Fuzzy Systems and Data Mining IX
A.J. Tallón-Ballesteros and R. Beltrán-Barba (Eds.)

doi:10.3233/FAIA231023

A Method to Explore the Synchronous Changes of High-Traffic Events Based on Dynamic Networks

Yu-Wen HUANG[a], Li-Na WANG[a,b,1] and Chen-Rui ZANG[c]
[a]*College of Sciences, Inner Mongolia University of Technology, Hohhot, 010051, China*
[b]*Inner Mongolia Key Laboratory of Statistical Analysis Theory for Life Data and Neural Network Modeling, Hohhot, 010051, China*
[c]*Inner Mongolia Branch, China Unicom, Hohhot, 010050, China*

Abstract. With the rapid development of mobile communication technologies, the mobile network has evolved into a highly heterogeneous network structure. Based on dynamic networks, we mainly investigated a method to explore the synchronous changes of high-traffic events. Event coincidence analysis is used to quantify the concurrency of high-traffic events. A variety of network measures are used to analyze the dynamic spatio-temporal characteristics of high-traffic. The static network is constructed to analyze the synchronous influence area and the temporal and spatial characteristics of high-traffic of base station. Taking hour as the time window, the dynamic network is constructed to study the dynamic spatio-temporal variation rule of high-traffic and the interactive relationship of traffic between base stations. It is found that static network is a small-world network. The spatial connectivity of high-traffic events at the base station is high, and the spatial connectivity is not sensitive to temporal changes. The traffic of different base stations has interactive relation at the same time in different days.

Keywords. Complex networks; Event coincidence analysis; Dynamic networks; Communication system

1. Introduction

With the rapid development of mobile communication and sensor technology, a large amount of spatiotemporal data emerges. Spatiotemporal data is spatial data that has a temporal element and changes with time. This information has important application value for urban traffic planning, cellular network management, air forecasting, disaster prediction, and other fields. Observe the spatial distribution of spatio-temporal data through visualization methods, and intuitively analyze requirements.

Event coincidence analysis (ECA) is often applied to analyze the statistical relation between event sequences in spatiotemporal data to study the spatiotemporal

[1] Corresponding Author, Li-Na WANG, Mathematics Department, Inner Mongolia University of Technology, AiMin Road #49, Hohhot, China; E-mail: wanglina@imut.edu.cn.
This work was supported by the Natural Science Foundation of Inner Mongolia (Grant Number 2022LHMS01005); and the Fundamental Research Funds for the Directly Affiliated Universities of Inner Mongolia (Grant Number JY20220095).

characteristics of event occurrence. ECA was originally used to study possible statistical correlations between mechanisms of paleoclimate variability in Africa and human evolution[1]. Until 2016, Donges et al.[2] described in detail the specific algorithmic process of event coincidence analysis, and provided a framework for quantifying the statistical correlation strength and the directional time lag between event sequences. At present, event coincidence analysis has been successfully applied in earthquake science, climatology, neuroscience, social science, epidemiology and other fields.

ECA is often used to study the temporal and spatial evolution of events. Frederik et al.[3] used the ECA to explored the temporal evolution of heavy precipitation patterns during the monsoon season. Sun et al.[4] studied the dynamic spatiotemporal covariation patterns of soil moisture and precipitation in India in four seasons. Ashish et al.[5] studied the dynamic relationship between soil moisture and precipitation in India from 2004 to 2020. Marc et al.[6] evaluated the dynamic changes in the possibility of synchronous occurrence of seasonal extreme precipitation on the global surface and El Niño and La Niña in different seasons. Fan et al.[7] quantified the probability of extreme temperature events under precipitation in different months and their comprehensive impact on crops. We use ECA to construct complex networks to study the dynamic characteristics of high traffic in mobile communication systems.

