Article

Building an Urban Spatial Structure from Urban Land Use Data: An Example Using Automated Recognition of the City Centre

Jie Zhu [1,2] and Yizhong Sun [1,2,*]

[1] Key Laboratory of Virtual Geographic Environment of Ministry of Education, Nanjing Normal University, Nanjing 210023, China; Chu_Je@163.com

[2] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

* Correspondence: sunyizhong_cz@163.com; Tel.: +86-137-7082-7090

Abstract: It has been suggested that the method of constructing an urban spatial structure typically follows a forward process from planning and design up to expression, as reflected in both graphic and text descriptions of urban planning. Although unorthodox, the original status structures can be extracted and constructed from an existing urban land-use map. This approach not only provides the methodological foundation for urban spatial structure evolution and allows for a comparative and quantitative analysis between the existing and planned conditions, but also lays a theoretical basis for failure in scientific decision making during the planning phase. This study attempts to achieve this by identifying the city centre (a typical element of the urban spatial structure) from urban land use data. The city centre is a special region consisting of several units with particular spatial information, including geometric attributes, topological attributes, and thematic attributes. In this paper, we develop a methodology to support the delineation of the city centre, considering these factors. First, using commercial land data, we characterise the city centre as units based on a series of indicators, including geometric and thematic attributes, and integrate them into a composite index of "urban centrality"; Second, a graph-based spatial clustering method that considers both topological proximity and attribute similarity is designed and used to identify the city centre. The precise boundary of the city centre is subsequently delimited using a shape reconstruction method based on the cluster results. Finally, we present a case study to demonstrate the effectiveness and practicability of the methodology.

Keywords: urban spatial structure; city centre; commercial land; spatial clustering; GIS

1. Introduction

Urban spatial structure is viewed as a generalized description of elements arranged in geographic space and the process of element interactions [1]. Many studies have been made on the spatial structure of cities, and we divide them into two main forms. One characterises the city as the node for the study of the urban external spatial structure in a "macro" depiction, including the urban system and urban agglomerations; the other identifies the city as the surface to study the urban internal spatial structure in a "micro" depiction, including urban land-use patterns, central business district (CBD), and urban image space. In the field of urban planning, the construction method of urban internal spatial structures typically follows a forward process from planning to expression; i.e., city planners aim to design a reasonable urban spatial structure to guide the development of a city, which is reflected in a graphic description of urban planning that emphasizes urban spatial structure. For example, according

to the Comprehensive Plan of Shenzhen City (2011–2020) in China, Shenzhen's urban spatial structure will be a polycentric structure (with seven planned city centres). City planners design this structure to guide the development of Shenzhen. It focuses on the future of urban planning. However, status structures (i.e., urban spatial structures at a certain time during the planning period) are significantly important to policymakers and urban planning. Status structures not only provide the methodological foundation of urban spatial structure evolution for comparative and quantitative analysis between existing and planned conditions, but also lay a theoretical basis for failure in scientific decision making during the planning phase. Can an urban spatial structure be analysed in the context of the status of a city? Most previous studies have focused specifically on forms of land use. The most early and well-known studies have formed the Agricultural Location Theory, Industrial Location Theory, and Central Place Theory; these theories have all successfully reflected the distribution and evolution of urban spatial elements, based on location models of land utilization. Many studies on urban land use, including urban morphology [2,3], city landscape pattern [4,5], urban land-use allocation [6,7], and urban expansion simulation [8,9], further confirm this fact, revealing hidden urban information from the distribution of various types of urban land. Thus, studying urban spatial structures using urban land use is an objective and feasible approach.

The Urban Planning Bureau in each city of China maintains and disseminates urban land datasets at a very fine scale. Designed as general purpose products, these datasets offer a wealth of (primarily geometry and land type) information about individual land parcels. However, they do not model higher-order geographic phenomena [10]. For example, we can easily obtain the spatial distribution information of various land types, but not the districts of functional areas (e.g., industrial areas) within cities; we can get the location and height of hills, but not the geographic extent of the hills. Providing more of the higher-level semantics in urban land datasets could allow the Urban Planning Bureau to respond better to user requirements. This allows the representation of city space to be closer to the way it is conceptualized by people [11] and not simply reflected in graphic descriptions of urban planning that emphasize the urban spatial structure.

As a developing country with a very large population, the change in central urban landscapes and functions is critical for promoting sustainable progress in China, because the city centre is considered to be the economic and social centre of the city [12]. In this paper, we use the city centre derived from urban land use as an example to attempt to build an urban spatial structure from urban land use. The city centre is a typical element of the urban spatial structure [13]. Moreover, as cities become more complex and some of cities are polycentric in design, the recognition of city centre areas is an important issue for urban planning, and draws the attention of policymakers and scholars [14,15]. Although human cognition might be used to sketch a city centre, many researchers have attempted to develop quantitative methods for its delineation (see Section 2). Most methods focus primarily on spatial patterns or non-spatial inequality. A city is usually divided into subareas, such as census tracts, and school zones, which usually emerge into irregular shapes and have different spatial information including geometric attributes, topological attributes, and thematic attributes [16]. Similarly, the city centre is a special region that consists of several units (commercial land in this paper) with different characteristics.

Based on the questions above, this study primarily aimed to establish an applicable methodology oriented towards urban land use for identifying city centre areas. In this context, we addressed the two following questions:

(1) How is the city centre characterised based on the existing urban land use?

The city centre is significantly influenced by economic aggregation, and its core function is commerce. The city centre is strongly relevant to commercial land. A city is always divided into a series of spatial statistical units, with different characteristics for analysing urban spatial structure. In this study, we characterise the city centre from commercial land data using a series of indicators, including geometric and thematic attributes, and integrate them into a composite index of "urban centrality".

(2) How can we delineate the city centre?

We proposed a spatial clustering method based on graph theory and attribute information, to delineate the city centre. Spatial clustering generally involves two steps. First, we constructed spatial proximity relationships among objects (commercial lands). A clustering result is obtained based on the geometric property of the objects, and we cluster these objects based on attribute similarity (i.e., composite index of urban centrality). This research contributes to determining the city centre, considering not only geometric and topological attributes, but also socio-economic attributes.

2. Related Work

The delineation of the city centre is strongly related to data sources [13,17] such as unit postcode [18], socio-economic data [16,19], geographic grid [20], mobility data (such as point of interest [21], travel flows [22], mobile phone positioning [23]), and remote sensing data [24]. Since the 1950s, many studies have been made available to quantitatively recognize the city centre using different data sources. These approaches can be divided into two primary categories: (1) index-based, and (2) clustering methods. The index-based method represents the concentration of city centre areas using a simple index or by constructing an evaluation index system. In the early stages, Murphy and Vance [13] proposed Central Business Height Index (CBHI) and Central Business Intensity Index (CBII) to gauge the degree of the CBD. Consecutive block units that satisfy the degree (CBII $\geq$ 50%, CBHI $\geq$ 1) are defined as central areas. Galster et al. [25] and Lee [26] used the Gini coefficient and delta index to measure the unequal distribution of population or employment based on spatial units in a city, which facilitated the understanding of monocentricity and polycentricity patterns. Pereira et al. [27] introduced the urban centrality index (UCI) based on the location coefficient [28] and proximity index [29], to measure the degree of urban centrality. The UCI is able to reflect spatial concentration, considers a centrality scale that varies from monocentricity to polycentricity, can quantify the change in urban spatial structure.

The clustering method recognizes city centre areas based on statistical analysis; that is, the objects in the same class are similar to one another, and dissimilar to those in different classes. One prevalent method for the use spatial analysis is in the identification of the city centre. Tsai [30] adopted the global Moran coefficient by population and employment, to characterise urban forms at the metropolitan level. The Moran coefficients of monocentric, polycentric, and decentralized sprawling forms are high, intermediate, and close to zero respectively. A similar study was performed by Zhou [23], who used mobile phone positioning data to identify the city centre. An alternative method is to use kernel density estimation (KDE) modelling to create continuous surface representations of indicators such as urban road network [31,32], housing price [33], geo-referenced images [34], central activities and functions [35] and economic activity [36]. The KDE method transforms these indicators from point object into continuous surfaces of spatial densities. Statistical analysis are then conducted on these surfaces to evaluate the borderline of the city centre with high density values. Spatial clustering is another primary technique for urban spatial structure analysis. Hu [37] defined urban areas of interest (AOI) as the areas within a city that attract the attention of people, such as prominent landmarks, commercial zones, and scenic views. In this study, the improved DBSCAN clustering algorithm (DBSC) was used to extract AOI from Flickr photo data. To identify multiple city centres (namely, polycentric pattern) and delineate their precise boundaries, Sun [38] sought to combine the DBSCAN clustering algorithm and Voronoi graph to achieve this by using location-based social network data.

Several advanced processing techniques, such as spatial cognition [10,39] and remote sensing [24,40], have also evolved to provide powerful tools that can be used in the quantitative city centre study. Montello [39] reported a study in which participants drew lines around the areas that they believed constituted downtown Santa Barbara. Vagueness in the boundaries was elicited in two ways: by comparing the variation in boundary locations across the participants, and by having participants draw different boundaries to indicate their varying confidence in regional membership for different parts of the area. The results provided evidence that the method is a viable approach

to externalizing people's representations of vague cognitive regions. Taubenboeck [24] presented a conceptual framework to define the CBD using physical and morphological parameters, and developed a transferable method to detect and delineate CBDs over larger areas from a combination of Cartosat-1 digital surface models and multispectral Landsat ETM+ imagery.

The extent of city centre areas derived by most of above-mentioned methods (e.g., [23,27,31,32]) is indeterminate. The identification of city centre that aims to find out accurate boundary of city centre becomes vital in many applications ranging from urban planning to epidemiology [38]. Moreover, the city centre is a special region that consists of several units (commercial land in this paper) with different spatial information. The present methods primarily focus on the spatial patterns (e.g., [37,38]) or non-spatial inequality (e.g., [33,36]). In this paper, we characterised the city centre from commercial land data relating to a series of socio-economic indicators, and propose a methodology considering spatial proximity and attribute similarity for delineating the city centre from commercial land use.

