Transnational Journal of Mathematical Analysis and Applications
Vol. 7, Issue 1, 2019, Pages 29-42
ISSN 2347-9086
Published Online on August 19, 2019

http://jyotiacademicpress.org

A RESEARCH ON $PM_{2.5}$ TRANSMISSION MODE BASED ON MOTIF ANALYSIS

DAN WANG[1], CUIPING LI[1] and YUFANG WANG[2]

[1]LMIB-School of Mathematics and Systems Science
Beihang University
Beijing, 100083
P. R. China
e-mail: 2587736495@qq.com
cuipingli@buaa.edu.cn

[2]Department of Statistics
Tianjin University of Finance and Economics
Tianjin, 300222
P. R. China

Abstract

Identification of $PM_{2.5}$ transmission mode is important for the government to take measures to effectively control the pollution of $PM_{2.5}$. We select 93 cities including the three provinces of Northeast China, Shandong Province, Henan Province, Shanxi Province and Jing-Jin-Ji region as examples. Then we apply higher-order organization of complex networks to study the transmission mode of $PM_{2.5}$ between these cities. The data we use to build the network contains meteorological conditions of wind speed and wind direction, social factor of region's GDP (Gross Domestic Product), as well as geographic distance and

2010 Mathematics Subject Classification: 94A, 94C, 92B.

Keywords and phrases: $PM_{2.5}$ transmission, network, motif, cluster.

This work is supported by the National Natural Science Foundation of China, No. 61134005.

Received August 1, 2019

$PM_{2.5}$ concentrations. By the motif analysis method, the major potential $PM_{2.5}$ victims are identified in each cluster.

1. Introduction

With the rapid development of economy, air pollution has become more and more serious. Especially, the $PM_{2.5}$ (particulate matter smaller than $2.5\mu m$) pollution occurred in many cities of China, arousing widespread attention in society.

In recent years, the related studies about $PM_{2.5}$ are generally concentrated on the pollution characteristics and pollution sources. For example, Wei et al. [1] collected $PM_{2.5}$ data from air quality environmental monitoring points in Nanjing. They analyzed the seasonal variation characteristics of $PM_{2.5}$ concentration and effects on atmospheric visibility. Xu et al. [2] applied principal factors analysis to study the chemical composition of $PM_{2.5}$ in Beijing. And five types of pollution sources of $PM_{2.5}$ in Beijing were identified.

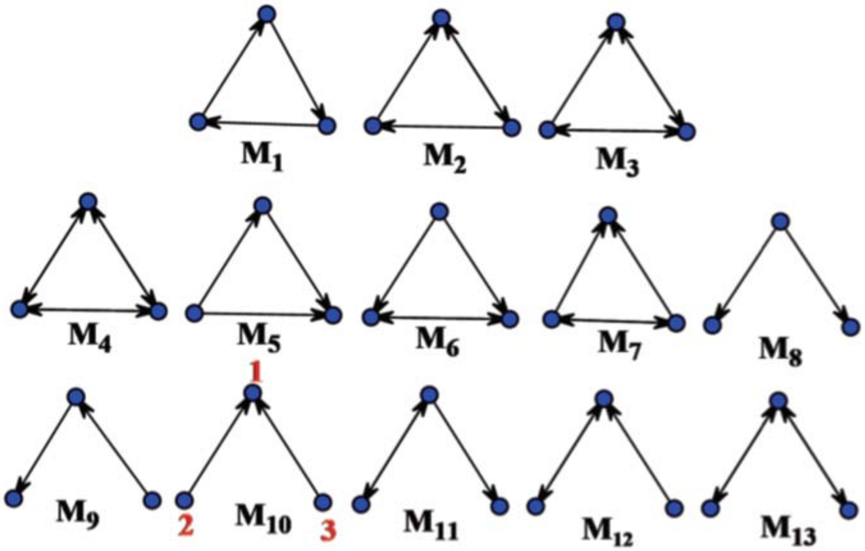

Figure 1. 13 motifs.

