IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 21, 2020, accepted March 5, 2020, date of publication March 16, 2020, date of current version March 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2980168

A Visualization Method for Mining Colocation Patterns Constrained by a Road Network

MENGJIE ZHOU[1,2], TINGHUA AI[3], GUOHUA ZHOU[1,2], AND WENQING HU[1]

[1]College of Resources and Environmental Science, Hunan Normal University, Changsha 410081, China
[2]Key Laboratory of Geospatial Big Data Mining and Applications, Changsha 410081, China
[3]School of Resource and Environment Sciences, Wuhan University, Wuhan 430072, China

Corresponding author: Tinghua Ai (tinghuaai@whu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 41901314, in part by the Open Research Fund Program of Key Laboratory of Digital Mapping and Land Information Application Engineering, MNR under Grant ZRZYBWD201902, and in part by the Scientific Research Program of Hunan Education Department under Grant 18C0049.

ABSTRACT Colocation mining is useful for understanding the interactions or dependencies that occur among geographic phenomena. Most colocation mining methods are based on planar space. However, in urban spaces, many human-related activities are constrained by a road network. Planar colocation mining methods are not suitable for studying the concerning geographic phenomena in an urban space. In this paper, we propose a visualization method to discover colocation patterns constrained by a road network. The method consists of two major components: network kernel density estimation and network colocation rule map construction. In the colocation rule map construction component, spatial interactions among spatial network geographic phenomena are modeled based on the idea of color mixing. We use simulated datasets with different spatial patterns, different sample sizes, and different maximum distances between road network events to test our method. The results show that our method is effective for mining colocation patterns in different situations. We also change the resolution of the network colocation rule maps, and the results show that the resolution has little influence on the results. In the case study, we apply our method to explore the spatial association between crimes and city facilities in the Loop and the Near North Side districts of Chicago.

INDEX TERMS Geographic information systems, visualization, geography, colocation, road network.

I. INTRODUCTION

Colocation mining is used to find subsets of geographic features whose instances are often close to each other [1], [2]. Colocation mining can be used to reveal interactions or dependencies that occur between geographic phenomena, for instance, to detect the association of land use facilities with crime patterns [3], to understand the correlation between the built environment and crashes [4], and to analyze the relationships among different urban facilities [5]. Moreover, the applications of colocation mining can be expanded to cases with different data formats, such as extract trajectory frequent pattern in outdoor environment and indoor environment [6], [7]

In urban spaces, human-related geographical phenomena, such as the locations of urban facilities, traffic jams, and street crimes, are usually constrained by a road network [8], [9]. These human-related geographical phenomena are also considered network spatial phenomena. Analyzing network spatial phenomena is suitable for applications associated with urban study [10]–[12], such as expanding the knowledge and understanding of street crime [13], [14] and optimizing the location of urban facilities [6], [8]. When analyzing network spatial phenomena, traditional studies regard the real world as a continuous and homogeneous space and use the Euclidean distance to measure the distance. However, a road network space is discrete and inhomogeneous, using network distance [15], [16]. Thus, conclusions that are drawn by traditional methods using the Euclidean distance are often flawed [17]. For mining the colocation patterns of network spatial phenomena, it is necessary to propose methods that can characterize the network space.

Spatial colocation mining approaches are divided into two classes: spatial statistics and spatial data mining approaches [2], [8]. Spatial statistics approach refers to use metrics of spatial correlation to characterize the relationship

The associate editor coordinating the review of this manuscript and approving it for publication was Laxmisha Rai.

between different types of spatial features, such as cross-K function [18], cross nearest distance [19], and Q test [20]. Spatial data mining approaches mainly refers to association rule-based approaches that adopt an approach similar to the Apriori algorithm [21]. Directly adopting planar colocation mining methods cannot be used to discover colocation patterns that are constrained by a road network [7], [22]. Therefore, some effort has been made to discover colocation patterns constrained by a road network through extended planar colocation mining methods. Based on the existing classification of spatial colocation mining approaches, we classify network colocation mining methods into two categories: extended spatial statistics methods and extended spatial data mining methods. In the extended spatial statistics methods, some spatial statistics methods for mining the planar colocation methods are extended to a network space. The cross-K function [18], cross nearest distance [19], and Q test [20] are extended to the network cross-K function [23], the network nearest distance [10], and the Q Statistic in the network space [6] by using the network distance. Statistical diagnostics [24] are extended to the network space based on the skeleton partitioning of a road network [25]; however, the partitioning may cause the loss of neighbor relationships among objects across the partitioning borders. In the extension, some characteristics of spatial statistics methods for mining planar colocation methods are also inherited, such as the cross-K function, which is suitable for two independent distributions and not suitable for categories in one population. In extended spatial data mining, Yu [26] extended the colocation mining study of Huang *et al.* [2] to a network space by defining neighborhoods that use the network distance rather than the Euclidean distance. Furthermore, Yu *et al.* [8] took the 'distance decay effect' of spatial interactions into account to improve the accuracy of modeling spatial dependency relationships between geographic features. Cai *et al.* [22] proposed a network-constrained spatial colocation pattern method with a nonparametric significance test that could detect colocation patterns with less priori knowledge.

In recent years, visualization has played an important role in spatial data mining [27], [28] and acts as representation tools or tools that facilitate spatial data mining. In colocation mining, some researchers apply visualization methods to improve colocation mining methods. Okabe and Yamada [23] used network graphs as visualization tools to represent the mining results, allowing users to find and explore colocated features. However, the network graphs cannot provide any information on the spatial configuration of the colocation. Flouvat *et al.* [29] proposed a method to represent colocation mining results on maps that are based on the clustering method. However, the locations of the colocation patterns are only approximately represented on the maps.

In the field of visualization, color mapping is an important and widely used technique [30]. Color mixing can be used for multivariate visualization. There are several techniques: multivariate color scale, color blending and color weaving [31], [32]. On a multivariate color scale, two or more data variables can be mapped to a single color. For example, three multispectral scanner bands are represented by red, green and blue using the RGB color model in Landsat "false color" images [33]. The color blending technique uses different colors to represent each variable and then blends the results [34], [35]. The color weaving technique weaves individual colors that represent multiple variables to form a fine-grained texture pattern [36], [37]. Some researchers have conducted studies to help users select the proper colors for multivariate representation to improve the perception of the original components that generate particular colors and identification of individual values [29], [38]. Inspired by these studies, we hypothesize that color-mixing strategies that are used for multivariate visualization may be applied to colocation mining. Before color mixing, a method should be used to visualize the spatial distribution patterns of spatial phenomena constrained by a road network. As an efficient spatial statistics tool in exploring and modelling spatial point patterns, kernel density estimation considers the distance decay impact of spatial point and generates a density surface from the input spatial points [39], [40]. To analyze network spatial phenomena, kernel density estimation is extended to network kernel density estimation [12], [41], [42]. We will use network kernel density estimation to visualize the distribution patterns of network spatial phenomena.

This paper focuses on the mining colocation pattern that is constrained by a road network through visualization. Much different than existing colocation mining methods that are based on abstract thinking and use spatial statistics or spatial data mining approaches to generate a textual form result, the visualization method presented herein is based on visual thinking and uses colors to represent the colocation rule and mining colocation patterns in a visual form. Our method can provide an intuitive approach to obtain cognition of colocation rules in a road network space. In the colocation mining process, visualization acts as a tool to mine geo-data and discovery knowledge rather than representation tools or tools that facilitate spatial data mining. This paper is a variation and extension of the visualization approach for colocation mining proposed by Zhou *et al.* [43]. Since the approach proposed by Zhou *et al.* [43] is a planar colocation mining method and is not suitable for mining colocation patterns constrained by a road network, we extend the approach to a road network space. In Section 2, we propose a visualization method for colocation mining in a network space. In Section 3, we use simulated data to test the performance of our method in different situations and analyze the influence of resolution on the results. Section 4 presents a case study that applies our method to analyze the association between crimes and city facilities, and Section 5 concludes this paper.

II. PROPOSED METHOD

This method can be used to visually explore whether two types of road network events are colocated. The basic method consists of two major components.

