Traffic Flow Characteristics mining of Time Series Data Based on Phase Space Reconstruction Using Complex Networks

Xue Xing [1], Yaqi Zhai [1], Zhongtai Jiang[2], Xiaoyu LI[1]

[1]*Dept. of College of Information and Control Engineering, Jilin Institute of Chemical Technology, China*

[2]*College of Transportation, Jilin University, China*

Abstract

Traffic flow time series is vital for mining the traditional statistical characteristics by using the theory of statistics and machine learning when its identity is a special time series. The network analysis of the traffic flow time series, who uses the complex network of time series analysis method, is designed to inquire into the special law of traffic flow time series which uses its visualization characteristics. Through the network analysis of traffic data flow, the connotation of traffic data flow can be revealed, and the relationship between all data and some data can be further studied. Therefore, it is constructed by combination with the phase space reconstruction theory. The phase space trajectory may be squeezed and the structure of attractor may change. We need to use C-C method to estimate the time delay according to the characteristics of integral parameters, and use G-P algorithm to estimate the embedding dimension to avoid it. This study can effectively reveal the motion law of the system. After constructing the complex network of traffic flow time series with various traffic parameters, the degree distribution, clustering coefficient and modularization of the representative critical threshold corresponding network are statistically analysed. The analysis results show that the new networked structure of traffic flow time series proposed in this study has strong advantages, and its core is phase space reconstruction, which can well reflect the information space of traffic dynamic fluctuation. The time series networking method based on phase space reconstruction has become a new approach to inquire into the characteristics of traffic flow time series. The degree distribution of the actual multi-traffic parameter time series construction network satisfies the characteristics of a Gaussian distribution. Their average clustering coefficients have attenuation characteristics, and their modularization degree is obvious.

Keywords: *Traffic flow analysis, phase space reconstruction, complex network; structural properties*

I. Introduction

The research of traffic flow time series has always been in the basic position in the dynamic analysis of traffic networks. With the increasing demand for transportation, the demand for intelligent transportation is also increasing. Therefore, it is necessary to deeply mine the temporal features in the traffic system. Previous studies on traffic flow analysis principally included traffic flow prediction as well as nonlinear analysis. In terms of prediction, we mainly focus on the application of statistical methods and machine learning methods to analyse traffic flow[1-3]. Previously, nonlinear analysis, researchers basically utilize chaos theory will investigation the nonlinear characteristics for traffic flow sequences. Traffic flow has many influencing factors and dynamic characteristics. How its nonlinear characteristics affect the traffic state has become a deep-seated problem in the field of transportation engineering. The research of time series tends to use spatial structure to analyze time series relationship. Reference [4, 5] lists three methods, which can construct complex networks from time series: cyclic time series construction method, visual method and phase space reconstruction (PSR) method. The method to construct the periodic time series is mainly applied in the field of life sciences. Reference[6] improved the method of mapping heart rhythm time series to complex networks and found the invisible sequence structure in traditional

ISSN: 0010-8189

algorithms. References[7] analysed the viewable algorithm of time series mapping into the network in detail and described the structure of the time series and its potential dynamic characteristics with graph theory. Reference[8] realized the purpose of reconstructing time series, which used phase space reconstruction to map them to complex networks. In summary, the statistical characteristics of complex networks provide an effective method for exploring the dynamic fluctuation of time series.

Traffic flow time series, as a special time series, have quasi-periodic and non-linear characteristics. The analysis of traffic flow time series mainly focuses on the state evaluation of traffic parameters and traffic flow prediction. Reference [9] presents a versatile parameter estimation methodology for univariate traffic condition forecasting using the Kalman filter. Reference [9] realized the goal of analyzing and predicting the average velocity of a given position, and a time series forecasting model is established and this model uses dynamic linear model (DLM).Reference [10-13] use a variety of machine learning methods to extract the features of traffic flow timing data. There are relatively few studies on the relationship between the structure of traffic timing data and the evolution of traffic state.

The network analysis of traffic data flow can present the connotation of traffic data flow, grasp the relationship between the whole and part of the data and excavate the hidden rules in big data to reduce the frequency of malignant traffic accidents and road congestion rates. Unlike the traditional methods of studying the traffic flow sequence, based on the complex network method, it opens a new perspective for exploring the characteristics of the traffic flow timing data. This paper analyzes the traffic flow in view of the aspect of phase space, and obtains the traffic characteristics from the characteristics of time sequence structure. Phase space reconstruction (PSR) based on time series provides a new analytical approach for traffic engineering.

II. Time Series Network Method Based on PSR Theory

As one of the methods to construct complex networks in time series, the complex network structured by view method retains some features of the original time series[14-16], but breaks the time series relationship of the original sequence. In addition, since the construction characteristics of the periodic method are mainly applied in the field of life sciences and the characteristics of various methods to construct complex networks in the time series are considered, this study selects the PSR theory to study the time series of traffic flow. PSR theory recommends how to recover the attractor of dynamic system to determine the time series, which stretches and folds the trajectory in high-dimensional space, and analyzes its dynamic properties [17-19]. The most widely used PSR method at this stage is the time delay state space reconstruction method. In a one-dimensional time series, the time-delay embedding method is used to construct a m-dimensional phase space by embedding dimension m and the delay time τ. Equation (1) is shown. Each point of the multidimensional state space is correspondingly mapped by a vector composed of delay points. A new approximate phase space can be rebuilded by repeating multiple times. The dimension of phase space is determined by the number of delay points of time series. This method is an extension of Takens theorem, i.e., that is to say, the condition of recovering the corresponding space is to find an appropriate embedding dimension m, which is also the premise of PSR.

$$Y_i = \{x_i, x_{i+\tau}, ..., x_{i+(m-1)\tau}\}, i = 1, 2, ..., N-(m-1)\tau \quad (1)$$

Selecting the appropriate delay time τ and embedding dimension m is the core of the method application. According to the assumption of Takens theorem, an infinite one-dimensional time series without noise-free leads to the arbitrariness of τ and m.The premise of the analysis is that the infinite length and noise-free time series are not consistent with reality. The key to the high quality and good application effect of chaotic time series is to apply reasonable time delay and embedding dimension. At present, the main methods to estimate the time delay parameters include the complex autocorrelation method, mutual information method and C-C method[16, 20]. For the embedding dimension m, if m is sufficiently large, the motion law of the chaotic system can be revealed. However, an m that is too large will cause many operations, so the optimal embedding dimension should be the minimum value that satisfies Takens theorem.