3. Methodology

Figure 1 shows an overview of the proposed methodology for delineating a city centre. The methodology involves three parts. (1) Detect clusters (see Section 3.1). We proposed a spatial clustering method that begins by constructing spatial proximity relationships among objects (commercial lands) and then clustering these objects based on attribute similarity (composite index of urban centrality). We could detect clusters of commercial lands that are adjacent to each other and have higher urban centrality; (2) Characterise the city centre (see Section 3.2). In this study, we characterised the city centre from commercial land data relating to a series of indicators, and these indicators are then integrated into a comprehensive index to measure the urban centrality of commercial lands; (3) Extract the city centre area (see Section 3.3). We could derive the city centre from the clustered commercial lands, and the city centre boundary was then obtained by applying the shape reconstruction method (triangle filtration) to discrete commercial lands with higher urban centrality.
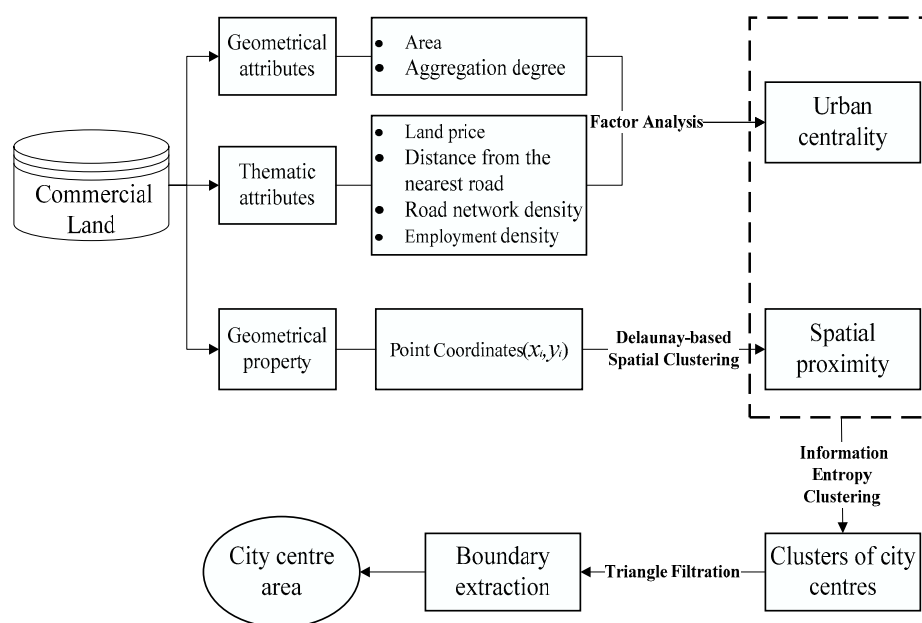

Figure 1. Overview of the procedure for delineating the city centre.

3.1. Detecting Clusters Using a Graph-Based Spatial Clustering Algorithm

The city centre is a special region that consists of commercial lands; if there is a spatially local cluster composed of commercial lands (1) which are adjacent to each another (i.e., spatial proximity) and (2) which have higher urban centrality (see Section 3.2), this cluster might exist in a city centre.

To detect this class of cluster, a spatial clustering algorithm should consider both spatial proximity and attribute similarity. We adopted a global and local constraint to model the spatial proximity relationships among objects. An adaptive spatial clustering method, based on attribute information entropy, was developed for clustering spatial objects with attribute similarity after constructing spatial proximity relationships.

3.1.1. Construction of Spatial Proximity Relationships

Delaunay triangulation is an effective way to express spatial proximity relationships [41]. The Delaunay-based algorithms can discover clusters of arbitrary shapes, and require few input parameters. However, they may not be reliable when the density varies between clusters [42]; namely, they are inaccurate near edges occurring in the gap between low- and high-density regions. In this paper, a two-level strategy algorithm was used to construct spatial proximity relationships among objects. First, we used the centroid to represent commercial land, and constructed Delaunay triangulation from these points. Second, the AUTOCLUST algorithm [41] was adopted to remove the long edges in Delaunay triangulation at the global level. Finally, inaccuracy near edges was further removed at the local level. After this two-level strategy, objects with the same edge were regarded as spatial neighbours. Two-level strategies are defined as follows:

Definition 1. *Global edge-length constraint [41]: Suppose D is a spatial database, and DT(D) is the Delaunay triangulation of D. For each point P ∈ D, the neighbourhood N(P) is the set of Delaunay edges incident to P. The global edge-length constraint can be represented as follows:*

$$Global\_Length\_Constraint(P) = Local\_Mean\_Length(P) + Global\_SD \tag{1}$$

$$Local\_Mean\_Length(P) = \frac{1}{d(P)} \sum_{i=1}^{d(P)} |e_i| \tag{2}$$

$$Local\_SD(P) = \sqrt{\frac{\sum_{i=1}^{d(P)} (Local\_Mean\_Length(P) - |e_i|)^2}{d(P)}} \tag{3}$$

$$Global\_SD = \frac{1}{N} \sum_{i=1}^{N} Local\_SD(P) \tag{4}$$

where *Local_Mean_Length(P)* is the mean length of edges directly incident to *P*, *Local_SD(P)* is the standard deviation of the length of edges directly incident to *P*, *Global_SD* is the standard deviation of the length of all edges in the Delaunay triangulation, $d(P)$ is the number of edges directly incident to *P*, $|e_i|$ is the length of edges directly incident to *P*, and *N* is the number of *D*.

If the length of an edge directly incident to *P* is larger than *Global_Length_Constraint(P)*, then the edge will be removed from the Delaunay triangulation (see Figure 2c). Through the global edge-length constraint, many inappropriate long edges were removed, as shown in Figure 2c. After filtering the long edges from the Delaunay triangulation, some inaccurate near edges (red circle regions in Figure 2c) remained in the local areas. A local edge-length constraint was developed to remove edges at the local level.

As observed from Figure 2, the length variations of inter-cluster edges incident to points in the cluster borders tended to be relatively larger because both short edges and long edges existed. To exploit this characteristic, we adopted a statistical variable-$F(P)$ to detect inter-cluster edges.

Definition 2. *Local edge-length constraint: In DT(S), for a point P ∈ S, the neighborhood N(P) is the set of edges incident to point p, and its local constraint is expressed as*

$$Sets = \{F(P)|F(P) \le \gamma\} \tag{5}$$

$$F(P) = Local\_SD(P) / Local\_Mean\_Length(P) \tag{6}$$

where γ denotes the threshold value of $F(P)$, and can be obtained using a heuristic method in [43]. The $F(P)$ value of each point inside the clusters was small because the length variations of their incident edges changed minimally. In contrast, the points on the cluster borders had large variations; thus, the $F(P)$ value was large. The final spatial clustering result (Figure 2d) was composed of all connected data points with $F(P) \leq \gamma$ and all data points belonging to their neighbourhoods.
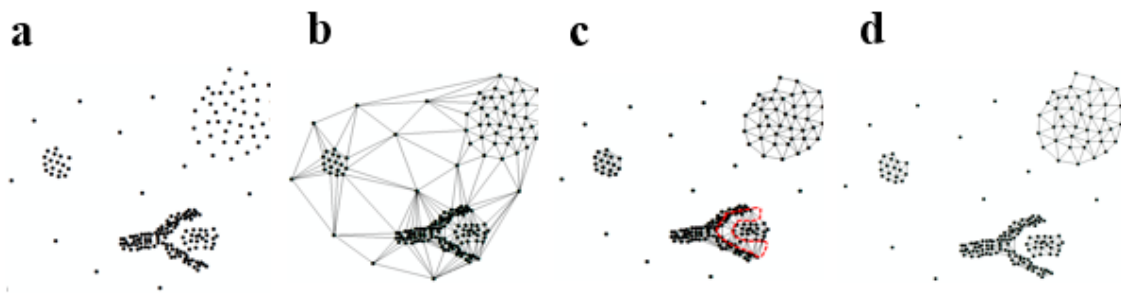

Figure 2. Construction of spatial proximity relationships: (**a**) simulated data set D [44]; (**b**) Delaunay triangulation of D; (**c**) global edge-length constraint; (**d**) local edge-length constraint ($\gamma = 0.507$).

3.1.2. Clustering Commercial Lands with Attribute Similarity

After global and local trimming operations, a modified Delaunay triangulation C-DT was obtained. Based on C-DT, attribute similarity was used to identify clusters with higher urban centrality.

Many studies have proposed the clustering of spatial objects with attribute similarity. Although DBSCAN, kernel density estimation and Local Getis-Ord are widely used for attribute inequality to derive the city centre, there are two common drawbacks: the first is that the clustering quality heavily depends on the user-defined parameters, which vary with the attribute distribution and cannot be assigned easily, and the second is that their similarity measure considers only the immediate similarity between two objects, which cannot differentiate their real differences, as reflected by ignoring the clustering tendency of geographical phenomena. Compared to the above methods, DBSC [42] does not require any parameters, and the similarity between an object and a set of objects can consider both local and global differences. The good performance of the DBSC algorithm has been demonstrated by both simulated and actual datasets. However, the threshold used to determine the similarity of attributes in DBSC is fixed, which directly leads to unsatisfactory clustering results when attributes are non-homogeneous in geographical space.

Considering these facts, we introduced information entropy [45] to measure attribute similarity among commercial lands, and a corresponding threshold considering local information was designed. The proposed method was derived from the DBSC algorithm, but used a different measurement strategy. Some basic concepts were defined and used to explain the proposed clustering algorithm.

Definition 3. *Attribute information entropy: For spatial objects (commercial lands) in $A = \{x_1, x_2, \ldots, x_n\}$, the attribute (composite index of urban centrality) of A is denoted as $R = \{r_1, r_2, \ldots, r_n\}$. The attribute probability of an object $x_k \in A$ in R can then be expressed as:*

$$P_k = r_k / \sum_{k=1}^{n} r_k \tag{7}$$

The attribute information entropy of *A* is expressed as:

$$H(M_r) = - \sum_{k=1}^{n} P_k \log_2 P_k \tag{8}$$

According to the Maximum Entropy Theory, when the attribute values of all objects in a dataset are equal, the information entropy of the data set will reach the maximum value, known as the principle of "equal probability maximum entropy". The more spatial objects in *A* that are similar to each another, the $H(M_r)$ value is higher. In contrary, when the spatial objects in *A* show great differences, the $H(M_r)$ value is lower. Clustering follows the principle of "Equal probability maximum entropy" such that a spatial object is partitioned into the most similar cluster.

Definition 4. ***Directly adjacent neighbour:*** *For an object P_1 in C-DT, the spatial neighbors of P_1 contain all the objects that directly link to P_1 denoted by* ***Directly adjacent neighbour*** *(P_1).*

Definition 5. ***Indirectly adjacent neighbour:*** *If a chain $P_1, P_2, \ldots, P_{n-1}, P_n$ meets the requirements that P_n only belongs to $DN(P_{n-1})$, P_k belongs to $\{DN(P_{k-1}) \cap DN(P_{k+1}), 1 < k < n\}$ and P_2 only belongs to $DN(P_1)$ then $P_3, P_4, \ldots, P_n$ will be the* ***Indirectly adjacent neighbours*** *of P_1.*

Definition 6. ***Information Entropy Measurement:*** *In C-DT, for spatial objects x_1 and $x_2 \in$* ***Directly adjacent neighbor*** *(x_1), x_i and x_j denotes their attributes, respectively. Therefore, the similarity between x_i and x_j is defined as*

$$S(x_1, x_2) = H_{x_1 x_2}(M_r) \tag{9}$$

$$H_{x_1 x_2}(M_r) = -\sum_{k=1}^{2} P_k \log_2 P_k, P_k = \frac{r_k}{\sum_{k=1}^{2} r_k} \tag{10}$$

Formula (9) considers only the similarity between two objects; it faces difficulties in obtaining their actual differences, because the binary relationship may conceal the tendency of attribute similarity in the spatial distribution. When comparing similarities between the object and cluster, many objects should be used in the similarity measure. The information entropy is introduced to overcome the defects of the similarity measure with the binary relationship, which can consider both local (two objects) and global (one object and a set of objects) differences.

