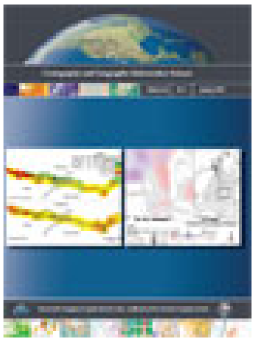

Cartography and Geographic Information Science

ISSN: 1523-0406 (Print) 1545-0465 (Online) Journal homepage: http://www.tandfonline.com/loi/tcag20

On the spatial distribution of buildings for map generalization

Zhiwei Wei, Qingsheng Guo, Lin Wang & Fen Yan

To cite this article: Zhiwei Wei, Qingsheng Guo, Lin Wang & Fen Yan (2018): On the spatial distribution of buildings for map generalization, Cartography and Geographic Information Science, DOI: 10.1080/15230406.2018.1433068

To link to this article: https://doi.org/10.1080/15230406.2018.1433068

Published online: 15 Feb 2018.

CARTOGRAPHY AND GEOGRAPHIC INFORMATION SCIENCE, 2018
https://doi.org/10.1080/15230406.2018.1433068

On the spatial distribution of buildings for map generalization

Zhiwei Wei [a], Qingsheng Guo [a,b], Lin Wang [a] and Fen Yan [a]

[a]School of Resource and Environment Science, Wuhan University, China; [b]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China

ABSTRACT

Information on spatial distribution of buildings must be explored as part of the process of map generalization. A new approach is proposed in this article, which combines building classification and clustering to enable the detection of class differences within a pattern, as well as patterns within a class. To do this, an analysis of existing parameters describing building characteristics is performed via principal component analysis (PCA), and four major parameters (i.e. convex hull area, IPQ compactness, number of edges, and smallest minimum bounding rectangle orientation) are selected for further classification based on similarities between building characteristics. A building clustering method based on minimum spanning tree (MST) considering rivers and roads is then applied. Theory and experiments show that use of a relative neighbor graph (RNG) is more effective in detecting linear building patterns than either a nearest neighbor graph (NNG), an MST, or a Gabriel graph (GG). Building classification and clustering are therefore conducted separately using experimental data extracted from OpenStreetMap (OSM), and linear patterns are then recognized within resultant clusters. Experimental results show that the approach proposed in this article is both reasonable and efficient for mining information on the spatial distribution of buildings for map generalization.

ARTICLE HISTORY
Received 20 May 2017
Accepted 23 January 2018

KEYWORDS
Map generalization; building clusters; classification; spatial clustering; spatial pattern

1. Introduction

Cartographic generalization is an abstraction process that aims to find an appropriate solution for information quantity given specific objectives, in particular scale and theme (Regnauld, 2001). Thus, in order to reduce map scale, it is necessary to convert data into a legible form while the spatial distribution of geographical elements, or phenomena, must also be well-expressed (He & Song, 2015). In this context, specific clusters into which buildings are gathered together perceptually, or are regularly arranged, are important. These clusters can be used to characterize built-up areas, encapsulate the historical, social, cultural, and functional implications of urban environments (Carmona, Heath, Oc, & Tiesdell, 2003), and must be preserved when developing building map generalizations. Although topographic data is geometrically rich, explicit relationships, structures, patterns, and the higher semantics of spatial objects are often absent (Zhang, Ai, & Stoter, 2010). It is therefore important to develop new techniques for detecting specific building clusters when developing map generalizations for buildings.

When viewing a map, observers are able to easily and conceptually group buildings together based on perceptual criteria mainly derived from Gestalt theory. This theory notes that if buildings are continuous, similar, close to one other, and orientated in the same direction, then they are more likely to form a group (Regnauld, 2001). Building clusters can therefore be detected based on their proximity, or similarities, to one another; some regularly arranged buildings will also form patterns, including linear relationships, which can be easily detected.

Although the proximity relationships between buildings have often been described using proximity graphs (Anders & Sester, 2001), it remains unclear which kind of method is most suitable for detecting specific patterns. Although the similarities between buildings can be measured on the basis of their shared characteristics (University of Zurich, 1999), and different parameters can be selected, or created, to describe these characteristics. It still remains unclear just how many are needed for accurate description and which are most representative. Although research to date has mostly focused on the similarities between two adjacent buildings, treating them as local structures, the map generalization approach aims to maintain a global spatial distribution. It is therefore important to express the similarities between all buildings within a large area.

CONTACT Qingsheng Guo guoqingsheng@whu.edu.cn

Motivated by these issues, we proposed a new approach to explore the spatial distribution of buildings based on their similarities as well as the proximity between entities using classification, clustering, and linear pattern recognition. To do this, we first selected a representative parameter set to describe building characteristics within a current system and verified this choice using PCA. Second, we separately applied our selected parameters and MST to divide buildings into clusters based on their similarities and proximity. Third, we applied RNG to detect linear building patterns within the groups obtained subsequent to clustering and classification, as this graph method has been shown to be the most effective of the four (i.e. NNG, MST, RNG, and GG). Application of this workflow enabled us to generate and explore detailed information about the spatial distribution of buildings across a large area. This kind of information is important for the subsequent development of automatic map generalization methods.

2. Related work

A number of approaches to explore the spatial distribution of buildings have been presented that mainly involve clustering and pattern recognition.

Two clustering strategies, divisive and agglomerative, are commonly applied in this context. The first of these initially considers all elements as a group, before dividing them into constituent parts step-by-step. In contrast, the agglomerative strategy starts with a single element, considers each entity as a distinct cluster, and merges components into higher-level groups step-by-step. In both two strategies, the distance between elements must be defined a priori; in the case of buildings, this distance is usually defined based on proximity. However, as buildings cannot simply be represented as points, clusters detected based on proximity may not be representative of actual perceptual building agglomerations and so similarity criteria generated from Gestalt theory are also considered (Fei, 2002).

The divisive strategy is mainly based on the application of graph-partitions. Thus, if just proximity is taken into account, Zhang, Ai, Stoter, Kraak, and Molenaar (2013) proposed a MST-pruning method that initially structures buildings. In this approach, edges that are defined as inconsistent can be deleted to obtain building clusters; an inconsistent edge has a length that is significantly longer than its close counterparts. However, the MST approach is only defined for the description of one level of proximity relationship and may be vulnerable for groups of buildings with different shapes. A number of other kinds of proximity graph-based approaches can also be applied, including the adaptive spatial clustering method that is based on Delaunay triangulation (ASCDT) (Cetinkaya, Basaraner, & Burghardt, 2015; Deng, Liu, Cheng, & Shi, 2011), and the graph-theoretical clustering method that utilizes two rounds of MSTs (Zhong, Miao, & Wang, 2010). In cases where proximity is described by density, a density-based spatial clustering application with noise (DBSCAN) method can be applied (Cetinkaya, Basaraner, & Burghardt, 2015). While in situations where similarity between buildings is also taken into account, Regnauld (2001) defined similarity between entities incorporating size and orientation. Similarly, Ai and Guo (2007) incorporated the concept of "visual distance" to encapsulate differences in proximity, size, and orientation, while Qi and Li (2008) also considered similarity in building shapes. It is noteworthy that MSTs are initially applied to structure buildings in all these methods; to determine whether, or not, clustering results conform to perceptual rules, additional cognitive experiments are required. And the concepts of similarity and proximity between buildings are intuitive to humans, which also remains a fuzzy question. In one attempt to address this issue, Wang, Du, Guo, and Luo (2015) proposed a multilevel graph-partition clustering method based on graph coarsening. Four similarities are defined in this approach, which initially incorporate distance, connectivity, size, and shape. The resultant polygonal graphs in which polygons are considered as nodes are then coarsened at multiple levels based on these four similarities to find maximally similar clusters among shapes in the same groups. Graph coarsening is then performed by collapsing a group of nodes into a coarse node (defined in practice) and then reconstructing their edges. Groups of buildings can then be identified based on the retained minimal similarity between different clusters.

Regarding the agglomerative strategy of building clustering, Karypis, Han, and Kumar (1999) proposed use of the CHAMELEON method when proximity alone is taken into account. This approach utilizes the number of links and their strength within a k-NNG in order to develop a dynamic model that determines similarity between two clusters. Building on this, Anders and Sester (2001) proposed a parameter-free method to detect building clusters based on four proximity graphs which takes "density," "distance," and "variance compatibility" into account. However, proximity is just one important factor that can affect clustering results, and building similarities also need to be taken into consideration. This agglomerative clustering method does not necessarily start with a single building, however, as groups can initially be produced; indeed, depending on process, the definitions of similarity between buildings can be adjusted. In one approach proposed by Yan, Weibel, and Yang (2008), two-building clusters are initially detected based on minimum distance and the area of visible scope between adjacent entities. Distance between groups of two-buildings can then be qualitatively defined as "strong," "average," or "weak," before they are merged into higher-level clusters. A similar triplet approach was also proposed by Boffet and Serra (2001). However, as buildings are just one class of features on a map, clustering must also consider the relationships between these entities and other classes; this was done by Li, Yan, Ai, and Chen (2004) who used roads as initial global constraints to partition building sets into parts before grouping them based on proximity and similarity.