There are many methods for studying $PM_{2.5}$. For instance, Support Vector Machine (SVM) [3] is used for the prediction of $PM_{2.5}$ in the atmosphere, and the Bayesian Maximum Entropy (BME) [4] method for the assessments of the space-time variability of $PM_{2.5}$ concentrations in machine learning tools.

Benson et al. [6] provided a generalized framework by 13 triangular motifs (Figure 1). Different motifs reveal different internal structures. According to motif selected, the graph is divided to find an optimal partition, then combining the motif to analyze the internal condition of this partition.

In [7], Wang et al. choose the triangular motif M_8 and M_{13} to study the $PM_{2.5}$ transmission in Yangtze River Delta. This method not only provides a global perspective on studying the transmission of $PM_{2.5}$, but also reveals the internal pollution structure between cities.

The three provinces of Northeast China, Shandong Province, Henan Province, Shanxi Province and Jing-Jin-Ji region, as densely-populated areas, are also the heavy industrial areas. In order to control the spread of pollution timely and to minimize the harm caused by $PM_{2.5}$ through effective measurements, it is of great significance to understand the transmission mode of $PM_{2.5}$ in the above regions.

In this paper, we focus on the triangular motif M_{10} to study the $PM_{2.5}$ transmission mode among above regions. First, we build a weighted network of $PM_{2.5}$ by adding the GDP (Gross Domestic Product) data into the network comparing with Wang el al. [7]. Second, the weights of motif instances are introduced for revealing the structure of $PM_{2.5}$ weighted networks. We apply the clustering algorithm to cluster cities into different groups. The major potential pollution victims are identified in each cluster lastly.

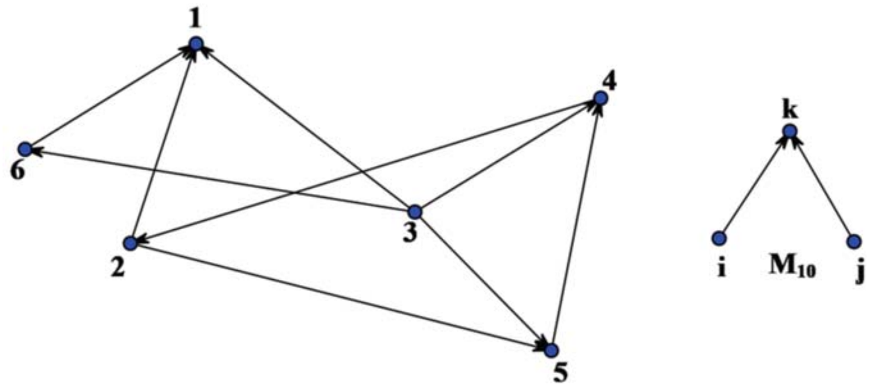

Figure 2. A network of motif M_{10}.

2. Data

The meteorological data that we use in this paper including wind speed and wind direction coming from National Meteorological Information Center, the $PM_{2.5}$ concentrations from the website of National Environmental Monitoring Centre and the GDP data of each city from the website of National Bureau of Statistics.

Assuming that meteorological conditions do not change much during a season, we take the data of February 2018 as a sample. In order to simplify the model, the data of wind speed, $PM_{2.5}$ concentration and GDP are the monthly average value. A prevailing wind direction is determined after counting the number of each wind direction in the month of February. We choose the prevailing wind direction as the wind direction in this paper.

3. Method

3.1. $PM_{2.5}$ network

In network theory, a network can be described by an adjacency matrix $B = (\omega_{ij})_{n\times n}$, where $\omega_{ij}(i, j = 1, 2, \cdots, n)$ express the weight information between the i-th site and the j-th site $(i \neq j)$.

Inspired by [7], we construct an adjacency matrix B by $\omega_{ij} = a_1(i, j) \times a_2(i, j) a_3(i, j)$, in which $a_k(i, j)(k = 1, 2, 3)$ are defined as the following:

- $a_1(i, j) = \frac{\gamma}{dist(i, j)}$, where $dist(i, j)$ indicates the geographic distance (kilometer) between the i-th city and the j-th city $(i \neq j)$, γ is an influence coefficient.