III. Complex Network Model of The Traffic Flow Timing Data based on PSR

3.1 Network model of traffic flow timing data

According to aforementioned PSR theory, a time series phase space is reconstructed to construct a network. The core of the construction is to determine the three elements of the network: the nodes of the network, edges between nodes, and connection relationship between nodes. Reconstructed sequence Y_i is used as the node, and the connection relationship between the nodes determined by reconstructed sequences Y_i and Y_j is determined by correlation coefficient C_{ij} between reconstructed sequences Y_i and Y_j. The correlation coefficient is defined in formula (2).

$$C_{ij} = \frac{\sum_{k=1}^{L}[Y_i(k) - \langle Y_i \rangle] \cdot [Y_j(k) - \langle Y_j \rangle]}{\sqrt{\sum_{k=1}^{L}[Y_i(k) - \langle Y_i \rangle]^2} \cdot \sqrt{\sum_{k=1}^{L}[Y_j(k) - \langle Y_j \rangle]^2}} \tag{2}$$

In the formula, L is the dimension of the vector, $\langle Y_i \rangle = \sum_{k=1}^{L} Y_i(k)/L$, and $\langle Y_j \rangle = \sum_{k=1}^{L} Y_j(k)/L$.

The connection relationship of network node is determined by the relationship between correlation coefficient C_{ij} and connection threshold r_c. If correlation coefficient C_{ij} is within the threshold control range, the nodes determined by reconstructed sequences Yi and Y_j have edge associations; otherwise, the node determined by reconstructed sequences Y_i and Y_j have no associated edges. Thus, network adjacency matrix M is determined, and the composition of the association matrix is shown in formula (3).

$$M_{ij} = \begin{cases} 1 & |C_{ij}| \geq r_c \\ 0 & |C_{ij}| < r_c \end{cases} \tag{3}$$

Considering that the main parameters of the estimated traffic state are the traffic flow, time occupancy rate and regional average speed obtained from traffic detection, the actual time series of the traffic flow, time occupancy time series, and regional average speed obtained through data preprocessing are used. The time series is networked. The construction process of traffic flow time series network will be illustrated by taking time series of traffic flow as an example. The following figure can directly and clearly illustrate the method of constructing the traffic flow time series network based on PSR, as shown in Fig.1.

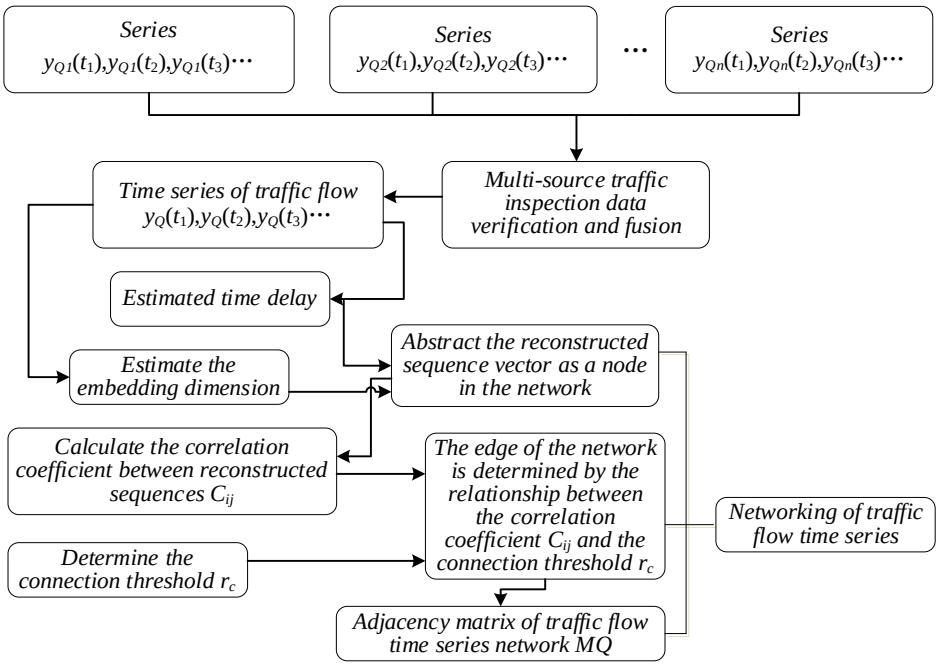

Fig. 1: Construction flowchart of the time series network of traffic flow based on phase space reconstruction.

3.2 Parameter estimation for PSR

Because the time series obtained in the actual detection must have errors and noise, the selection of the time delay must prevent the phase space trajectory from being too small and avoid structural changes of the attractor caused by a time delay that is too large. Among various methods to estimate the delay-time, the C-C method proposes the concept of related integrals and estimates the time delay based on the integral parameters. It is vital to determine the parameters of phase space reconstruction of series, which will be determined by C-C method.

Based on the principle that the divided time series cannot intersect, t time series will be divided, denoted by $\{x_i, i=1,2,...,N\}$. Compute the test statistics for $S(m,N,r,t)$ to estimate time delay t. First, define $C(m,N,r,t)$, which mainly counts the number of phase space vector distances in the vicinity of the radius, as shown in formula (4).

$$C(m,N,r,t)=\frac{2}{M(M-1)}\sum_{1\le i<j\le M}\Theta(r-\left\|Y_i-Y_j\right\|) \tag{4}$$

In the formula, $\Theta(\cdot)$ is the Heaviside step function, which is $\Theta(a)=\begin{cases}0 & if \quad a<0\\ 1 & if \quad a\ge 0\end{cases}$;*m* represents the embedding dimension under the premise of phase space; *N* denotes the length of time series; The number of vectors in phase space is represented by $M=N-(m-1)t$; *R* denotes the neighborhood radius.We define $S(m,N,r,t)$ as the test statistic of the time series, which is shown in formula (5). The first zero point of $S(m,N,r,t)\sim t$, which can intuitively reflect some characteristics of sequence or time point, primarily autocorrelation with the smallest amount of appropriate distance within radius r, is taken as the most time-delayed point. At present, the points representing the rebuilded phase space are nearest to a symmetric probability distribution, that is, uniform distribution. Therefore, the difference between maximum and minimum radii is selected as the statistical quantity, as shown in formula (6). Additionally, we define indicator $S_{cor}(t)$, and its

minimum value is the embedded window width of the time series, as shown in formula (7). $S(m,N,r,t)$, $\Delta S(m,t)$ and $S_{cor}(t)$ show three evaluation indicators, which are used to determine t as time delay, are the reference indicators of C-C method.

$$S(m,N,r,t)=\frac{1}{t}\sum_{s=1}^{t}[C_s(m,N/t,r,t)-C_s^m(1,N/t,r,t)] \tag{5}$$

In the formula, $C_S(\cdot)$ denotes the correlation score obtained after the S-th subsequence of the experiment.

$$\Delta S(m,t)=\max\{S(m,r_i,t)\}-\min\{S(m,r_j,t)\},i\neq j \tag{6}$$

$$S_{cor}(t)=\Delta\bar{S}(t)+\left|\bar{S}(t)\right| \tag{7}$$

In the formula, the mean value of $\Delta S(m,t)$ is represented by $\Delta\bar{S}(t)$, and the average value of $S(m,r,t)$ is represented by $\bar{S}(t)$.