For a set of spatial objects B (the number of objects is greater than or equal to two) in C-DT and $R = \{r_1, r_2, \ldots, r_m\}$, the similarity between an object Q ($R = r_Q$ and cluster B can be denoted by $S(Q, B)$, and is expressed as follows:

$$S(Q, B) = H_{QB}(M_r) \tag{11}$$

$$H_{QB}(M_r) = -\sum_{k=1}^{m+1} P_k \log_2 P_k, P_k = \frac{r_k}{\sum_{k=1}^{m+1} r_k} \tag{12}$$

Definition 7. ***Neighbourhood Entropy:*** *For an object P in C-DT, all the objects of its directly adjacent neighbours, including itself, are defined as Sub_SN(P). The neighborhood entropy of P is is expressed as*

$$Hnear(P) = H_{n=\|Sub\_DN(P)\|}(M_r) / \| Sub\_DN(P) \| \tag{13}$$

where $H_{n=\|Sub\_DN(P)\|}(M_r)$ denotes the information entropy of *Sub_DN(P)* and $\| Sub\_DN(P) \|$ is the number of $Sub\_DN(P)$ Neighborhood entropy measures the similarity between *P* and directly adjacent neighbors, and if the object *P* is similar to its directly adjacent neighbors, the value of $Hnear(P)$ is high.

Definition 8. ***Clustering Centre Point:*** *In C-DT, Cluster_Point expresses the clustering centre point, defined as*

$$Cluster\_Point(P) = max(Hnear(P)) \tag{14}$$

By computing the $Hnear(P)$ of all objects sorted in descending order, the maximum value is extracted as the clustering centre point. The clustering centre point is not fixed; when a cluster is formed, the clustering centre point will be updated among the objects that are not clustered.

Definition 9. *Cluster: For a Cluster_Point(P) breadth-first search [46] is used to visit its directly and indirectly adjacent neighbours. One cluster is formed if they satisfy the threshold (see Section 3.1.3) and no new object is added to the cluster.*

Definition 10. *Noise: Given an object P, if P does not belong to any cluster, P will be identified as noise.*

3.1.3. Algorithm Description

Determination of the Attribute Clustering Threshold

From Definitions 3 and 6, when attribute values in one cluster are equal, $H(M)$ reaches the maximum value, denoted by $H_{max}(M)$, and the cluster achieves the maximum similarity. If an object Q is similar to one known object or cluster, then the difference between the information entropy of the new cluster including Q and $H_{max}(M)$ of the new cluster will be small, according to the monotonicity of information entropy. In contrast, the actual difference will be great. Therefore, an object Q is similar to one object or cluster if the following is satisfied:

$$H(M') / H_{max}(M') \geq \theta \tag{15}$$

where $H(M')$ is the information entropy of a new cluster and $H_{max}(M')$ is the maximum information entropy of a new cluster. An intermediate value for the parameter, $0 < \theta < 1$, is defined a priori, which can adapt to a range of different clusterings.

The values in the same cluster should be similar to one another, and dissimilar to those in different clusters, in order to obtain a suitable θ The partitioning best method (PBM) index [47] provides a measure of how "neatly split" the clusters are, and is expressed as:

$$\mathrm{PBM} = \left(\frac{1}{N_c} \frac{E_1}{E_{N_c}} D_{N_c} \right)^2 \tag{16}$$

$$E_{N_c} = \sum_{i=1}^{N_c} E_i, \ E_i = \sum_{j=1}^{N_i} \| x_j - v_i \|, \mathrm{D}_{N_C} = max \sum_{i,j=1}^{N_C} (\| v_i - v_j \|) \tag{17}$$

where N_c is the number of clusters, N_i is the total number of points in C_i cluster, v_i is the centre of the C_i cluster, and D_{N_c} measures the maximum separation between a pair of clusters. When this index is large, the rule of k standard deviations obtains a better result.

Detection of Clusters Using Our Spatial Clustering Algorithm

The proposed methodology for detecting clusters involves four distinct stages.

Step 1. Construct spatial proximity relationships among commercial land blocks. This step can be implemented by the following operations:

① Convert commercial land blocks into discrete points, and then construct Delaunay triangulation for these points.
② Remove the edges from the Delaunay triangulation using global edge-length constraints.
③ Remove the edges from the Delaunay triangulation using local edge-length constraints.

Step 2. Compute the neighbourhood entropy. For each commercial land, calculate its neighbourhood entropy and then sort the commercial land blocks in descending order of neighbourhood entropy.

Step 3. Implement attribute clustering. This step involves the following operations:

①   Select the maximum value of the neighbourhood entropy as $Cluster\_Point(P)$.

②   Use the breadth-first search to visit directly and indirectly adjacent neighbours of P in descending order of their neighborhood entropy. The cluster is formed if they satisfy Formula (15) and no new object is added to the cluster, thus identifying them as clustered.

③   Traverse all points that are not clustered by iterating operations (1)–(2). When clustering is finished, any point that does not belong to a cluster will be identified as noise.

3.1.4. Algorithm Analysis

Implementation Procedure of the Algorithm

The implementation procedure of our algorithm is further illustrated through the simulated data in Figure 3. The attribute values (i.e., comprehensive index) of objects (i.e., centroids of commercial lands) were labelled in Figure 3a. Delaunay triangulation with global and local edge-length ($\gamma = 0.2365$) constraints was utilised to model spatial proximity relationships, and the clustering result is shown in Figure 3b. The next step computed the neighbourhood entropy of all objects based on the spatial proximity relationships. P1 was first selected as the clustering centre point (see Figure 3c), and the neighbours of P1 were detected as the first cluster C1 ($\theta = 0.98$). P2 was then identified as the new clustering centre point, and the neighbours of P2 formed the second cluster C2. Finally, P3 was selected as the Clustering Center Point, and the neighbours of P3 were classified as cluster C3. The clustering results of simulated datasets (see Figure 3d) illustrated that objects in the same cluster are indeed similar in both the spatial and attribute domains.
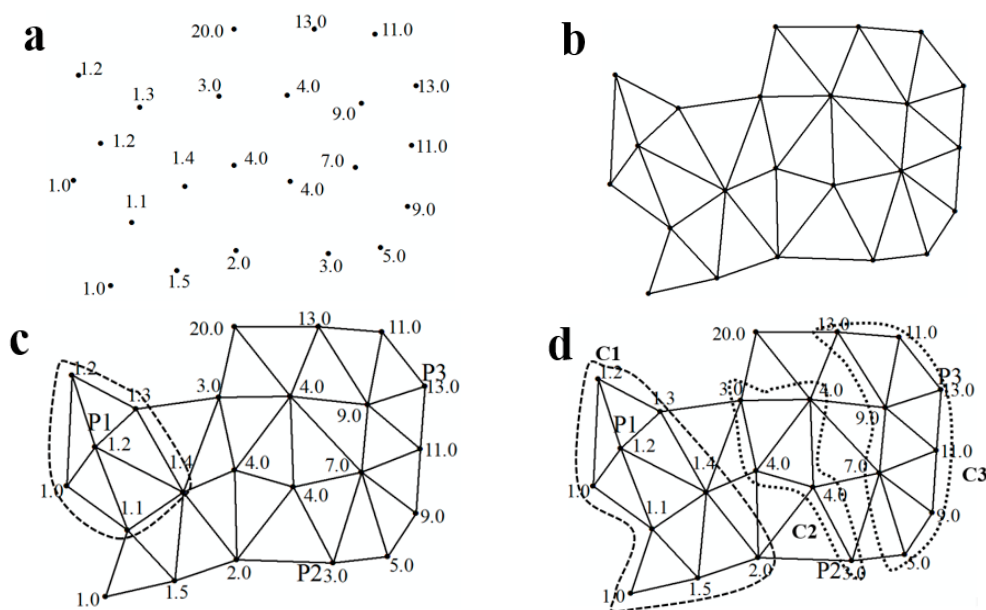

Figure 3. Spatial clustering implemented using our algorithm: (**a**) simulated data with attributes; (**b**) spatial clustering considering spatial proximity; (**c**) the clustering centre point; (**d**) spatial clustering considering spatial proximity and attribute similarity.

Algorithm Validation

To validate the proposed algorithm in this paper, a 3-D (1-D attribute) simulated dataset was utilised to test our algorithm. The attribute value (AV) range of objects in each predefined cluster (C1–C5) is labelled in Figure 4a. The attribute values of objects in each cluster were randomly assigned within a certain range. To ensure randomization, for each object, 20 replications of the attribute value

were generated, and the average value of these 20 replications for each object was set as the final attribute value.
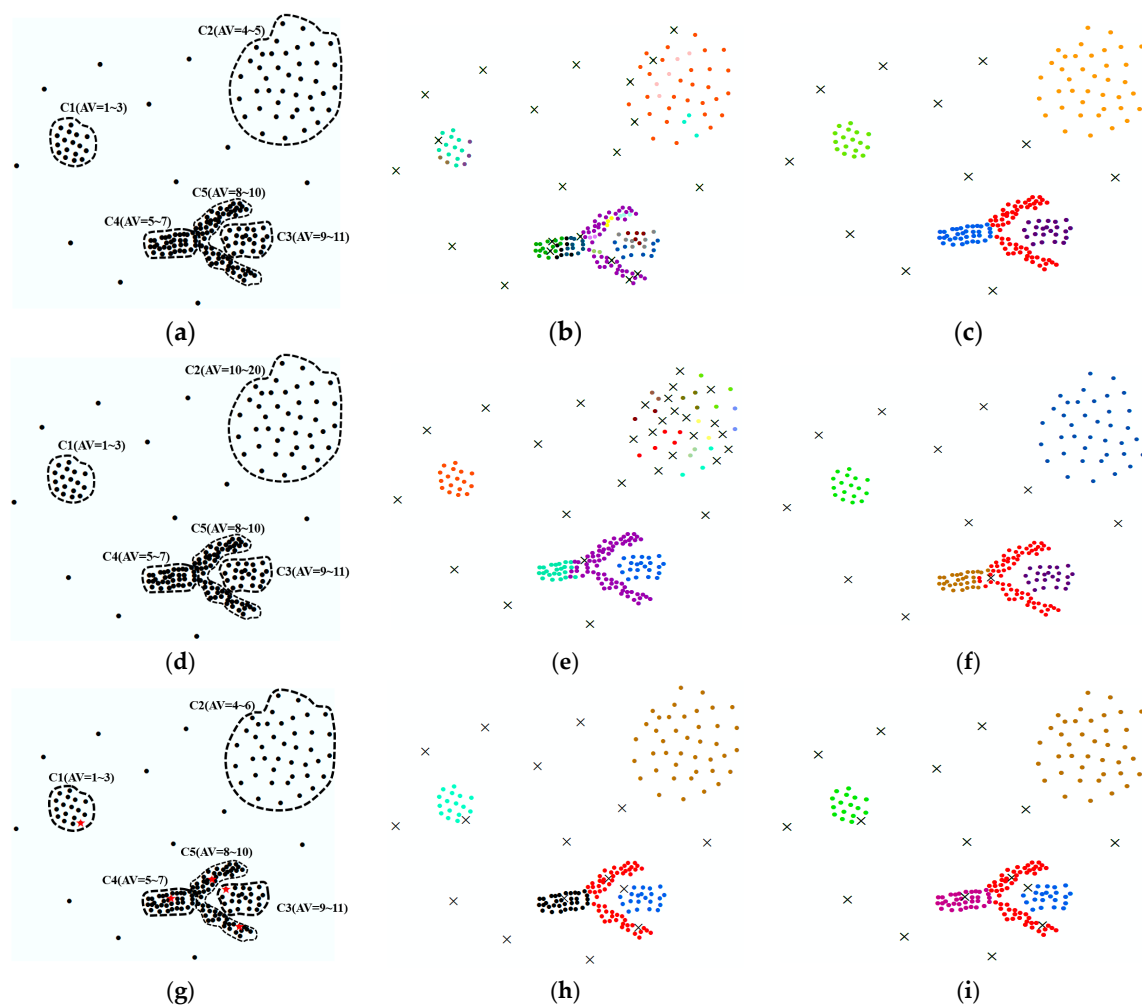

Figure 4. Clustering results from our method and DBSC algorithm (× represents noise). (**a**) Simulated dataset 1; (**b**) DBSCAN algorithm (DBSC) clustering ($T = 1.035$); (**c**) Our result ($\theta = 0.982$); (**d**) Simulated dataset 2; (**e**) DBSC clustering ($T = 2.216$); (**f**) Our result ($\theta = 0.97$); (**g**) Simulated dataset 3; (**h**) DBSC clustering ($T = 0.916$); (**i**) Our result ($\theta = 0.984$).