Developments in computer science have led to the introduction of algorithms that originate from artificial intelligence. In this context, Allouche and Moulin (2005) pioneered the use of Kohonen-type neural networks for amalgamation, a variant of the k-means method that was initially designed mainly for use in dense areas. Support vector machine (SVM)-based methods were first applied by Zhang, Deng, Chen, and Wang (2013) for automated clustering of urban buildings. This method applies three underlying principles of Gestalt theory, proximity, similarity, and orientation, defined on the basis of seven indicators. Building clusters are then manually labeled as either "positive" or "negative" in order to train the SVM model, which can then be applied to identify final clusters. However, to make a SVM model as accurate as possible, more effective indicators like size and texture should also be taken into account.

A range of comparative analyses have also been performed to identify the strengths and weaknesses of available clustering methods. Cetinkaya, Basaraner, and Burghardt (2015) compared four algorithms for urban building clustering and found that the DBSCAN and ASCDT methods were most efficient. Deng, Tang, Liu, and Wu (2017) carried out a comparative analysis of nine clustering methods used for recognizing building groups in map generalization and showed buffer analysis to be the most efficient approach when just proximity is considered. Local constraint-based approaches are the most effective when multiple grouping principles are incorporated.

Building patterns are also types of clusters in which buildings are regularly arranged. Definitions are required before patterns can be recognized, and previous workers have proposed some matching methods if the definitions are to be based on a template. Rainsford and Mackaness (2002) used this approach to detect linear patterns, for example, while Yang (2011) introduced a series of stair-, Z-, and H-pattern templates defined by structural parameters. As building patterns are varied, however, the use of template matching methods can be inflexible if common rules are not put in place.

A number of mathematical pattern definitions have been introduced to address this issue; the results of Christophe and Ruas (2002), for example, contained a number of false positives when a straight-line-based approach was applied to detect buildings along a road taking into account their proximity and similarity. While Anders (2006) used an RNG to detect grid-like building patterns for typification. The definition of a grid-like pattern applied in this research remained vague, however. Building on MST approach, Zhang, Ai, et al. (2013) recognized a series of more refined patterns, including collinear and curvilinear examples, initially detecting potentially Gestalt groups. Patterns were then recognized in these groups by applying seven algorithms based on perceptual rules and MST. However, this method may not be appropriate for detecting linear patterns in dense areas because MSTs usually comprise trees with short branches. Gong, Wu, Qian, and Ma (2014) proposed an approach for detecting multi-connected linear and grid-like patterns by applying a DT-like proximity graph. Edges are then deleted from this DT-like proximity graph using structural parameters to find multi-connected linear patterns. Grid-like ones are then recognized. However, each building is abstracted to its smallest minimum bounding rectangle (SMBR) in this approach without taking shape into account. Pattern recognition is a top-down process in all these methods, it also can be bottom-up; Du, Shu, and Feng (2015) developed a three-level relational approach based on spatial reasoning whereby patterns including collinear ones were detected based on 169 sets of relative relations, qualitative angle descriptions, and qualitative size descriptions defined between two buildings.

In order to obtain building clusters or patterns, it is also necessary to measure inherent characteristics, and this is done in many of the methods discussed above. For example, area can be used to describe the size of a building, while the number of edges can be used to describe shape. In collaborative research in Europe, the Multi-Agent system applied for generalization enabled a detailed analysis of the parameters used to describe characteristics of buildings. Duchêne et al. (2003), for example, performed an analysis to determine the absolute orientation of buildings, while Steiniger, Lange, Burghardt, and Weibel (2008) used PCA to study the relationships and interdependencies between parameters. Building classifications in remote sensing field, building characteristics can also be described based on shape index, and fractal dimension (Hermosilla, Ruiz, Recio, & Cambra-López, 2012; Lu, Im, Rhee, & Hodgson, 2014). The most recent analysis of this type was carried out by Basaraner and Cetinkaya (2017) who studied the efficiency of different parameters at characterizing the perceptual shape complexity of buildings.

3. Classifying buildings based on similarity

In order to classify buildings on maps based on similarities between their characteristics, principal components and the geographical information content of relevant features must first be determined before representative parameters can be selected. Thus, research has also been undertaken on the parameters used to describe the characteristics of building, emphasized here from the perspectives of size, shape, and orientation. Shape, for example, can be measured based either on boundary information or on boundary plus interior content, including shape signatures, histograms, invariants, moments, curvature, context, matrices, and spectral features (Peura & Iivarinen, 1997; Zhang & Lu, 2004). These parameters are computationally expensive, however, and lack information about geographical context. As the aim of this study is for map generalization, and buildings on maps also tend to have simple shapes, we summarized 24 parameters used in previous research (Tables 1–3).

The parameters utilized in a statistical clustering method for building classification must be as independent as possible from one other, some applied in this study (Tables 1–3) are actually interrelated; for example, ShapeIndex and IPQCom are reciprocal, which reflects correlation between the two. "Elongation" describes the degree to which a building is extended, "ellipticity" describes its flatness; a more extended building will also appear flatter. While a building with greater ellipticity will more closely resemble a circle, as described by DCircularity.

Table 1. Parameters used to describe building size.

Parameter	Definition	Source
Area	Building area	–
Perimeter	Building perimeter	–
SArea	Building SMBR area	–
CArea	Building convex hull area	–
MeanR	Based on a set of equally spaced radials from a given location within a shape to its perimeter, MeanR denotes the average distance value of n radials	Peura and Iivarinen (1997)
LChord	Length between the two most distant points in a building	–

Table 3. Parameters used to describe building orientation.

Parameter	Definition	Source
MWO	Orientation of the longest edge of a building	Ruas & Holzapfel (2003)
SWWO	Average orientation value of each edge weighted by length	Zhang, Ai, et al. (2013)
SMBRO	Orientation of building SMBR	Yan, Weibel, and Yang (2008)

We therefore used PCA to investigate the correlations between parameters as this approach transforms a set of variables that may be correlated into a new set of uncorrelated linear variables via orthogonal transformation. As shown in Figure 1, a PCA performed on area and perimeter of several buildings will generate a PC (size); if two parameters have the same PC coefficient, they can be used interchangeably to describe the size of buildings within the group. This approach can therefore be used to identify dependencies between parameters and to reduce the number of most important components. Representative parameters can then be selected according to the coefficient of a given parameter.

We used the Beijing OSM map as our study area, and chose 1,390 buildings within Haidian District for PCA

Table 2. Parameters used to describe building shape.

Parameter	Definition	Source
ShapeIndex	$ShapeIndex = Perimeter/2\sqrt{\pi Area}$	Burghardt and Steiniger (2005)
Concavity	$Concavity = Area/CArea$	Peura and Iivarinen (1997)
IPQ compactness	$IPQCom = 4\pi Area/Perimeter^2$	MacEachren (1985)
Ritter's compactness	$RCom = Perimiter/Area$	MacEachren (1985)
Richardson's compactness	$RicCom = 2\sqrt{\pi Area}/Perimeter$	MacEachren (1985)
Gibbs's compactness	$GibCom = 4Area/LChord^2$	MacEachren (1985)
DCM compactness	$DCMCom = Area/Asc$ Asc denotes the area of a minimum circumscribed circle around the building to be mapped	Li, Goodchild, and Church (2013)
Bottema compactness	$BotCom = 1 - Intersect(Area, Ao)/Area$ Intersect denotes the intersection area between polygons, while Ao refers to a circle with the same area and center	Li et al. (2013)
Boyce-Clark compactness	$BoyClaCom = 1 - \frac{\sum\left\lvert \frac{Ri}{\sum Ri}*100 - \frac{100}{n} \right\rvert}{200}$ In this expression, n equally spaced radials are drawn from the center of the building to be mapped within a shape to its perimeter, while R_i denotes the length of each radial	Li et al. (2013)
Elongation	Ratio between the length of the maximum chord, A, to the length of the maximum chord, B, perpendicular to A. Chord A and chord B are derived from the building SMBR	University of Zurich (1999)
Ballard & Brown eccentricity	$BBEccentricity = 4Area/(LChord \times Perimeter)$	University of Zurich (1999)
Davis circularity	$DCirularity = Perimeter^2/Area$	Basaraner and Cetinkaya (2017)
Ellipticity	$Ellipticity = w/l$ In this expression, w denotes the longest axis of a building to be mapped, while l is the width of the building perpendicular to its long axis	Burghardt and Steiniger (2005)
Fractal dimension	$Fd = 2\log(Perimter)/\log(Area)$	Basaraner and Cetinkaya (2017)
EdgeCount	Number of building edges	Yan, Weibel, and Yang (2008)

as this region comprises similar building types to the overall city dataset (Figure 2, Table 4). All parameters were standardized in advance to enable more accurate parameter comparisons using Equation (1), as follows:

$$X_{ij'} = \left(X_{ij} - \overline{X_j}\right)/\delta_j \tag{1}$$

where X_{ij} refers to the numerical value of parameter j for building i, $\overline{X_j} = \frac{1}{N}\sum_{i=1}^{N} X_{ij}$, $\delta_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_{ij} - \overline{X_j}\right)^2}$. After this conversion, PCA is carried out; PC contribution rates and the coefficient matrix are presented as Table 5 and Table 6.

Our experimental results (Table 5) show that the eigenvalues of the first four PCs are greater than 1, and their cumulative percentage reaches 88.791%; this suggests that the first four PCs contain almost the same total information as all 24 parameters. In other words, the first four PCs in this dataset can be used to effectively describe the characteristics of buildings. Similarly, the coefficient matrix (Table 6) derived from this analysis illustrates correlations between PCs and parameters, which enables the selection of representative parameters. Results show PC 1 is mainly determined by ShapeIndex, IPQCom, RicCom, GibCom, DCMCom, BotCom, Ellipticity, BBEccentricity, DCircularity, and Fd, it reflects shape, while PC 2 is mainly determined by Area, SArea, and CArea, it reflects size. The PC 3 reflects orientation as it is mainly determined by MWO, SWWO, and SMBRO, while PC 4 also reflects shape because it is mainly determined by EdgeCount and Fd.

