Journal of Geological Resource and Engineering 4 (2016) 151-159
doi:10.17265/2328-2193/2016.04.001

Constrained Delaunay Triangulation for Grouping Functional Areas by Land Use

Jirong Gu[1], Xianwei Cheng[2] and Zhi Dou[1]

1. Key Laboratory of Land Resources Evaluation and Monitoring in Southwest (Ministry of Education), Sichuan Normal University, Chengdu 610068, China

2. The Bureau of Land and Resources Chengdu, Chengdu 610072, China

Abstract: Functional land use maps are used for land evaluation, environmental analysis, and resource conservation. Spatial data clustering identifies the sparse and crowded places, thus discovering the overall distribution pattern of the dataset. Some clustering methods represent an attribute-oriented approach to knowledge discovery. Other methods rely on natural notions of similarities (e.g., Euclidean distances). These are not appropriate for constructing functional areas. We propose a similarity value to evaluate the closeness between a pair of points based on the total functional area and the proportion of the main land use type for the entire functional area. We develop constrained attributes employing this similarity value and a DT (Delaunay triangulation) criterion function when merging clusters. Four thresholds are set to ensure that functional areas have acceptable proportions, regular shapes, and no overlap. An experimental study was conducted with cadastral data for Chengdu, China, from 2009. The results show the advantages for objectivity and efficiency in using the proposed algorithm to define functional areas. The areas are created dynamically at any convenient time.

Key words: Constrained Delaunay triangulation, functional areas, land use, similarity, constrained clustering.

1. Introduction

The objective of urban planning is to integrate the four basic societal functions (living, working, recreation, and movement) and to make plans that provide for their interrelationship and growth [1]. Planners have subdivided cities into sectors or components for centuries in much of the world. Thus, analytical processes for subdividing cities were pursued at the expense of an organic urban order.

Urban functional areas are defined as groupings of individuals on the basis of the function that each performs within the organization (such as accounting, marketing, or manufacturing) and groupings of activities or processes on the basis of their necessity in accomplishing one or more tasks. Here, our specific definition of urban functional areas allows us to carry out zoning on the basis of land use types such as residential, commercial, and industrial. This allows the creation of rational zones with mutual contact and a reasonable layout that can act as the basis for research on urban land use and expansion.

The division of functional areas is the process of dividing urban land into separate parts. The parts should not merely have the same attribute or the closest attributes but should have roughly the same appropriate size and should have an acceptable minimum ratio of land parcels with the same price, land use capabilities, intensity, and direction. That is, a residential functional area does not mean an area that has only residential parcels. The area may include parcels with different land uses, such as transportation, residences, industry, and commerce, but the proportion of residential use is much higher than that of any other use. The ratio of the main land use type must be greater than a set threshold, and the total area of the functional area must be within a certain range.

Cluster analysis divides data into groups that are

Corresponding author: Jirong Gu, professor, research field: geographic information system.

meaningful, useful, or both [2]. Cluster analysis has long played an important role in a wide variety of fields such as social sciences, biological sciences, and earth sciences. It is used for numerical taxonomy, typological analysis, and partitions [3].

The traditional DT (Delaunay triangulation) uses distances between points for clustering but is not appropriate for constructing functional areas. We develop a constrained DT employing a similarity value and a DT criterion function when merging clusters. The similarity value evaluates the closeness between a pair of points based on the total functional area and the proportion of the main land use type for the entire functional area. The rest of this paper is organized as follows: Section 2 summarizes the relevant research. Section 3 discusses our constrained DT algorithm, including the definition of similarity and constrained attributes. Section 4 describes the steps in the algorithm. Section 5 describes the experiments and shows the results. Section 6 presents the conclusion.

2. Summary of Relevant Research

Clustering algorithms are to search for hidden patterns that may exist within a database. Spatial clustering algorithms in particular are for the discovery of interesting relationships and characteristics that may exist implicitly within spatial databases. A rough but widely agreed framework is to classify clustering techniques as hierarchical clustering or partitional clustering based on the properties of the clusters generated. Specifically, the DT algorithm is a clustering technique based on graph theory and is an important graphical representation for hierarchical clustering analysis.

