clear clustering phenomenon. The authors note that the maxima value is reached in station A1. The small value of density (about 0.01) show that a node is related with few other nodes, which is about seven by the value of the

average degree (about 7).

The authors note from Table 1 that most extreme values are reached in the Huairou Town station (A1). Therefore, the authors will analyze in the next subsection the VGN of AQI of the Huairou Town station in a longer time period.

3.3. Important Nodes

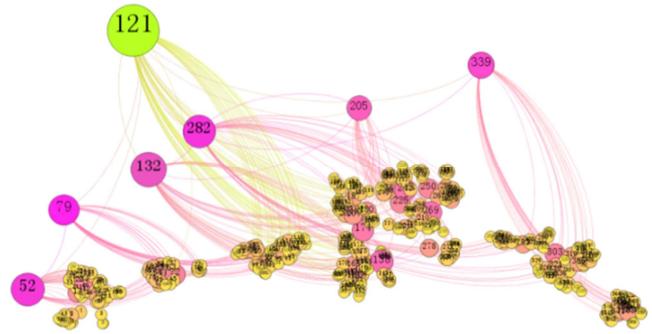
The authors construct visibility graph networks from the Huairou Town in years 2011-2016. Figure 3 visualizes the six networks. The circles represent the nodes with the number inside it as the index of the date. The size of the circle stands for the degree of node. The greater the circle, the higher the degree. The edges are represented by the arced segments.

From Figure 3, the authors can find the most important date are in spring for the previous year while in winter for the later years. The node 121 has the largest degree in 2011. The date of the original data corresponding to it is 1st May 2011, a date in spring. In Spring, Beijing always suffers from sandstorms. Thus the appearance of the big node may be caused by natural laws. In 2012 and 2013, nodes 120 and 126 have the largest degree respectively. The corresponding dates are 29th April, 2012 and 6th May, 2013. This tells us that the laws of environmental change in the past three years are still within the adjustable range of nature. However, the authors still cannot ignore the narrowing of the gap between the 2013 annual value gap and the 2011 and 2012 annual values (this can be concluded based on the diameter variation of nodes of adjacent sizes in the Gephi diagram).

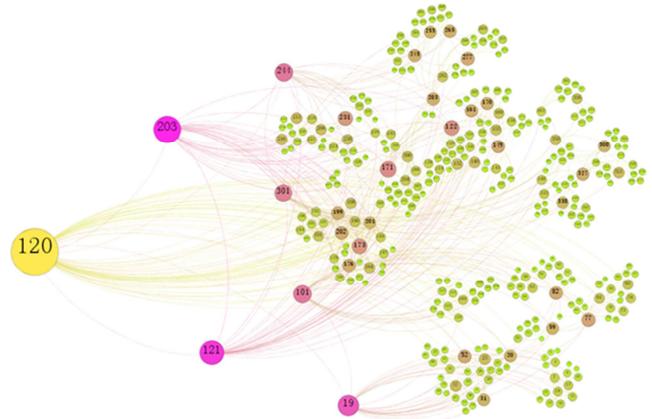
Figure 3 (d) shows a different pattern than others. There is no obvious big node in year 2014. The difference of the degree of the nodes is small, and the distribution from the big to the small nodes is relatively uniform. This indicates that the AQI in 2014 changed steadily, and there were no serious environmental pollution factors. There seems to be some connection between the emergence of this steady state in 2014 and the decrease in the gap between the node degrees in 2013. The reduction of the gap of the annual value is a sign of the steady state.

The important nodes correspond to dates in the winter for the later sample period. The date of the original data corresponding to this node is 1st December, 2015. It is a day in winter, which is the season of coal-fired heating. This is the first time rebound after the occurrence of steady state in 2014. In 2016, a node 64 with a large degree of value also appeared. The date of the original data that is in correspondence with the node is 4th March, 2016, which belongs to the end of the winter and early spring.