These PCA results show that building characteristics need only be described using size, shape (two PCs), and orientation. Although building size is usually described by Area, the results of this analysis also suggest that SArea and CArea are actually more effective, we selected the latter. As for PC 2, IPQCom is selected for computational simplicity, as this is one of the most highly correlated parameters on PC 2 as well as a commonly applied measure for polygon compactness (Li et al., 2013). It is noteworthy, however, that both RicCom or Fd could also be used with equal validity in this context; Moser et al. (2002) noted that these parameters are the most effective for describing the characteristics of polygons with more complex shapes. Thus, they may need to be selected a priori when mapping buildings with complex shapes. PC 4 is determined by Concavity and EdgeCount. As buildings are artificial objects, they tend to be square. In other words, if a building has more edges, it is also likely to be more complex, and so we utilized EdgeCount even though Concavity could also be used. Finally, although any one of three equally

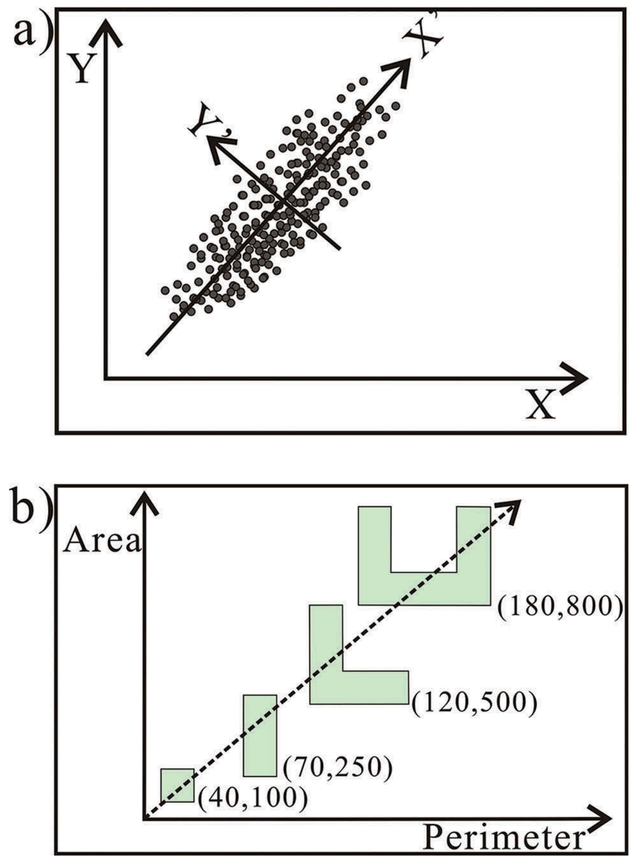

Figure 1. PCA (refer to Burghardt and Steiniger (2005)): (a) "Principal axis transformation" with simple rotation to create a better adapted coordinate system; (b) Reduction of two correlated parameters (area, perimeter) into one PC (size). The two parameters in this case will have the same PC coefficient.

correlated parameters could be selected from PC 3 to describe building orientation, we selected SMBRO from a visual cognition perspective. An earlier comprehensive analysis also showed that this parameter is the most effective at describing the orientation of any shape (Duchêne et al., 2003). We therefore consider the parameters CArea, IPQCom, EdgeCount, and SMBRO to comprise a representative set for describing building characteristics (Table 7).

Nevertheless, and as shown above, the parameters in this set can also be replaced; SArea can also be utilized for PC 1, for example, while RicCom or Fd can also be selected for PC 2. Numerous other parameters are also introduced to describe polygon characteristics including Compl and NCSP, utilized to describe the complexity of polygons by Brinkhoff, Kriegel, Schneider, and Braun (1995) and Moser et al. (2002). Although we used Fd and EdgeCount to describe complexity in this study, if Compl is selected to replace Fd and EdgeCount and another PCA is completed. Four PCs are still detected and the cumulative percentage reaches 89.488%. Indeed, if the same analytical procedure

Figure 2. The building dataset used for PCA in this study. (Data copyright OpenStreetMap contributors, CC BY-SA)

for parameter selection is applied as above, then CArea, IPQCom, Compl, and SMBRO can also be representative for describing building characteristics.

A number of statistical methods are also available for building classification. Of these, hierarchical clustering methods are the most widely applied which seek to build a hierarchy of clusters. In these methods, classification is mainly performed based on two "distances." "Distance" between elements, in this article, is defined as the square of Euclidean distance. "Distance" between two clusters, can be defined according to single- or complete-linkage; we chose to use ward-linkage in this study as it seeks a minimum increase of sum-of-squares after the merging of two clusters. Although parameter weights in building classification should be varied for different purposes or study areas, we consider all parameters equally and with the same weight in this study. At the same time, all parameters should be standardized in advance (Equation 1) prior to classification, and a terminal condition must also be set depending on the definition of similarity between building characteristics, based on cartographical and perceptual rules. The similarity of building characteristics between two adjacent entities was defined by Yan, Weibel, and Yang (2008); on this basis, if two buildings are similar, then the ratio between their areas will be greater than 60%, the ratio between their edges will be

Table 4. Building types.

	Rates (%)	
Type	Whole dataset	Selected dataset
Apartment	9.76	11.53
Commercial	6.44	3.84
House	32.83	26.92
Residential	38.06	30.76
Total	87.09	73.05

Table 5. PC contribution rates.

	Component					
Initial eigenvalue	1	2	3	4	5	6
Eigenvalue	11.871	5.204	2.592	1.643	0.928	…
% of variance	49.464	21.682	10.800	6.845	3.865	…
% of cumulative	49.464	71.146	81.946	88.791	92.656	…

Table 6. PC coefficient matrix. Underlined values mean that a component is mainly determined by the noted parameters.

	Component					Component			
Parameter	1	2	3	4	Parameter	1	2	3	4
Perimeter	−0.185	0.326	−0.008	0.018	DCMCom	0.265	0.145	−0.017	−0.142
Area	−0.055	0.390	−0.053	0.201	BotCom	−0.271	−0.128	0.023	0.110
SArea	−0.079	0.395	−0.033	0.076	BoyClaCom	0.232	0.046	−0.057	0.183
CArea	−0.075	0.397	−0.039	0.114	Elongation	0.206	0.179	0.030	−0.392
LChord	−0.208	0.264	−0.022	0.151	Ellipticity	0.253	0.134	−0.014	−0.171
MeanR	−0.205	0.259	−0.025	0.168	BBEccentricity	0.275	0.122	−0.026	−0.064
ShapeIndex	0.263	0.012	0.057	−0.194	DCircularity	−0.263	−0.012	0.057	−0.193
Concavity	0.145	−0.087	−0.108	0.580	Fd	0.280	−0.064	−0.036	0.081
IPQCom	0.279	0.090	−0.040	0.044	EdgeCount	−0.099	0.225	0.055	−0.394
RCom	−0.055	−0.290	0.035	−0.090	MWO	0.055	0.030	0.572	0.010
RicCom	0.281	0.075	−0.044	0.078	SWWO	0.062	0.050	0.580	0.087
GibCom	0.264	0.148	−0.014	−0.153	SMBRO	0.059	0.041	0.545	0.083

Table 7. Representative parameters used to describe building characteristics.

Spatial feature	Parameter	Definition
Size	CArea(or SArea)	Area of building convex hull
Shape	IPQCom	$IPQCom = 4\pi Area/Perimeter^2$
	EdgeCount	Number of building edges
Orientation	SMBRO	Building SMBR orientation

greater than 60%, and the gap between their orientations will be less than 40%. Thus, using these criteria for reference, constraints on the similarity of building characteristics over large areas can be relaxed to a certain extent. However, as similarity between clusters is defined as "distance" in hierarchical clustering methods, the terminal condition should be set as this variable; any clustering process will terminate if the threshold "distance" is Td, and if the "distance" between two clusters is greater than Td.

4. Clustering buildings based on proximity

From a first view on the buildings, we may divide buildings into parts, building patterns are then detected (Fei, 2002). Utilizing this visual process, we initially obtain building clusters in which structures are close to each other using a simple but practical MST-based method that generates groups via a pruning process (Zhang, Ai, et al., 2013). In cases where buildings are related to specific geographical environments, roads and rivers have already naturally divided these entities into parts.

A DT is therefore initially built using the central points of buildings to model their contextual and proximity relationships. Edges that connect across buildings or roads are then deleted, as the buildings they connect are not actually close to one another (Figure 3); the new DT generated after this deletion is considered representative of the proximity relationships between buildings and can therefore be used to derive MSTs.

A new DT does not always have to be connected, however, in order to do this; after one MST is built in the connected DT, and a new building is randomly selected as a starting point to build another. The edge of each MST is therefore weighted using the nearest distance between two buildings, and some edges are additionally deleted to generate building clusters. Although a number of constraints can be applied for edge deletion, since our aim is to find building clusters for map generalization, just longer ones were removed. According to Liu, Guo, Sun, and Ma (2014), an edge should be deleted if its length is greater than four times the positional accuracy threshold (0.5 mm) value of a building.