The two-level strategy algorithm was first used to construct spatial proximity relationships among 2-D planar points (the result can be seen in Figure 2), and attribute clustering was then implemented. The attributes were non-homogeneous in geographical space—whereas the attribute value range in C2 was different from the others. The clustering results obtained using the DBSC algorithm and our method are shown in Figure 4b,c. Our method separated the five clusters and the isolated noise very well, but the DBSC algorithm detected 17 clusters. To observe this type of non-homogeneous effect, we broaden the attribute value range in C2 (see Figure 4d). As shown in Figure 4e,f, our method still performed well, while the DBSC algorithm misclassified some objects in C2 as noise. The DBSC algorithm could not provide satisfactory results, primarily because the global parameter could not adapt to the non-homogeneous phenomena. In addition, Figure 4g shows the sensitivity test to the outliers, where the attributes were homogeneous in geographical space (the attribute difference in each cluster is similar). Some clusters were assigned as outliers (denoted by ★), with the attribute values of outliers randomly set from 40 to 60. As seen in Figure 4h, i, our method and the DBSC algorithm could separate five clusters and outliers very well.

3.2. Characterising the City Center

Delineating the city centre by using spatial information requires not only geometric and topological attributes (spatial proximity relationships in Section 3.2.1), but also socio-economic attributes. According to a related analysis of the city centre for China's urban study, we suggested three key factors for characterising the city centre: geometry (area and aggregation degree), economy (employment density [48], and land price [12,49]), and traffic accessibility (road network density and distance from the nearest road [50]). These were then combined into a comprehensive index to measure the urban centrality of commercial lands.

3.2.1. Extraction of Indicators

In this study, several indicators, such as aggregation degree, employment density, road network density, and distance from the nearest road could not be obtained directly because commercial land data features high spatial resolution. Thus, we used geographic information system (GIS) to support the analysis.

(1)   Aggregation degree of commercial land. The distribution of commercial land in city centre areas tends to be clustered, and the area of each commercial plot cannot measure the unequal distribution of commercial land. Therefore, we considered the area and space occupied by commercial land, to measure the aggregation degree of commercial land. We used an approach provided by Ai and Van Oosterom [51] to compute the space occupied by commercial land. A geometric construction similar to a Voronoi diagram was created based on the skeleton of the Delaunay triangulation built using a polygon (commercial land) boundary, as shown in Figure 5a. The aggregation degree of commercial land is expressed as

$$Den(s_i) = AP_i / AV_i \tag{18}$$

where s_i denotes a commercial plot, $Den(s_i)$ denotes the aggregation degree of s_i, AP_i denotes the area of s_i, and AV_i denotes the area of space occupied by s_i. The range of $Den(s_i)$ is 0 to 1. If $Den(s_i)$ tends to 1, then the distribution of commercial land is clustered. The higher clustering of commercial plots was in the city centre areas, as shown in Figure 5b (Region 1). In contrast, when $Den(s_i)$ tended to 0, the distribution of commercial land was decentralized, as shown in Figure 5b (Region 2).
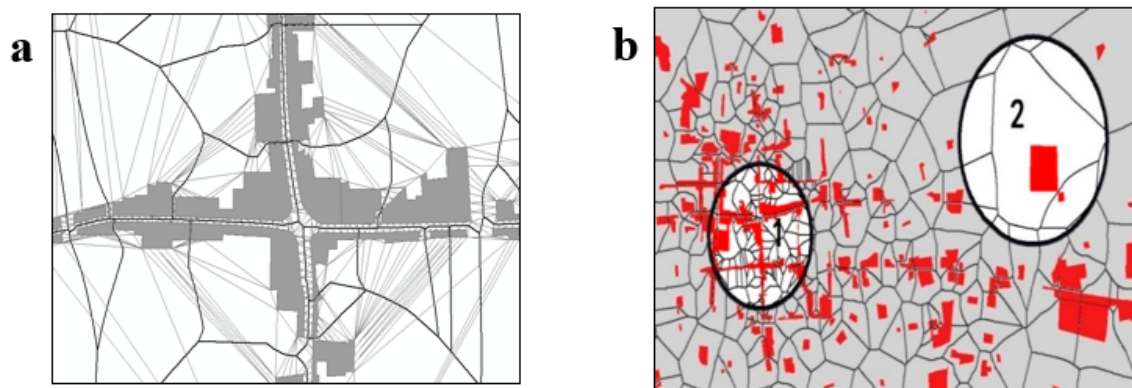

Figure 5. (**a**) Similar Voronoi polygon (black bold line) of the surface entity (gray block) built using Delaunay triangulation (black line); (**b**) aggregation degree of commercial land.

(2) Road network density and distance from the nearest road. Many urban districts with high commercial functions have been formed around the road network. Those regions feature high accessibility, namely, the comercial lands located in the city centre areas show higher accessibility. Road network density and distance from the nearest trunk road are important indicators for measuring accessibility. To compute the road network density for commercial land, the buffer operation of GIS was used to determine for how long the roads are located in a circular area with a radius distance of 1 km within the commercial land, as shown in Figure 6. Therefore,

$$RoadDen(s_i) = LP_i / AB_i \qquad (19)$$

where s_i denotes a commercial plot, $RoadDen(s_i)$ denotes the road network density of s_i, AB_i denotes the buffer area of s_i (the buffer radius is usually set to 1 km), and LP_i denotes the sum length of roads that are located inside the circular area of the commercial plot.
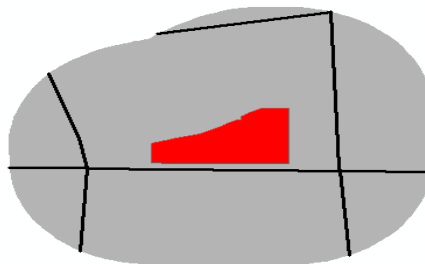

Figure 6. Road network density obtained from the buffer operation of GIS.

We also adopted the nearest neighbour analysis of GIS to compute the distance from the commercial land to the nearest road, including trunk roads, road intersections, and subways.

(3) Employment density. Broadly, the distribution of employment data was used to identify city centres. Conventionally, employment density obtained from the census tract failed to meet our statistical zone (i.e., commercial land). We computed the employment density of commercial land as the density of the census tract if the primary part of the commercial land fell into this census tract. As shown in Figure 7, the employment density of commercial land 1, 2 and 3 was, respectively, the density of CT-1, CT-2 and CT-3. The value of commercial land 4 should have been the same as commercial land 2 because the primary part of the commercial land fell into CT-2.
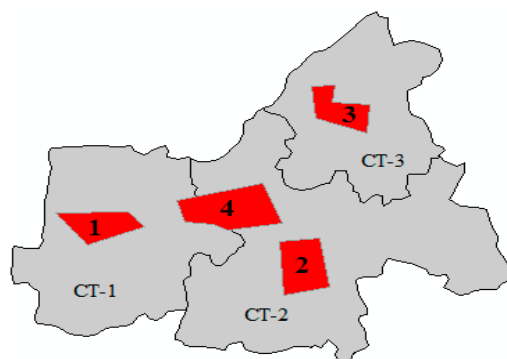

Figure 7. Employment density of commercial land (1, 2 ,3 and 4 represent different commercial lands).

3.2.2. Index of Urban Centrality

The next step combined these indicators into a comprehensive index to represent the "urban centrality" of each commercial land. Factor analysis [52] was originally developed by psychologists to reduce many variables to a smaller number of underlying factors, dimensions, or components. In this study, we adopted this method to reduce the dimensions of multidimensional attribute data, aggregating geometric and thematic attributes into a comprehensive index of "urban centrality". Using the variance contribution rate of the factors as weight, factor analysis was applied to formulate the comprehensive index value F, expressed as

$$F = \frac{\omega_1}{\sum_{i=1}^{m} \omega_i} F_1 + \frac{\omega_2}{\sum_{i=1}^{m} \omega_i} F_2 + \cdots + \frac{\omega_m}{\sum_{i=1}^{m} \omega_i} F_m \qquad (20)$$

where F_m mdenotes the extracted factors, ω_i denotes the variance contribution rate of the factors, and $\sum_{i=1}^{m} \omega_i$ denotes the total variance contribution rate.

3.3. Extracting the City Centre Area

After spatial clustering considering both spatial proximity and attribute similarity operations, we could detect clusters with different urban centrality values. The next step was to derive the city centre from these clustered commercial lands. First, we computed the mean value of each cluster and used positive standard deviation values [12,35,53] to delimit the city centre areas showing higher urban centrality. Second, polygons were used to determine the precise boundaries of the city centre areas consisting of commercial lands. In this study, we utilised Delaunay-based shape reconstruction to generate the city centre area from a set of commercial land clusters.

Many studies have been proposed in the literature for boundary representation. The Delaunay-based method captures geometric and topological information of the objects well, and it is simpler and more systematic than other methods. As depicted in Figure 8b, we initially constructed the Delaunay triangulation of the boundary points for commercial lands. Boundary points should be used to increase density according to the Gestalt law of proximity (Figure 8a). Triangle filtration using Peethambaran and Muthuganapathy [54] was subsequently employed to extract the boundaries of the city centres (Figure 8c). This method was selected because it can derive an acute and unique shape without any external parameters. Triangle filtration proceeded by iteratively removing all thin boundary triangles if the CIRCUMCENTER and REGULARITY constraints are satisfied (for additional details, see inside triangle filtration).
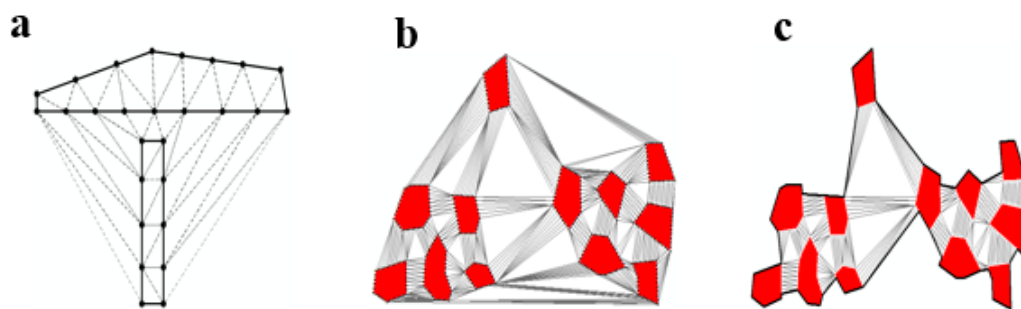

Figure 8. Extracting the city centre areas from commercial lands using triangle filtration: (**a**) boundary points with increased density; (**b**) Delaunay triangulation of boundary points for commercial lands; (**c**) boundary generated using triangle filtration method

4. Application

The proposed methodology was applied to delineate city centres in two cities from China, Yangzhou and Wenshan. The two cities are markedly different cities. First, Yangzhou is a city located in the Yangtze River Delta, the largest economic zone in China. With its development in recent years, Yangzhou has evolved into a polycentric city. However, Wenshan is a new and small city located in the southwest basin region of China. The two cities represent two typical urban spatial structures in China, i.e., polycentric and monocentric cities. The analysis of the city centre could be broadened to other cities in China.