Generally, the data to be clustered involve two types of attributes: metric and nonmetric. Some studies propose an attribute-oriented approach to knowledge discovery. Most of these studies are concerned with knowledge discovery for nonspatial data [4-13]. The algorithms most relevant to our focus here are CLARANS and BIRCH, which are used in studies of spatial data. More specifically, CLARANS includes a spatial data-dominant knowledge-extraction algorithm and a nonspatial data-dominant algorithm, both of which aim to extract high-level relationships between spatial and nonspatial data.

The nonspatial attributes used in CLARANS influence the clustering. As inputs the method takes relational data, generalization hierarchies for attributes, and a learning query specifying the focus of the mining task to be carried out. Then, based on the generalization hierarchies of the attributes, it iteratively generalizes the tuples. For example, suppose that the tuples relevant to a certain learning query have attributes (major, ethnic group). Further, assume that the generalization hierarchy for ethnic groups has Indian and Chinese generalized to Asians. Then, a generalization operation on the ethnic group attribute causes all tuples of the form (major, Indian) and (major, Chinese) to be merged into the tuple (major, Asians). This merging has the effect of reducing the number of (generalized) tuples. Finally, DT is applied to the spatial attributes and the clusters are found [14-16]. CLARANS relies on geostatistical methods. It fails to discover geometric information like the shapes of clusters. However, the information that spatial data mining discovers should include not only statistical regrouping but also geometric characteristics [10].

In land use databases, the nonspatial attributes used are land use type, land price, and block. These attributes have no meaning for the data clustering. The data space is usually not uniformly occupied. Data clustering identifies the sparse and crowded places, thus discovering the overall distribution pattern of the dataset.

For nonspatial data mining, we extract a set of relevant tuples by SQL queries. Then, based on the generalization hierarchies of the attributes, we iteratively generalize the tuples. For spatial data mining, our approach here is to apply cluster analysis only to the spatial attributes, for which there exist natural notions of similarities (e.g., Euclidean distances). We

eliminate all edges whose distance is greater than a threshold. Then, we identify the remaining connected components that satisfy the attribute constraints, and those are functional areas.

3. Constrained Delaunay Triangulation

Because of the vast amounts of spatial data that are obtained from a vector database, it is costly and often unrealistic for users to examine spatial data in detail. We need to extract interesting spatial features and capture intrinsic relationships between spatial and nonspatial data.

In conventional DT, the nodes V of a weighted graph G correspond to data points in the pattern space, and the edges reflect the proximity between each pair of data points. The dissimilarity matrix is defined as

$$D_{ij} = \begin{cases} 1 & if\ D(x_i, x_j) < d_0 \\ 0 & otherwise \end{cases} \quad (1)$$

where $D(x_i, x_j)$ is the Euclidean distance between x_i and x_j while d_0 is a user-defined threshold value. This matrix is the most commonly used criterion function for Euclidean space. Thus, intuitively, the criterion function attempts to minimize the distance between each point and the mean position of the cluster to which the point belongs [3]. To preserve the directions and shapes of the polygons, some researchers use the adjacency distance as the distance metric for the single-link clustering algorithm [17].

In this paper, we still regard the Euclidean distance as the clustering criterion. We define an algorithm based on DT. Fig. 1 shows an example on which to apply the method. The region in the figure has three residential parcels (R1, R3 and R3), two commercial parcels, two industrial parcels, and two road parcels. We have to decide whether R1, R2 and R3 are in a cluster. Some questions must be considered:

- Is residential the main land use type?
- Are the parcels R1-R3 in the same district?
- Do the parcels R1-R3 have the same land price?
- Is the size of the functional area within the set thresholds?
- Does the proportion of residential parcels exceed the set threshold?

Conventional DT: If we consider none of the above questions, then all of the parcels will belong to a cluster when the distance is less than d_0.