For the embedding dimension m, if m is sufficiently large, the motion law of the chaotic system can be revealed. However, if m is too large, it will cause many calculations.Therefore, the determination of optimal embedding dimension is consistent with the purpose of Grassberger and procaceia G-P algorithm. Inspired by this, this method is used to obtain the the minimum value satisfying Takens theorem.

Before calculating the correlation dimension, first, we define the correlation integral, as shown in formula (8).

$$C(r)=\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=i+1}^{M}\Theta(r-\left\|Y_i-Y_j\right\|) \tag{8}$$

In the formula, $\Theta(\cdot)$ is the Heaviside step function, which is $\Theta(a)=\begin{cases}0 & if \quad a<0\\ 1 & if \quad a\geq 0\end{cases}$; M is the number of phase space vectors.

The relationship between linked points $C(r)$ and r is shown in formula (9).

$$\lim_{r\to 0}C(r)\propto r^D \tag{9}$$

$$D=\lim_{r\to 0}\frac{\ln[C(r)]}{\ln r}$$

We simplify correlation dimension *D* as by the above approximation and calculate *D (m)* for each m. When m increases, *D (m)* tends to be stable. According to correlation dimension *D* and Takens theorem, embedding dimension m is determined.

IV. Case Analysis

The data source of the experiment was selected from the traffic demonstration area of Xiamen City on January 24, 2015. The multisource traffic detection data (including induction coil detection data, geomagnetic detection data, and bayonet video detection data) on typical roads in the period from 00:00 to 23:00. The time series of multiple traffic parameter flows is shown in Fig. 2.

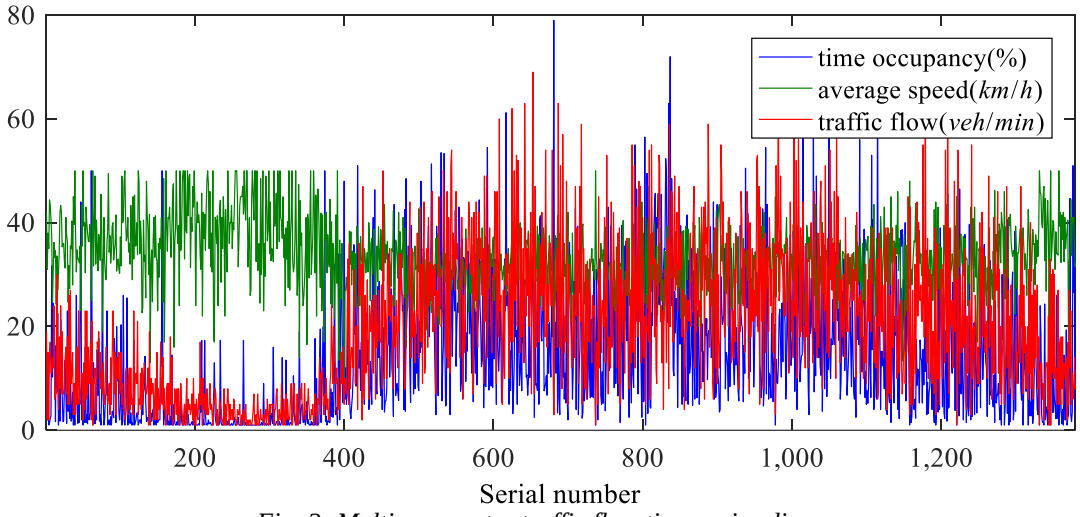

Fig. 2: Multi-parameter traffic flow time series diagram

4.1 Estimated time delay τ

Extract the data of the time occupancy of the traffic flow parameters, the data of the average speed of the traffic flow parameter area, and the data of the traffic flow parameters from the data set. The C-C algorithm is used to calculate the relevant integrals under different time delays for the three types of traffic parameter time series. Below is an application example to solve the optimal time delay.

Using the obtained time occupancy time series, calculate $S(m,N,r,t)$, $\Delta S(m,t)$ and $S_{cor}(t)$ of the phase space reconstruction sequence that traverses during the time delay. The index curve is shown in Fig.3 (a). Compare the corresponding $S(m,N,r,t)$, $S_{cor}(t)$, and $S_{cor}(t)$ curves in the figure, and select the optimal value as $\tau_O=8t$. Using a time series of traffic flows, calculate $S(m,N,r,t)$, $\Delta S(m,t)$ and $S_{cor}(t)$ of the phase space reconstruction sequence under traversal during time delay. The index curve is shown in Fig.3 (b). Compare the corresponding $S(m,N,r,t)$, $\Delta S(m,t)$, and $S_{cor}(t)$ curves in the figure and select the optimal value as $\tau_Q=11t$. Using a time series of regional average speed, calculate the $S(m,N,r,t)$, $\Delta S(m,t)$ and $S_{cor}(t)$ of the phase space reconstruction sequence under traversal during time delay. The index curve is shown in Fig. 3 (c). Compare the corresponding $S(m,N,r,t)$, $\Delta S(m,t)$, and $S_{cor}(t)$ curves in the figure and select the optimal value as $\tau_V=4t$.

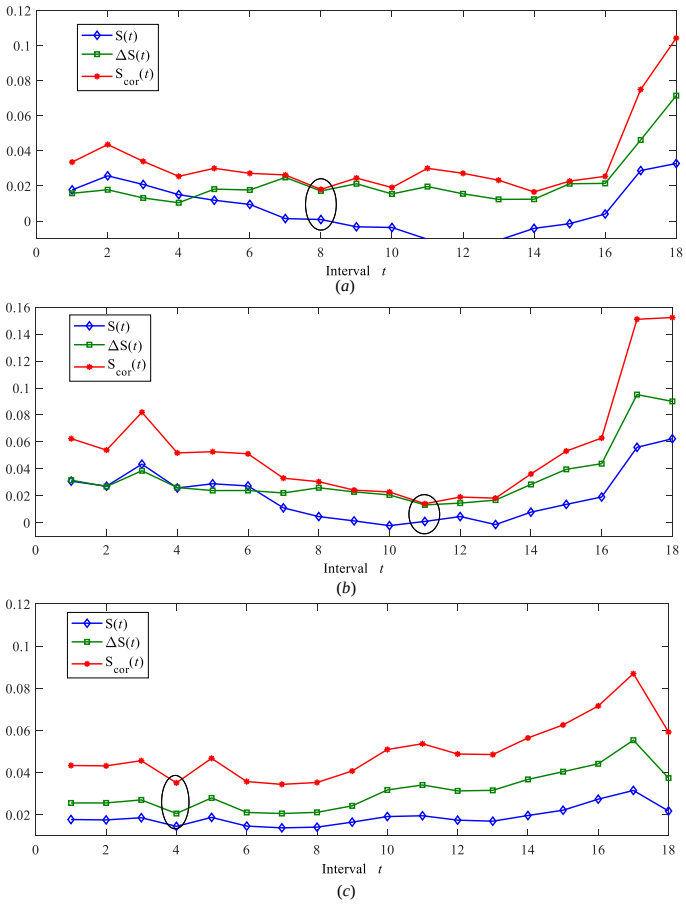

Fig. 3: Comparison of the indicator curves of the three parameter time series in Fig. 2.