With the development of mobile Internet and mobile communication, traffic data of mobile communication services has attracted more attention. Using complex networks to study mobile communication complex systems and analyzing the time and space requirements of mobile users for data services will help operators plan base stations. Wang et al.[8] used the event synchronization method to study the similarity between high traffic base stations to visualize the system congestion. To optimize network resource allocation Zhang et al.[9] proposed an optimization algorithm for complex network communication resource allocation based on an improved neural network. Zhao et al.[10] studied the spatio-temporal similarity of cellular traffic between base stations, and conducted a detailed spatial study of traffic and base station deployment. Mining the spatio-temporal data of mobile communication analyzed the evolution law of communication data so that the utilization rate of network resources improved.

The ECA method is used to identify high-traffic event synchronization and construct a complex network, which is used to study the spatio-temporal characteristics of high-traffic dynamic changes. A static network is constructed to analyze the spatial connectivity of high-traffic events and the synchronous influence area of high-traffic occurrence in base stations. The dynamic network is constructed by dividing the time window by an hour to explore the dynamic changes of high-traffic synchronization spatio-temporal characteristics. Excavate the interactive relationship between base stations and the law of dynamic synchronization of high-traffic base stations.

2. Data and Methods

2.1 Data

The data comes from hourly traffic data automatically collected by a certain operator's base station in a certain city. From 0:00 on February 22, 2017, to 23:00 on February 27, 2017, a total of nearly 90,000 pieces of spatio-temporal data (including time, longitude, latitude, and traffic) were collected. The study area contains a total of 171 base stations, and the length of the time series at each base station is 144 hours.

The research steps include identifying high traffic events, extracting events, quantifying the similarity between event sequences, and constructing complex networks. The threshold method is applied to identify high-traffic events. The 90th percentile of global spatio-temporal data is taken as the threshold[2], and events above the threshold are considered as high-traffic events.

2.2 Event Coincidence Analysis

Event coincidence analysis (ECA) is used to quantify the statistical correlation between event sequences. ECA defines a coincidence interval $\triangle T$ to identify event synchronization. If the time interval of event occurrence is within the coincidence interval $\triangle T$, the event is considered to occur synchronously, i.e., $0 <= |t_i^l - t_j^m| <= \triangle T$. where $t_i^l - t_j^m$ is the time interval between the l-th occurrence of the event at geographic location i and the m-th occurrence of the event at geographic location j. In order to be statistically significant, the coincidence interval $\triangle T$ needs to be smaller than T/N_{max} [2]. N_{max} is the maximum value of the length of the event sequence, which means that the event sequence has N_{max} times of high-traffic events. In order to quantify the synchronicity between two event sequences, the event coincidence rate is defined as [3]

$$r(i \mid j;\Delta T) = \frac{1}{N_i}\sum_{l=1}^{N_i}\Theta\{\sum_{m=1}^{N_j}1_{[0,\Delta T]}(t_i^l - t_j^m)\}, \tag{1}$$

where $r(i|j;\triangle T)$ represents the fraction of events in event sequence i which are preceded by at least one event in event sequence j within $\triangle T$. l = 1, 2, ⋯, N_i, m = 1,2, ⋯, N_j indicates that N_i and N_j high-traffic events occurred at geographic location i and geographic location j respectively. $\Theta(\cdot)$ is the Heaviside function ($\Theta(x) = 0$ when $x <= 0$, $\Theta(x) = 1$ when x is other). The Heaviside function prevents double counting [3], enabling subsequent calculations to be properly normalized. $1_I(\cdot)$ indicates the indicator function ($1_I(x) = 1$ when $x \in I$, $1_I(x) = 0$ when x is other).

Similarly, $r(j|i; \triangle T)$ represents the fraction of events in event sequence j which are preceded by at least one event in event sequence i within $\triangle T$. ECA is used toconstruct the adjacency matrix $A = (a_{ij})_{n\times n}$. $a_{ij} = [r(i|j; \triangle T) + r(j|i; \triangle T)] / 2$, n is the total number of geographical locations in the study area. Adjacency matrix A is a symmetrical matrix, i.e., $a_{ij} = a_{ji}$.