- $a_2(i, j) = \omega_i \cos(\theta_{ij})$, where ω_i indicates the wind speed (meter per second) of i-th city, θ_{ij} is the angle between the wind direction of i-th city and the directional line segment from i-th city to j-th city in the planimetric map $(i \neq j)$.

- $a_3(i, j) = g(i)\frac{c(i)}{c(j)}$, where $g(i)$ indicates the GDP of i-th city, $c(i)$ and $c(j)$ are the $PM_{2.5}$ concentration (microgram per cubic meter) of i-th city and j-th city, respectively $(i \neq j)$.

For the simplicity, we make the following assumptions in this paper:

$$a_1(i, j) = \begin{cases} \frac{0.8}{dist(i, j)}, & 0 < dist(i, j) \leq 300\text{km}; \\ \frac{0.2}{dist(i, j)}, & 300 < dist(i, j) \leq 500\text{km}; \\ 0, & \text{else.} \end{cases}$$

$$a_2(i, j) = \begin{cases} \omega(i) \cos \theta_{ij}, & \theta_{ij} \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \text{ and } \omega_i < 2\text{m/s}; \\ 0, & \text{else.} \end{cases}$$

$$a_3(i, j) = \begin{cases} g(i)\frac{c(i)}{c(j)}, & c(i) > c(j); \\ 0, & \text{else.} \end{cases}$$

Then a $PM_{2.5}$ network has been built.

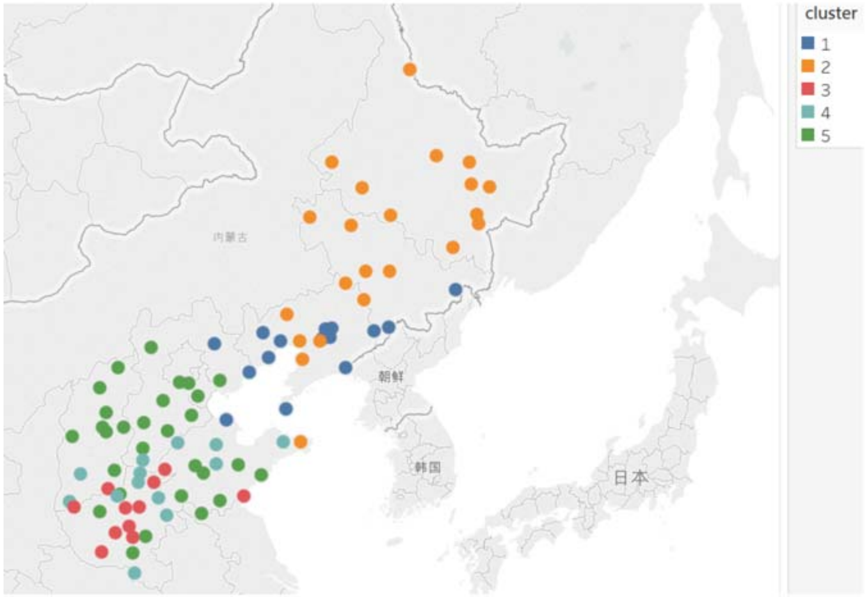

Figure 3. 5 clusters obtained by M_{10}-analysis.

3.2. Motif analysis

Higher-order structures are captured by network motifs. They are patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks [5].

As defined in [7], considering motifs to be a pattern of edges on a small number of nodes. A triangular motif on three nodes can be defined by a tuple $(B, \mathscr{A})$, where B is a 3×3 adjacency matrix and $\mathscr{A} \subset \{1, 2, 3\}$ is a set of anchor nodes. In a triangular motif, there are two anchor nodes and each triangular motif is anchored by the two nodes.