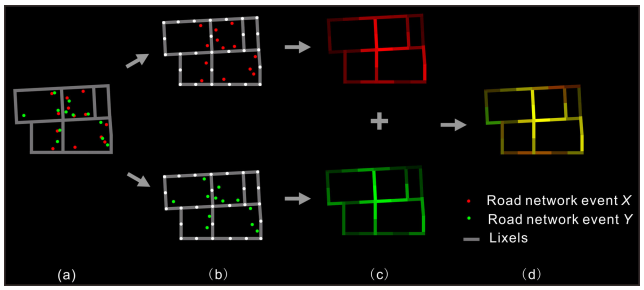

FIGURE 1. Illustration of the visualization method for network colocation mining: (a) a road network and road network events X and Y, (b) network kernel density estimation, (c) density surfaces of road network events X and Y, and (d) network colocation visualization based on color mixing.

(1) Network kernel density estimation. Network kernel density estimation is used to calculate the distribution density of two types of road network events, as shown in Fig. 1(b).

(2) Network colocation rule map construction. The spatial interaction of the two types of road network events are modeled. Two density network surfaces of the two types of road network events are produced. One type of event is represented in red, and the other is represented in green, as shown in Fig. 1(c). Based on the law of additive color mixing, the colors of the two density network surfaces are mixed to produce a **network colocation rule map**(a visualization result that uses colors to represent colocation rules at each network location), as shown in Fig. 1(d).

A. NETWORK KERNEL DENSITY ESTIMATION

Planar kernel density estimation considers the distance decay impact of a point event and can generate a smooth density surface that is easy to understand. This method is commonly used to analyze the spatial distribution of a point event [39], [40]. The kernel density estimator is defined as follows:

$$f(s) = \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{s - s_i}{\tau}\right) \tag{1}$$

where s represents a location, $f(s)$ represents the density value at s, $k()$ represents the kernel weighting function that is centered on s and spreads according to the bandwidth τ, $s_1, \ldots, s_n$ are the locations of the n event points, and $s - s_i$ is the distance between s and s_i. Kernel density estimation requires two parameters: the bandwidth τ and the kernel weighting function $k()$. It has been reported that $k()$ has little influence on the results and the τ is a critical parameter [39], [40], [44]. When the value of τ increases, the spatial variation in density is smoother. When the value of τ reduces, the estimation become increasingly 'spiky'. Different values of τ can be used to explore variation in density value at different scales.

In analyzing the distribution of road network events, using planar kernel density estimation may produce a bias [41]. Thus, based on the characteristics of road networks and road network events, some researchers have extended planar kernel density estimation and have proposed network kernel density estimation algorithms [12], [41], [42]. Network kernel density estimation uses the same form of a kernel density estimator (Eq. (1)), although the distance is measured by the network distance. Due to the complexity of a road network, compared with planar kernel density estimation, network kernel density estimation could be relatively time consuming, especially at the stage of distance calculations. To reduce the computational complexity, we chose the network kernel density estimation algorithm proposed by Yu *et al.* [12]. The idea of the algorithm is inspired by water flow extension. The algorithm is briefly described here. Readers are encouraged to refer to a more complete description (see the reference [12]).

(1) Divide the road network into a set of linear units (Lixels) with a defined road network length l. The intersection point of Lixels is called an lxnode. Construct the network topological relationship between Lixels and between Lixels and lxnodes.

(2) Project the road network events to the nearest Lixels.

(3) Calculate the density value at each Lixel based on an operator of the 1-D sequential expansion operator that is extended from the dilation operator in mathematical morphology [45]. Then, the bandwidth τ is measured with the steps of expansion. Because it has been reported that the kernel weighting function $k()$ has little influence on the results, the commonly used quartic kernel function (Eq. (2)) is selected.

$$k\left(\frac{s - s_i}{\tau}\right) = \frac{3}{4}\left(1 - \frac{(s - s_i)^2}{\tau^2}\right) \tag{2}$$

For road network events X and Y, network kernel density estimation [12] is used to analyze their spatial distribution. Step (1) mentioned above is executed, where the i th Lixel is labeled as $L_i (i = 1 \ldots n$, n is the total number of Lixels). Then, steps (2) and (3) mentioned above are executed, and the density value of spatial events X and Y at L_i are labeled KL_i^X and KL_i^Y, respectively.

B. CONSTRUCTION OF A COLOCATION RULE MAP BASED ON COLOR MIXING

To construct a colocation rule map, we set the color of the L_i from two local network density values KL_i^X and KL_i^Y of the road network events X and Y based on additive color mixing. Here, the RGB color is used to represent the color. The RGB color model is an additive color model that has three color components, R, G, and B, representing red, green, and blue, respectively. The values of the three color components range from 0 to 255.

The network colocation rule map is constructed based on additive color mixing. Additive color mixing conforms to the following basic rules:

- Zero intensity for each component produces black, and the full intensity for each component produces white.
- The sum of two components of equal intensity forms a secondary color: C(cyan) = G + B, Y(yellow) = R + G, M(magenta) = R + B.
- When one of the components has the strongest intensity, the color hue is close to the primary color.

We use red and green to represent the network density of X and Y, and the colored L_i is represented by $C_i^X(R, 0, 0)$ and $C_i^Y(0, G, 0)$ in the network density surfaces of X and Y, respectively, as shown in Fig. 1(c). Based on the law of additive color mixing, the colors of the network density surface of X and Y are mixed. $C_i^{X\&Y}(R, G, 0)$ represents the color of the ith linear unit after color mixing, as shown in Fig. 1(d). The function to determine the color components is as follows:

$$KL_i^{X'} = KL_i^X \Big/ \max(KL_i^X), \quad KL_i^{Y'} = KL_i^Y \Big/ \max(KL_i^Y) \tag{3}$$

$$R = g(KL_i^{X'}), \quad G = g(KL_i^{Y'}) \tag{4}$$

Here, g is a monotonically increasing function. Indeed, g can be regarded as a many-to-one operator that assigns the kernel density values (the 'many') to an interval.

Small number of intervals produces classed map. Whereas, big number of intervals, such as 256, produces a continuous map. Classed map and continuous map have their own advantages and disadvantages. Choosing number of classes should also depends on different users and applications. Classed map may be easy to interpret, as people can detect and discriminate different classes in it. On the other hand, classification method can affect the results, and larger numbers of classes are much less affected than smaller numbers of classes.

C. INTERPRETATION OF THE RESULT

To interpret the results of our method, a rotation function is used to transform two local network density variables to variables that represent the network colocation and pseudo intensity of the joint distribution in a road network space:

$$Co = \arctan\left(\frac{KL_i^{Y'}}{KL_i^{X'}}\right) - \frac{\pi}{4} \tag{5}$$

$$I = \sqrt{KL_i^{X'2} + KL_i^{Y'2}} \tag{6}$$

where Co represents the network colocation and I represents the pseudo intensity of the joint distribution of the road network events X and Y in a network space. The value of Co ranges from $-\pi/4$ to $\pi/4$, and the value of 0 indicates the strongest colocation. The value of I ranges from 0 to $\sqrt{2}$, and the larger the value is, the larger the pseudo intensity [43].

How can the color of road segments be associated with the measurements of colocation? We take an unclassed network colocation rule map as an example to illustrate the association, as shown in Fig. 2. The first row represents $C_i^X(R, 0, 0)$, and the first column represents $C_i^Y(0, G, 0)$. The color components R and G are calculated as follows: $R = 255KL_i^{X'}$ and $G = 255KL_i^{Y'}$. The remaining part represent the corresponding color mixing result $C_i^{X\&Y}(R, G, 0)$. At the main diagonal (top left to bottom right), the color is yellow and the value of Co reaches 0, which indicates the strongest network colocation. The larger the angle from the main diagonal, the weaker the network colocation pattern. The color above the main diagonal is reddish, which indicates that the road network event X is 'dominant' (the value of Co is negative). The color below the main diagonal is greenish, which indicates that the road network event Y is 'dominant' (the value of Co is positive). From the top left to bottom right in Fig. 2, the pseudo intensity of the joint distribution in a network space increases. The stronger intensity of yellow indicates a stronger network colocation. The global colocation is estimated based on the proportion of yellow in the entire network colocation rule map.