4.2 Estimated embedding dimension m

The factors that decide to use the C-C algorithm to obtain the delay of the corresponding sequence are as follows: the time series of parameters with occupancy characteristics, the time series of parameters with the average speed, and the time series of parameters with traffic flow characteristics and traffic flow.

Calculate the associated integral $C(r)$ for the corresponding sequence according to the determined corresponding time delay. Compare the correlation curves of $\ln C(r)$ and $\ln r$ and determine the correlation dimension by formula $D = \lim_{r \to 0} \frac{\ln[C(r)]}{\ln r}$. The optimal embedding dimension m can make $D(m)$ converge to a constant, which is the saturated embedding dimension.

The experiment is accompanied by the change *D (m)* of the embedding dimension m of the time occupancy data. Here we use G-P algorithm to observe it. Fig. 4 shows the correlation curves of dimension m from 1 to 30 during phase space reconstruction of the sequence. The embedded graph in Fig. 4 represents the change curve of *D* value with the change in dimension m. Position character *D* in the circle in the figure is generally stable, so the saturated embedded dimension $m_O = 7$ is determined.

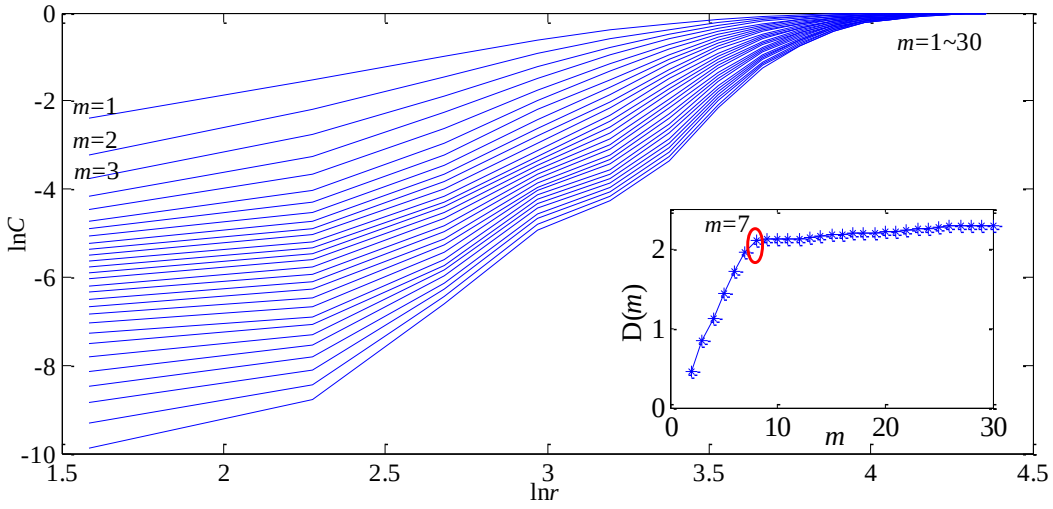

Fig. 4: $\ln C(r)$ *and* $\ln r$ *correlation plots for the time occupancy series.*

Fig. 5 shows the distribution of $\ln C(r)$ and $\ln r$ and their correlation curve under the condition of rebuilding the phase space of traffic flow time series, where the dimension of m is 1-30. The embedded figure in Fig. 5 shows the inflection point in *D* with dimension m. Position character *D* in the circle in the figure is generally stable, so the saturated embedding dimension $m_Q = 6$ is determined.

Fig. 6 shows the distribution of $\ln C(r)$ and $\ln r$ and their correlation curve when m is 1-30 dimension. At the same time, these distributions occur in the process of phase space reconstruction of regional average velocity time series. The embedded figure in Fig. 6 shows the change in *D* with dimension m. Position character *D* in the circle in the figure is generally stable to determine the saturation embedded dimension $m_V = 6$.

4.3 Characteristic analysis of experimental data

The main function of phase space reconstruction is to recover the attractor of dynamic system in high-dimensional space, so it is also a key method to analyze nonlinear time series. With the participation of time delay estimation τ and embedding dimension m, the actual traffic flow sequence can be reconstructed in phase space.According to the phase-space reconstructed traffic flow time-series network construction process, connection threshold r_c determines the network structure to be constructed. By comparing the network matrices with different thresholds, it can be intuitively found that a higher the threshold value corresponds to more prominent highly associated nodes, and the research finds that a reasonable connection threshold helps analyse the original clustering characteristics of the network.