The research was performed by considering two cities: Yangzhou and Wenshan. Since the algorithm steps were the same in our cases, the first application will be presented as the primary example. The urban land use dataset was provided by the Yangzhou Planning Bureau for the year 2012 in the first application. The dataset is composed of 735 records of commercial lands, and their spatial distributions are shown in Figure 9. Different stages of the recognition process and their outputs for the presented case study are illustrated in the following.
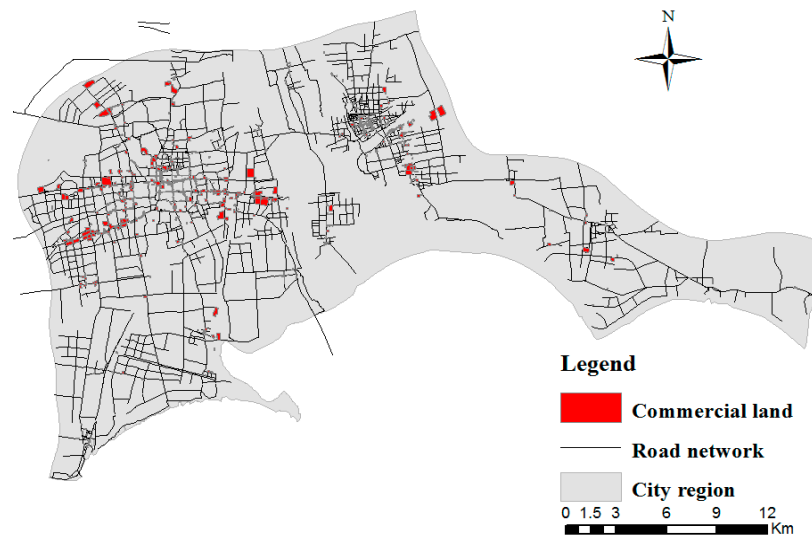

Figure 9. Spatial distribution of commercial land in Yangzhou.

4.1. Construction of Spatial Proximity Relationships

As shown in Figure 10, the centroid was used to represent commercial land and the Delaunay triangulation is introduced for the construction of spatial proximity among these points. Delaunay triangulation with global and local edge-length ($\gamma = 0.4638$) constraints was utilised to model the spatial proximity relationships, and the clustering results are shown in Figure 10a–d. Delaunay triangulation clustering can discover clusters with different shapes, and is robust with outliers. The final result reflected the spatial distribution pattern of commercial land in Yangzhou.
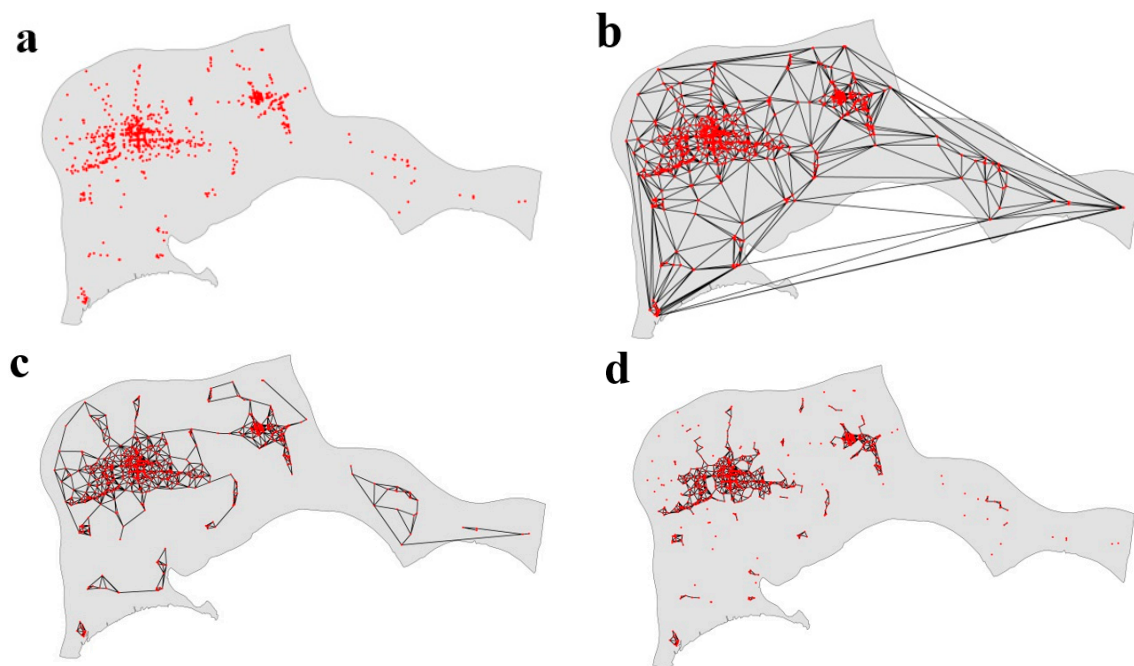
Figure 10. The process of spatial clustering considering spatial proximity: (**a**) the centroids of commercial land; (**b**) Delaunay triangulation of centroids; (**c**) result using the *Global constraint*; (**d**) result using the *Local constraint* (γ = 0.4638).

4.2. Clustering Commercial Lands with Attribute Similarity

4.2.1. Using the Factor Analysis Method for the Index of Urban Centrality

Factor analysis can be implemented through several methods; the principal axis factor method was applied in this study. We considered the geometric and thematic attributes, as discussed above in the analysis, and a 6 × 6 matrix of correlation coefficients was created. Three primary factors with eigenvalues over 1.0 were extracted using the principal axis factor and maximum variance rotation methods. Table 1 provides the factor loadings, and the accumulation variance reached 91.957%. According to Formula (3), the comprehensive index of urban centrality for each commercial land is represented as:

$$F = 46.726\% * (f1 * value) + 28.190\% * (f2 * value) + 17.041\% * (f3 * value)$$

where $f1$ denotes the first factor score, and *value* is the indicator value. To facilitate the calculation, we normalize the F value to the range of (0, 10).

Table 1. Factor loading matrix.

Item	Variables	Factor		
		1	**2**	**3**
Thematic attributes	Land price	0.728	−0.556	0.202
	Distance from the nearest road	0.627	−0.683	0.019
	Road network density	0.736	0.650	0.098
	Employment density	0.908	−0.114	0.006
Geometrical attributes	Aggregation degree	0.694	0.686	0.140
	Area	−0.183	0.095	0.976

4.2.2. Clustering Commercial Land with Attributes

After the global and local trimming operations, we applied the clustering method based on information entropy to the comprehensive index of urban centrality. Clusters were identified, and noises were removed. Figure 11a, b show the clustering results using our method and the DBSC algorithm (different colours represent different clusters). Basic statistical information (see Figure 12 and Table 2) from the clustering results obtained using the two approaches is also provided, including the number of clusters and noises, the mean of each cluster, the standard deviation of each cluster, the trend of mean values of clusters and the coefficient of variation (CV) value of clusters. By a simple comparison, both algorithms presented a general attribute distribution with similar patterns, as reflected by the trend of mean values of clusters. However, some cluster details from our method and DBSC were rather different. As shown in Table 2, there were 72 clusters and 26 noise points discovered by our method. Figure 12c shows that there was a significant difference between the adjacent clusters, and that the variation in each cluster was small (see Figure 12a). Compared to the results obtained using DBSC, the number of clusters and noise points generated using DBSC algorithm were higher than the number generated using our method, leading directly to a smaller difference between adjacent clusters (i.e., the CV value of mean values of clusters was smaller than ours). The DBSC algorithm could not provide satisfactory results. This outcome was primarily due to the global parameter used for clustering commercial lands with attributes not adapting to the local variation, which failed to recognize similar clusters; therefore, we handled this local variation by setting the relative variation proportion θ, with the PBM index being used to find a suitable θ value. Figure 13 shows the resulting curve plot, with the x-axis set as the θ value and the y-axis set as the PBM index. The maximum value of the PBM index was achieved when θ equals 0.986. Notably, the PBM index began to change when θ equalled 0.96, because the clustering results of spatial proximity and attribute similarity were considered the same in the interval (0, 0.958].
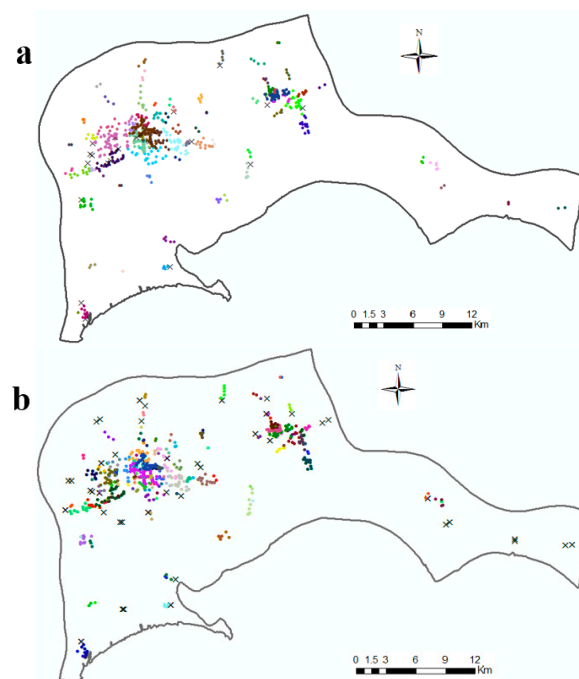