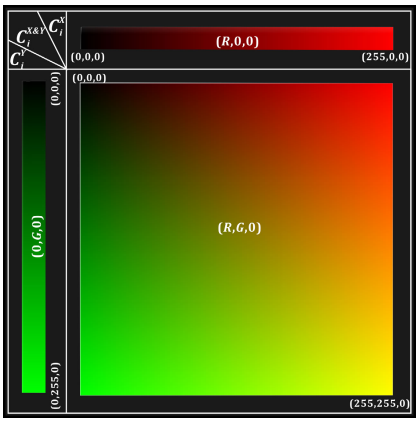

FIGURE 2. Interpretation of the results (using an unclassed map as an example).

III. SIMULATED DATA EXPERIMENTS

We use simulated data to test whether our method can effectively discover network colocation patterns in different situations. The different situations are as follows: (1) datasets with different patterns (section 3.1), (2) datasets with different sample sizes (section 3.2), and (3) datasets with the same sample size but different maximum distances between the road network events X and Y (section 3.3). In sections 3.1-3.3, we divide the road network into a set of Lixels with a network distance of 10. We also conduct experiments to analyze how the method is affected by the length of the Lixels (the resolution) (section 3.4). The idea for generating the simulated data is similar to that in the work in [46]. The interaction distance is set to 100 (network distance) in datasets 1-6. In this section, color components R and G are set as follows: $R = 255KL_i^{X'}$ and $G = 255KL_i^{Y'}$; thus, the legend of network colocation maps refers to Fig. 2.

A. DATASETS WITH DIFFERENT PATTERNS

We use simulated data to test the method in the situation of spatial autocorrelation or abundance of feature instances. Road network events in datasets 1 and 2 are colocated, whereas road network events in datasets 3 and 4 are not colocated. Random road network events in the experiment are generated using Sanet (http://sanet.csis.u-tokyo.ac.jp/).

Dataset 1: Road network events X and Y are colocated. Road network event X has 50 instances and is distributed randomly in the network space. Road network event Y also has 50 instances, and each instance of Y can be found within a network distance of 100 of an instance of X, as shown in Fig. 3(a).

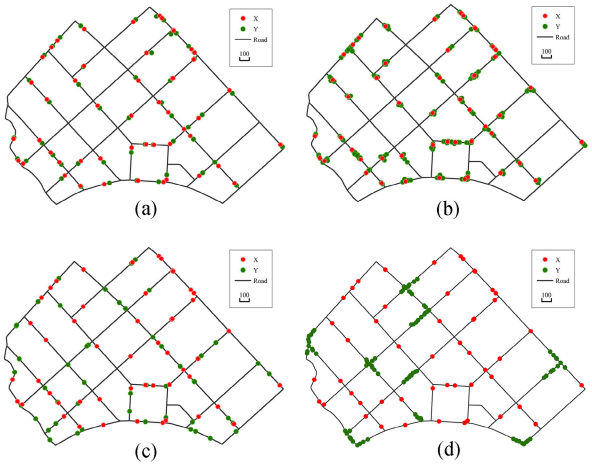

FIGURE 3. Simulated dataset with different patterns. In datasets 1 and 2, road network events X and Y are colocated. In datasets 3 and 4, road network events X and Y are not colocated. (a) Dataset 1, (b) dataset 2, (c) dataset 3, and (d) dataset 4.

Dataset 2: Road network event X and Y are colocated. Road network event X has 50 instances and is distributed randomly in the network space, as shown in Fig. 3(b). Road network event Y has 160 instances and is spatially autocorrelated. Each instance of Y can be found close to an instance of X within a network distance of 100.

Dataset 3: Road network events X and Y are not colocated. As shown in Fig. 3(c), road network events X and Y have 50 instances each and are distributed randomly in the network space, and road network events X and Y are independent of each other.

Dataset 4: Road network events X and Y are not colocated. As shown in Fig. 3(d), the road network event X is distributed randomly in the network space and has 50 instances. Road network event Y is spatially autocorrelated and has 90 instances. No two instances of road network instances X and Y are seen within a network distance of 100.

Result 1: The bandwidth range is set from 100 to 400; Fig. 4 shows the results. At a bandwidth of 100, we observe some yellow segments on the network colocation rule map. When the bandwidth increases, more road segments turn yellow on the network colocation rule map, and yellow accounts for all proportions of the colored road, which indicates that road network event Y is colocated with road network event X.

Result 2: The bandwidth range is set from 100 to 400; Fig. 5 shows the results. When the bandwidth is 100, we observe yellow or orange segments on the colocation rule map. When the bandwidth increases, the proportion of yellow increases. At a bandwidth of 200-400, the yellow area accounts for most of the colored area, which indicates that road network event Y is colocated with road network event X.

Result 3: The bandwidth range is set from 100 to 400, and Fig. 6 shows the results. When the bandwidth is 100, we only observe red or green segments on the colocation rule map. When the bandwidth increases, the red, green and yellow regions expand in the network space, although the proportion

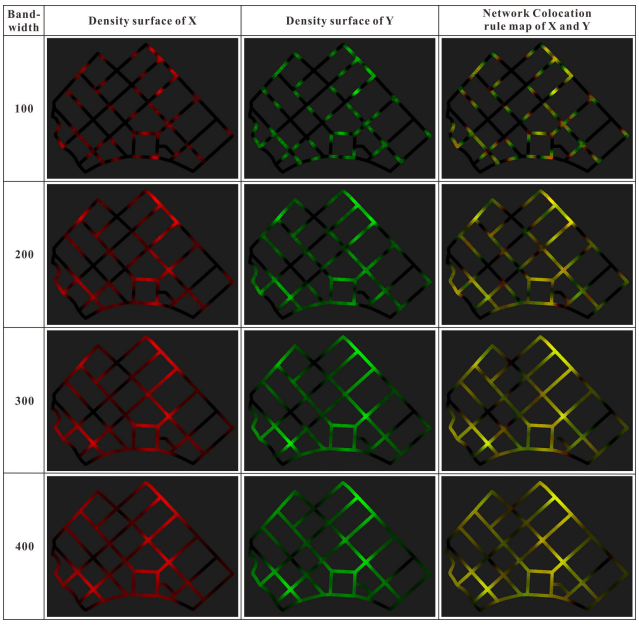

FIGURE 4. Visualization results of dataset 1.

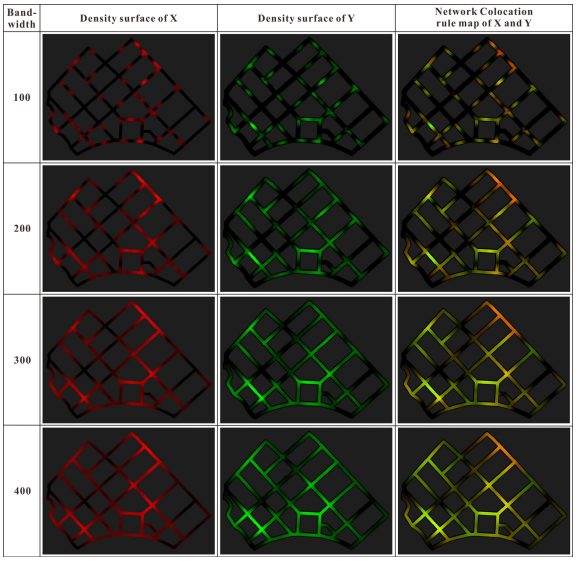

FIGURE 5. Visualization results of dataset 2.

of yellow is still very small, which indicates that road network event X is not colocated with road network event Y.

Result 4: The bandwidth range is set from 100 to 400; Fig. 7 shows the results. When the bandwidth is 100, the colocation rule map contains red and green line segments. When the bandwidth increases, the red and green line segments expand, and the colocation rule maps are full of red or green lines, which indicates that road network event X is not colocated with road network event Y.