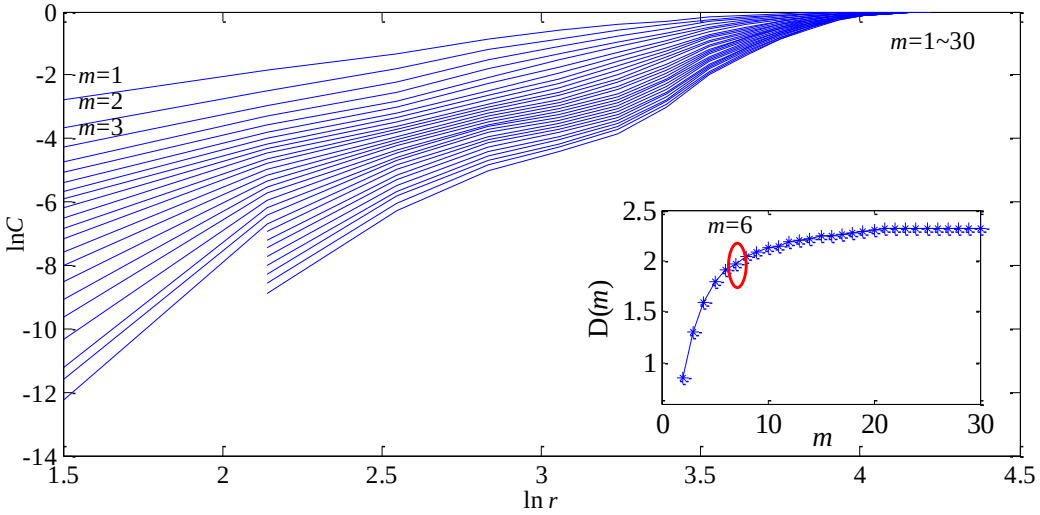

Fig. 5: $\ln C(r)$ *and* $\ln r$ *correlation curves for traffic flow series.*

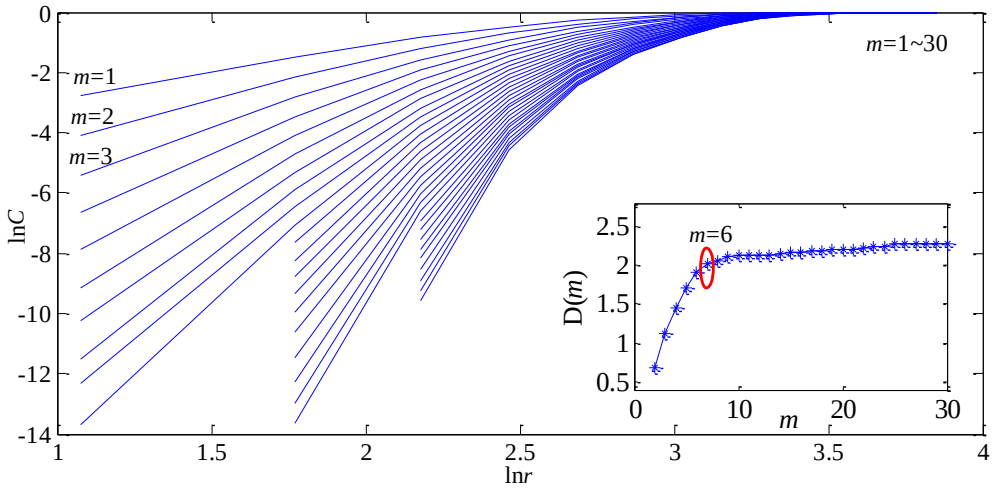

Fig. 6: $\ln C(r)$ *and* $\ln r$ *correlation curves for the regional average velocity series.*

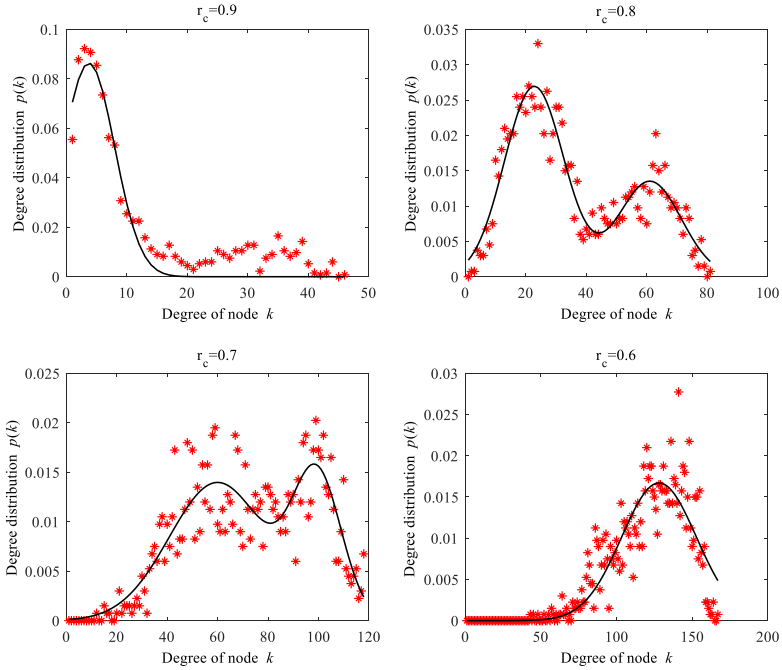

Fig. 7: Distribution of network node degrees corresponding to the time occupancy series under different connection thresholds r_c.

The degree distribution of network nodes is analysed for each traffic flow time series to determine the threshold range of each type of traffic parameter time series network feature analysis. The network node degree distribution map of the traffic flow time occupancy sequence with different thresholds r_c is selected, as shown in Fig. 7. Select the network node degree distribution map of the traffic flow time occupancy sequence with different threshold r_c, as shown in Fig. 8. Select the node speed distribution of the average speed sequence network with different thresholds r_c, as shown in Fig. 9.

The degree distribution of network nodes is analysed for each traffic flow time series to determine the threshold range of each type of traffic parameter time series network feature analysis. The network node degree distribution map of the traffic flow time occupancy sequence with different thresholds r_c is selected, as shown in Fig. 7. Select the network node degree distribution map of the traffic flow time occupancy sequence with different threshold r_c, as shown in Fig. 8. Select the node speed distribution of the average speed sequence network with different thresholds r_c, as shown in Fig. 9.

In the traffic flow time occupancy sequence network, the fitting of the node degree distribution of the network with the threshold r_c = 0.60 conforms to the Gaussian distribution; the fitting line of the node degree distribution is y_1=0.0167*exp(-((d_1-128.1)/35)^2)); the fiducial intervals of the fitted estimation parameters are [0.01565, 0.01776], [126.2, 130] and [32.12, 37.88]. The node degree distribution curve with threshold r_c = 0.70 has two peaks and no longer conforms to a single Gaussian distribution. Hence, a double superposition fit from the Gaussian distribution is selected, and the node degree distribution fitting line is y_2=0.01408*exp(-((d_2-99.53)/13.63)^2)+ 0.01397*exp(-((d_2-59.98)/27.02)^2); the fiducial intervals of the fitted estimation parameters are [0.01201, 0.01615], [97.74, 101.3], [10.99, 16.26], [0.01275, 0.01519], [56.93, 63.04] and [22.58, 31.47]. The network node degree distribution with threshold r_c = 0.80 is also selected for the double superposition of the Gaussian distribution for fitting. The fitting line of the node degree distribution is obtained by formula $y3$=0.02693*exp(-((d_3-22.77)/13.93)^2)+0.01351*exp(-((d_3-61.03)/14.72)^2), and the fiducial intervals for the fitted estimation parameters are [0.02525, 0.02861], [22.02, 23.53], [12.77, 15.1], [0.01185, 0.01516], [59.47, 62.58], and [12.23, 17.21]. The fitting degree of the node degree distribution of threshold r_c = 0.9 fits the Gaussian

distribution, the fitting line of the node degree distribution is $y4$=0.0864*exp(-((d$_4$-3.723)/6.045)^2), and the fiducial intervals of the fitted estimation parameters are [0.07862, 0.09418], [2.933, 4.512], and [4.937, 7.152]. All these estimates have 95% confidence.