Figure 11. Spatial clustering results for the urban centrality index obtained using: (**a**) our method ($\theta = 0.986$) and (**b**) the DBSC method ($T = 0.3965$).
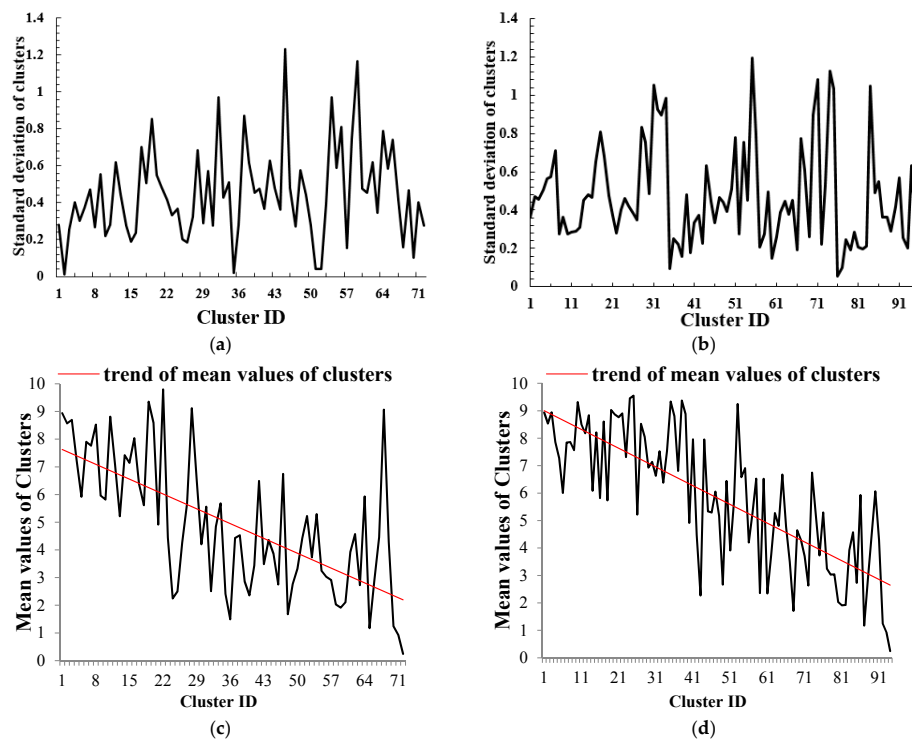
Figure 12. Comparison of spatial clustering results between the DBSC algorithm and our method. (**a**) Standard deviation of each cluster using our method; (**b**) standard deviation of each cluster using DBSC; (**c**) mean value of each cluster using our method (▬ represents the trend of mean values of clusters); (**d**) mean value of each cluster using DBSC (▬ represents the trend of mean values of clusters).

Table 2. Statistics for clustering results in Figure 10.

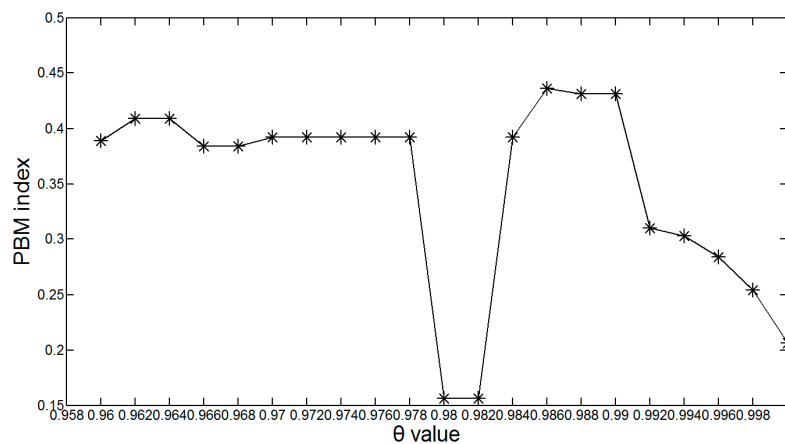|	Our Method	DBSC Algorithm
Number of Clusters	72	95
Number of Noise Points	26	54
CV value of mean values of clusters	0.50	0.46

Figure 13. Curve plot of the θ value and the PBM index.

4.3. Results and Validation

To delimit the geographical extent of the city centres, we further examined the tail values of the cluster results, using positive standard deviation values. According to a study on delimitating the city centre, a value of three standard deviations is often suggested [35,53]. Figure 14a, c show the urban centrality classification of the commercial lands using a standard deviation of 3 with our cluster and the DBSC cluster results (i.e., mean value of each cluster). The red commercial lands in Figure 14 were concentrated in two areas of the city, forming a polycentric structure. We confirmed that the two areas with an urban centrality value over 7.28 (tail value of our method) or 7.38 (tail value of DBSC) had a larger population density, road network density, and a higher land price. As polygon clusters had been identified from commercial lands, we used the method discussed above (see Section 3.3) to derive the boundaries of the city centres from the clustered polygons. Figure 14b, d map the identified two city centres (centre 1 and centre 2) with precise boundaries using two alternative methods. We validated the boundaries produced using computational models to compare them to city centres from prior available knowledge. According to the Comprehensive Plan of Yangzhou City (1996–2010), the functional core was primarily located in two regions: the traditional Wen Changge centre, and the new He Dong centre. The comparative city centre representations provided narrative descriptions of the Yangzhou city centres extents. These descriptions are mapped with precise boundaries (See the blue areas in Figure 15a). Figure 15a, b show the computed city centres (purple areas) using two alternative methods versus the comparative city centres (blue areas).
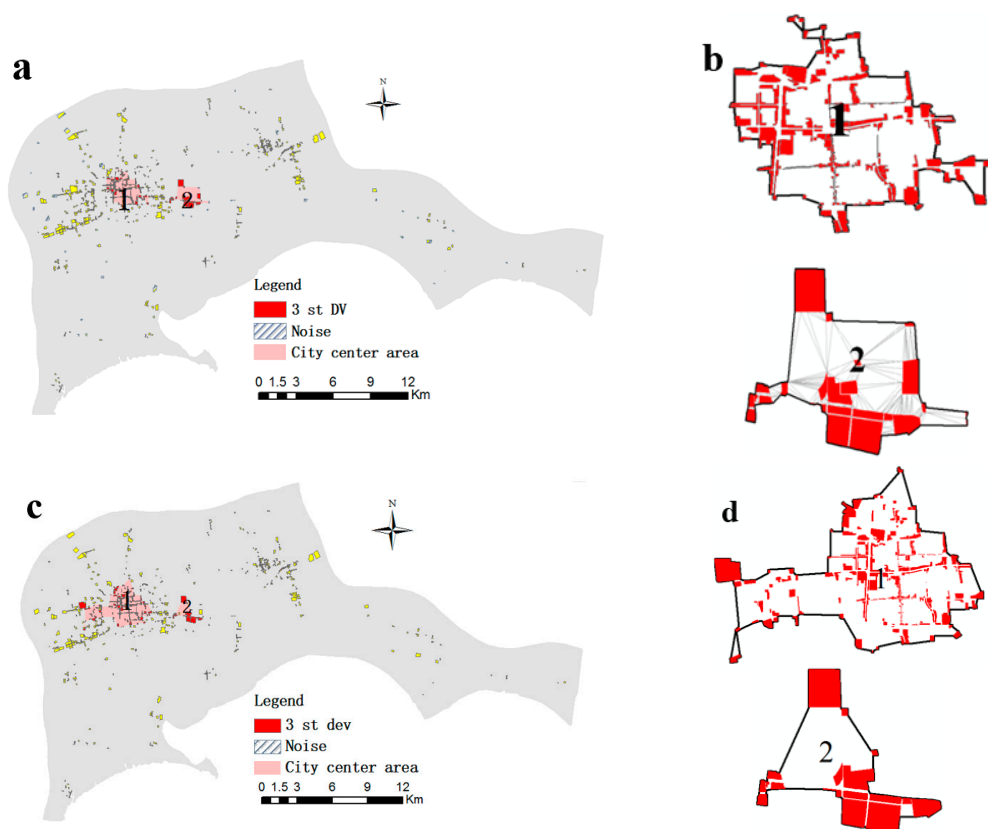

Figure 14. Extraction results of city centre areas in Yangzhou using urban land as the primary data source: (**a**) city centres derived using our method; (**b**) geographical extent of city centres using triangle filtration; (**c**) city centres derived by DBSC; (**d**) geographical extent of city centres using triangle filtration.

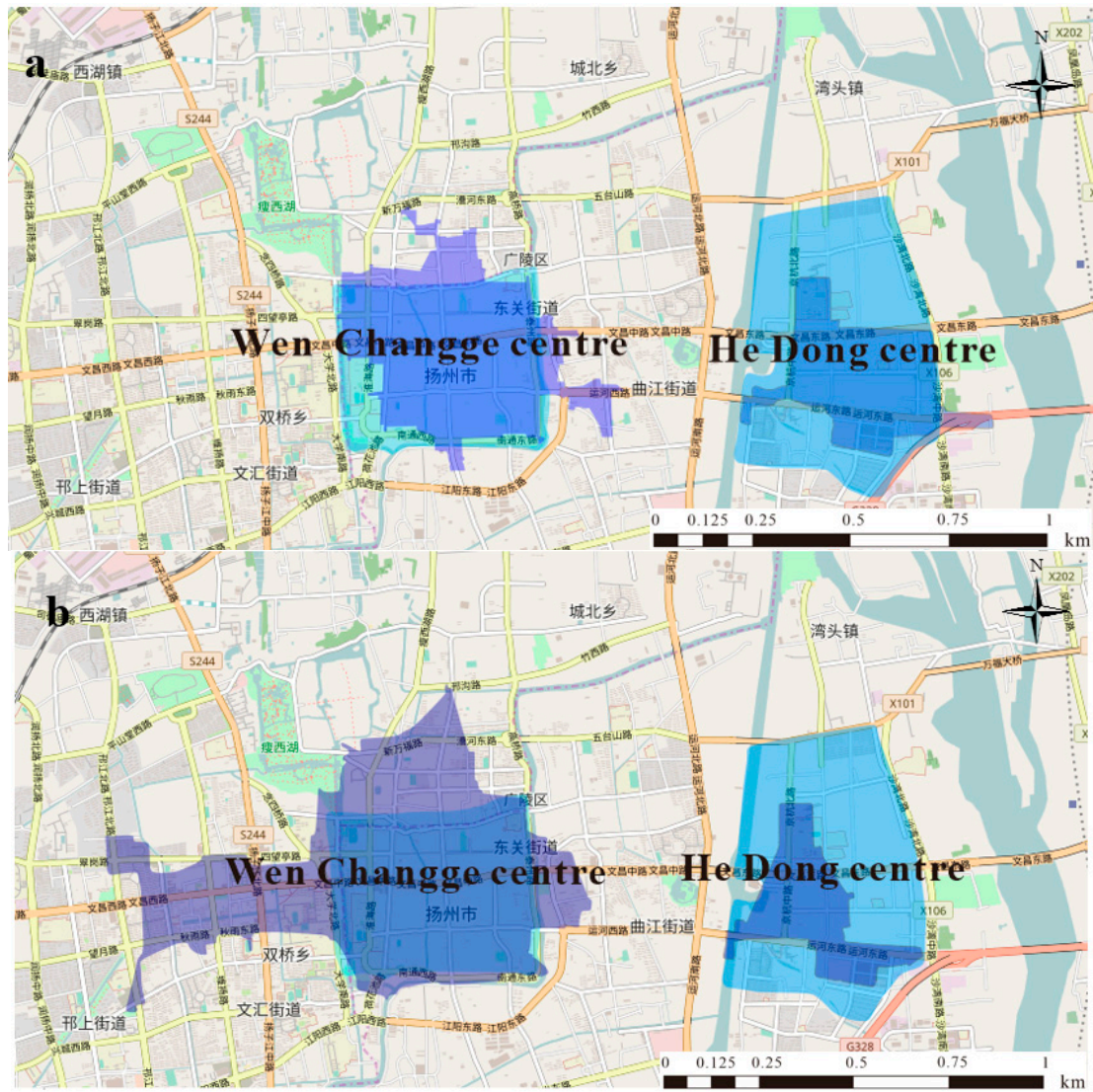**Figure 15.** Comparison of the boundaries of city centres delineated by the two algorithms (in purple) with comparative city centres (in blue): (**a**) our result; (**b**) DBSC result. Background mapping from OpenStreetmap.