Constrained DT: We use GIS (geographic information system) methods to extract the centroids of a set $\{V\}$, which belongs to the cadastral parcels. We extract a set of relevant tuples by SQL queries. All centroids in tuples have the same land use type, are in the same district, and have the same land price. If R1, R2 and R3 satisfy these conditions and are in a DT, they are in a cluster and may be a functional area. We add up the areas of R1, R2 and R3. If the area is between the thresholds, we find the next parcel R4 in the sub-centroids. If the area is greater than the upper threshold, we stop and then proceed to find the next

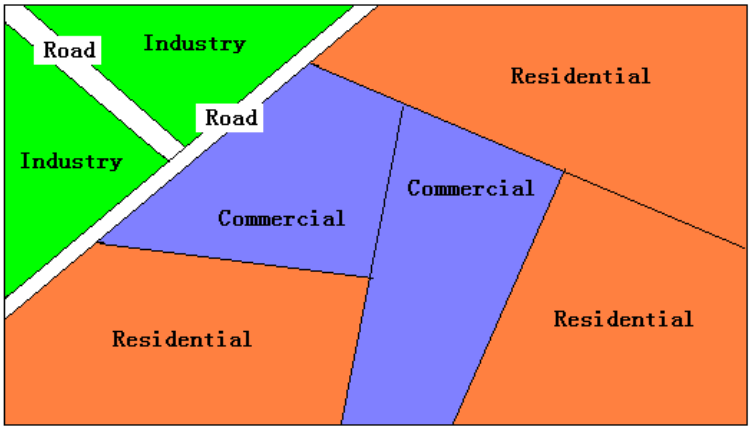


Fig. 1 Example of regional land use.

parcel. We calculate the ratio of residential parcels. If the ratio exceeds the threshold, the parcels belong to a residential cluster and this is a functional area. We use other points to repeat these steps.

In this paper, we present a concept of clustering that is based on similarity between data points rather than on distances alone. Let a pair of points be neighbors if their similarity exceeds a certain threshold. The similarity values for pairs of points must be based on Euclidean distances and on nonmetric similarity functions obtained from the centroid attribute table. Points belonging to a single cluster will be neighbors.

3.1 Constrained Attributes

To simplify the procedure, we filter the centroid points $\{V\}$ with the following attributes:

- Land use type (to filter out the centroid points with the same land use type).
- District (to filter out the centroid points with the same land use type in the same district).
- Block (to filter out the centroid points with the same land use type in the same district and in the same block).

If we filter the centroid points using only the first two constraints, the output DTs may overlap. By adding the block constraint, the shapes of the output DTs are regulated with less overlap (Fig. 3).

Chengdu has five districts, containing 96 blocks in total. We mainly consider three land use types: residential, commercial and industrial. Thus, we obtain 15 × 96 categories for the centroid point sets $\{V_{ijk}\}$, where i denotes the land use type, j denotes the district, and k denotes the block.

3.2 Similarity

Let the points p_i represent the centroids of the parcels of land. Each clustering constructs a functional area by merging the parcels (Fig. 2). Let SIM (p_i, p_j) be a similarity function that is normalized and captures the closeness between the pair of points p_i and p_j. The function SIM can be one of the well-known distance metrics or it can even be nonmetric. We assume that SIM is a Boolean value (0 or 1), where the value 1 indicates that the points are similar.

We use two characteristics in defining the similarity. SUM Area represents the total area of the functional area. RATIO represents the proportion of the main land use type for the entire functional area. According to the "Evaluation Rules of the Ministry of Land and Resources" (2009), the total area of the functional area is between 40,000 m^2 and 300,000 m^2, and the proportion of the main land use type is over 40%.