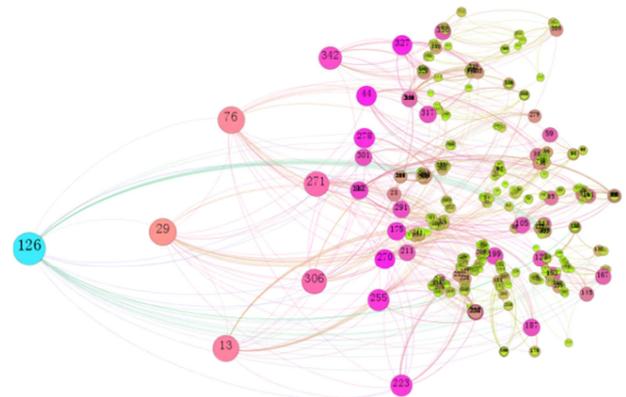
Figure 3 shows two things. On the one hand, the AQI in Huairou Town has seasonal property, and the maximum value of the node generally appears in spring or winter. On the other hand, the AQI has a comprehensive six-year stationary change, and the small difference in the adjacent nodes may indicate the occurrence of a steady state.



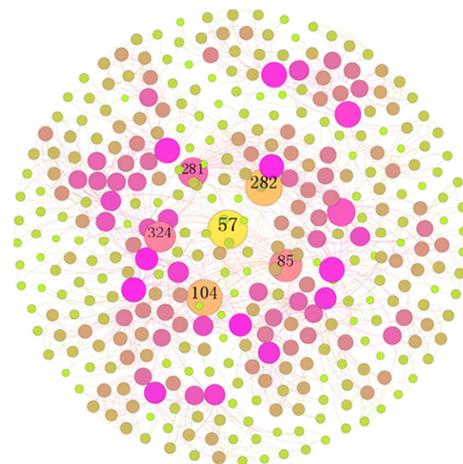
(a) Year 2011. Node 121 corresponds to May 1.



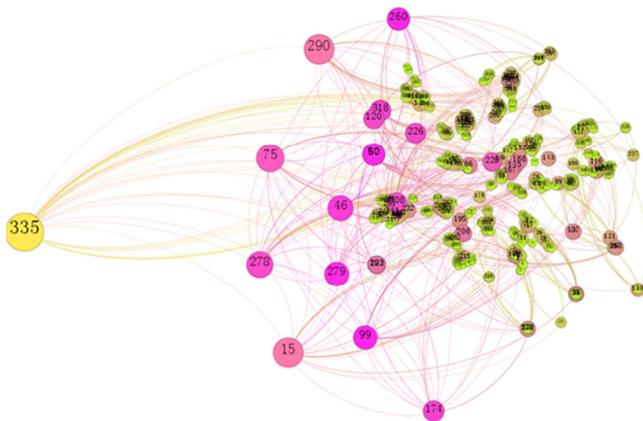
(b) Year 2012. Node 120 corresponds to April 29.



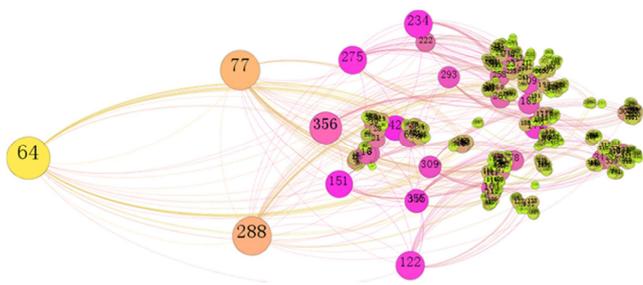
(c) Year 2013. Node 126 corresponds to May 6.



(d) Year 2014. There is no nodes with extra large degree.



(e) Year 2015. Node 335 corresponds to December 1.



(f) Year 2016. Node 64 corresponds to March 4.

Figure 3. Visible graph networks for air quality data of Huairou Town station in years 2011-2016.

3.4. The Small-World Character of Network

Small-world network model is different from the regular networks and random networks in that it has small average path length and high clustering coefficient. It is a kind of network with good properties [21, 22]. This subsection study the small-world property of the visibility graph constructed from the data of the Huairou Town monitoring station. The time period is from 2011-2016.

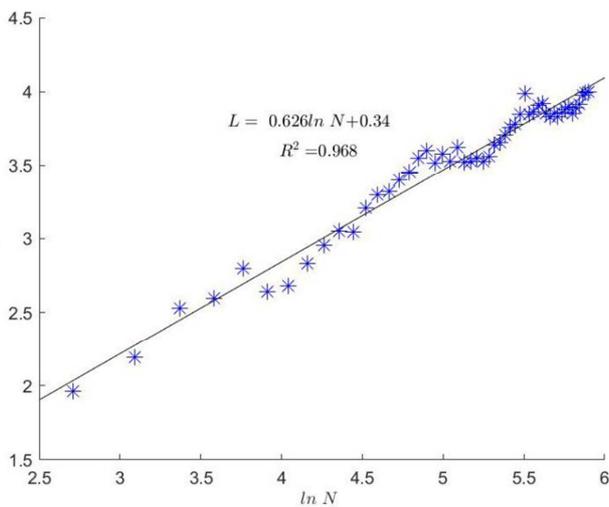


Figure 4. The change of average path with the growth of network nodes in year 2014 for the Huairou Town station.