We further used the *F1-Score*, which is the harmonic mean of precision and recall for a quantitative comparison of the overlap between the computed and comparative city centre areas. The evaluation indicator is calculated as follows:

$$precision = \frac{a_{overlap}}{a_{computed}}, \; recall = \frac{a_{overlap}}{a_{comparative}} \tag{21}$$

$$F_1 - score = 2 \times \frac{precision \times recall}{precision + recall} \tag{22}$$

where $a_{computed}$ is the city centre area as delimited by the algorithm, $a_{comparative}$ is the area of the comparative city centre, and $a_{overlap}$ is the area where the computed and comparative city centres overlap.

The evaluation results in Figure 15 are presented in Table 3. We found that the *F1-Score* of our result reached 82.19% and 54.84%, and all indicators were larger than the corresponding indicators from the DBSC results, i.e., 64.38% and 41.11%. The He dong centre computed using the algorithms

was smaller than the comparative city centre. The computational models omitted the north part in Figures 14 and 15, due to many instances of residential land in these areas, and a lack of commercial land. Given the results from the year 2012, the urban development and its functions evolved as originally planned. Such evaluation illustrates that our methodology based on commercial land has the potential to yield satisfactory results, possibly leading to the development of a new method of confirming the recognition of the city centre.

Table 3. Comparison of overlap between the computed and comparative city centres.

City Centre	Algorithm	Precision	Recall	F1-Score
Wen Changge	Our method	0.8053	0.8395	0.8219
	DBRS	0.9606	0.3837	0.5484
He Dong	Our method	0.4819	0.9756	0.6438
	DBRS	0.9973	0.2587	0.4111

4.4. Case study of Wenshan

In this section, we verified the feasibility of the proposed method using a different city in China, i.e., Wenshan. Figure 16a presents the spatial distribution of commercial lands for the year 2013. Delaunay triangulation with global and local edge-length constraints was utilised to model the spatial proximity relationships, and the clustering results are shown in Figure 16b, c. Factor analysis was then applied for the comprehensive index to represent the "urban centrality" of each commercial land, and the accumulation variance reached 86.764%. Similarly, two different methods were used and compared to detect clusters with higher urban centrality. Figures 17a and 18a illustrate the cluster results detected by the methods. To verify the feasibility of our method, the final city centre was computed using the classification of a standard deviation of 3 (See Figures 17b and 18b). The red commercial land areas in Figures 17b and 18b were concentrated in one area of the city, forming a monocentric structure. According to the Comprehensive Plan of Wenshan City (2006–2025), we mapped the geographic extent of the planned city centre (See Figures 17c and 18c). Based on the results, we compared our method and the DBSC algorithm by *F1-Score*. Table 4 presents the evaluation results. Compared to our method, the DBSC method generated a larger space for the city centre, as reflected by lower precision. Through the computation of the evaluation indictor, the proposed method was feasible for the different cities.
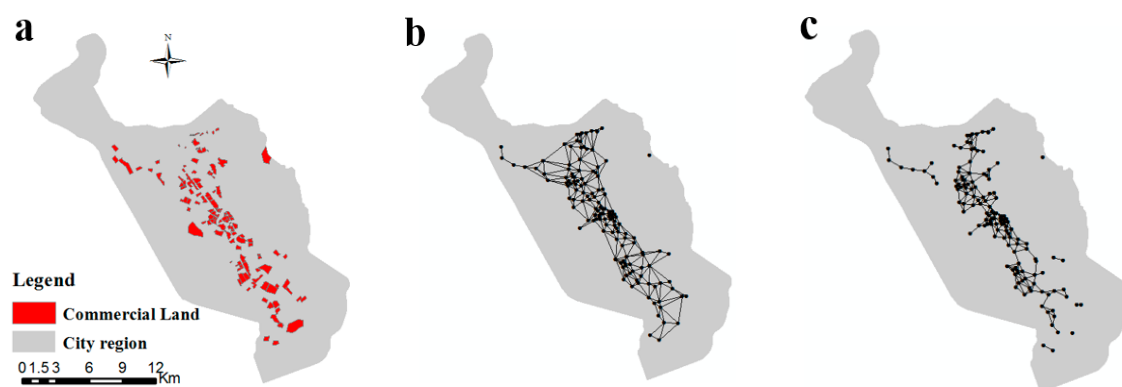

Figure 16. The process of spatial clustering considering spatial proximity:(**a**) the distribution of commercial land; (**b**) result using the *Global constraint*; (**c**) result using the *Local constraint* ($\gamma = 0.4213$).
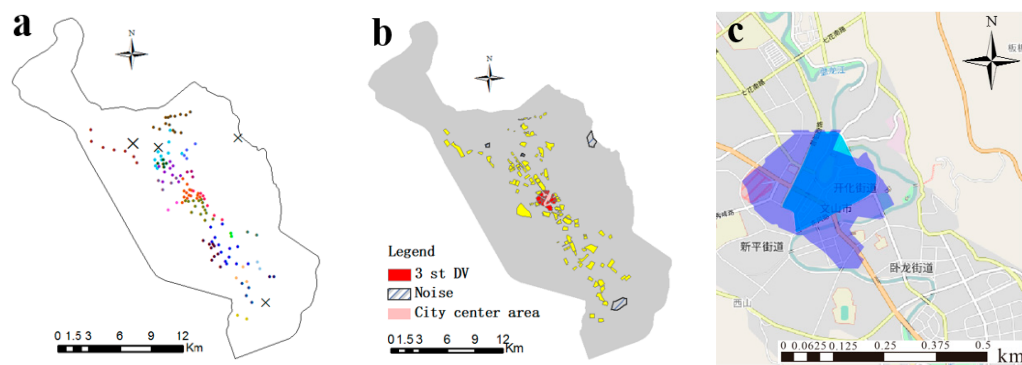
Figure 17. The identified city centres generated using our method: (**a**) spatial clustering results for the urban centrality index ($\theta = 0.98$); (**b**) the derived city centre areas; (**c**) the computed city centre (in purple) versus the comparative city centre (in blue). Background mapping from OpenStreetmap.

Figure 18. The identified city centres generated using the DBSC method: (**a**) spatial clustering results for urban centrality index (T = 0.436); (**b**) the derived city centre areas; (**c**) the computed city centre (in purple) versus the comparative city centre (in blue). Background mapping from OpenStreetmap.
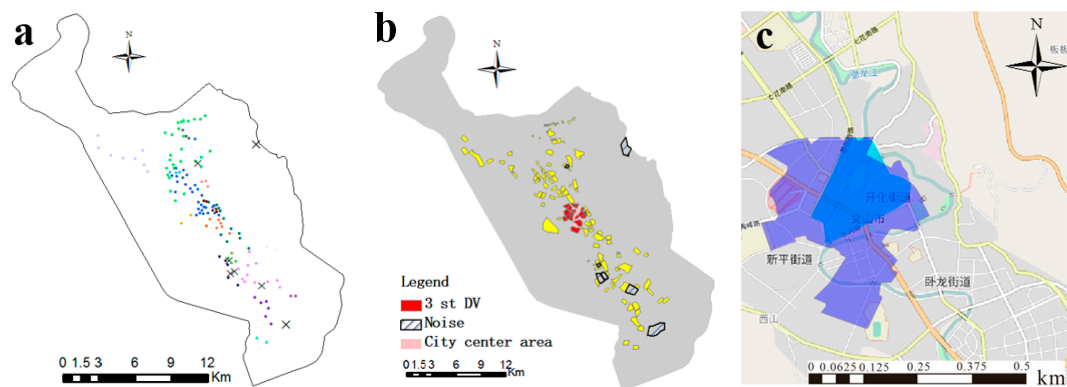
Table 4. Comparison of the overlap between the computed and comparative city centres.

Algorithm	Precision	Recall	F1-Score
Our method	0.4130	0.9744	0.5801
DBRS	0.3220	0.9590	0.4821

5. Conclusions and Outlooks

Considering that the construction of the urban spatial structure in the planning and design stages follows the forward process from planning to expression, this research contributes to providing a new method for the construction of an urban spatial structure via the urban land use. Using the city centre as an example, this study presents a methodology to delineate the city centre from commercial land data. The advantage of using urban land as a primary data source is that it has a much higher spatial resolution and smaller space granularity than conventional statistical units. This advantage affords the possibility of gaining an immediate impression of the distribution of geographic phenomena, and high precision in the statistical agglomeration of the city centre. The research considers the city centre to be a special region consisting of several commercial lands with different characteristics. The proposed methodology contributes to determining the city centre with the consideration of not only geometric and topological attributes, but also socio-economic attributes. Two types of cities were used to validate the effectiveness and practicability of our methodology, and the results showed that

constructing an urban spatial structure from urban land use is feasible, and that our method provides the methodological foundation of city centre evolution between the existing and planned conditions.

In this paper, we characterised the city centre using commercial land data as units relating to a series of indicators, and integrated these into a composite index of "urban centrality". Factor analysis was applied to formulate such a comprehensive index, and it was applied through statistical testing, i.e., an accumulation variance $\geq$80%. When different indicators or cities are considered, the question is whether this formulated method of comprehensive index is appropriate for characterising the city centre. The weighted sum calculation model is another method type for determining the comprehensive index in urban analysis [16,31]. This method determines the weight for each indicator, and these weighted indicators are then summed into a single index. Compared to factor analysis, the weight determined is a user-driven process, not a data-driven process. A previous study [16] showed that there is little difference between the two methods in urban structure analysis. Further research might suggest that when the experiment cannot satisfy the factor analysis method, the weighted sum calculation model can be used.

To delineate the city centre with a precise boundary, we proposed an adaptive spatial clustering method considering spatial proximity and attribute similarity. Although DBSCAN, kernel density estimation and Local Getis-Ord are widely used to derive city centre, the results are highly influenced by their parameters, especially when there is a little knowledge about the cities. In addition, their similarity measurement considers only the immediate similarity between two objects, which cannot differentiate their real differences. The DBSC algorithm was proposed to overcome these drawbacks, however, it suffered the limit of global parameter. The final city centre boundaries related directly to the process of clustering. Our proposed method emphasizes the constraints of DBSC algorithm for obtaining a more precise clustering result. To better test the validity, we compared our method and DBSC algorithm to identify simulated datasets and city centres. Our method obtained a more accurate clustering result, and was more suitable for cities with different urban spatial structures.

As urban land use is not available in China, we used two cities of different structure (i.e., polycentric and monocentric) to validate our methodology. We see two main extensions of our work in future research. One is that the same experiments should be carried out for cities in China to better represent and confirm our work. The other is that additional or different thematic attributes should be considered when indicating urban centrality of other cities.