Two user-defined thresholds, α and β, are set between 40,000 m^2 and 300,000 m^2 based on the empirical values. A pair of points p_i and p_j is defined to be neighbors if the following holds:

$$\alpha \le \text{SUM Area} \le \beta. \quad (2)$$

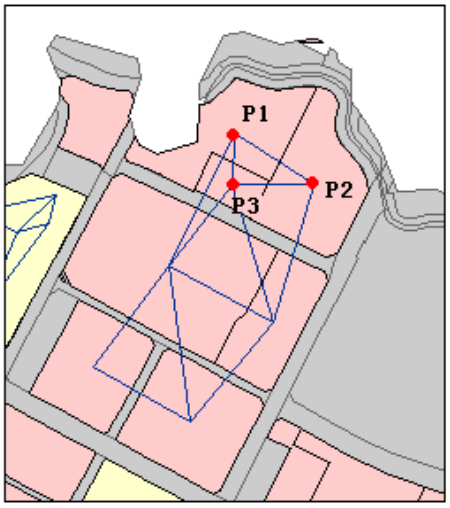


Fig. 2 Example of points p_i and parcels.

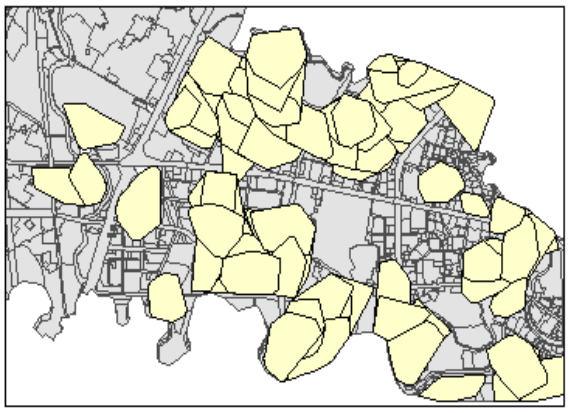


Fig. 3 Overlapping functional areas.

A user-defined threshold θ is set between 0 and 100. A pair of points p_i and p_j is defined as neighbors if the following holds:

$$\text{RATIO} \geq \theta \quad (3)$$

With the above, we define the similarity function as follows:

$$SIM(pi, pj) = \begin{cases} 0 & \text{SUMArea} \geq \beta \quad or \quad \text{SUMArea} \leq \alpha \; or \quad RAITO \leq \theta \\ 1 & \alpha \leq \text{SUMArea} \leq \beta \; and \quad RAITO \geq \theta \end{cases} \quad (4)$$

SIM equals unity when p_i belongs to the cluster and the parcel associated with p_i belongs to the functional area.

3.3 Forming the DT

From $\{V_{ijk}\}$, the points p_1 and p_2 that are nearest to each other are used to initiate clustering. We find p_{mid}, which is the midpoint between p_1 and p_2. Next, p_3 is found on the basis of the shortest distance to p_{mid}. If p_3 satisfies the DT criterion defined in Eq. (1), it will be added to the cluster that includes p_1 and p_2.

When p_1, p_2 and p_3 are in the cluster, we assess the SUM Area and RATIO using Eqs. (2) and (3). If the SIM from Eq. (4) equals unity, we record the points as neighbors. We find the other points by using the same steps, until SIM equals zero or no more points are added.

4. The Constrained DT Algorithm Proposed

The algorithm generates centroids by using the constraints to divide the centroid point set $\{V\}$ into a number of partitions such that the straddled partitions are minimized. The partitioning algorithm minimizes the relationship among the data points across the resulting partitions and minimizes the edge-cut effectively. The inputs to the clustering algorithm are the set $\{V_{ijk}\}$ from the partitioning algorithm and the thresholds d_0, α, β and θ (Fig. 4).

The procedure begins by computing the distances between pairs of points in $\{V_{ijk}\}$. Initially, each point is in a separate cluster. We build a local heap $q[i]$ that contains every distance between two points in $\{V_{ijk}\}$, and we maintain the heap during the execution of the algorithm. The entries of $q[i]$ are in order of increasing distance.