We compare our method with network cross-K function. The network cross-K results of datasets 1-4 are shown in Fig 8. The blue curve indicates the observed curve; the red curve indicates the mean value under the CRS hypothesis; the green and pink curves are, respectively, the upper and

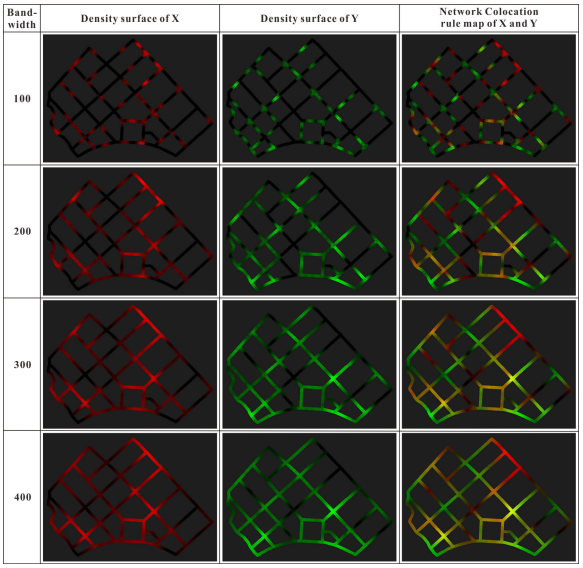

FIGURE 6. Visualization results of dataset 3.

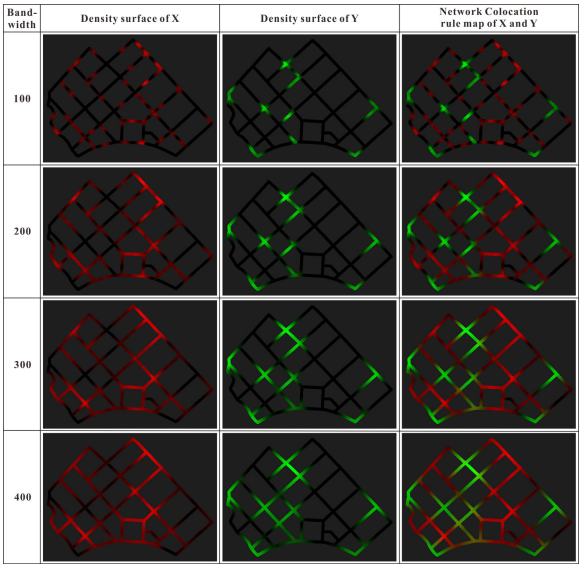

FIGURE 7. Visualization results of dataset 4.

lower envelop curves under the CSR hypothesis. For datasets 1 and 2, the observed network cross-K value is above the upper envelop curve, which indicates a significant pairwise aggregation tendency. For datasets 3, the observed network cross-K value is in the simulation envelope, indicates the mean value under the CRS hypothesis and road network event Y is not colocated with road network event X. For dataset 4, the observed network cross-K value is under the lower envelop curve, indicating that road network event Y is tend to be dispersed from road network event X. The network cross-K results of datasets 1-4 are consistent with our results.

B. DATASETS WITH DIFFERENT SAMPLE SIZES

The datasets in Section 3.1 are not large. People may doubt that if the sample sizes increase, especially for two large randomly distributed road network events, the density of each

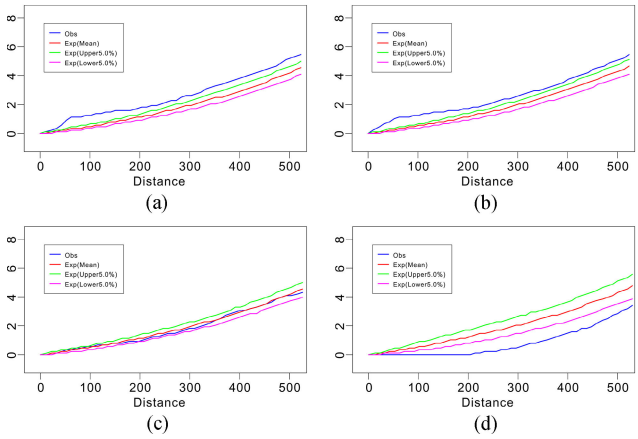

FIGURE 8. Network cross-k results of datasets 1-4. (a) Dataset 1, (b) dataset 2, (c) dataset 3, and (d) dataset 4.

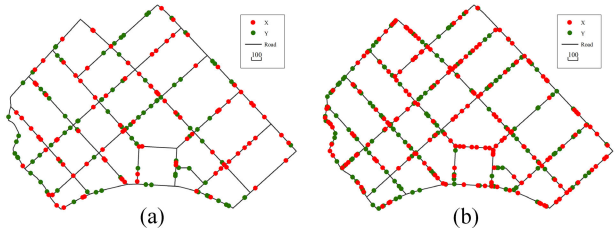

FIGURE 9. Simulated dataset with different sample sizes. (a) Dataset 5, road network events X and Y have 100 instances each. (b) Dataset 6, road network events X and Y have 200 instances each.

location might be similar, and then the correlation will be incorrect. We use larger datasets to test the method in the situation that road network events are randomly distributed and not colocated. We change the sample size of dataset 3. In datasets 5 and 6, road network events X and Y have 100 instances each and 200 instances each, respectively. Road network events X and Y are distributed randomly in the network space and are independent of each other, as shown in Fig. 9.

Results 5 and 6: The bandwidth range is set from 100 to 400, and Fig. 10 shows the results. For the initial bandwidth, red and green occupy nearly all the colocation rule maps of datasets 5 and 6. With increasing bandwidth, although the proportion of yellow increases, it remains substantially smaller than that of red or green. The colocation rule maps indicate that road network event X is not colocated with road network event Y in datasets 5 and 6. At the same bandwidth in datasets 3, 5 and 6, the density surfaces and network colocation rule maps are smoother in the larger dataset. We also compared our method to the network cross-K function. The results of the network cross-K function are shown in Fig. 11. The observed network cross-K value is in the simulation envelope, which is consistent with our results.

C. DATASETS WITH DIFFERENT INTERACTION DISTANCES

In datasets 7-9, the interaction distance is set to 50, 100, and 150. Road network events X and Y are colocated and have 50 instances each. The maximum network distances between

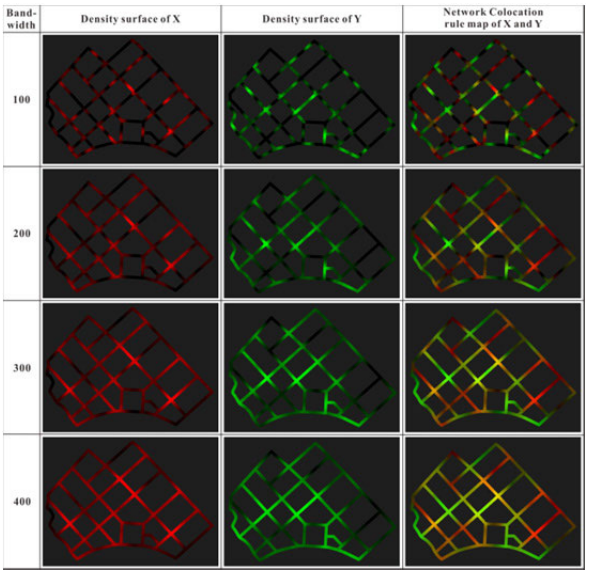

FIGURE 10. Visualization results of dataset 5 (road network events X and Y have 100 instances each).

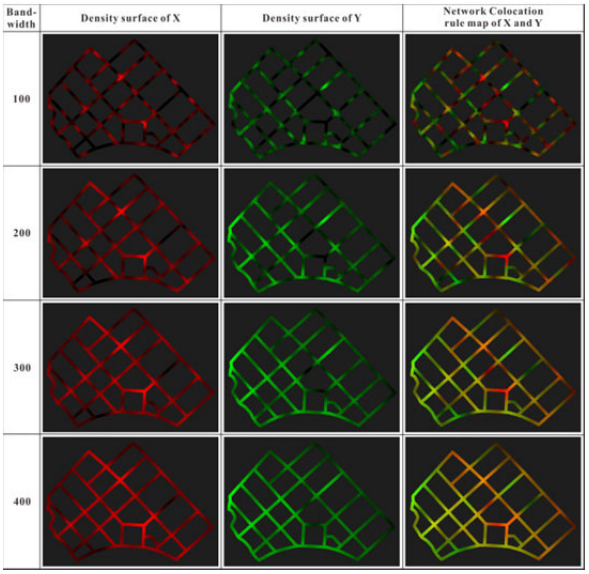

FIGURE 11. Visualization results of dataset 6 (road network events X and Y have 200 instances each).