A number of specific patterns can be detected subsequent to the generation of building clusters. Linear patterns are most common with buildings

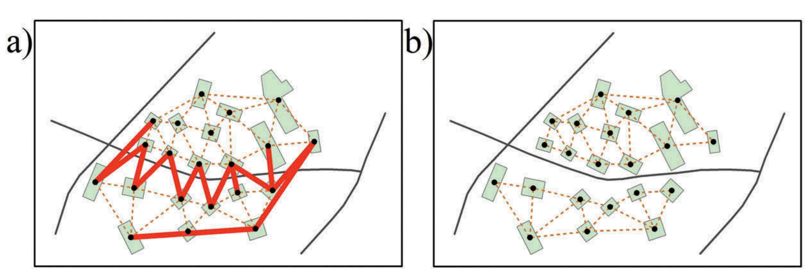

Figure 3. Building clustering taking obstacles into account: (a) Edges marked as heavy lines need to be deleted; (b) Result after deletion.

evenly distributed along straight lines. As discussed, MST has been widely utilized in this context to detect building patterns based on proximity relationships, a number of other proximity graphs are also available, including NNG, RNG, and GG. Definitions for these four proximity graphs were presented in the earlier work of Guo, Zheng, and Hu (2008); in a two-dimensional Euclidean space, a point set can be defined as $V = \{v_i|\mathrm{i} = 1, 2, \ldots \mathrm{n}\}$, and the distance between two points denoted as $D(v_i,v_\mathrm{j})$. Thus, if a DT is built based on this point set, then E denotes all edges; it therefore follows that $E = \{\mathrm{e}(v_i, v_j), v_i \in V, v_j \in V\}$ and can be used to define NNG, MST, RNG, and GG.

(1) NNG: If an edge $e(A,B)$ has a length denoted as $D(A,B)$, then if $D\mathrm{is}\ (A,B) \leq Min(Dis(v_i,A)), v_i, A, B \in V, v_i \neq A \neq B$, $e(A,B)$ is an edge of NNG.
(2) MST: This comprises an acyclic subset $T \subseteq E$ that connects all points in V, and has a total minimum length.
(3) RNG: If an edge $e(A,B)$has a length denoted as $D(A,B)$, then if $D\mathrm{is}(\mathrm{A},\mathrm{B}) \leq \mathrm{Max}(\mathrm{Dis}(\mathrm{v_i},\mathrm{A}), Dis(v_i,B)), v_i, A, B \in V, v_i \neq A \neq B$, $e(A,B)$ is a RNG edge.
(4) GG: If an edge $e(A,B)$ has a length denoted as $D(A,B)$, then if $(D\mathrm{is}(A,B))^2 \leq Min((Dis(v_i,A))^2 + (Dis(v_i,B))^2), v_i, A, B \in V, v_i \neq A \neq B, (A,B)$, is an edge of GG.

The results of four proximity graphs built from a point set are shown in Figure 4, and show that $NNG \subseteq MST \subseteq RNG \subseteq GG \subseteq DT$. This result can also be proven based on definitions and means that different graphs can be used to describe different proximity relationship levels. However, while the use of a specific graph may enable the discovery of a specific pattern, this does not always have to be a MST.

In cases where just proximity relationships are considered, linear patterns can be recognized based on differences in the orientation (D_o) and length (DL) of graph edges. Thus, D_o can be expressed as the angle between two adjacent edges, while DL between two adjacent edges can be expressed via multiples of the length of the longer edge divided by the length of the shorter edge. Suppose that the

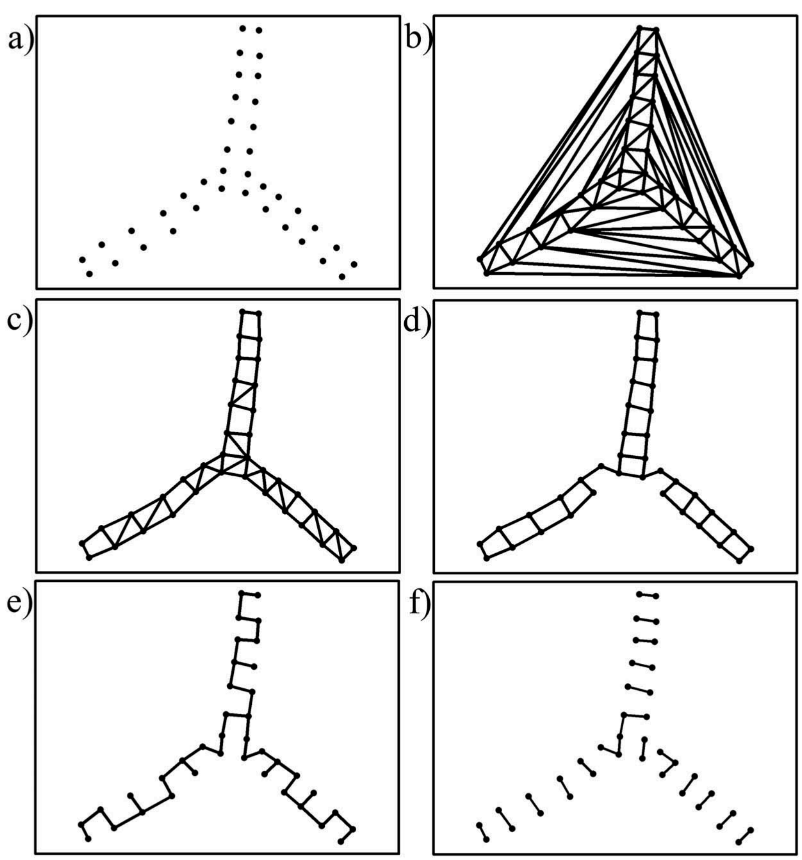

Figure 4. Four kinds of proximity graphs: (a) Point set; (b) DT; (c) GG; (d) RNG; (e) MST; (f) NNG.

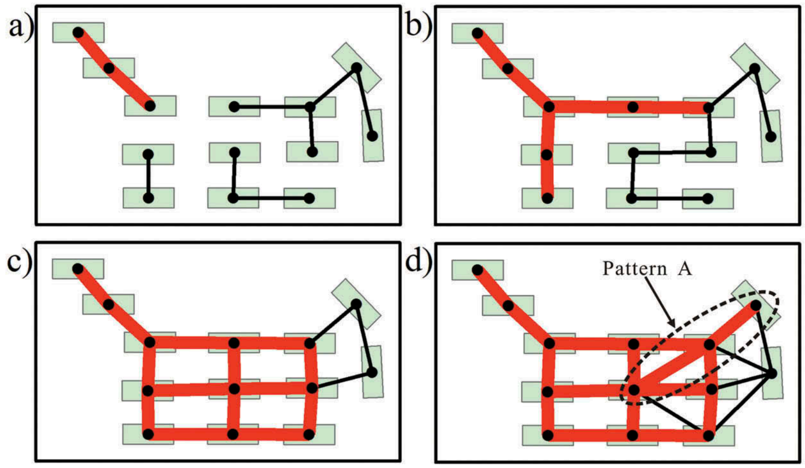

Figure 5. Linear patterns recognized using four different kinds of proximity graphs and marked as heavy lines: (a) NNG; (b) MST; (c) RNG; (d) GG.

thresholds of the two parameters D_o and DL are δ_o and δL; it therefore follows that if $D_o \geq \delta_o$, and $DL \leq \delta L$, then the three buildings that are connected by these two edges comprise a linear pattern. Thus, in the set $\delta_o = 165°$, $\delta L = 3$, linear patterns can be recognized using four proximity graphs (Figure 5). We adopted the method proposed by Du, Shu, and Feng (2015) for statistical analysis. Recognized linear patterns are correct if they are equal to the ones identified by users, denoted *tp*. Recognized linear patterns are considered incorrect if they are not equal to the ones identified by users, denoted *fp*, while those missed by this approach are denoted *fn*. *Correctness* is therefore defined as $tp/(tp + fp)$ and *Completeness* as $tp/(tp + fn)$. The results of linear pattern recognition are shown in Table 8.

Results show that NNG, MST, RNG all have similar *Correctness* (100%). It means that the linear patterns recognized using the three graphs can all be identified by users. While some linear patterns recognized based on GG may be denied by users (e.g. Pattern A in Figure 5), potentially leading to a confusion in understanding the spatial distribution of buildings. In addition, as the RNG method recovers a higher *Completeness* than either NNG and MST, more linear patterns can be recognized based on RNG as GG and RNG both recover the same *Completeness*. It means use of RNG is sufficient to detect linear patterns. The ability of a RNG to detect perceptually meaningful structures from a point set has also been shown more effective by Toussaint (1980), while linear patterns recognized based on RNG might also be helpful for subsequent grid-like pattern recognition (Figure 5). The RNG is therefore the most effective at detecting linear patterns in the four proximity graphs.

For a more convinced conclusion, linear patterns are also recognized under constraints ($\delta_o = 170°$, $\delta L = 3$) in three areas separated by roads in our building dataset for PCA (Figure 6, Table 9). We can easily get the similar conclusion as above in Area A and Area B. While in Area C, GG recovers a higher *Completeness* than RNG, it means GG enable the discovery of additional linear patterns than RNG. But GG is also with a lower *Correctness*.