For a complex network, let L be the average shortest

path length and N be the number of nodes. If there is a linear relationship between L and the logarithm of N , that is $L = \alpha + \beta \ln N$, the network is called a small-world network. On the contrary, a network may be considered as a big world network when L has a linear relationship with N .

Table 2. Fitting results of $L = \alpha + \beta \ln N$ for the VGN of Huairou Town station from 2011-2016.

	2011	2012	2013	2014	2015	2016
α	0.958	0.816	0.156	0.34	0.43	0.0254
β	0.413	0.451	0.633	0.626	0.583	0.629
R^2	0.859	0.899	0.927	0.968	0.883	0.939

Figure 4 shows the relationship between L and N in 2014. Fitting curves are similar as that in 2014. For the length of the paper, the authors list the fitting results in Tab 2. The authors see a clear linear relationship as the fitting goodness are all greater than 0.85. This means that the six complex networks are small-world. The authors also notice that the maximum of R^2 is obtained in year 2014, indicating that the complex network of 2014 has the most obvious characteristics of small-world. This means that the air quality data for 2014 is more stable without extreme fluctuations and all the nodes have almost the same degree. This phenomenon has been displayed in Figure 3 (d) where the nodes are in near same sizes.

4. Conclusion

In this paper, the authors map the air quality of eight monitoring points in Beijing basing on visual algorithms to complex networks and study the topological features of complex networks. The authors map the six-year air quality time series of Huairou Town to complex networks basing on complex network view algorithms. The authors use Gephi to map the view network sequence of Huairou Town’s air quality time series for six years (2011-2016). From the perspective of complex network diagrams, the reasons for the formation of larger nodes in the network and some rules that can be revealed by the view are analyzed. In addition to 2014, there are one important big node in the other five networks. In 2014, there are more nodes with similarity level, and nodes between adjacent similarity levels. The difference in diameter is also smaller. The authors also find that the six-year air quality time series in Huairou Town has a good Small-world feature through the linear relationship between the average path and the growth of network nodes. In 2014, the air quality corresponding network has the most obvious Small-world characteristics. It is important to analyze the environmental quality of Beijing by analyzing the air quality time series of air quality in the view network of air quality, which can help to excavate the inherent characteristics of time series more deeply.

This study finds that most important days are the end of April and beggaring of May in the previous half sample period. It confirms the significant effect of synoptic meteorological conditions on air pollution [23]. However, the

important days transfer to the coal-heat seasons for the later half period, indicating the contribution of human activities on air quality. Thus, practical measures should be adopted and strengthened to decrease coal consumption both in Beijing and its neighbour provinces. Power plants and factories violating emissions standards should be fined severely even shutdown. It is still a long run for the Beijingers to breathe in a green and comfortable way, although the air quality in the Chinese capital is actually improving gradually in the studying period.

Acknowledgements

Research is supported by the National Natural Science Foundation of China (No.71673116) and the Humanistic and Social Science Foundation from Ministry of Education of China (Grant 16YJAZH007).