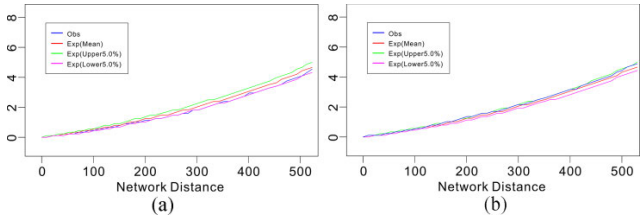

FIGURE 12. Network cross-k results of datasets 5 and 6. (a) Dataset 5 and (b) dataset 6.

road network events X and Y are 50, 100, and 150. Road network event X is randomly distributed in the network space, as shown in Fig. 12.

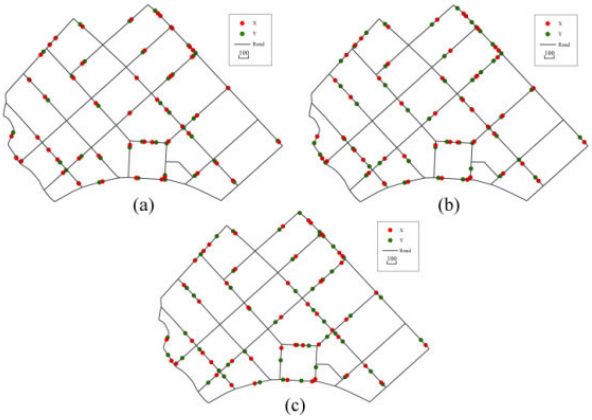

FIGURE 13. Simulated dataset with different maximum distances between road network events. (a) Dataset 7, maximum distance: 50, (b) dataset 8,maximum distance: 100, and (c) dataset 9, maximum distance: 150.

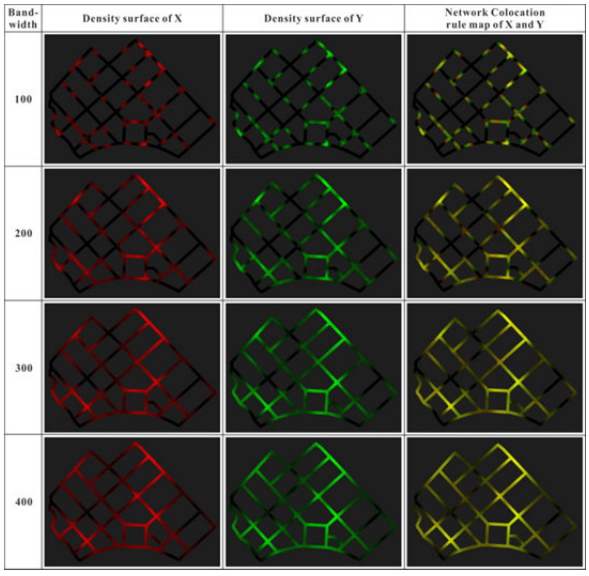

FIGURE 14. Visualization results of dataset 7 (maximum distance: 50).

For datasets 7-9, the bandwidth range is set from 100 to 400; Fig. 13 shows the results. When the bandwidth increases, more road segments turn yellow on the network colocation rule map. Yellow occupies a large proportion in the colocation rule maps for datasets 7-9. From a bandwidth of approximately 100 in dataset 7, 200 in dataset 8, and 300 in dataset 9, yellow accounts for nearly all of the colored area. The colocation rule maps indicate that road network event Y is colocated with road network event X in datasets 7-9. The differences in the datasets with different maximum distances can also be detected from the colocation rule map. We also compared our method to the network cross-K function. The results of the network cross-K function are shown in Fig. 14. The observed network cross-K value is always greater than the upper bound of the simulation envelope, which indicates a significant pairwise aggregation tendency. However, the results for the different datasets are very similar and do not reflect the different patterns of datasets 7-9.

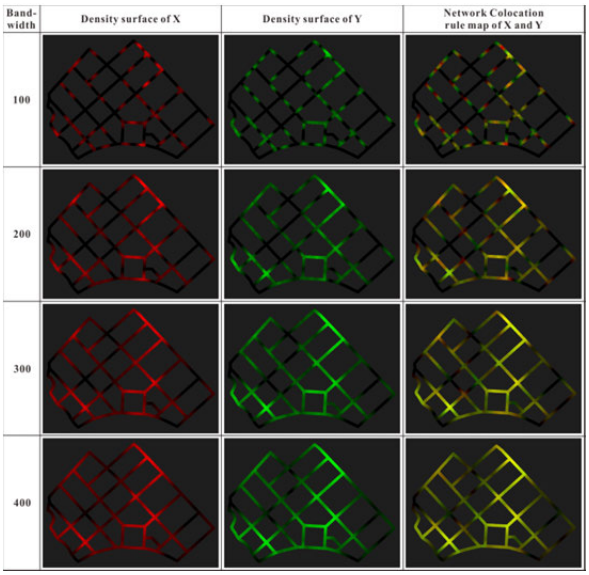

FIGURE 15. Visualization results of dataset 8 (maximum distance: 100).

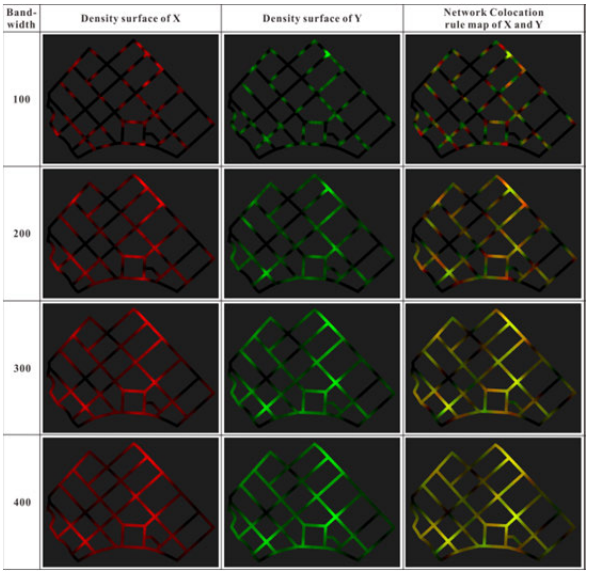

FIGURE 16. Visualization results of dataset 9 (maximum distance: 150).

D. ANALYSIS ON THE INFLUENCE OF THE LENGTH OF THE LIXELS (THE RESOLUTION) ON THE RESULTS

For datasets 1 and 3, we changed the length of the Lixels (the resolution) and reperformed the experiments to analyze how the length of the Lixels affects the results. Fig. 18 shows the results of dataset 1, and Fig. 19 shows the results of dataset 3. Results using Lixels with a network distance of 5 (1/4 of the original Lixel length) are shown in the left column, and results using Lixels with a network distance of 20 (two times the original Lixel length) are shown in the left column. Comparing the results of different lengths of Lixels, the results are extremely similar. However, the color transition in the results with a high resolution is smoother than that of the results with a low resolution. The length of the Lixels (the resolution) has little influence on the results.

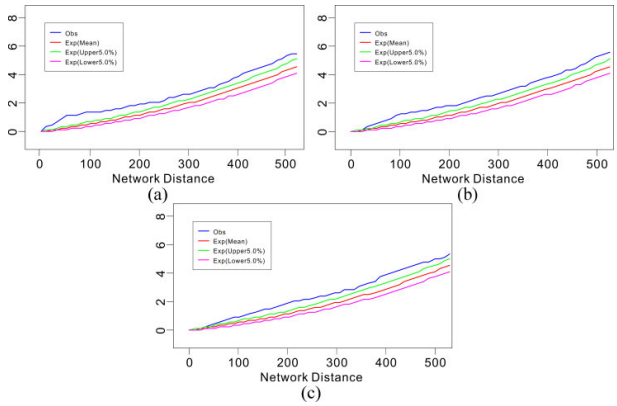

FIGURE 17. Network cross-k results of datasets 7-9. (a) Dataset 7, (b) dataset 8, and (c) dataset 9.