As our linear patterns are only detected based on proximity graphs using nearest distances, those recognized based on RNG using gravity distance can also be a supplement of the result of linear pattern recognition. As shown in Figure 7(a), missed linear patterns (e.g. Pattern A, Pattern B, and Pattern C in Area C) can be recognized by RNG using gravity distance. Thus, linear patterns recognized using GG in Area C can also be recognized using RNG. Different constraints can also influence the results of linear pattern recognition; as shown in Figure 7(b), linear patterns are recognized based on RNG using nearest distance and constraints ($\delta_o = 165°, \delta L = 3$). Thus, linear patterns (e.g. Pattern A and Pattern B in Figure 7(b)) can also be recognized in Area B.

5. Analysis of building spatial distribution

5.1. Data handling procedure

The aim of this study was to separately take proximity and similarity between buildings into account for the identification of differences in classes within clusters and patterns, and clusters and patterns within classes. The data handling procedure applied is shown in Figure 8.

Table 8. Evaluating recognized linear patterns using four proximity graphs under the same constraints.

Graph	tp	fp	fn	Correctness	Completeness
NNG	1	0	6	100%	14.3%
MST	3	0	4	100%	42.9%
RNG	7	0	0	100%	100%
GG	7	1	0	87.5%	100%

buildings were classified within a single category are considered extraordinary within the context of this dataset, for example class 5 and class 7. Although in some cases where the numerical difference between the orientation of two is more than 90°, the actual difference is a supplementary angle.

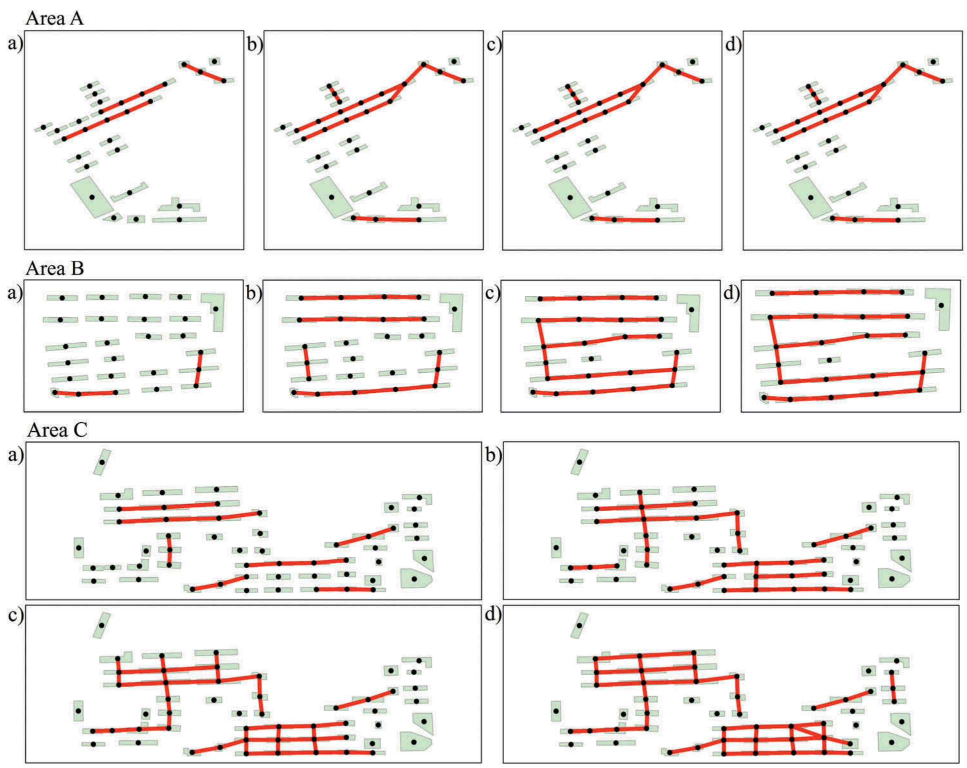

Figure 6. Linear patterns recognized using four different kinds of proximity graphs and marked as heavy lines: (a) NNG; (b) MST; (c) RNG; (d) GG.

Table 9. Evaluating recognized linear patterns using four proximity graphs under the same constraints.

	Area A		Area B		Area C	
Graph	Correctness	Completeness	Correctness	Completeness	Correctness	Completeness
NNG	100%	20%	100%	50%	100%	17.4%
MST	100%	50%	100%	83.3%	100%	47.8%
RNG	100%	70%	100%	83.3%	100%	82.4%
GG	100%	70%	100%	83.3%	94.4%	100%

5.2. *Classification, clustering, and pattern recognition*

We tested our approach using a sample of 208 buildings at a 1:10,000 scale. Buildings were classified using the method described in Section 3, and with the terminal condition between classes set at a "distance" of no more than 150. Buildings were therefore divided into seven classes (Figure 9). Buildings within the same category are similar in size, orientation, and shape. Cases where no more than 10 And we used the first value for our classification. It may lead a result when buildings differ greatly in orientation, they will be divided into two classes. Furthermore, given a practical building classification, terminal conditions can also be adjusted; for example, if a stringent terminal condition is set to classify buildings into more classes, the buildings in each class will be more visually similar to one another.

Applying a MST-based building clustering approach to our dataset (Section 4) led to the generation of 29

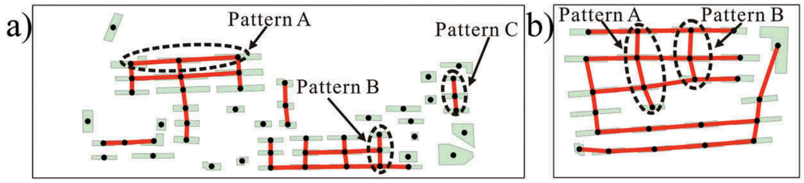

Figure 7. Linear patterns recognized using RNG and marked as heavy lines: (a) Gravity distance and the same constraints ($\delta o = 170°, \delta L = 3$); (b) Nearest distance and loose constraints ($\delta o = 165°, \delta L = 3$).

sub-clusters (Figure 9), and also enabled clusters of buildings within each class to be detected (Figure 10).

It is possible to detect specific patterns subsequent to the generation of building clusters. One approach to achieve this is to detect linear patterns using a RNG incorporating the constraints discussed above (Section 4, $\delta o = 165°, \delta L = 3$). A range of different distance definitions can be applied to build RNGs, including gravity and the nearest distance. However, applying these different distance definitions can result in varying pattern recognition (Figure 11); observations show that different linear patterns can be recognized using two kinds of distance, and that these patterns can meet structural maintenance requirements when developing map generalizations. As two distances will describe different aspects of the proximity relationship between buildings, linear patterns recognized using these variable approaches can also supplement one another. A fused set of two outcomes can then be used as a final result to express linear patterns within a building cluster.

This approach also enables building patterns within each class to be detected as linear relationships that correspond to similarities in building characteristics (Figure 12). Traditional linear pattern recognition approaches aim to identify these kinds of patterns using perceptual rules, and are successful in most cases (Figure 12). However, some linear patterns remain invisible because of stringent building classification constraints even though they may actually comprise a class if looser constraints are applied. Comparing the results in Figure 11 and Figure 12 reveals a number of additional meaningful linear patterns (circled in Figure 12); of these, the buildings that comprise linear patterns 1, 6, 7, and 8 are not actually very close to one another. This means that it is possible to detect patterns in cases where buildings are not close

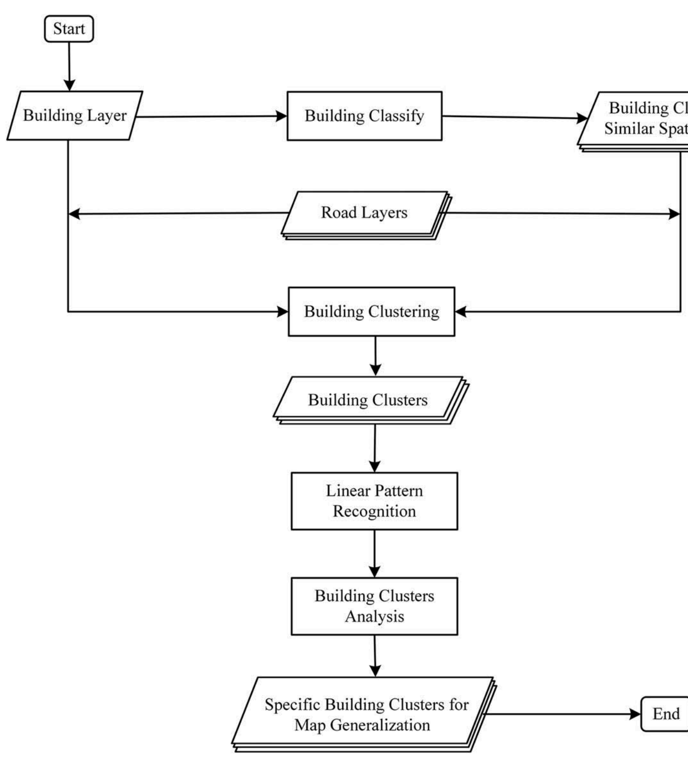

Figure 8. The data handling procedure applied in this study.