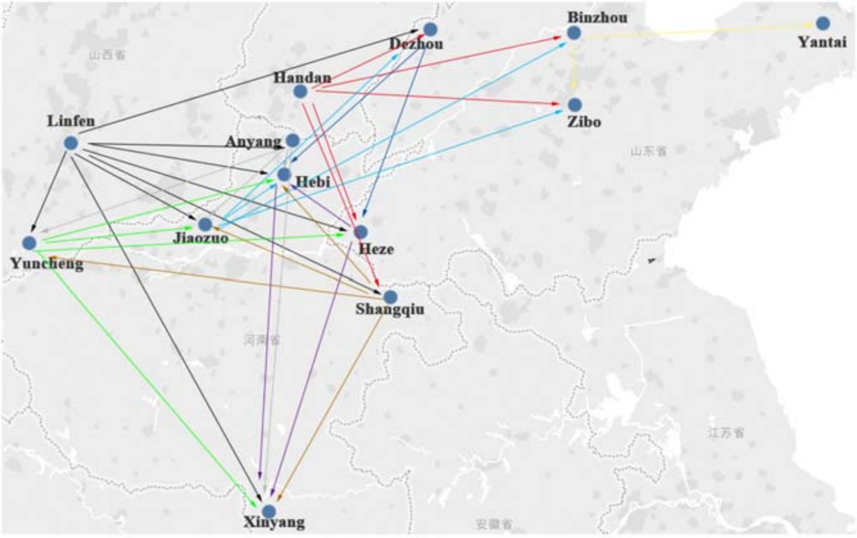

Figure 4. Cluster 4 based on M_{10}.

Each triangular motif can be considered as a small network.

For motif M_{10} in Figure 1, if we consider it as an unweighted and directed network, then $\mathscr{A} = \{2, 3\}$, B is a binary matrix and B is $\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.

If we consider it as a weighted and directed network, then $\mathscr{A} = \{2, 3\}$, B is a weighted matrix, $B = \begin{pmatrix} 0 & 0 & 0 \\ \omega_{21} & 0 & 0 \\ \omega_{31} & 0 & 0 \end{pmatrix}$, where $\omega_{ij}(i, j = 1, 2, 3; i \neq j)$ express the weight information between node i and node j.

A selection function $\chi_{\mathscr{A}}(v)$ is defined as the following:

$$\chi_{\mathscr{A}}(v) = \{\{v_i, v_k\} | v_i \in \mathscr{A}, \text{ and there is a directed edge from } v_i \text{ to } v_k\}.$$

The set of motifs in a weighted and directed network is defined by $M(B, \mathscr{A})$,

$$M(B, \mathscr{A}) = \{(v, \chi_{\mathscr{A}})|v \in V^3, v_i, v_j, v_k \text{ distinct}\},$$

where $v = (v_i, v_j, v_k)$ is an any three-node tuple.

For example, for motif M_{10} in Figure 2, we have

$$M(B, \mathscr{A}) = \{(\{1, 2, 3\}, \{\{2, 1\}, \{3, 1\}\}), (\{5, 2, 3\}, \{\{2, 5\}, \{3, 5\}\}), (\{1, 2, 6\}, \{\{2, 1\}, \{6, 1\}\})\}.$$

Furthermore, each $(v, \chi_{\mathscr{A}}(v))$ is called a motif instance of the three-node motif M_{10}.

According to [8], the weight of the motif instance $(v, \chi_{\mathscr{A}})$ can be defined as

$$\omega(v, \chi_{\mathscr{A}}(v)) = (\prod_{(i, j) \in \chi_{\mathscr{A}}(v)} \omega_{ij})^{\frac{1}{2}}.$$

