

Typology based on three density variables central to Spacematrix using cluster analysis

Meta Berghauser Pont, Jesper Olsson

Department of Architecture and Civil Engineering, Chalmers University of Technology,
Gothenburg, Sweden

E-mail: meta.berghauserpont@chalmers.se, gusjespol@student.gu.se

Abstract. *Since the publication of the book ‘Spacematrix. Space, density and urban form’ (Berghauser Pont and Haupt, 2010), the Spacematrix method has been linked back to its theoretical foundations by Steadman (2013), is further developed using the measure of accessible density to arrive at a density measure that more closely relates to the environment as experienced by people moving through the city (Berghauser Pont and Marcus, 2014) which then is used to arrive at a multi-scalar density typology (Berghauser Pont et al. 2017). This paper will take yet another step in the development of the Spacematrix method by including the measure of network density in the classification which until now was not used to its full potential. Important for successful classification is the ability to ascertain the fundamental characteristics on which the classification is to be based where the work of Berghauser Pont and Haupt (2010) will be followed addressing three key variables: Floor Space Index (FSI), Ground Space Index (GSI) and Network density (N) where especially the last was not fully included in the earlier work. Besides a typology based on these three variables, this paper will also result in a robust statistical method that can later be used on larger samples for city-scale comparisons. Two statistical methods are tested: hierarchical clustering and centroid-based clustering and besides a general discussion about their strong and weak points, the paper shows that the hierarchical method is more convincing in distinguishing differences in both building type and street pattern that is especially captured with Network density (N). As this method is not useful for large datasets we propose a combination of the two clustering methods as the way forward.*

Keywords: Typology, classification, cluster analysis, density, Spacematrix

Introduction

Since the publication of the book ‘Spacematrix. Space, density and urban form’ (Berghauser Pont and Haupt, 2010), the Spacematrix method has been linked back to its theoretical foundations by Steadman (2013), is further developed using the measure of accessible density to arrive at a density measure that more closely relates to the environment as experienced by people moving through the city (Berghauser Pont and Marcus, 2014) which then is used to arrive at a multi-

scalar density typology (Berghauser Pont et al. 2017). This paper will take yet another step in the development of the Spacematrix method by including the measure of network density in the classification which until now was not used to its full potential.

Berghauser Pont and Haupt (2010) showed that a multivariable density concept consisting of four density variables offers a method to define building types, which these variables separately were incapable of doing. The variables used are well-known indicators of density where

Ground Space Index (GSI) describes the use of the ground in two dimensions and Floor Space Index (FSI)¹ describes the intensity of the use of the ground floor by stacking floor space in the third dimension which brings us to the third variable: the average height (i.e. amount of floors) of the buildings (L). A fourth variable describes the spaciousness (also called Open Space Ratio, OSR) and provides an indication of the intensity of use of the non-built space. It was introduced in Germany in the 1920 to avoid overcrowding and can be derived from FSI and GSI using the following expression $(1-GS)/FSI$. All four variables can be represented simultaneously in a scatter graph named 'Spacemate' with FSI on the y-axis and GSI on the x-axis; OSR and L are gradients that fan out over the diagram. By plotting a large number of observations (i.e. neighbourhoods) on the Spacemate, Berghauser Pont and Haupt (ibid.) showed convincingly that building types cluster. High-rise strip types and mid-rise block types with similar FSI position at distinct location on the graph due to differences in GSI, OSR and L (see figure 1). However, one important characteristic of the urban fabric, that is the grain of the street layout, that is street-to-street distance which directly relates to the size of the islands or urban blocks², is not captured with these four variables. This, Berghauser Pont and Haupt (ibid.) propose, could be captured by measuring network density (N), that is, the length of street network per area unit. Two Amsterdam examples with identical built densities, that is FSI, GSI, OSR and L values, can illustrate this (see figure 1). The Grachtengordel is composed of large islands and thus low N (0.012), while De Pijp with the same built density, has smaller islands and thus a high N (0.023). One building type, in this case the mid-rise block type, can thus be realized with different street layouts, in the same manner as the same FSI can be realised with very different solutions of GSI, OSR and L (recall the example discussed earlier).

Berghauser Pont and Haupt (2010) convincingly argue that network density is needed to develop a typology that goes beyond building types and includes characteristics of the street layout. They show the three-dimensional scatter graph 'Spacematrix' with FSI on the

y-axis, GSI on the x-axis and N on the z-axis and discuss such a typology in concept, but do not use it to develop it. An important reason is that clustering using such a three-dimensional scatter graph is complex and demands "further development of computerized software to manage and analyse data, and to represent the outcome" (ibid., page 190).

The aim of this paper is to do just that and develop a typology of urban fabrics based on the full set of variables discussed above: FSI, GSI, OSR, L and N using statistical methods that can deal with this complexity. Further, the purpose is to develop a method that also can be used in comparative studies of whole cities, which puts requirements to the capability of the statistical method to handle large datasets. In the following sections the two statistical methods used for this purpose and the resulting clustering will be presented, but before that we will shortly discuss the importance of network density and the possibilities of cluster analysis in more general terms.