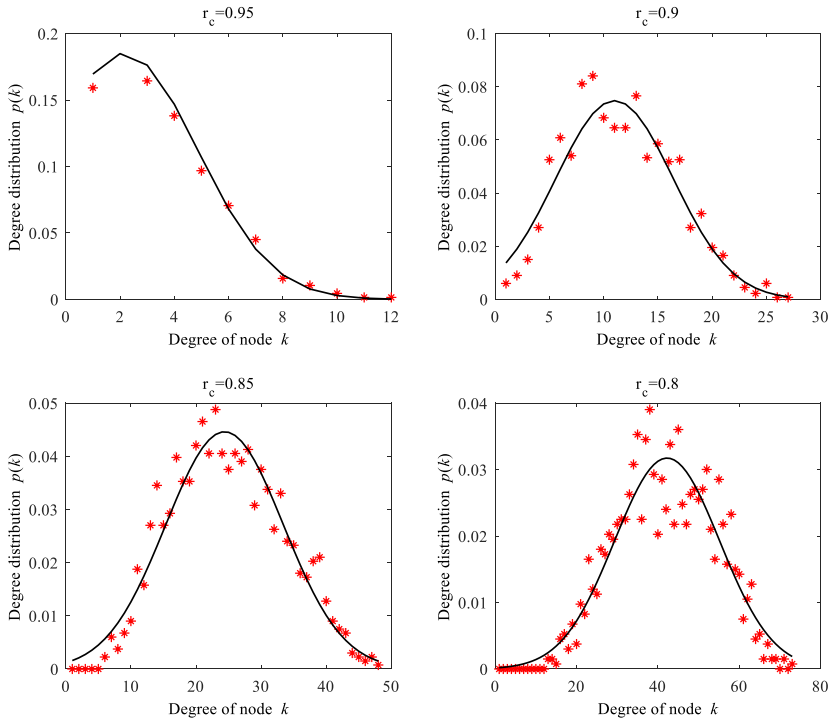

Fig. 8: Distribution of network node degrees corresponding to traffic flow series under different connection thresholds r_c.

In the traffic flow sequence network, the fitting of the node degree distribution of the network with threshold r_c = 0.80 conforms to the Gaussian distribution, the fitting line of the node degree distribution is y_1=0.03174*exp(-((d_1-42.25)/18.29)^2), and the fiducial intervals of the fitted estimation parameters are [0.02983, 0.03365], [41.36, 43.15] and [17.01, 19.57]. The fitting of the node degree distribution of the network with threshold r_c = 0.85 conforms to the Gaussian distribution, the fitting line of the node degree distribution is y_2=0.04463*exp(-((d_2-24.39)/12.8)^2), and the fiducial intervals for the fitted estimation parameters are [0.0418, 0.04746], [23.73, 25.05], and [11.86, 13.75]. The network node degree distribution with threshold r_c = 0.90 is also selected to fit the Gaussian distribution. The fitting line of the nodal degree distribution is y_3=0.07478*exp(-((d_3-11)/7.679)^2), and the fiducial intervals of the fitted estimation parameters are [0.0678, 0.08175], [10.41, 11.59], and [6.811, 8.548]. The fitting of the node degree distribution of the network with threshold r_c = 0.95 conforms to the Gaussian distribution. The fitting line of the node degree distribution is y_4=0.1852*exp(-((d_4-2.144)/3.854)^2). The confidence interval of the fitted estimation parameter is [0.1762, 0.1943], [1.802, 2.486] and [3.413, 4.295]. All these estimates have 95% confidence.

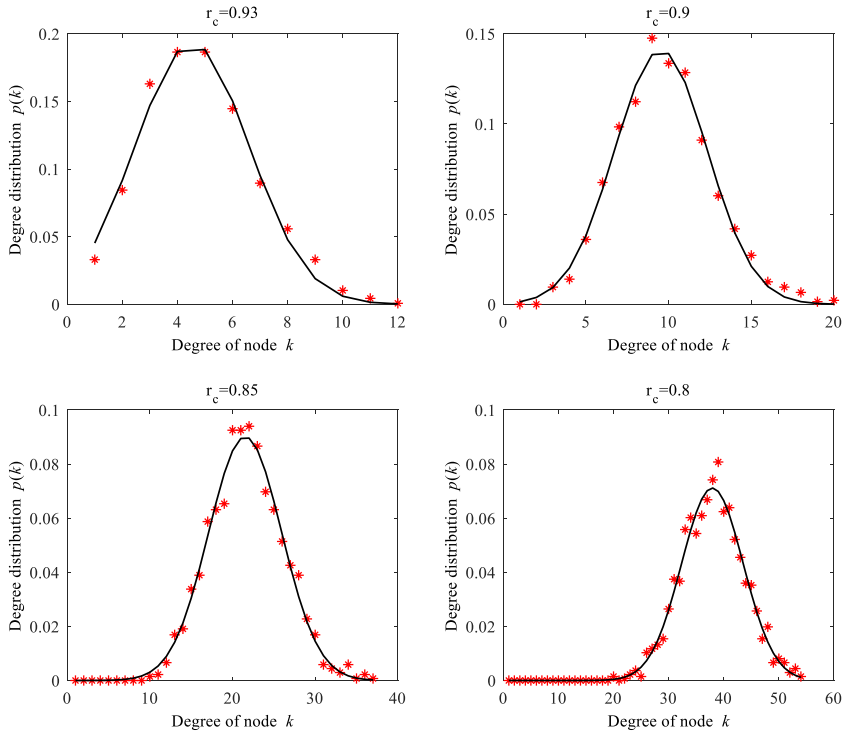

Fig. 9: Distribution of network node degrees corresponding to the average velocity series under different connection thresholds r_c.

In the traffic area average speed sequence network, the fitting of the node degree distribution of the network with threshold r_c = 0.80 conforms to the Gaussian distribution, and the fitting line of the node degree distribution is y_1= 0.07115*exp(-((d_1- 37.94)/7.923)^2). The fiducial intervals of the fitted estimation parameters are [0.06866, 0.07365], [37.71, 38.16], and [7.602, 8.244]. The fitting of the network node degree distribution with threshold r_c = 0.85 conforms to the Gaussian distribution, and the fitting line of the node degree distribution is y_2= 0.0901*exp(-((d_2-21.53)/6.267)^2). The fiducial intervals of the fitted estimation parameters are [0.08679, 0.09342], [21.35, 21.72], and [6.001, 6.533]. The network nodes with threshold r_c = 0.90. Gaussian distribution is also used to fit the degree distribution. The fitting line of the node degree distribution is y_3=0.141*exp(-((d_3-9.534)/3.965)^2). The confidence interval to fit the estimated parameters is [0.1348, 0.1472]. [9.392, 9.676] and [3.764, 4.166]; the network node degree distribution corresponding to threshold r_c = 0.93 in Fig. 8 fits the Gaussian distribution, and the fitting line of the node degree distribution is y_4=0.1932*exp(-((d_4-4.533)/2.933)^2). The fiducial intervals of the fitted estimation parameters are [0.1792, 0.2071], [4.36, 4.706], and [2.678, 3.189]. All these estimates have 95% confidence.