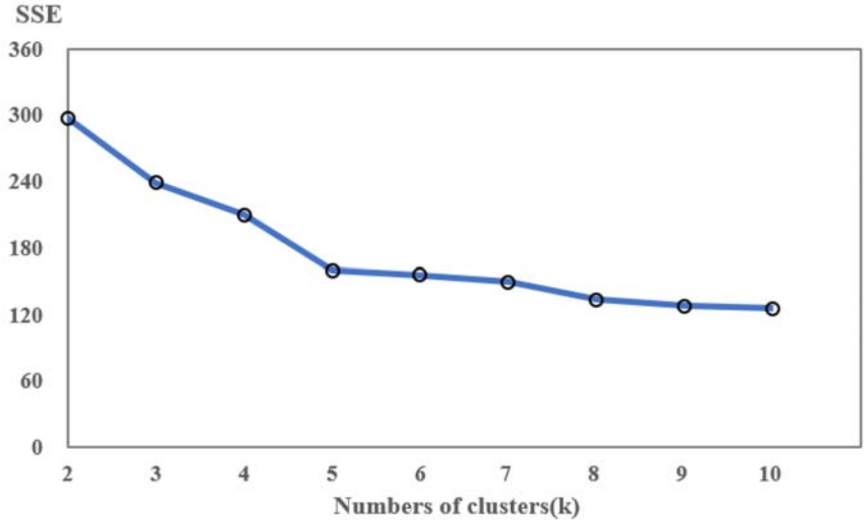

Figure 5. $SSE - k$ for motif M_{10}.

For example, the weight of the motif instance $(\{1, 2, 3\}, \chi_{\mathscr{A}}(v))$ is

$$\omega(v, \chi_{\mathscr{A}}) = (\omega_{21}\omega_{31})^{\frac{1}{2}}.$$

A network can be expressed by a weighted adjacency matrix A.

For example, the network in Figure 2, we have

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \omega_{21} & 0 & 0 & 0 & \omega_{25} & 0 \\ \omega_{31} & 0 & 0 & \omega_{34} & \omega_{35} & \omega_{36} \\ 0 & \omega_{42} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_{54} & 0 & 0 \\ \omega_{61} & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

By [6], based on motif M_{10}, adjust the motif adjacency matrix $W_M = (a_{ij})_{n \times n} (i \neq j)$ as follows:

$$a_{ij} = \sum_{(v, \chi_{\mathscr{A}}(v)) \in M} \omega(v, \chi_{\mathscr{A}}(v)) \mathbb{1}(\{i, j\} \subset v).$$

Therefore, for the network of motif M_{10} in Figure 2, we have

$$W_M = \begin{pmatrix} 0 & a + c & a & 0 & 0 & c \\ a + c & 0 & a + b & 0 & b & c \\ a & a + b & 0 & 0 & b & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & b & b & 0 & 0 & 0 \\ c & c & 0 & 0 & 0 & 0 \end{pmatrix},$$

where $a = (\omega_{21}\omega_{31})^{\frac{1}{2}}$, $b = (\omega_{25}\omega_{35})^{\frac{1}{2}}$, and $c = (\omega_{21}\omega_{61})^{\frac{1}{2}}$.

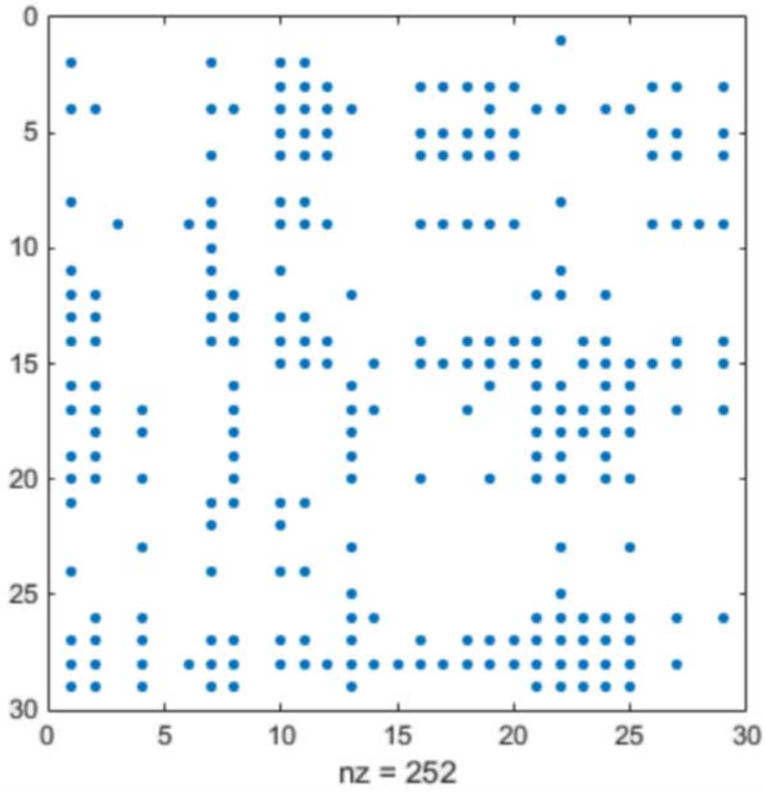

Figure 6. Spy plot of cluster 5 in Figure 2.