We find the shortest distance in $q[i]$ to obtain the points p_1 and p_2. We then calculate the midpoint between p_1 and p_2, saving that as p_{mid}. Next, p_3 is found on the basis of the shortest distance to p_{mid}. This distance must be less than d_0. The total area enclosed

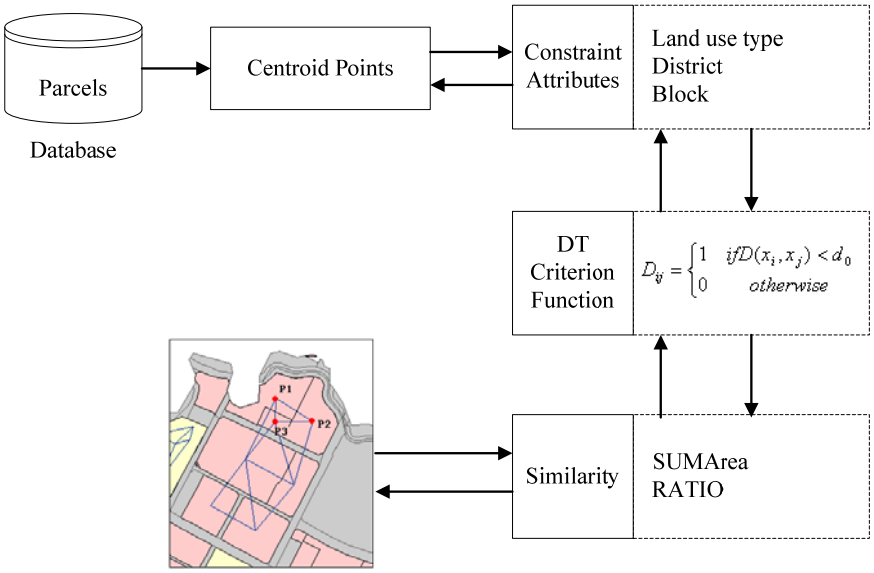


Fig. 4 The refined DT algorithm.

by p_1, p_2 and p_3 must be less than β. If the points satisfy the thresholds, we add the points to the cluster $\{c_i\}$ and delete them from $\{V_{ijk}\}$ and $q[i]$.

The while-loop iterates until the following two conditions are met:

- No point remains in $\{V_{ijk}\}$. The points in $\{c_i\}$ must satisfy the thresholds for SUM Area and RATIO to form a cluster that can be a functional area. If any of the thresholds is not satisfied, the cluster cannot be formed.
- SIM equals zero. When $\{c_i\}$ satisfies the thresholds for SUM Area and RATIO, it will be defined as a functional area. The points remaining in $\{V_{ijk}\}$ will be subjected to the steps again, until no point is left.

5. Experimental Results

Chengdu is the main city in the southwest of China. The information center of the Chengdu Land Management Bureau has been building a cadastral database since 2007. It covers five districts, 96 blocks, and 34,291 parcels of land. The parcel data contain polygons that have been assigned land use types and areas. The boundary data contain the block polygons that are assigned the district name and block name. The land price data contain polygons that are assigned the land price. These are vector polygon data and are stored in the cadastral database. We overlaid the three layers and assigned all the attributes to the parcels. Then, we extracted the centroids and stored those in a set {V}. Each point is associated with a land use type, district name, block name, and land price. There are 52 land use types, including residential, industrial, commercial, and transportation.

Several experiments were performed to validate the algorithm. The main purpose was to determine the effect of changes in constraints on the correctness of the results and the runtime needed to create the triangles. All the experiments presented in this paper were performed on land use data for 2009 from the Bureau of Land Resources in Chengdu, China. The bureau also invited 50 experts to define the functional areas. This paper compares the results of the experts with those of the proposed algorithm.

5.1 Experiment 1

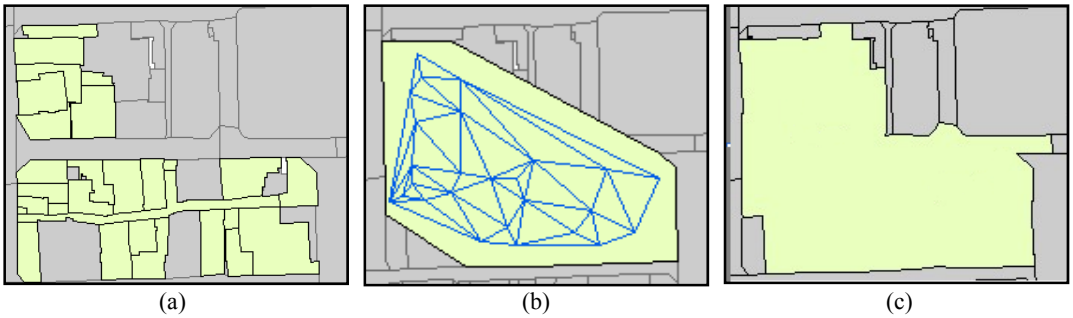
The purpose of the first experiment was to estimate the correctness of the functional areas created by the algorithm and by the experts. Table 1 lists the characteristics used.