In three types of traffic parameter time series networks, if threshold r_c is too large, the Gaussian distribution characteristics of the original time series network degree distribution will be destroyed. When the appropriate threshold range increases, the peaks of the network degree distribution of each parameter traffic flow time series gradually approach the y-axis. The specific time series characteristics of each type of traffic parameter are as follows: The network degree distribution of the traffic flow time occupancy sequence under threshold $r_c \in (0.65, 0.80)$ has an obvious double-Gaussian distribution overlap feature. The double peak gradually moves away from the y-axis when the threshold decreases, and the high peak of the double peak moves away from y when the threshold decreases. The speed of the shaft is faster. However, the single-Gaussian distribution peak asymptotic degree of the network degree distribution at threshold $r_c \in [0.80, 1)$ is zero. When the threshold is $r_c \in [0.90, 1)$, the complete Gaussian distribution feature is not satisfied. The network degree distribution of the

traffic flow sequence has obvious Gaussian distribution characteristics, the peak value decreases with the decrease in threshold r_c, and the complete Gaussian distribution characteristics are not satisfied at threshold $r_c \in [0.9,1)$. The network degree distribution of the average speed sequence in the traffic area has obvious Gaussian distribution characteristics, the peak value decreases with the decrease in threshold r_c, and the complete Gaussian distribution characteristics are not satisfied at threshold $r_c \in [0.93,1)$.

The local characteristics of complex networks are quantitatively measured by the clustering coefficient of each node. The value of each clustering coefficient depends on the ratio of the number of adjacent nodes and the maximum possible connection points. Fig. 10 shows that the average clustering coefficient of the three types of traffic parameter time series networks decreases when the threshold increases. The overall trend of the average clustering coefficient of the three types of networks in the example can be described as a steep decline after a certain steady trend. The specific performance of the curve is that when the time occupancy time series network is at threshold $r_c = 0.82$, the traffic flow time series network is at threshold $r_c = 0.93$, and the regional average speed time series network is at threshold $r_c = 0.92$, the average clustering coefficient sharply decreases. Thus, the correlation coefficient of the reconstructed sequence accounts for the largest proportion at these thresholds. In addition, before the average clustering coefficient of the traffic flow time series network and the curve of the traffic flow time series network suddenly decrease, a pick-up process occurs, as shown by the circle in Fig. 10. The threshold in this range only limits the nodes with a degree of 1 (reconstructed sequence). Removing these nodes whose clustering coefficients tend to infinity makes the curve rise instead.

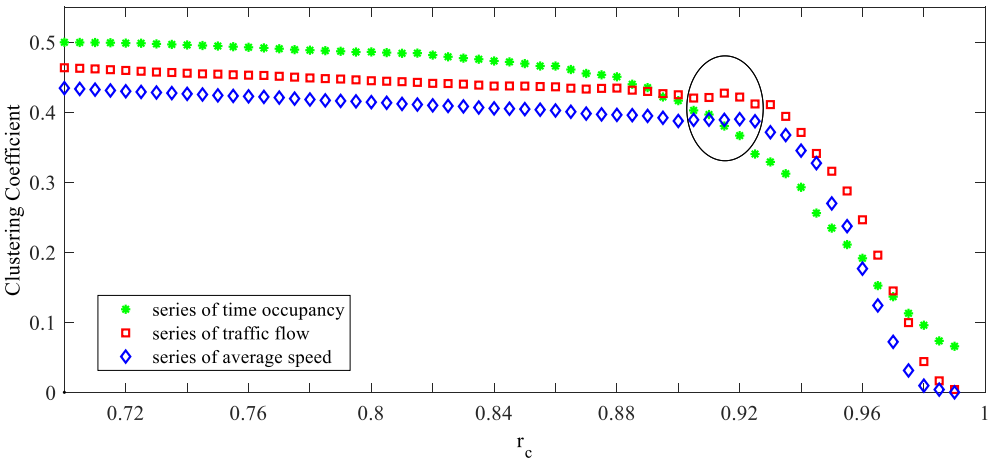

Fig. 10: Curve graph of average clustering coefficients of multiple traffic parameter time series networks under different thresholds

By combining the change law of the degree distribution under different connection thresholds and the change law of the average clustering coefficient under different connection thresholds, we select a suitable threshold for the instance time occupancy time series network [0.84, 0.87], a suitable threshold for the example traffic flow time series network [0.90,0.93], and a suitable threshold for the traffic flow time series network [0.89, 0.92]. We discuss the modularity of the traffic flow parameter time series. Taking the traffic flow time occupancy series network as an example, the connection thresholds r_c are selected to be 0.81, 0.84, 0.87, and 0.90. The corresponding network presents obvious modular characteristics, as shown in Fig. 11, and the corresponding modular coefficients of the network structure are 0.744, 0.765, 0.793, and 0.830. The modular structure of space shows that the time series of the traffic flow parameter time occupancy becomes 7 modules according to the spatial structure, which has a high

modularity feature. Since each phase space vector represents the time series of the traffic parameters in the unit of delay time, the 7 types of phase space vectors in the network structure have 7 types of corresponding modes for the time series of the traffic parameters, i.e., 7 types of modes in which the traffic state changes. The traffic congestion status in the traditional traffic flow time series analysis generally reflects the subjective feelings of traffic participants or satisfies the decision-making needs of management agencies. The specific classification is described by common index variables or corresponding segments of traffic parameters. The state of traffic congestion objectively exists, but due to different parameters selected for measurement or different subjects described, there is no uniform standard to classify the current road traffic conditions. PSR network structure analysis based on the actual traffic flow parameter time series can provide the information space division criteria for traffic state division.

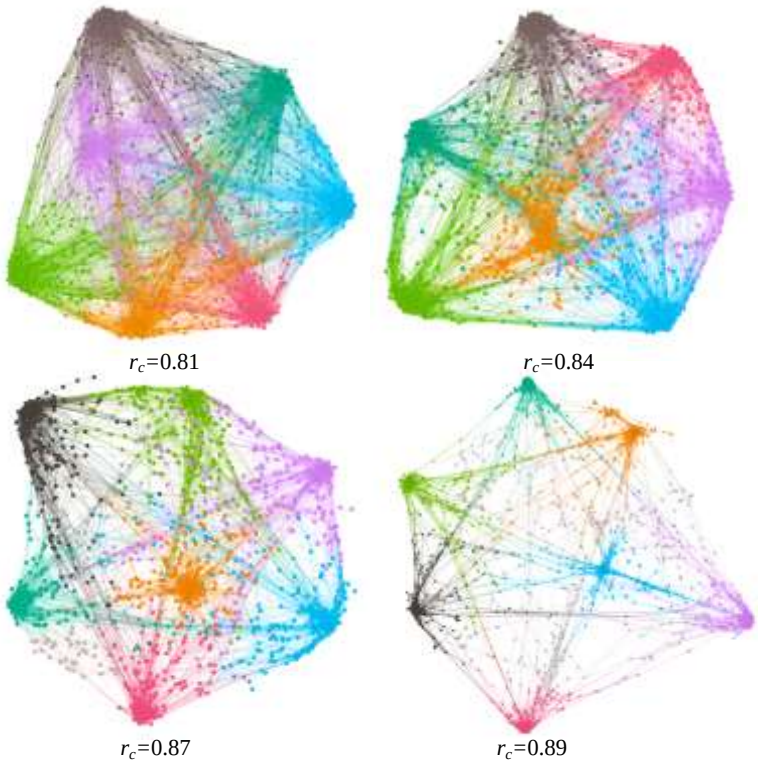

Fig. 11: Modular structure of a traffic flow time occupancy sequence network with different connection thresholds rc

V. Conclusions

The complex network is one of the most vital ways to analyze traffic flow sequence, which is different from the traditional traffic flow sequence analysis method. The major challenge of this study is how to deconstruct the time series to get its statistical characteristics. We use the phase space method to solve this problem. The chief idea of this method is to analyze the traffic time series network under different thresholds from the actual road network. The experiment provides a valuable perspective for studying the characteristics of traffic flow time series. This paper aims to reconstruct the time series to explore the attributes of traffic flow and determine the optimal critical threshold.