References

- [1] S. Zheng, M. E. Kahn, Understanding China's urban pollution dynamics, *Journal of Economic Literature* 51 (3) (2013) 731–72.
- [2] M. Greenstone, R. Hanna, Environmental regulations, air and water pollution, and infant mortality in India, *American Economic Review* 104 (10) (2014) 3038–72.
- [3] R. A. Rohde, R. A. Muller, Air pollution in China: Mapping of concentrations and sources, *PLOS ONE* 10 (8) (2015) 1–14.
- [4] Z. Yang, M. Tang, Does the increase of public transit fares deteriorate air quality in Beijing?, *Transportation Research Part D: Transport and Environment* 63 (2018) 49 – 57.
- [5] X. Liu, J. Li, Y. Qu, T. Han, L. Hou, J. Gu, C. Chen, Y. Yang, X. Liu, T. Yang, et al., Formation and evolution mechanism of regional haze: a case study in the megacity Beijing, China, *Atmospheric Chemistry & Physics* 13 (9).
- [6] S. Wang, M. Zhao, J. Xing, Y. Wu, Y. Zhou, Y. Lei, K. He, L. Fu, J. Hao, Quantifying the air pollutants emission reduction during the 2008 Olympic Games in Beijing, *Environmental science & technology* 44 (7) (2010) 2490–2496.
- [7] Z. Chen, B. Xu, J. Cai, B. Gao, Understanding temporal patterns and characteristics of air quality in Beijing: A local and regional perspective, *Atmospheric Environment* 127 (2016) 303–315.
- [8] M. Elangasinghe, N. Singhal, K. Dirks, J. Salmond, S. Samarasinghe, Complex time series analysis of PM10 and PM2.5 for a coastal site using artificial neural network modelling and k-means clustering, *Atmospheric Environment* 94 (0) (2014) 106 – 116.
- [9] Y. Li, T. Lin, F. Wang, T. Ji, Z. Guo, Seasonal variation of polybrominated diphenyl ethers in PM2.5 aerosols over the East China Sea, *Chemosphere* 119 (2015) 675 – 681.
- [10] J. Chen, J. Lu, J. C. Avise, J. A. DaMassa, M. J. Kleeman, A. P. Kaduwela, Seasonal modeling of PM2.5 in California's San Joaquin Valley, *Atmospheric Environment* 92 (2014) 182 – 190.
- [11] R. Zhang, J. Jing, J. Tao, S.-C. Hsu, G. Wang, J. Cao, C. S. L. Lee, L. Zhu, Z. Chen, Y. Zhao, et al., Chemical characterization and source apportionment of PM 2.5 in Beijing: seasonal perspective, *Atmospheric Chemistry and Physics* 13 (14) (2013) 7053–7074.
- [12] X. Li, J. Wu, M. Elser, F. Tian, J. Cao, I. El-Haddad, R. Huang, X. Tie, A. S. Prévôt, G. Li, Contributions of residential coal combustion to the air quality in Beijing–Tianjin–Hebei (BTH), China: a case study, *Atmospheric Chemistry and Physics* 18 (14) (2018) 10675–10691.
- [13] Q. Liu, J. Baumgartner, Y. Zhang, Y. Liu, Y. Sun, M. Zhang, Oxidative potential and inflammatory impacts of source apportioned ambient air pollution in Beijing, *Environmental science & technology* 48 (21) (2014) 12920–12929.
- [14] F. Huang, X. Li, C. Wang, Q. Xu, W. Wang, Y. Luo, L. Tao, Q. Gao, J. Guo, S. Chen, et al., PM2.5 spatiotemporal variations and the relationship with meteorological factors during 2013–2014 in Beijing, China, *PloS one* 10 (11) (2015) e0141642.
- [15] D. Wu, Y. Xu, S. Zhang, Will joint regional air pollution control be more cost-effective? An empirical study of China's Beijing–Tianjin–Hebei region, *Journal of Environmental Management* 149 (2015) 27 – 36.
- [16] X. Fan, L. Wang, H. Xu, S. Li, L. Tian, Characterizing air quality data from complex network perspective, *Environmental Science and Pollution Research* 23 (4) (2016) 3621–3631.
- [17] E. Zhuang, M. Small, G. Feng, Time series analysis of the developed financial markets' integration using visibility graphs, *Physical A: Statistical Mechanics and its Applications* 410 (2014) 483 – 495.
- [18] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J. C. Nuo, From time series to complex networks: The visibility graph, *Proceedings of the National Academy of Sciences* 105 (13) (2008) 4972–4975.
- [19] M. Ahmadlou, H. Adeli, A. Adeli, New diagnostic EEG markers of the Alzheimer's disease using visibility graph, *Journal of neural transmission* 117 (9) (2010) 1099–1109.
- [20] D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks, *nature* 393 (6684) (1998) 440.
- [21] M. Sun, Y. Wang, C. Gao, Visibility graph network analysis of natural gas price: The case of north American market, *Physical A: Statistical Mechanics and its Applications* 462 (2016) 1–11.
- [22] A. Vázquez, R. Pastor-Satorras, A. Vespignani, Large-scale topological and dynamical properties of the internet, *Physical Review E* 65 (6) (2002) 066130.
- [23] J. Liu, W. Li, J. Wu, Y. Liu, Visualizing the intercity correlation of PM2.5 time series in the Beijing–Tianjin–Hebei region using ground-based air quality monitoring data, *PloS one* 13 (2) (2018) e0192614.