The three types of differences between the two sets of results are as follows:

(1) A functional area is created by the algorithm but not by the experts (Fig. 5). The experts judged that the area cannot be a functional area because it is too small or not important. In the algorithm, all functional areas that satisfy the constraints and the thresholds will be created.

Table 1 Characteristics used.

Characteristic	Value
Land use type	Residential
District	Qing Yang
Block	The same block
d_0 (m)	< 400
RATIO	0.5
SUM Area (m^2)	40,000-300,000



(a) (b) (c)

Fig. 5 (a) Residential land use, (b) the refined DT, and (c) the functional area.

(2) A functional area created by the algorithm partly overlaps the functional area defined by the experts. In this case, the area defined by the experts is smaller and more regular (Fig. 6).

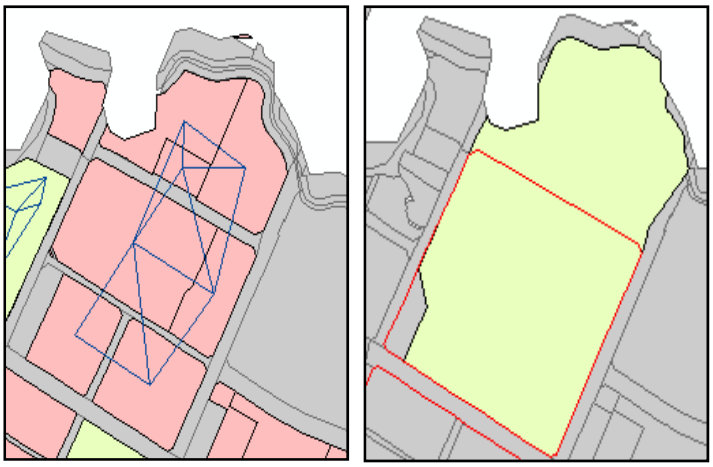
(3) A functional area created by the algorithm almost completely overlaps that created by the experts (Fig. 7). As shown by the results in Table 2, all the functional areas created by the experts are covered by the functional areas created by the algorithm.

5.2 Experiment 2

The purpose of the second experiment was to determine how the number of functional areas and the runtime of the algorithm change with the constraints.

To test the speed of the algorithm, we fixed RATIO (0.5) and SUM Area (40,000-300,000 m^2) but varied d_0 (Table 3). We then compared the runtime under different distance constraints. The changes did not affect the efficiency of the algorithm.

Then, we fixed SUM Area (40,000-300,000 m^2) but varied RATIO and d_0. The changes are shown in Table 4. These two variables have a significant impact on the generation of functional areas. With reductions in d_0 and RATIO, more functional areas are generated.



(a) (b)

Fig. 6 (a) Residential land use (red area) and refined DT (blue line). (b) Functional area (green area created by the algorithm; red line created by the experts).

(a) (b)

Fig. 7 (a) Residential land use (red area), and (b) the functional area (green area created by the algorithm; red line created by the experts).

Table 2 Number of functional areas identified by the algorithm and by experts.

Type	Number of functional areas
Total by the algorithm	86
Total by the experts	26
Almost overlap	8
Partly overlap	18

Table 3 Comparison of runtime under different constraints.

Constraints	I	II	III
Shortest-distance constraint (m)	1-200	1-400	100-400
Runtime (s)	2	3	2

Table 4 Comparison of the number of functional areas created under different constraints.

Constraints	1	2	3
RATIO	0.5	0.5	0.4
Shortest-distance constraint (m)	1-400	1-350	1-350
Number of functional areas	52	63	86

Fig. 8 The points define either a line or segments with very slight angles between them.