(1) Considering the phase space reconstruction requirements of the single-parameter traffic flow time series, the time delay and embedding dimension m of the three types of traffic parameter time series in the example are estimated. The network characteristics statistics (such as the modularity of the network structure, the average clustering coefficient, and the degree distribution) are performed on the PSR network of the three types of traffic

parameters with reasonable estimated parameters.

(2) The new structure which is reconstructed by the network phase space of traffic flow time series is the information space in traffic flow dynamics. The case analysis shows that the degree distribution of the traffic flow time series network with different parameters has Gaussian distribution characteristics, and the average clustering coefficient has attenuation characteristics and high modularity characteristics. At this stage, phase-space reconstruction is used to perform network analysis on the traffic flow time series. Subsequent research will explore the relationship between the visual method and this method to build a traffic flow time series network to obtain a deeper understanding of the traffic flow state.

Acknowledgements

This paper is funded by Projects of Jilin Provincial Education Department (JJKH20200253KJ).

References

[1] D.W. Xu, Y.D. Wang, Peng Peng, Shen Beilun, Zhang Deng and Haifeng Guo, “Real-Time Road Traffic State Prediction Based on Kernel-KNN”, in Transportmetrica (Abingdon, Oxfordshire, UK), vol. 16, no. 1, Jan. pp. 104-118, 2020.

[2] Doğan Erdem, “Analysis of The Relationship Between LSTM Network Traffic Flow Prediction Performance and Statistical Characteristics of Standard and Nonstandard Data”, in Journal of Forecasting, vol. 39, no. 8, pp. 1213-1228, 2020.

[3] T.Ma, Constantinos Antoniou and Tomer Toledo, “Hybrid Machine Learning Algorithm and Statistical Time Series Model for Network-Wide Traffic Forecast”, in Transportation research. Part C, Emerging technologies, vol. 111, pp. 352-372, 2020.

[4] A.A. Snarskii and I. V. Bezsudnov, “Phase transition in the parametric natural visibility graph”, in Physical Review. E, vol. 94, no. 4-1, pp. 42137, 2016.

[5] Somaye Karimi and Amir H. Darooneh, “Measuring persistence in a stationary time series using the complex network theory”, in Physica A-Statistical Mechanics and Its Applications, vol. 392, no. 1, pp. 287-293, 2013.

[6] I. V. Bezsudnov and A. A. Snarskii, “From The Time Series to The Complex Networks: The Parametric Natural Visibility Graph”, in Physica A-Statistical Mechanics and Its Applications, vol. 414, pp. 53-60, 2014.

[7] I. V. Bezsudnov and A. A. Snarskii, “Phase Transition in The Parametric Natural Visibility Graph.”, in Physical Review E, vol. 94, no. 4-1, pp. 042137, 2016.

[8] J.L. Li, Q.J Zhang, Y.M Zhang, X.J Wu, X.M Wang and Y.P Su, “Hidden Phase Space Reconstruction: A Novel Chaotic Time Series Prediction Method for Speech Signals”, in Chinese Journal of Electronics, Vol. 27, no. 6, pp.1221-1228, 2018.

[9] Shashank Shekhar and Billy M. Williams, “Adaptive Seasonal Time Series Models for Forecasting Short-Term Traffic Flow”, in Transportation Research Record, vol. 2024, no. 1,2007, pp.116-125, 2007.

[10] T. Ma, Constantinos Antoniou and Tomer Toledo, “Hybrid Machine Learning Algorithm and Statistical Time Series Model for Network-Wide Traffic Forecast”, in Transportation Research Part C, vol. 111, pp.352-372, 2020.

[11] W.D Zou and Y. Q Xia, “Back Propagation Bidirectional Extreme Learning Machine for Traffic Flow Time Series Prediction”, Neural Computing and Applications, vol. 31, no. 11, pp. 7401-7414, 2019.

[12] M. V. Peppa, D. Bell, T. Komar and W. Xiao, “Urban traffic flow analysis based on deep learning car detection From CCTV Image Series”, in ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-4, pp. 499-506, 2018.

[13] L. S Yin, Y. G He, X.P Dong and Z.Q Lu, “Multi-step Prediction Algorithm of Traffic Flow Chaotic

Time Series Based on Volterra Neural Network", in Journal of Computers, vol. 8, no. 6, pp 1480-1487, 2013.

[14] W.W Cao, X.N Feng, J.M Jia and H. Zhang, "Characterizing the Structure of the Railway Network in China: A Complex Weighted Network Approach", Journal of Advanced Transportation, vol. 2019-02-03 2019, pp. 1-10, 2019.

[15] Jacques J. F. Commandeur, Paul Wesemann, Frits Bijleveld, Voun Chhoun and Socheata Sann, "Setting Road Safety Targets in Cambodia: A Methodological Demonstration Using the Latent Risk Time Series Model", in Journal of Advanced Transportation, vol. 2017, pp. 1-9, 2017.

[16] A. Guillen, L. J. Herrera, G. Rubio, H. Pomares, A. Lendasse and I. Rojas, "New method for instance or prototype selection using mutual information in time series prediction", in Neurocomputing, vol. 73, no. 10,2009.

[17] A.G Zhang and Z Xu, "Chaotic Time Series Prediction Using Phase Space Reconstruction Based Conceptor Network", in Cognitive Neurodynamics, vol. 14, no. 6, pp. 849-857, 2020.

[18] Javier de Pedro-Carracedo, David Fuentes-Jimenez, Ana María Ugena and Ana Pilar Gonzalez-Marcos, "Phase Space Reconstruction from a Biological Time Series: A Photoplethysmographic Signal Case Study", in Applied Sciences, vol. 10, no. 4, Feb, pp. 1430, 2020.

[19] Z.Y Pu, "Network Security Situation Analysis Based on A Dynamic Bayesian Network and Phase Space Reconstruction", in The Journal of Supercomputing, vol. 76, no. 2, pp. 1342-1357, 2020.

[20] F Yuan, J Guo, Z.H Xiao, B. Zeng, W.Q Zhu and S.X Huang, "An Interval Forecasting Model Based on Phase Space Reconstruction and Weighted Least Squares Support Vector Machine for Time Series of Dissolved Gas Content in Transformer Oil", in Energies, vol. 13, no. 7,pp. 1687, 2020.

ISSN: 0010-8189