3. Results

3.1 Static Networks

The purpose of this study is to explore the dynamic characteristics of high-traffic events. Firstly, the static network across the study area is analyzed to explore the spatiotemporal characteristics of high-traffic events. All locations share one threshold (determined by the 90th percentile for all sequences). The maximum event sequence length is $N_{max} = 109$. The coincidence interval $\triangle T$ must be less than T / N_{max}, so the $\triangle T$ is taken as one hour. Treat the base station as a node and build an undirected and unweighted network. The existence of edges between nodes indicates that there are events occurring synchronously between nodes. Build static networks with high-traffic event sequences. The network has 104 nodes and 3839 edges. The average degree is 73.827, the average clustering coefficient is 0.864 and the average path length is 1.283. A static network is a small-world network. According to Clauset, Shalizi and Newman method, the cumulative degree distribution is Weibull distribution, and the parameters are alpha = 4.000 and scale = 81.715.

The degree represents the synchronous influence area of high-traffic events occurring in the base station. The degree of a node is defined as the number of edges linked with it. The existence of an edge between two nodes indicates that there is an event synchronously occurring between the two base stations. The base stations are distributed in different geographical locations. Therefore, base stations with high degrees have synchronized events with a wide area of base stations, that is, base stations with a high degree have a large area of synchronous influence. The base station with a high degree in the southwestern region and some northern regions are large, as shown in Figure 1(a). High-traffic events in these regions exhibit a wide synchronous influence area. Base stations with high degrees exhibit high synchronization and a wide area of synchronization influence.

The clustering coefficient reflects the spatial connectivity of high-traffic e vents at the base station. The node clustering coefficient can be used to measure the possibility of synchronous events occurring among neighboring base stations. A base station with a high clustering coefficient indicates that there is a high probability of connecting edges between neighboring base stations. That is, the possibility of event synchronization between neighboring base stations is high. Therefore, the clustering coefficient can reflect the spatial connectivity of high-traffic in the base station. The base stations all exhibit high clustering coefficients, as shown in Figure 1(b). The high-traffic events of the base stations in the study area exhibit high spatial connectivity.

Average link distance (LD) is the average geographical distance between a node and its connected neighbor nodes, which is normalized by node degree centrality. A base station with a high-valued LD indicates that the base station has high synchronization with the base station at a long distance. The base stations in the southwestern region have a low-valued LD, which shows that the synchronization between the base stations in this region and the nearby base stations is high, as shown in Figure 1(c). With the southwest region as the center, the LD of base stations increases from south to north, and the base stations in the northern region show the characteristics of high synchronization with long-distance base stations.

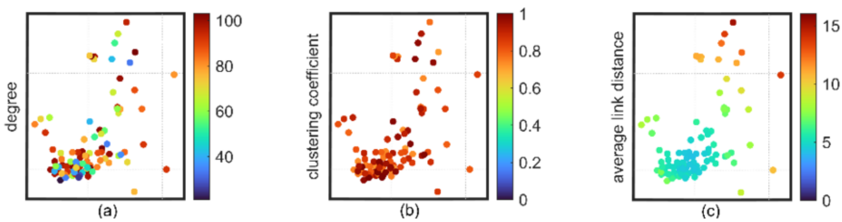

Figure 1. The geographic distribution of (a) node degree, (b) node clustering coefficient, (c) node average link distance in the static network.

3.2 Dynamic Network

The fluctuation of the hourly data flow of the base station is constantly changing with the usage of the users. The time series data (the length is 6 hours) of six days at the same hour were used to construct 24 dynamic networks, which are undirected and unweighted. The maximum value of the event sequences length is 6, and the conformance interval is still one hour. If there is a connection between nodes, it means that event synchronization occurs between base stations at the same time on different days. There are obvious changes in the number of dynamic network nodes, the number of connected edges, and the degree at different times, as shown in Figure 2(a)(b)(c). Excluding points 2am to 7am (rest time), the clustering coefficient and the average link distance fluctuated similarly in the dynamic network, as shown in Figure 2(d)(e). The high-traffic event synchronization of the base station presents regular fluctuations at different hours.