4. Identifying Potential Pollution Victims by Motif M_{10}

By Subsections 3.1, 3.2 and the data observed, we construct a directed, weighted network and its motif adjacency matrix $(W_M)_{93\times 93}$.

Next, we use the motif-based higher-order spectral clustering algorithm [6] to cluster the 93 cities into k different clusters.

We calculate the sum of squared errors (*SSE*) [10] to obtain an optimal clustering value.

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

where *x* is one of the city in cluster C_i; C_i is the *i*-th cluster, *dist* is the standard Euclidean distance, and c_i is the centroid of cluster C_i. k is the number of clusters, and too larger k has not much meaning for clustering 93 cities, so the optimal value is chosen from 2 to 10.

As the number of clusters k increases, the *SSE* will gradually become smaller. When k reaches some value, the change of *SSE* will be sharply. Then *SSE* tends to be at as the value of k continues to increasing, which means that the relationship between *SSE* and k is the shape of an elbow (Figure 5). We choose corresponding k as the number of clusters.

The major steps of the motif-based higher-order spectral clustering algorithm are listed below.

(1) Calculating the eigenvectors: $z_1, z_2, \cdots, z_k$ of $k\,(2 \leq k \leq 10)$ smallest eigenvalues for $L_M = I - D_M^{-\frac{1}{2}} W_M D_M^{-\frac{1}{2}}$, where D_M is a diagonal matrix with $(D_M)_{ii} = \sum_{j=1}^{93} (W_M)_{ij}$;

(2) Letting $Z = (z_1, z_2, \cdots, z_k) = (z_{ij})_{93 \times k}$, then calculating matrix

$$Y = (Y_{ij})_{93 \times k}, \text{ where } Y_{ij} = z_{ij} \Big/ \sqrt{\sum_{j=1}^{k} z_{ij}^2},\ z_j = \begin{pmatrix} z_{1,i} \\ z_{2,i} \\ \vdots \\ z_{93,i} \end{pmatrix};$$

(3) Taking each row of matrix *Y* as a sample, we have 93 samples;

(4) Running k-means algorithm [9] on these samples.

For the $PM_{2.5}$ network with adjacency matrix *B*, we use motif M_{10} to identify the potential pollution victims.

After performing the motif-based higher-order spectral clustering algorithm, the cities are clustered into 5 clusters (Figure 3). In Figure 3, each dot indicates a city and the dots with the same color belong to the same cluster.

We have selected two representative clusters, cluster 4 and cluster 5 as illustrated below:

In cluster 4, there are 13 cities. From the Figure 4, a city with more in-direction arrows which means it is potential pollution victim. This reveals Xinyang is vulnerable to pollution in other cities in $PM_{2.5}$ transport.

In cluster 5, there are 29 cities which covering most of Jing-Jin-Ji region. They are shown in Figure 6. We can observe that some cities of x-axis direction correspond to more dots in the vertical line, which indicates that they have more in-direction arrow lines than other cities in the network subgraph of the cluster. The ID number is 1, 8, 11, 13, 22, they are potential pollution victims in cluster 5. In the research of $PM_{2.5}$ transmission, they play the role of pollution victims and suffering pollution transmission from surrounding cities, which is related to the geographical location of the city and the meteorological conditions in the vicinity and the industrial conditions of the city [11].

Therefore, the motif can help us better identify the potential victims. For example, we have drawn the Jing-Jin-Ji region, and Beijing is vulnerable to other cities in the region, resulting in a widespread concentration of $PM_{2.5}$ in winter in Beijing. In addition, we also found that Tangshan, Handan, Baoding as the gathering place of large-scale industries are the main contributors of pollution transmission, which provides us with a strong guidance to solve the pollution.