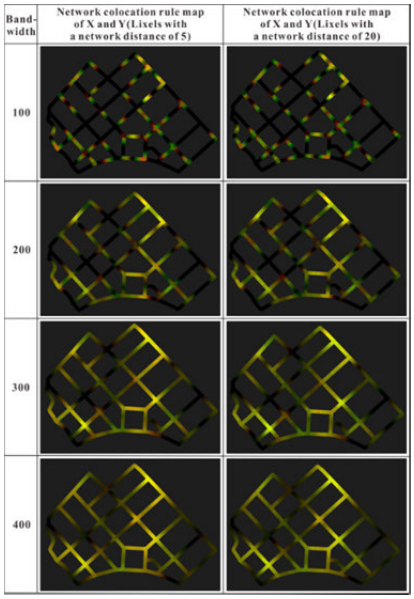

FIGURE 18. Visualization results of dataset 1 using different lengths of Lixels (resolution).

IV. CASE STUDY-EXPLORING THE SPATIAL ASSOCIATION BETWEEN CRIME AND FACILITIES

A. STUDY AREA AND DATA

Understanding the spatial association between crime and facilities may help us know why crime incidents tend to happen in some areas and thus allocate police resources and help reduce crime [47]–[49]. In this section, our method is applied to analyze the colocation of two types of crimes and two types of city facilities in the Loop and the Near North Side districts of Chicago. These two districts belong to central Chicago. The Loop is the central business district and downtown area of Chicago. The Magnificent Mile is a stretch of shops in the Near North Side. The crime and street centerline data were provided by the Chicago Data Portal, and the city facilities data were obtained from Open Street Map. Two types crimes studied from 2017 include theft (a total of 11754 records) and robbery (a total of 804 records).

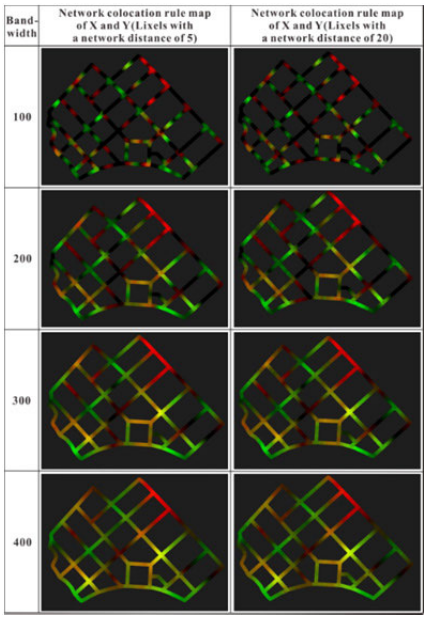

FIGURE 19. Visualization results of dataset 3 using different lengths of Lixels (resolution).

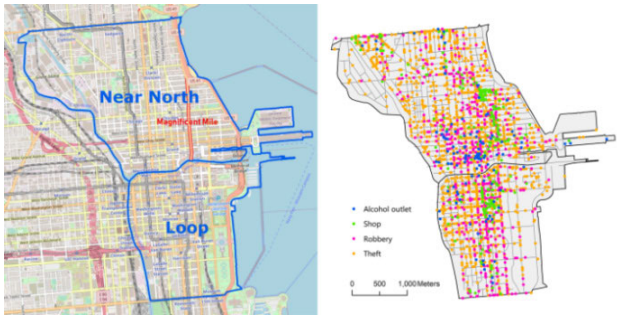

FIGURE 20. Study area.

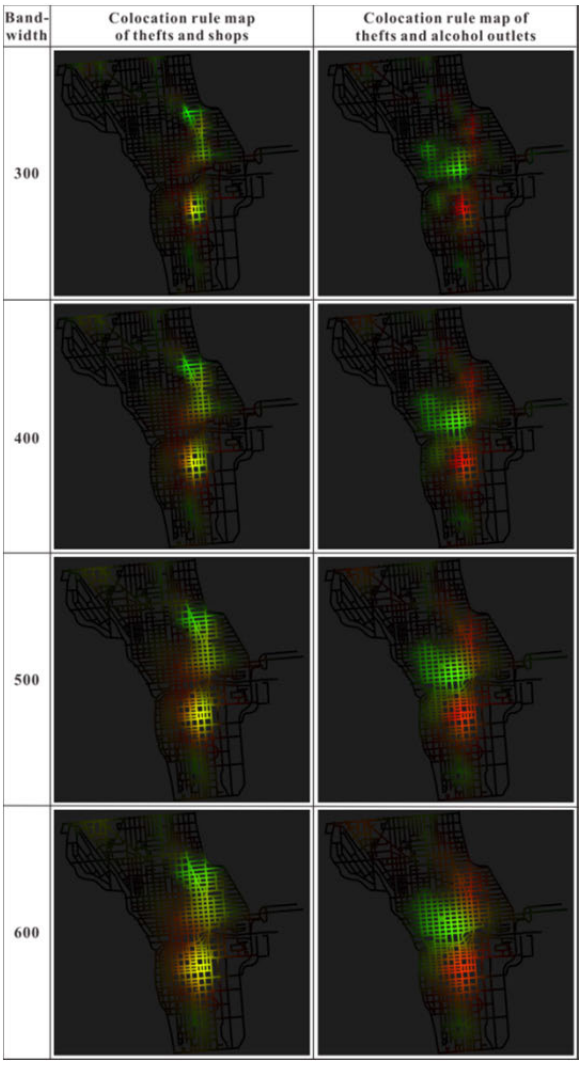

FIGURE 21. Network colocation rule maps of thefts and shops as well as thefts and alcohol outlets at different bandwidths.

The two types of city facilities studied include shops (a total of 245 instances) and alcohol outlets (alcohol consumption place) (a total of 113 instances), as shown in Fig. 20.

B. RESULTS

We divide the road network into a set of *Lixels* with a network length of 20 m. Our method comprises a network kernel density estimation step, and the bandwidth is a critical parameter in kernel density estimation [44]. In the experiment, we analyze the colocation patterns of crime and facilities at different bandwidths from 300 to 600 m. The network density of theft and robbery crimes is represented by red, and the shops and alcohol outlets are represented by green. Then, the color network density surfaces of the crimes and facilities are mixed to produce network colocation rule maps. Fig. 21 shows the generated network colocation rule maps of thefts and the two facilities (shops and alcohol outlets). In the network colocation rule map of thefts and shops (Fig. 21, first column), at a bandwidth of 300 m, we can observe green, greenish, yellow, and reddish colors in the colocation rule map. With increasing bandwidth, some areas become yellow or yellowish at a bandwidth of 400 m. At bandwidths of 500 and 600 m, yellow and yellowish areas account for a large proportion of the colored area, whereas the area around the northern area of the Magnificent Mile is greenish. Network colocation maps detect that theft incidents co-locate with shops. Shops act as ''crime generators'' that have higher associated volumes of crime where people go to perform legal daily activities [50]. Shops create opportunities for theft to those who intend to commit a crime. People gather at shopping precincts, providing interactions between potential offenders and potential targets. When shopping, people mainly focus on commodities, and this may weaken some people's sense of prevention and give some potential offenders chances to commit theft crimes. In the network colocation rule map of thefts and alcohol outlets (Fig. 21, second column), the red and green areas take up almost all of the colored area on the map. With increasing bandwidth, the colored area expands, and a small part of area turns greenish or reddish. However,