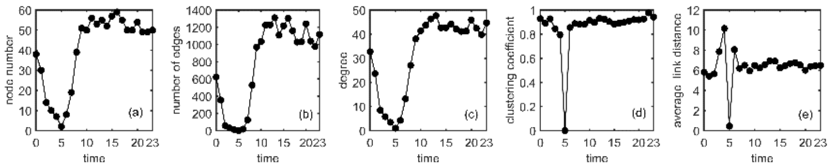

Figure 2. The topology characteristics of dynamic network, (a) number of nodes, (b) number of edges, (c) degree, (d) clustering coefficient, (e) average link distance.

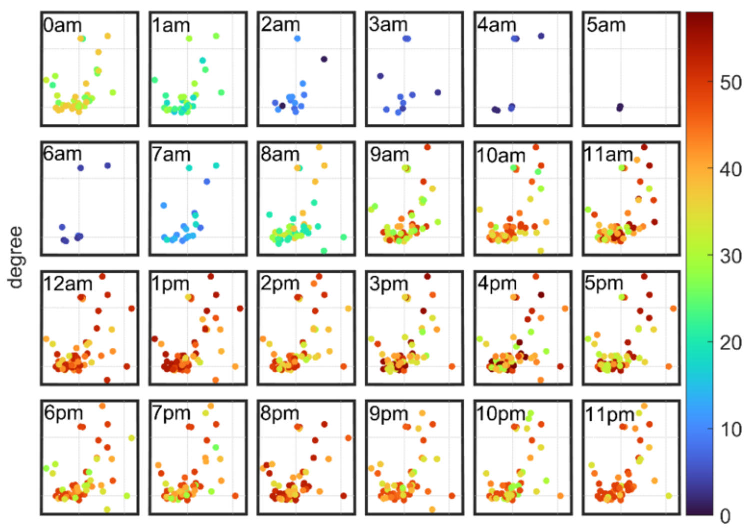

Figure 3. Geographic distribution of node degrees in dynamic network.

The synchronous influence area of high-traffic events in base stations is affected by time and geographical location. The synchronization of high-traffic events at a large number of base stations at 9am to 11pm has a wide synchronous influence area, as shown in Figure 3. From 2am to 7am, there are few base stations where event synchronization occurs, and the synchronous influence area of high-traffic event synchronization of base stations is narrow. It shows the dynamic spatio-temporal changes of user traffic demands. The location of high-traffic events varies during different time periods.

The node degree of dynamic network reflects the interaction of the traffic of the base station at the same time on different days. The event synchronization between base stations includes the synchronization of high traffic events between base stations at the same time on different days. The traffic between a base station with a high degree and other base stations has a high number of interactions.

High-traffic events have high spatial connectivity and are insensitive to temporal changes, as shown in Figure 4. Although, the number of base stations with high-traffic events at 4am and 5am is small. Most of the nodes in the dynamic network show the characteristics of high clustering coefficients and are less affected by time.

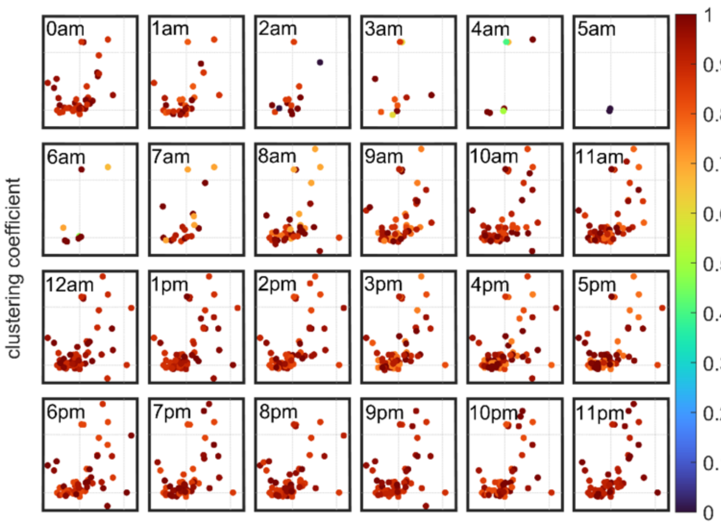

Figure 4. Geographic distribution of node clustering coefficient in dynamic network.