5. Discussion

In this paper, we apply the motif-based clustering algorithm to cluster 93 cities into different groups. We used motif M_{10} to determine potential pollution victims, especially in the Jing-Jin-Ji region where $PM_{2.5}$ pollution frequently occurs. We find Beijing, Tianjin and Tangshan are potential pollution victims. Meanwhile, we found that Handan, Shijiazhuang and Baoding are playing the role of contributors in the process of $PM_{2.5}$ transmission. This provides a favorable guide for the government to deal with pollution.

$PM_{2.5}$ transmission between cities is a complex process. In this paper, we consider some meteorological conditions and social factor, such as wind speed, wind direction and GDP and so on. In the future research work, we will add more meteorological factors such as air pressure, relative humidity, and also more social factors, such as growth of population, vehicle exhaust etc.

Acknowledgement

We are grateful for discussion with Professor Haiyan Wang who have read the whole manuscript and suggested numerous improvement.

References

[1] Yuxiang Wei, Yan Yin, Weifen Yang, Dongmei Rui and Weiqi Hang, Analysis of the pollution characteristics & influence factors of $PM_{2.5}$ in Nanjing area, Journal of Environmental Science and Management 34(9) (2017), 29-34.

[2] Jing Xu, Guoan Ding, Peng Yan, Shufeng Wang, Zhaoyang Meng, Yangmei Zhang, Yuche Liu, Xiaoling Zhang and Xiangde Xu, Componential characteristics and sources identification of $PM_{2.5}$ in Beijing, Journal of Applied Meteorological Science 18(5) (2007), 645-654.

[3] Weizhen Hou, Zhengqiang Li, Yuhuan Zhang, Hua Xu, Ying Zhang, Kaitao Li, Donghui Li, Peng Wei and Yan Ma, Using support vector regression to predict PM_{10} and $PM_{2.5}$, IOP Conference Series: Earth and Environmental Science 17(1) (2014), 1-6.

DOI: https://doi.org/10.1088/1755-1315/17/1/012268

[4] Yong Yang and George Christakos, Spatiotemporal characterization of ambient $PM_{2.5}$ concentrations in Shandong province (China), Environmental Science and Technology 49(22) (2015), 13431-13438.

DOI: https://doi.org/10.1021/acs.est.5b03614

[5] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, Network motifs: Simple building blocks of complex networks, Science 298(5594) (2002), 824-827.

DOI: https://doi.org/10.1126/science.298.5594.824

[6] Austin R. Benson, David F. Gleich and Jure Leskovec, Higher-order organization of complex networks, Science 253(6295) (2016), 163-166.

DOI: https://doi.org/10.1126/science.aad9029

[7] Yufang Wang, Haiyan Wang and Shuhua Zhang, A weighted higher-order network analysis of fine particulate matter $(PM_{2.5})$ transport in Yangtze river delta, Physica A: Statistical Mechanics and its Applications 496 (2018), 654-662.

DOI: https://doi.org/10.1016/j.physa.2017.12.096

[8] Tore Opsahl and Pietro Panzarasa, Clustering in weighted networks, Social Networks 31(2) (2009), 155-163.

DOI: https://doi.org/10.1016/j.socnet.2009.02.002

[9] Andrew Y. Ng, Michael I. Jordan and Yair Weiss, On spectral clustering: Analysis and an algorithm, Advances in Neural Information Processing Systems (2002), 849-856.

[10] Lisa K. Baxter and Jason D. Sacks, Clustering cities with similar fine particulate matter exposure characteristics based on residential infiltration and in-vehicle commuting factors, Science of the Total Environment 470-471 (2014), 631-638.

DOI: https://doi.org/10.1016/j.scitotenv.2013.10.019

[11] Shudong Zhou, Weiqing Ouyang and Jihong Ge, Study on the main influencing factors and their intrinsic relations of $PM_{2.5}$ in Beijing-Tianjin-Hebei, China Population, Resources and Environment 27(4) (2017), 102-109.

■