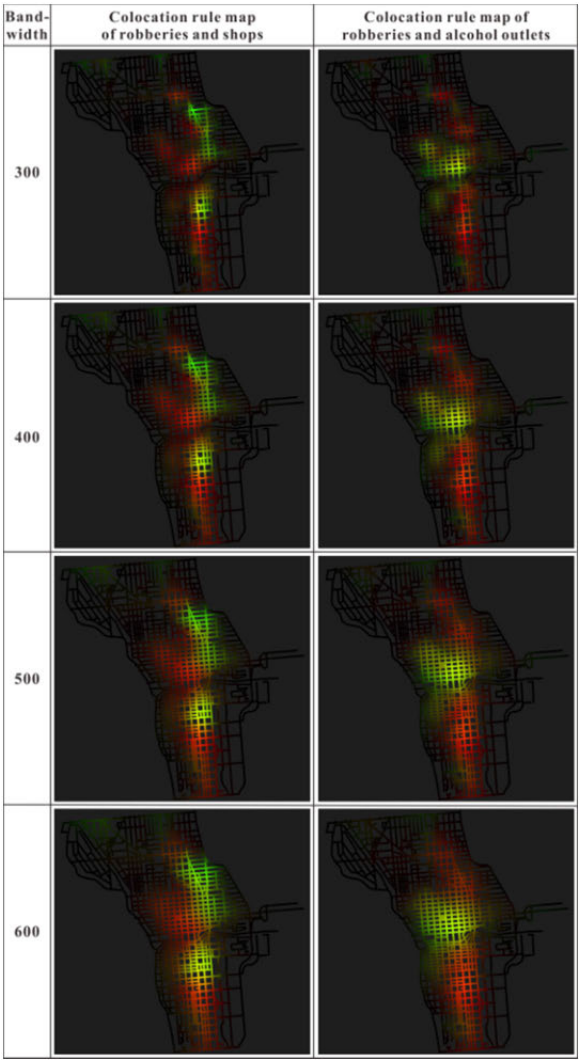

FIGURE 22. Network colocation rule maps of robberies and shops as well as robberies and alcohol outlets at different bandwidths.

yellow or yellowish areas can barely be observed at the different bandwidths, which indicates that theft incidents do not colocate with alcohol outlets.

Fig. 22 shows the generated network colocation rule maps of robberies and the two facilities (shops and alcohol outlets). In the network colocation rule map of robberies and shops (Fig. 22, first column), red, reddish, green, and greenish areas are present in almost all the colored area in the colocation rule maps at all bandwidths, which indicates that robbery incidents do not colocate with shops. On the network colocation rule map of robberies and alcohol outlets (Fig. 22, second column), green, greenish, red, and reddish regions comprise nearly all of the colored area at a bandwidth of 300 m. With increasing bandwidth, the yellowish area occurs and expands in the southern area of the Near North Side district. However, red and green regions still comprise most of the colored area, which indicates that robbery incidents do not colocate with alcohol outlets. The result is consisting with the findings [51] that alcohol stores and bars are not associated with higher street robbery counts.

V. CONCLUSION

In this study, we proposed a visualization method to discover colocation patterns that are constrained by a road network. The method consists of two major components: network kernel density estimation and network colocation rule map construction. Users can discover global colocation through the perception of the color of the network colocation rule map. We use simulated data with different patterns and sample sizes to test our method, and the results show that our method can effectively discover network colocation patterns. We use datasets with different maximum distances between road network events to test our method, and the results show that the network colocation rule maps with different bandwidths can reflect the differences in the maximum distance. We use simulated data to analyze how the method is affected by the length of the Lixels, and the results show that the length of the Lixels has little influence. We also provide a case study to illustrate how to use our method to analyze the spatial association between crimes and city facilities. Our method can be applied to analyze human activities, such as urban facilities, street crimes, traffic jams and crashes, that are constrained by a road network.

This work has several perspectives. (1) We use simulated datasets with given patterns to test the method. However, Users detect colocation patterns and their distribution based on the color perception of a network colocation map. Thus, the work needs to be evaluated in the future for end-user assessment. (2) The method is suitable for people with normal color vision. However, color blindness affects approximately 8% of men and 0.5% of women around the world, most often in the red-green form. In the future, the visualization method can be improved for those who are color blind. (3) This method is suitable for measure pairwise network colocation pattern. How to mine colocation patterns of more than two road network events is still a remaining problem for us.

REFERENCES

[1] S. Shekhar and Y. Huang, "Discovering spatial co-location patterns: A summary of results," in *Proc. 7th Int. SSTD*. Redondo Beach, CA, USA: Springer, 2001, pp. 236–256.
[2] Y. Huang, S. Shekhar, and H. Xiong, "Discovering colocation patterns from spatial data sets: A general approach," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 12, pp. 1472–1485, Dec. 2004.
[3] F. Wang, Y. Hu, S. Wang, and X. Li, "Local indicator of colocation quotient with a statistical significance test: Examining spatial association of crime and facilities," *Prof. Geographer*, vol. 69, no. 1, pp. 22–31, Jan. 2017.
[4] Y. Hu, Y. Zhang, and K. S. Shelton, "Where are the dangerous intersections for pedestrians and cyclists: A colocation-based approach," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 431–441, Oct. 2018.
[5] Y. Rui, Z. Yang, T. Qian, S. Khalid, N. Xia, and J. Wang, "Network-constrained and category-based point pattern analysis for Suguo retail stores in Nanjing, China," *Int. J. Geographical Inf. Sci.*, vol. 30, no. 2, pp. 186–199, Feb. 2016.
[6] W. Yu, "Discovering frequent movement paths from taxi trajectory data using spatially embedded networks and association rules," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 855–866, Mar. 2019.
[7] Y. Chen, P. Yuan, M. Qiu, and D. Pi, "An indoor trajectory frequent pattern mining algorithm based on vague grid sequence," *Expert Syst. Appl.*, vol. 118, pp. 614–624, Mar. 2019.
[8] W. Yu, T. Ai, Y. He, and S. Shao, "Spatial co-location pattern mining of facility points-of-interest improved by network neighborhood and distance decay effects," *Int. J. Geographical Inf. Sci.*, vol. 31, no. 2, pp. 280–296, Feb. 2017.

[9] D. Zhu, N. Wang, L. Wu, and Y. Liu, "Street as a big geo-data assembly and analysis unit in urban studies: A case study using Beijing taxi data," *Appl. Geography*, vol. 86, pp. 152–164, Sep. 2017.

[10] A. Okabe and K. Sugihara, *Spatial Analysis Along Networks: Statistical and Computational Methods*. Hoboken, NJ, USA: Wiley, 2012.

[11] B. Jiang and A. Okabe, "Different ways of thinking about street networks and spatial analysis," *Geographical Anal.*, vol. 46, no. 4, pp. 341–344, Oct. 2014.

[12] W. Yu, T. Ai, and S. Shao, "The analysis and delimitation of central business district using network kernel density estimation," *J. Transp. Geography*, vol. 45, pp. 32–47, May 2015.

[13] S. Shiode, "Street-level spatial scan statistic and STAC for analysing street crime concentrations," *Trans. GIS*, vol. 15, no. 3, pp. 365–383, Jul. 2011.

[14] V. Spicer, J. Song, P. Brantingham, A. Park, and M. A. Andresen, "Street profile analysis: A new method for mapping crime on major roadways," *Appl. Geography*, vol. 69, pp. 65–74, Apr. 2016.

[15] A. Okabe, K.-I. Okunuki, and S. Shiode, "SANET: A toolbox for spatial analysis on a network," *Geographical Anal.*, vol. 38, no. 1, pp. 57–66, Jan. 2006.

[16] G. Borruso, "Network density estimation: A GIS approach for analysing point patterns in a network space," *Trans. GIS*, vol. 12, no. 3, pp. 377–402, Jun. 2008.

[17] I. Yamada and J.-C. Thill, "Comparison of planar and network K-functions in traffic accident analysis," *J. Transp. Geography*, vol. 12, no. 2, pp. 149–158, Jun. 2004.

[18] B. D. Ripley, "The second-order analysis of stationary point processes," *J. Appl. Probab.*, vol. 13, no. 2, pp. 255–266, Jun. 1976.

[19] A. Okabe and F. Miki, "A conditional nearest-neighbor spatial-association measure for the analysis of conditional locational interdependence," *Environ. Planning A, Economy Space*, vol. 16, no. 2, pp. 163–171, Feb. 1984.

[20] M. Ruiz, F. López, and A. Páez, "Testing for spatial association of qualitative data using symbolic dynamics," *J. Geographical Syst.*, vol. 12, no. 3, pp. 281–309, Sep. 2010.

[21] C.-F. Tsai, Y.-C. Lin, and C.-P. Chen, "A new fast algorithms for mining association rules in large databases," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 1994, pp. 487–499.