Author Contributions: Yizhong Sun principally conceived the idea for the study and provided financial support. Jie Zhu was responsible for the design of the study, setting up experiments, completing most of the experiments and he also wrote the initial draft of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bourne, L.S. Internal structure of the city: Readings on urban form, growth, and policy. *Historian* **1963**, *26*, 1–18.
2. Yoshida, H.; Omae, M. An approach for analysis of urban morphology: Methods to derive morphological properties of city blocks by using an urban landscape model and their interpretations. *Comput. Environ. Urban Syst.* **2005**, *29*, 223–247. [CrossRef]
3. Ariza-Villaverde, A.B.; Jiménez-Hornero, F.J.; Ravé, E.G.D. Multifractal analysis of axial maps applied to the study of urban morphology. *Comput. Environ. Urban Syst.* **2013**, *38*, 1–10. [CrossRef]
4. Wang, L. Object-based spatial cluster analysis of urban landscape pattern using nighttime light satellite images: A case study of China. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 2328–2355.
5. Zhang, L.Q.; Shu, J. A GIS-based gradient analysis of the urban landscape pattern of shanghai metropolitan region. *Acta Polym. Sin.* **2004**, *28*, 78–85.

6. Faris, J.M.; Beever, L.B.; Brown, M. Geography Information System (GIS) and Urban Land Use Allocation Model (ULAM) techniques for existing and projected land use data. In Proceedings of the seventh National Conference on Transportation Planning for Small and Medium-Sized Communities, Little Rock, AR, USA, 28–30 September 2000.

7. Haque, A.; Asami, Y. Optimizing urban land use allocation for planners and real estate developers. *Comput. Environ. Urban Syst.* **2014**, *46*, 57–69. [CrossRef]

8. Feng, Y.; Liu, Y.; Tong, X.; Liu, M.; Deng, S. Modeling dynamic urban growth using cellular automata and particle swarm optimization rules. *Landsc. Urban Plan.* **2011**, *102*, 188–196. [CrossRef]

9. Jafari, M.; Majedi, H.; Monavari, S.M.; Alesheikh, A.A.; Kheirkhah Zarkesh, M. Dynamic simulation of urban expansion based on cellular automata and logistic regression model: Case study of the Hyrcanian Region of Iran. *Sustainability* **2016**, *8*, 810. [CrossRef]

10. Lüscher, P.; Weibel, R. Exploiting empirical knowledge for automatic delineation of city centres from large-scale topographic databases. *Comput. Environ. Urban Syst.* **2013**, *37*, 18–34. [CrossRef]

11. Egenhofer, M.J.; Mark, D.M. Naive geography. In *Spatial Informat-Ion Theory. A Theoretical Basis for GIS. Proceeding of the Sinternational Conference COSIT '95*; Frank, A.U., Kuhn, W., Eds.; Springer: Berlin, Germany, 1995; pp. 21–23.

12. Yu, W.; Ai, T.; Shao, S. The analysis and delimitation of Central Business District using network kernel density estimation. *J. Transp. Geogr.* **2015**, *45*, 32–47. [CrossRef]

13. Murphy, R.E.; Vance, J.E. Delimiting the CBD. *Econ. Geol.* **1954**, *30*, 189–222. [CrossRef]

14. Eubank, S.; Guclu, H.; Kumar, V.S.; Marathe, M.V.; Srinivasan, A.; Toroczkai, Z.; Wang, N. Modelling disease outbreaks in realistic urban social networks. *Nature* **2004**, *429*, 180–184. [CrossRef] [PubMed]

15. Wang, P.; González, M.C.; Hidalgo, C.A.; Barabási, A.-L. Understanding the Spreading Patterns of Mobile Phone Viruses. *Science* **2009**, *324*, 1071–1076. [CrossRef] [PubMed]

16. Zhang, C. An analysis of urban spatial structure using comprehensive prominence of irregular areas. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 675–686. [CrossRef]

17. Alonso, W. A theory of the urban land market. *Pap. Pro. Reg. Sci. Assoc.* **1960**, *6*, 149–157. [CrossRef]

18. Martin, D.; Higgs, G. Population georeferencing in England and Wales: Basic spatial units reconsidered. *Environ. Plann. A* **1997**, *29*, 333–347. [CrossRef]

19. Baumont, C.; Ertur, C.; Gallo, J. Spatial analysis of employment and population density: The case of the agglomeration of Dijon 1999. *Geogr. Anal.* **2004**, *36*, 146–176. [CrossRef]

20. Yang, X.; Huang, Y.; Dong, P.; Jiang, D.; Liu, H. An updating system for the gridded population database of China based on remote sensing, GIS and spatial database technologies. *Sensors* **2009**, *9*, 1128–1140. [CrossRef] [PubMed]

21. Ord, J.K,; Getis, A. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geogr. Anal.* **1995**, *27*, 286–306. [CrossRef]

22. Roth, C.; Kang, S.M.; Batty, M.; Barthélemy, M. Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. *PLoS ONE* **2010**, *6*, e15923. [CrossRef] [PubMed]

23. Zhou, X.; Liu, J.; Yeh, A.G.O.; Yue, Y.; Li, W. The uncertain geographic context problem in identifying activity centers using mobile phone positioning data and point of interest data. In *Advances in Spatial Data Handling and Analysis*; Springer: Berlin, Germany, 2015; pp. 107–119.

24. Taubenböck, H.; Klotz, M.; Wurm, M.; Schmieder, J.; Wagner, B.; Wooster, M.; Esch, T.; Dech, S. Delineation of Central Business Districts in Mega city regions using remotely sensed data. *Remote Sens. Environ.* **2013**, *136*, 386–401. [CrossRef]

25. Galster, G.; Hanson, R.; Ratcliffe, M.R.; Wolman, H.; Coleman, S.; Freihage, J. Wrestling sprawl to the ground: Defining and measuring an elusive concept. *Hous Policy Debate* **2001**, *12*, 681–717. [CrossRef]

26. Lee, B. Urban Spatial Structure, Commuting, and Growth in US Metropolitan Areas. Ph.D. Thesis, University of Southern California, Los Angeles, CA, USA, 2006.

27. Pereira, R.H.M.; Nadalin, V.; Monasterio, L.; Albuquerque, P.H. Urban centrality: A simple index. *Geogr. Anal.* **2013**, *45*, 77–89. [CrossRef]

28. Berrill, K.; Florence, P.S.; Baldamus, W. Investment, Location, and Size of Plant. *Econ. Stat.* **1948**, *58*, 411–414.

29. Midelfart-Knarvik, K.H.; Venables, A.J. Integration and industrial specialization in the European Union. *Rev. Econ.* **2002**, *53*, 469–482.

30. Tsai, Y.-H. Quantifying urban form: Compactness versus sprawl. *Urban Stud.* **2005**, *42*, 141–161. [CrossRef]

31. Thurstain-Goodwin, M.; Unwin, D. Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations. *Trans. GIS* **2000**, *4*, 305–317. [CrossRef]

32. Borruso, G. Network density and the delimitation of urban areas. *Trans. GIS* **2003**, *7*, 177–191. [CrossRef]

33. Battaglia, F.; Borruso, G.; Porceddu, A. Real Estate values, urban centrality, economic activities. A GIS Analysis on the City of Swindon (UK). In Proceedings of the 10th Computational Science and Its Applications–ICCSA, Fukuoka, Japan, 23–26 March 2010.

34. Hollenstein, L.; Purves, R. Exploring place through user-generated content: Using Flickr to describe city cores. *J. Spat. Inf. Sci.* **2010**, *1*, 21–48.

35. Borruso, G.; Porceddu, A. A Tale of Two Cities: Density Analysis of CBD on Two Midsize Urban Areas in Northeastern Italy. In *Geocomputation and Urban Planning*; Springer: Berlin, Germany, 2009; pp. 37–56.

36. Battino, S.; Borruso, G.; Donato, C. Analyzing the central business district: The case of sassari in the sardinia island. In Proceedings of the 12th Computational Science and Its Applications–ICCSA, Salvador de Bahia, Brazil, 18–21 June 2012.

37. Hu, Y.; Gao, S.; Janowicz, K.; Yu, B.; Li, W.; Prasad, S. Extracting and understanding urban areas of interest using geotagged photos. *Comput. Environ. Urban Syst.* **2015**, *54*, 240–254. [CrossRef]

38. Sun, Y.; Fan, H.; Li, M.; Zipf, A. Identifying the city center using human travel flows generated from location-based social networking data. *Environ. Plann. B* **2016**, *43*, 480–498. [CrossRef]

39. Montello, D.R.; Goodchild, M.F.; Gottsegen, J.; Fohl, P. Where's downtown: Behavioral methods for determining referents of vague spatial queries. *Spat. Cognit. Comput.* **2003**, *3*, 185–204. [CrossRef]

40. Henderson, M.; Yeh, E.T.; Gong, P.; Elvidge, C.; Baugh, K. Validation of urban boundaries derived from global night-time satellite imagery. *Int. J. Remote. Sens.* **2003**, *24*, 595–609. [CrossRef]

41. Estivill-Castro, V.; Lee, I. Argument free clustering for large spatial point-data sets via boundary extraction from Delaunay Diagram. *Comput. Environ. Urban Syst.* **2002**, *26*, 315–334. [CrossRef]

42. Liu, Q.; Deng, M.; Shi, Y.; Wang, J. A density-based spatial clustering algorithm considering both spatial proximity and attribute similarity. *Comput. Geosci.* **2012**, *46*, 296–309. [CrossRef]

43. Liu, D.; Sourina, O. Free-parameters clustering of spatial data with non-uniform density. In Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems, Singapore, 1–3 December 2004.

44. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm fordiscovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA; 1996; pp. 226–231.

45. Shannon, C.E. A mathematical theory of communication. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **2001**, *5*, 3–55. [CrossRef]

46. Cormen, T.H.; Leiserson, C.E.; Rivest, R.L. *Introduction to Algorithms*; MIT Press, Inc.: Cambridge, MA, USA, 1990.

47. Pakhira, M.K.; Bandyopadhyay, S.; Maulik, U. Validity index for crisp and fuzzy clusters. *Pattern Recogn.* **2004**, *37*, 487–501. [CrossRef]

48. Gordon, P.; Richardson, H.W.; Wong, H.L. The distribution of population and employment in a polycentric city: The case of Los Angeles. *Environ. Plann. A* **1986**, *18*, 161–173. [CrossRef]

49. Yan, X.; Zhou, C.; Leng, Y.; Chen, H. Functional features and spatial structure of CBDs in Guangzhou. *Acta Geogr. Sin.* **2000**, *55*, 475–486.

50. Wu, M.W. *Urban Central District Planning*; Southeast University Press: Nanjing, China, 2009. (In Chinese)

51. Ai, T.; Van Oosterom, P. GAP-tree extensions based on skeletons. In *Advances in Spatial Data Handling*; Springer: Berlin, Germany, 2002; pp. 501–513.

52. Kline, P. *An Easy Guide to Factor Analysis*; Routledge: London, UK, 2014.

53. Chainey, S.; Reid, S.; Stuart, N. When is a hotspot a hotspot? A procedure forcreating statistically robust hotspot maps of crime. In *Socio-Economic Applications of Geographic InformationScience*; Kidner, D., Higgs, G., White, S., Eds.; Taylor and Francis: London, UK, 2002; pp. 21–36.

54. Peethambaran, J.; Muthuganapathy, R. A non-parametric approach to shape reconstruction from planar point sets through Delaunay filtering. *Comput. Aided Des.* **2015**, *62*, 164–175. [CrossRef]