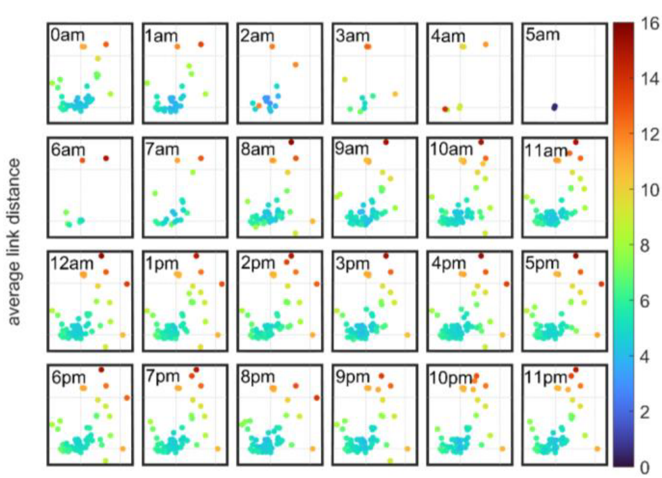

Figure 5. Geographic distribution of node average link distance in dynamic network.

The geographical distribution of the LD of base stations at different times shows a similar pattern. The color division of scatter points in different areas at different times is similar, as shown in Figure 5. The pattern of LD is insensitive to temporal and geographic changes.

4. Conclusions

We identify high-traffic events in base station traffic data. ECA is used to quantify the statistical correlation between high traffic event sequences and construct complex networks to analyze network topology characteristics. Visualize dynamic changes in high-traffic features. Mining the rules of high-traffic changes can help prevent problems such as network congestion and communication delays caused by traffic surges.

High traffic has high spatial connectivity, and it changes regularly. There is an interaction of high traffic between base stations. The high traffic synchronization distance mode of the base station is relatively fixed. The concurrency of high-traffic events is often related to the flow of people and economic development. In future research, crowd activity information can be combined to mine and predict traffic.

References

[1] Donges JF, Donner RV, Trauth MH, et al. Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution. PNAS, 2011, 108(51): 20422-20427.
[2] Donges JF,Schleussner CF, Siegmund JF,et al. Event coincidence analysis for quantifying statistical interrelationships between event time series. EPJ ST, 2016, 225(3): 471-487.
[3] Wolf F, Ozturk U, Cheung K, et al. Spatiotemporal patterns of synchronous heavy rainfall events in East Asia during the Baiu season. Earth System Dynamics, 2021, 12(1): 295–312.
[4] Sun AY, Xia Y, Todd GC, et al. Patterns of precipitation and soil moisture extremes in Texas, US: a complex network analysis. Advances in Water Resources, 2018, 112(2): 203-213.
[5] Ashish MJ, Ravi KG, Ankit A. Spatiotemporal dependence of soil moisture and precipitation over India. Journal of Hydrology, 2022, 610(7): 127898.
[6] Wiedermann M, Siegmund JF, Donges JF, et al. Differential imprints of distinct ENSO flavors in global patterns of very low and high seasonal precipitation. Frontiers in Climate, 2021, 3(1): 618548.
[7] Fan X, Zhu D, Sun X, et al. Impacts of extreme temperature and precipitation on crops during the growing season in South Asia. Remote Sensing, 2022, 14(23): 6093.
[8] Wang LN, Tan GM, Zang CR. Identifying the spatiotemporal organization of high-traffic events in a mobile communication system using event synchronization and complex networks. Chaos, 2022, 32(9): 093122.
[9] Zhang H, Liu Q. Optimization algorithm of communication resource allocation in a complex network based on an improved neural network. Journal of Function Spaces, 2022, 2022(7): 7309612.
[10] Zhao Z, Li M, Li R, et al. Temporal-spatial distribution nature of traffic and base stations in cellular networks. IET Communications, 2017, 11(16): 2410-2416.