[22] J. Cai, M. Deng, Q. Liu, Z. He, J. Tang, and X. Yang, "Nonparametric significance test for discovery of network-constrained spatial co-location patterns," *Geographical Anal.*, vol. 51, no. 1, pp. 3–22, Jan. 2019.

[23] A. Okabe and I. Yamada, "The K-function method on a network and its computational implementation," *Geographical Anal.*, vol. 33, no. 3, pp. 271–290, 2001.

[24] R. Sierra and C. R. Stephens, "Exploratory analysis of the interrelations between co-located Boolean spatial features using network graphs," *Int. J. Geographical Inf. Sci.*, vol. 26, no. 3, pp. 441–468, Mar. 2012.

[25] J. Tian, F.-Q. Xiong, and F. Yan, "Mining co-location patterns between network spatial phenomena," in *Advances in Spatial Data Handling and Analysis*, F. Harvey and Y. Leung, Eds. Berlin, Germany: Springer, 2015, pp. 123–142.

[26] W. Yu, "Spatial co-location pattern mining for location-based services in road networks," *Expert Syst. Appl.*, vol. 46, pp. 324–335, Mar. 2016.

[27] M. C. F. de Oliveira and H. Levkowitz, "From visual data exploration to visual data mining: A survey," *IEEE Trans. Vis. Comput. Graphics*, vol. 9, no. 3, pp. 378–394, Jul. 2003.

[28] J. Mennis and D. Guo, "Spatial data mining and geographic knowledge discovery—An introduction," *Comput., Environ. Urban Syst.*, vol. 33, no. 6, pp. 403–408, 2009.

[29] F. Flouvat, J. F. N. Van Soc, E. Desmier, and N. Selmaoui-Folcher, "Domain-driven co-location mining," *Geoinformatica*, vol. 19, no. 1, pp. 147–183, 2015.

[30] L. Wang, J. Giesen, K. T. McDonnell, P. Zolliker, and K. Mueller, "Color design for illustrative visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 14, no. 6, pp. 1739–1754, Nov. 2008.

[31] S. Silva, B. S. Santos, and J. Madeira, "Using color in visualization: A survey," *Comput. Graph.*, vol. 35, no. 2, pp. 320–333, Apr. 2011.

[32] H. Hagh-Shenas, S. Kim, V. Interrante, and C. Healey, "Weaving versus blending: A quantitative assessment of the information carrying capacities of two alternative methods for conveying multivariate data with color," *IEEE Trans. Vis. Comput. Graphics*, vol. 13, no. 6, pp. 1270–1277, Nov./Dec. 2007.

[33] P. K. Robertson, "Visualizing color gamuts: A user interface for the effective use of perceptual color spaces in data displays," *IEEE Comput. Graph. Appl.*, vol. 8, no. 5, pp. 50–64, Sep. 1988.

[34] N. Gossett and B. Chen, "Paint inspired color mixing and compositing for visualization," in *Proc. IEEE Symp. Inf. Visualizat.* Austin, TX, USA: IEEE, 2004, pp. 113–118.

[35] A. Debiasi, B. Simões, and R. De Amicis, "Depiction of multivariate data through flow maps," in *Service-Oriented Mapping* (Lecture Notes in Geoinformation and Cartography), J. Döllner, M. Jobst, and P. Schmitz, Eds. Berlin, Germany: Springer, 2019, pp. 125–141.

[36] T. Urness, V. Interrante, I. Marusic, E. Longmire, and B. Ganapathisubramani, "Effectively visualizing multi-valued flow data using color and texture," in *Proc. IEEE Trans. Ultrason., Ferroelectr., Freq. Control*. Washington, DC, USA: IEEE Computer Society, 2003, pp. 121–151.

[37] M. Luboschik, A. Radloff, and H. Schumann, "A new weaving technique for handling overlapping regions," in *Proc. Int. Conf. Adv. Vis. Inter. (AVI)*. New York, NY, USA: ACM, 2010, pp. 25–32.

[38] S. Gama and D. Gonçalves, "Guidelines for using color blending in data visualization," in *Proc. Int. Work. Conf. Adv. Vis. Inter. (AVI)*. New York, NY, USA: ACM, 2014, pp. 363–364.

[39] T. C. Bailey and A. C. Gatrell, *Interactive Spatial Data Analysis*, vol. 413. Essex, U.K.: Longman Scientific and Technical, 1995.

[40] B. W. Silverman, "Density estimation for statistics and data analysis," in *Monographs on Statistics and Applied Probability*, vol. 26. London, U.K.: Chapman & Hall, 1986.

[41] A. Okabe, T. Satoh, and K. Sugihara, "A kernel density estimation method for networks, its computational method and a GIS-based tool," *Int. J. Geographical Inf. Sci.*, vol. 23, no. 1, pp. 7–32, Jan. 2009.

[42] Z. Xie and J. Yan, "Kernel density estimation of traffic accidents in a network space," *Comput., Environ. Urban Syst.*, vol. 32, no. 5, pp. 396–406, Sep. 2008.

[43] M. Zhou, T. Ai, C. Wu, Y. Gu, and N. Wang, "A visualization approach for discovering colocation patterns," *Int. J. Geographical Inf. Sci.*, vol. 33, no. 3, pp. 567–592, Mar. 2019.

[44] T. K. Anderson, "Kernel density estimation and K-means clustering to profile road accident hotspots," *Accident Anal. Prevention*, vol. 41, no. 3, pp. 359–364, May 2009.

[45] H. J. A. M. Heijmans and C. Ronse, "The algebraic basis of mathematical morphology I. Dilations and erosions," *Comput. Vis., Graph., Image Process.*, vol. 50, no. 3, pp. 245–295, Jun. 1990.

[46] S. Barua and J. Sander, "Mining statistically significant co-location and segregation patterns," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1185–1199, May 2014.

[47] P. Day, G. Breetzke, S. Kingham, and M. Campbell, "Close proximity to alcohol outlets is associated with increased serious violent crime in New Zealand," *Austral. New Zealand J. Public Health*, vol. 36, no. 1, pp. 48–54, Feb. 2012.

[48] W. A. Pridemore and T. H. Grubesic, "A spatial analysis of the moderating effects of land use on the association between alcohol outlet density and violence in urban areas," *Drug Alcohol Rev.*, vol. 31, no. 4, pp. 385–393, Jun. 2012.

[49] E. R. Groff and B. Lockwood, "Criminogenic facilities and crime across street segments in philadelphia: Uncovering evidence about the spatial extent of facility influence," *J. Res. Crime Delinquency*, vol. 51, no. 3, pp. 277–314, May 2014.

[50] P. Brantingham and P. Brantingham, "Criminality of place," *Eur. J. Criminal Policy Res.*, vol. 3, no. 3, pp. 5–26, Sep. 1995.

[51] C. P. Haberman and J. H. Ratcliffe, "Testing for temporally differentiated relationships among potentially criminogenic places and census block street robbery counts," *Criminology*, vol. 53, no. 3, pp. 457–483, Aug. 2015.

MENGJIE ZHOU was born in Xiaogan, China, in 1990. She received the B.S. degree in geographic information systems, the M.S. degree in surveying and mapping engineering, and the Ph.D. degree in cartography and geographic information systems from Wuhan University, China. She is currently a Lecturer with the College of Resources and Environmental Science, Hunan Normal University. Her research interests include cartographic visualization and spatial data mining.

TINGHUA AI was born in Yidu, China, in 1969. He received the B.S. and M.S. degrees in cartography and the Ph.D. degree in cartography from the Wuhan University of Surveying and Mapping Science and Technology, Wuhan, in 1994 and 2000, respectively. He is currently a Professor with the School of Resource and Environmental Sciences, Wuhan University. His research interests include the multiscale representation of spatial data, map generalization, spatial cognition, and spatial big data analysis.

GUOHUA ZHOU was born in 1965. He is currently a Professor with the College of Resources and Environmental Science, Hunan Normal University. His research interests include urban and rural planning, land use planning, agricultural geography, and rural development analysis.

WENQING HU was born in Changsha, China. She is currently pursuing the master's degree with the College of Resources and Environmental Science, Hunan Normal University.