6. Conclusions

In this paper, we proposed a new concept to define the similarity between a pair of data points. We then described a constrained Delaunay triangulation algorithm that employs this similarity and distances for merging clusters. Our methods extend DT to group the functional areas.

The algorithm still has shortcomings. When the points define either a line or segments with very slight angles between them, the algorithm cannot create the DT (Fig. 8). We removed such points from the set $\{V\}$ as outliers. This technique may influence the results of the clustering.

The results of our experimental study are very encouraging. The functional areas identified by our algorithm overlap completely or partly with those identified by experts. The runtime of the algorithm is not greatly influenced by the constraints. The advantages of using a computer for triangulation are objectivity and increased speed. The functional areas can be created dynamically at any convenient time. Thus, when the land use parcels change, the areas can be revised.

Acknowledgments

They are also thankful to the support from the Ministry of Land and Resources of People's Republic of China (2006S01).

References

[1] Universidad Nacional Federico Villarreal. 1978. *The Charter of Machu Picchu*. Washington, DC: American Institute of Architects.

[2] Tan, P., Steinbach, M., and Kumar, V. 2006. "Cluster Analysis: Basic Concepts and Algorithms." In *Introduction to Data Mining* (Chapter 8). Accessed June 6, 2009. http://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap8_basic_cluster_analysis.pdf.

[3] Xu, R. 2005. "Survey of Clustering Algorithms." *IEEE Transactions on Neural Networks* 16 (3): 645-78.

[4] Han, J., Cai, Y., and Cercone, N. 1992. "Knowledge Discovery in Databases: An Attribute-Oriented Approach." In *Proceedings of the 18th International Conference on Very Large Data Bases (VLDB '92)*, 547-59.

[5] Han, J., Kamber, M., and Tung, A. 2001. "Spatial Clustering Methods in Data Mining: A Review." In *Geographic Data Mining and Knowledge Discovery*, edited by Miller, H. J., and Han, J. London, UK: Taylor and Francis, 188-217.

[6] Hastie, T., Tibshirani, R., and Friedman, J. H. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer.

[7] Jain, A. K., and Dubes, R. C. 1988. *Algorithms for Clustering Data*. London, UK: Prentice Hall.

[8] Jain, A. K., Murty, M. N., and Flynn, P. J. 1999. "Data Clustering: A Review." *ACM Computing Surveys* 31 (3): 264-323.

[9] Jardine, N., and Sibson, R. 1971. *Mathematical Taxonomy*. New York, NY: Wiley.

[10] Kang, I., Kim, T., and Li, K. 1997. "A Spatial Data Mining Method by Delaunay Triangulation." In *Proceedings of the 5th ACM International Workshop on Advances in Geographic Information Systems (GIS'97)*, 35-9.

[11] Karypis, G., Han, E. H., and Kumar, V. 1999. *Multilevel Refinement for Hierarchical Clustering.* Technical report for University of Minnesota, Minneapolis, MN.

[12] Leach, G. 1992. "Improving Worst-Case Optimal Delaunay Triangulation Algorithms." Presented at the Fourth Canadian Conference on Computational Geometry, Newfoundland, Canada.

[13] Lu, W., Han, J., and Ooi, B. C. 1993. "Discovery of General Knowledge in Large Spatial Databases." In *Proceedings of the Far East Workshop on Geographic Information Systems*, 275-89.

[14] MacQueen, J. 1967. "Some Methods for Classification and Analysis of Multivariate Observations." In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 281-97.

[15] Mitchell, T. 1997. *Machine Learning*. Boston, MA: McGraw-Hill.

[16] Ng, R. T., and Han J. 1994. "Efficient and Effective Clustering Methods for Spatial Data Mining." In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB'94)*, 144-55.

[17] Yang, L., Zhang, L., Ma, J., Xie, J., and Liu, L. 2011. "Interactive Visualization of Multi-resolution Urban Building Models Considering Spatial Cognition." *International Journal of Geographical Information Science* 25 (1): 5-24.