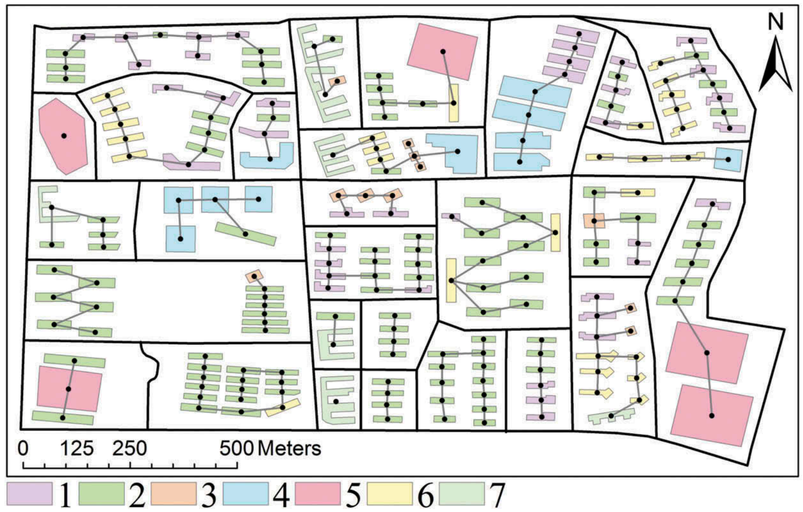

Figure 9. Result of building classification and clustering using based on MST. Buildings connected by MST edges are regarded as a cluster.

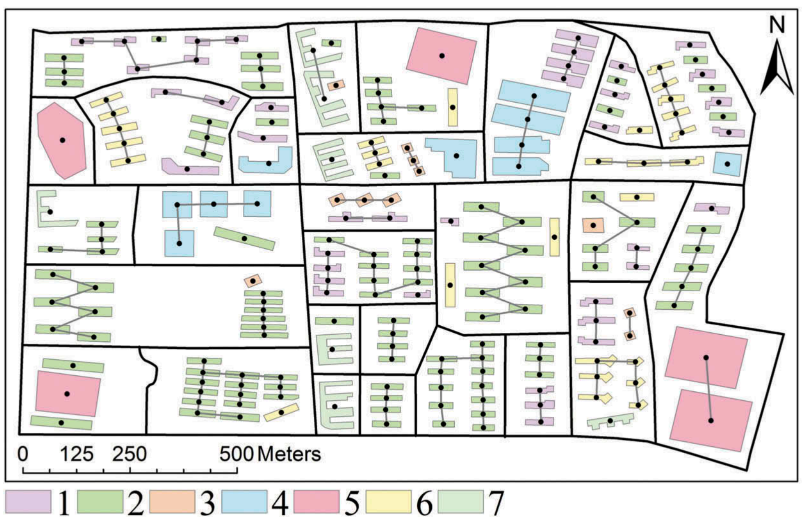

Figure 10. Result of clustering within each class of buildings based on MST. Buildings connected by MST edges are regarded as a cluster.

to one another. While pattern 1 to pattern 13 actually encompass structure within patterns and may further aid our understanding of encapsulated linear relationships.

5.3. Analysis of spatial distribution

Building classes based on similarities in characteristics were obtained via classification, while clusters of buildings based on proximity were obtained using clustering. We then detected linear patterns within these building clusters or classes, and combined this information together to enable a better understanding of spatial building distributions. This approach enabled us to evaluate differences in both classes and building characteristics within each cluster; indeed, since linear patterns are also kind of building cluster, we also evaluated differences in classes and building characteristics within each pattern, selecting seven typical building clusters and seven typical linear patterns for further analysis by visual inspection (Figure 13).

In order to measure class differences within each cluster, it is first necessary to determine the membership of each and count the number of buildings in each class. This was done using the ratio between buildings in each class versus specific clusters. Differences between building characteristics were then measured using variation in relevant parameters, specifically the

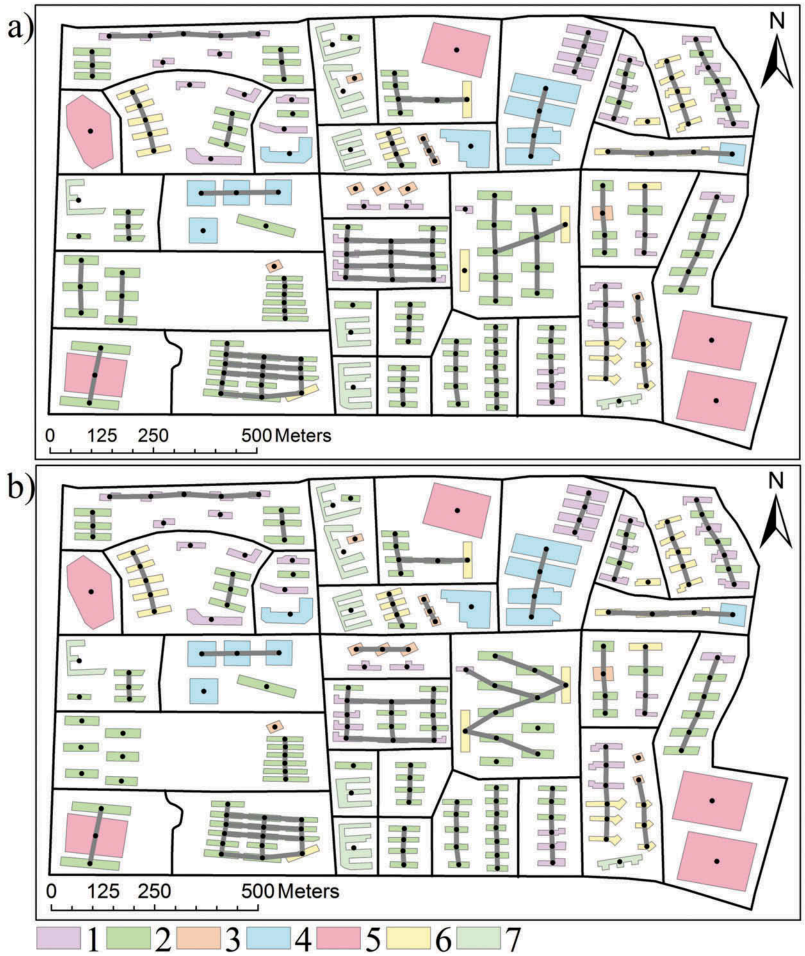

Figure 11. Linear pattern recognition using RNG for each cluster of buildings: (a) Gravity distance; (b) Shortest distance. Linear patterns are marked as heavy lines.

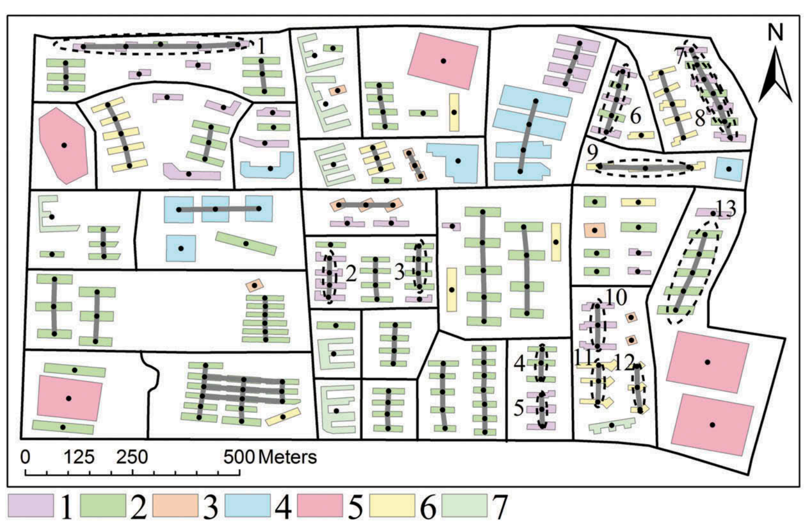

Figure 12. Linear pattern recognition using RNG for each class of buildings marked as heavy lines.

four outlined in Section 3. A coefficient of variation is also adopted to measure the difference in building characteristics and to mitigate the influence of different parameter dimensions, as follows:

$$C.v = (Std/M) \times 100\% \tag{2.}$$

In this expression, M denotes the average parameter value, and the differences between each building cluster are summarized in Table 10. These results show that just one class of buildings are grouped within Cluster A that are all similar in size, shape, and orientation, while at the same time Cluster B and Cluster C both contain two building classes. The buildings in class 2 within Cluster B account for 93.3% of variation and are the major class within this cluster, while the other set comprises a minor class. Similarly, results show that the buildings in these two classes differ mainly in CArea and SMBRO, while those in both the two classes in Cluster C encompass 50% of variation and differ mainly in CArea and EdgeCount. Cluster D, Cluster E, Cluster F, and Cluster G all comprise two more classes; of these, the buildings in class 2 of Cluster D accounting for almost 71.4% of variation; buildings in Cluster E are all similar in size and shape, but differ markedly in SMBRO.

As a linear pattern is just a special kind of building cluster, the same approach to analysis can also be applied in this case (Table 11). Comparing linear patterns based on building characteristics shows that buildings in pattern 1 are all similar in size, shape, and orientation, while those that make up pattern 2, pattern 3, pattern 5, and pattern 7 are markedly different in SMBRO. The buildings that comprise pattern 4 all show marked differences in EdgeCount, while those in pattern 6 exhibit marked differences in CArea and SMBRO. Adding class information to these patterns shows that buildings within some linear patterns comprise the same class; this is the case, for example, for pattern 1, a traditional linear arrangement. In contrast, some examples are from two classes that are nevertheless arranged in a specific order (e.g. pattern 2, pattern 3, pattern 4, and pattern 7). These buildings are arranged in the form aab, abba, abababa, and aaabbb, which can be meaningful for specific generalization operators such as typification. The buildings that make up pattern 5 also derive from two classes, while the one building in class 3 can be considered an extraordinary one. Although the buildings that comprise some linear patterns may encompass a variety of classes, they tend to share the same orientation (e.g. pattern 4), size (e.g. pattern 2), or shape (e.g. pattern 6).

The clustering results and linear patterns recognized within each class are illustrated in Figure 9 and Figure 11. Although some spatial statistical parameters

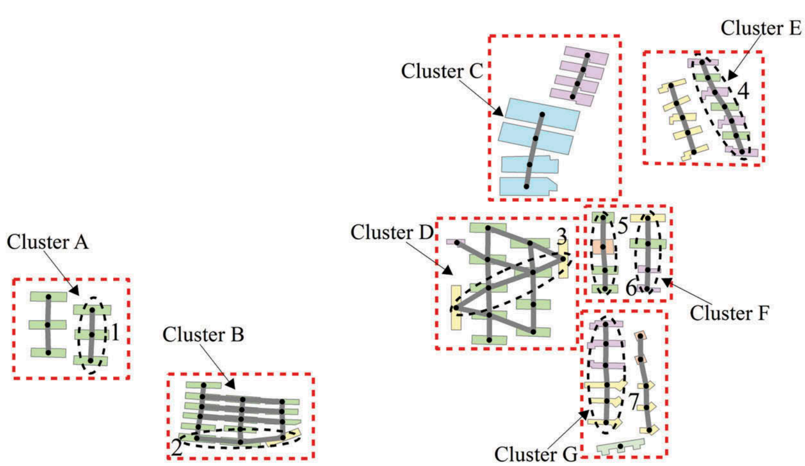

Figure 13. Building clusters and linear patterns selected for further analysis. Clusters are set in rectangles and labeled as Cluster A to Cluster G, while linear patterns are circled and labeled 1 to 7.

Table 10. Differences between building clusters.

			C.v			
Cluster	Number	Class specification and rate	CArea	IPQCom	EdgeCount	SMBRO
A	6	$2^{(100\%)}$	6.7%	4.0%	0%	0.3%
B	15	$2^{(93.3\%)}$, $6^{(6.7\%)}$	17.2%	6.5%	6.3%	24.2%
C	8	$1^{(50\%)}$, $4^{(50\%)}$	52.4%	11.4%	21.2%	2.8%
D	12	$1^{(8.3\%)}$, $2^{(75\%)}$, $6^{(16.7\%)}$	22.6%	9.5%	13.8%	21.0%
E	12	$1^{(33.3\%)}$, $2^{(25\%)}$, $6^{(41.7\%)}$	18.2%	15.2%	32.3%	90.7%
F	8	$1^{(25\%)}$, $2^{(50\%)}$, $3^{(12.5\%)}$, $6^{(12.5\%)}$	32.1%	23.0%	19.8%	61.3%
G	12	$1^{(25\%)}$, $3^{(16.7\%)}$, $6^{(50\%)}$, $7^{(8.3\%)}$	52.8%	38.3%	40.0%	124.4%

Table 11. Differences between linear building patterns.

Pattern	Number	Class order	C.*v*			
			CArea	IPQCom	EdgeCount	SMBRO
1	3	2, 2, 2	8.7%	3.0%	0%	0.4%
2	3	2, 2, 6	7.2%	2.9%	0%	72.1%
3	3	6, 2, 2, 6	13.3%	4.1%	0%	38.2%
4	4	1, 2, 1, 2, 1, 2, 1	17.2%	14.6%	33.3%	0.5%
5	4	2, 3, 2, 2	19.2%	14.7%	0%	65.8%
6	4	6, 2, 1, 1	46.1%	10.5%	18.2%	66.5%
7	4	1, 1, 1, 6, 6, 6	16.1%	4.5%	7.3%	104.8%

could be used to describe the features of these clusters (e.g. Morans I), these may not be reliable in cases where few buildings are present. We therefore provide qualitative descriptions (Figure 9); the buildings in the two classes of Cluster C tend to group together separately, it is also the same for buildings in class 2 of Cluster D and in class 6 of Cluster E. Similarly, pattern 4 of Cluster E comprises two inner classes where buildings are not adjacent to one other and form a sub-linear pattern (Figure 11), while pattern 7 in Cluster G contains two inner classes with buildings toward the sides of the pattern also arranged sub-linearly.

5.4. Discussion

Building generalization aims to solve conflicts while clearly maintaining the reality of buildings via operators including selection, aggregation, and typification (Schmid, 2008). In most cases, building reality refers to the characteristics of individual buildings, or their density, the boundaries of groups, and specific patterns (Fei, 2002; Guo, Wei, Wang, & Wang, 2017), while maintenance should also take into account map space capacity. As our method also enables information about spatial distributions, it should allow for more reasonable generalizations. Considering the buildings in Cluster B, for example, which comprises two classes; as those in class 6 only account for 6.7% of variation, and can be regarded as special cases, these need be selected in building generalization. While the same is true for buildings in class 7 in Cluster G. In terms of the generalization of building patterns, those in pattern 4, for example, derive from two classes arranged sub-linearly. Thus, the application of a typification operator should maintain these internal relationships, and the same will also be true for buildings in pattern 7; these will also need to be maintained as they belong to two sub-linear arrangements along two sides.

Some limitations to our method should also be highlighted. We have only summarized 24 parameters from related works for the PCA in this study, but numerous additional factors were introduced to accurately measure polygon characteristics. Further work will therefore be required to determine whether, or not, these variables are effective and reasonable for the measurement of building characteristics. Similarly, further enhancements in geographic information content, especially building heights, may need to be incorporated into future building classification. It is also noteworthy that map generalization is a data simplification process involving all kinds of feature classes, not just buildings. Thus, in cases where other classes are to be considered, their relationships with buildings is a further key issue which needs to be taken into account in classification and pattern recognition. As our approach detected only linear patterns, more kinds should also be addressed in future works. The constraints used in this study will also require adjustment when geographical regions or types of building patterns change, while additional factors, such as building density, should be taken into account.

The use of spatial distributions is just one important aspect that needs to be taken into account when developing more accurate building generalization. Additional features, including map space capacity and the maintenance of building group boundaries, also need to be taken into consideration. And further analyses will also be required to address the question of how to combine results following clustering, classification, and pattern recognition to enable an enhanced understanding of building spatial distribution based on generalization requirements. More detailed research will also be necessary to determine how information can be applied more directly to building generalization while at the same time taking all relevant factors into account.

6. Conclusion

We have separately considered the proximity and similarity between buildings in this study to explore the spatial distributions within building groups. In order to improve the efficiency of building classification, we try to only select some representative parameters for this process. Thus, four parameters were selected and verified using PCA, enabling us to utilize them for subsequent building classification. The spatial environment has also been taken into account when applying a clustering method to partition buildings. We also conducted a study on the abilities of different proximity graphs to detect linear building patterns, and showed that the use of a RNG is the most efficient. Application of the method outlined in this article will be helpful for identifying simple building groups given a map generalization operator, while maintaining and clearly expressing spatial distributions. Our method can also be applied for the identification of other building pattern types and spatial point groups.

Future research should aim to take more information into account during the process of building classification, including heights, relationships between buildings and other classes of features. It will also be necessary to determine which parameters are the most representative when measuring the specific characteristics of buildings. Finally, the use of proximity graphs for buildings that take into account additional factors such as density in the recognition of different and complex spatial patterns should also be explored.

Acknowledgments

The authors would like to thank the editors and the anonymous reviewers for their helpful and constructive comments that greatly contributed to improve the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China [41471384].

ORCID

Zhiwei Wei http://orcid.org/0000-0002-3494-3686
Qingsheng Guo http://orcid.org/0000-0001-5863-1946
Lin Wang http://orcid.org/0000-0002-4045-5398
Fen Yan http://orcid.org/0000-0003-3835-9249

References

Ai, T., & Guo, R. (2007). Polygon cluster pattern mining based on Gestalt principles. *Acta Geodaetica et Cartographica Sinica*, 36(3), 303–308. Retrieved from http://en.cnki.com.cn/Article_en/CJFDTOTAL-CHXB200703012.htm

Allouche, M., & Moulin, B. (2005). Amalgamation in cartographic generalization using Kohonen's feature nets. *International Journal of Geographical Information Science*, 19(8–9), 899–914. doi:10.1080/13658810500161211

Anders, K. H. (2006). Grid typification. In A. Riedl, W. Kainz, G. A. Elmes (Eds), *Progress in Spatial Data Handling* (633–642). Berlin: Springer. doi:10.1007/3-540-35589-8_40

Anders, K. H., & Sester, M. (2001). Parameter-free cluster detection in spatial databases and its application to typification. *International Archives of Photogrammetry and Remote Sensing*, 33(B4/1; PART 4), 75–83. Retrieved from http://www.isprs.org/proceedings/XXXIII/congress/part4/75_XXXIII-part4.pdf

Basaraner, M., & Cetinkaya, S. (2017). Performance of shape indices and classification schemes for characterising perceptual shape complexity of building footprints in GIS. *International Journal of Geographical Information Science*, 31(10), 1952–1977. doi:10.1080/13658816.2017.1346257

Boffet, A., & Serra, S. (2001, August). Identification of spatial structures within urban blocks for town characterization. In *Proceedings of the 20th International Cartographic Conference*, Beijing. Retrieved from: http://icaci.org/files/documents/ICC_proceedings/ICC2001/icc2001/file/f13016.pdf

Brinkhoff, T., Kriegel, H. P., Schneider, R., & Braun, A. (1995, December). Measuring the complexity of polygonal objects. In *Proceedings of the 3rd ACM International Workshop on Advances in Geographic Information Systems, Baltimore, MD.* Retrieved from: https://static.aminer.org/pdf/PDF/000/232/833/measuring_the_complexity_of_polygonal_objects.pdf

Burghardt, D., & Steiniger, S. (2005, July). Usage of principal component analysis in the process of automated generalisation. In *Proceedings of the 22nd International Cartographic Conference*, A Coruña, Spain. Retrieved from: http://www.geo.uzh.ch/~sstein/manuscripts/pca_settlement_icc2005.pdf

Carmona, M., Heath, T., Oc, T., & Tiesdell, S. (2003). *Public places, urban spaces: The dimensions of urban design*. Oxford: Architectural Press.

Cetinkaya, S., Basaraner, M., & Burghardt, D. (2015). Proximity-based groupingof buildings in urban blocks: A comparison of four algorithms. *Geocarto International*, 30 (6), 618–632. doi:10.1080/10106049.2014.925002

Christophe, S., & Ruas, A. (2002). Detecting building alignments for generalisation purposes. In D. E. Richardson, & P. van Oosterom (Eds) *Advances in Spatial Data Handling* (pp. 419–432). Berlin: Springer. doi:10.1007/978-3-642-56094-1_31

Deng, M., Liu, Q., Cheng, T., & Shi, Y. (2011). An adaptive spatial clustering algorithm based on Delaunay triangulation. *Computers, Environment and Urban Systems*, 35(4), 320–332. doi:10.1016/j.compenvurbsys.2011.02.003

Deng, M., Tang, J., Liu, Q., & Wu, F. (2017). Recognizing building groups for generalization: A comparative study. *Cartography and Geographic Information Science*, advance online publication. doi:10.1080/15230406.2017.1302821

Department of Geography, University of Zurich. (1999). *Selection of Basic Measures* [Report]. Retrieved from http://agent.ign.fr/deliverable/DC1.html.

Du, S., Shu, M., & Feng, C. C. (2015). Representation and discovery of building patterns: A three-level relational approach. *International Journal of Geographical Information Science*, 30(6), 1161–1186. doi:10.1080/13658816.2015.1108421

Duchêne, C., Bard, S., Barillot, X., Ruas, A., Trévisan, J., & Holzapfel, F. (2003, April). Quantitative and qualitative description of building orientation. In *5th Workshop on Progress in Automated Map Generalization*, Paris.

Fei, L. (2002). *A method of automated cartographic displacement: On the relationship between streets and buildings* (Doctoral dissertation). University of Hanover, Germany. Retrieved from http://hdl.handle.net/10068/134828

Gong, X., Wu, F., Qian, H., & Ma, K. (2014). The parameter discrimination approach to multi-connected linear pattern recognition in building groups [in Chinese]. *Geomatics and Information Science of Wuhan University*, 39(3), 335–339. doi:10.13203/j.whugis20120708

Guo, Q., Wei, Z., Wang, Y., & Wang, L. (2017). The method of extracting spatial distribution characteristics of buildings combined with feature classification and proximity graph. *Acta Geodaetica et Cartographica Sinica*, 46(5), 631–638. http://xb.sinomaps.com/EN/Y2017/V46/I5/631

Guo, Q, Zheng, C, & Hu, H. (2008). Hierarchical clustering method of group of points based on the neighborhood graphs. *Acta Geodaetica et Cartographica Sinica*, 37(2), 256–261. http://xb.sinomaps.com/CN/Y2008/V37/I2/0

He, Z., & Song, Y. (2015). *General Map Compilation* [in Chinese]. Wuhan: Wuhan University Press.

Hermosilla, T., Ruiz, L. A., Recio, J. A., & Cambra-López, M. (2012). Assessing contextual descriptive features for plot-based classification of urban areas. *Landscape and Urban Planning*, 106(1), 124–137. doi:10.1016/j.landurbplan.2012.02.008

Karypis, G., Han, E., & Kumar, V. (1999). Chameleon: Hierarchical clustering using dynamic modeling. *Computer*, 32(8), 68–75. doi:10.1109/2.781637

Li, W., Goodchild, M., & Church, R. (2013). An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. *International Journal of Geographical Information Science*, 27(6), 1227–1250. doi:10.1080/13658816.2012.752093

Li, Z., Yan, H., Ai, T., & Chen, J. (2004). Automated building generalization based on urban morphology and Gestalt theory. *International Journal of Geographical Information Science*, 18(5), 513–534. doi:10.1080/13658810410001702021

Liu, Y., Guo, Q., Sun, Y., & Ma, X. (2014). A combined approach to cartographic displacement for buildings based on skeleton and improved elastic beam algorithm. *PLoS One*, 9(12), e113953. doi:10.1371/journal.pone.0113953

Lu, Z., Im, J., Rhee, J., & Hodgson, M. (2014). Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data. *Landscape and Urban Planning*, 130(1), 134–148. doi:10.1016/j.landurbplan.2014.07.005

MacEachren, A. (1985). Compactness of geographic shape: Comparison and evaluation of measures. *Geografiska Annaler, Series B, Human Geography* 67(1), 53–67. doi:10.1080/04353684.1985.11879515

Moser, D., Zechmeister, H. G., Plutzar, C., Sauberer, N., Wrbka, T., & Grabherr, G. (2002). Landscape patch shape complexity as an effective measure for plant species richness in rural landscapes. *Landscape Ecology*, 17(7), 657–669. doi:10.1023/A:1021513729205

Peura, M., & Iivarinen, J. (1997). Efficiency of simple shape descriptors. In L. P. Cordella, C. Arcelli, & G. Sanniti di Baja (Eds), *Advances in Visual Form Analysis: Proceedings of the 3th International Workshop on Visual Form* (pp. 443–451). Singapore: World Scientific.

Qi, H., & Li, Z. (2008). An approach to building grouping based on hierarchical constraints. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B2), 449–454. Retrieved from: http://www.isprs.org/proceedings/XXXVII/congress/2_pdf/3_WG-II-3/13.pdf

Rainsford, D., & Mackaness, W. (2002). Template matching in support of generalization of rural buildings. In D. E. Richardson, & P. van Oosterom (Eds) Advances in Spatial Data Handling (pp. 137–151). Berlin: Springer. doi:10.1007/978-3-642-56094-1_11

Regnauld, N. (2001). Contextual building typification in automated map generalization. *Algorithmica*, 30(2), 312–333. doi:10.1007/s00453-001-0008-8

Ruas, A., & Holzapfel, F. (2003, August). *Automatic characterisation of building alignments by means of expert knowledge*. Proceedings of 21st international cartographic conference (pp. 1604–1615), Durban, South Africa. Retrieved from: https://icaci.org/files/documents/ICC_proceedings/ICC2003/Papers/208.pdf

Schmid, S. (2008). *Automated constraint-based evaluation of cartographic generalization solutions* (Master's Thesis), University of Zurich. Retrieved from http://www.geo.uzh.ch/dam/jcr:a1dfc9a8-dfc8-49d3-9dc0-f519490836c9/msc_stefan_schmid_2008.pdf

Steiniger, S., Lange, T., Burghardt, D., & Weibel, R. (2008). An approach for the classification of urban building structures based on discriminant analysis techniques. *Transactions in GIS*, 12(1), 31–59. doi:10.1111/j.1467-9671.2008.01085.x

Toussaint, G. (1980). The relative neighbourhood graph of a finite planar set. *Pattern Recognition*, 12(4), 261–268. doi:10.1016/0031-3203(80)90066-7

Wang, W., Du, S., Guo, Z., & Luo, L. (2015). Polygonal clustering analysis using multilevel graph-partitions. *Transactions in GIS*, 19(5), 716–736. doi:10.1111/tgis.12124

Yan, H., Weibel, R., & Yang, B. (2008). A multi-parameter approach to automated building grouping and generalization. *Geoinformatica*, 12(1), 73–89. doi:10.1007/s10707-007-0020-5

Yang, W. (2011). Identify building patterns. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37 (B2), 391–398. Retrieved from http://www.isprs.org/proceedings/XXXVII/congress/2_pdf/3_WG-II-3/04.pdf

Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, 37(1), 1–19. doi:10.1016/j.patcog.2003.07.008

Zhang, L., Deng, H., Chen, D., & Wang, Z. (2013). A spatial cognition-based urban building clustering approach and its applications. *International Journal of Geographical Information Science*, 27(4), 721–740. doi:10.1080/13658816.2012.700518

Zhang, X., Ai, T., & Stoter, J. (2010). Characterization and detection of building patterns in cartographic data: Two algorithms. In A.Yeh, W. Shi, Y. Leung, & C. Zhou (Eds), *Advances in Spatial Data Handling and GIS* (pp. 93–107). Berlin: Springer. doi:10.1007/978-3-642-25926-5_8

Zhang, X., Ai, T., Stoter, J., Kraak, M. J., & Molenaar, M. (2013). Building pattern recognition in topographic data: Examples on collinear and curvilinear alignments. *Geoinformatica*, 17(1), 1–33. doi:10.1007/s10707-011-0146-3

Zhong, C., Miao, D., & Wang, R. (2010). A graph-theoretical clustering method based on two rounds of minimum spanning trees. *Pattern Recognition*, 43(3), 752–766. doi:10.1016/j.patcog.2009.07.010