Network density

The street system is one of the main entities of the town plan (or ground plan) besides the lot pattern and the building configuration (Conzen, 1960; Moudon, xx; Whitehand, 2001). These three entities street, lot and building are similar to the distinction made by Jan Heeling, et al. (2002) between the public streets (street system) and private islands (consisting of lots and buildings). The combination of streets and a series of islands constitute the urban fabric, or tissue. The main task of a designer, according to Heeling et al., is to combine these two in the best possible manner.

Street space, to use the words of Stephen Marshall (2005), constitutes "the basic core of all urban public space forming a contiguous network by which everything is linked to everything else. This continuum is punctured by lots of private land." For one, built floor space generates movements and causes flows (people, cars, etcetera) that need to be facilitated by the network. In addition, the open space of the network enables light to access the buildings and influences privacy, depending

on the profile width and the size of the islands (Martin and March, 1972). Further, following the argument of Manuel de Solà-Morales (1978), the grain of the urban fabric is decisive in establishing the relationship between general form and built content where smaller blocks (i.e. islands) provide the greatest proportion of public ways and overall exposure ratio (façade length to area). “This whole interrelatedness of network, islands and building bulk should thus be at the core of a new definition of density” (Berghauser Pont and Haupt, 2010, p. 97) and following this, the development of an urban fabric typology.

Cluster analysis

Cluster analysis is a generic name for a large set of statistical methods aiming at the detection of groups (i.e. clusters) also called automatic classification, numerical taxonomy, and typological analysis. Tryon published a book on cluster analysis in 1939 the method of clustering can be traced back to the major work of Carl von Linné (*Systema Naturae*, 1735), although at that time not using statistics (Wilmink and Uytterschaut, 1984). Still, even when using statistical method today, cluster analysis is not an automatic task, but an iterative process of knowledge discovery that involves trial and failure. In very general terms we can say that clusters show internal cohesion and external isolation (ibid.). Berghauser Pont and Haupt (2010) did not use statistical methods either to arrive at their building typology based on the four built density variables. Instead of statistical methods, they present a building typology based on the visual interpretation of the scatter graph ‘Spacemate’. This might be the reason the last variable, network density (N), was not included in the presented typology and the urban fabric typology, including N, is merely presented as concept. In this paper, we will test the use of two distinct statistical methods for clustering to develop a typology including N: hierarchical clustering and centroid-based clustering. The use of two methods is partly driven by the fact that centroid-based clustering is better suited for

large datasets (>200 observations) and therefore fits our purpose to also use the method to study whole cities while hierarchical clustering has the practical advantage of visualising results using a dendrogram that allows to follow all steps in the clustering procedure. We will discuss this more extensively in the method section.

Methodology

As discussed above, two distinct statistical methods for clustering are used for his paper which will be discussed in order below.

Hierarchical clustering

In hierarchical clustering analysis, clusters are defined based on the distance between the data points. The core idea is that objects, in this case neighbourhoods, are more related to nearby objects, in terms of data, than to objects farther away (Wilmink and Uytterschaut, 1984). In the first step, objects are grouped that are nearest, that is, neighbourhoods with most similarity in terms of FSI, GSI, OSR, L and N; the second step joins new objects to these groups formed in the first step, again based on distance, etc. In the last step all objects will be grouped in one cluster. A so called dendrogram provides an overview of this extensive hierarchy of clusters that merge with each other in each step. The y-axis marks the distance at which the clusters merge, while the objects (neighbourhoods) are placed along the x-axis. It is possibly this representation (i.e. dendrogram) that makes this method very popular as “humans can assimilate information from pictures better than from tables of numbers” (Dubes and Jain 1976 quoted in Wilmink and Uytterschaut, 1984). A drawback from hierarchical methods is that these maximize the clustering criterion in a suboptimal way where the best fusion of objects in the first step results in so called N-1 clusters, followed by the fusion into N-2, N-3 and N-4 clusters which does not necessarily lead to the best fusion of N objects into for instance 4 clusters directly. Such grouping of all samples in a pre-defined amount of clusters

is what centroid-based clustering is doing and will be discussed below. Another limitation of this method is that it cannot handle large datasets.

Centroid-based clustering

Centroid-based clustering, also called k-group clustering, groups objects in a number of clusters based on the distance of the data points to the clusters centre (i.e. centroids). The k indicates that the algorithm can identify any number of clusters which in most algorithms needs to be specified in advance. This is considered to be one of the biggest drawbacks of these algorithms, because it can be problematic to know how many clusters a particular set of data preferably is divided into. The analysis usually results in a clustering where each object (in our case neighbourhoods) belongs to a cluster or not (i.e. hard clustering), but some methods also allow a clustering where each object belongs to each cluster to a certain degree (i.e. soft clustering or fuzzy clustering). The latter will be used in this paper to be able to see whether neighbourhoods end up in between two or more clusters. In the first step every data point is assigned a random membership grade for each cluster, then an iterative process begins: (1) the membership grades are used to calculate cluster centres, (2) the cluster centres are used to update the membership grades and (3) the sum of the distance between every data point and every cluster centre with the

membership grades as weights is calculated. The closer a data point is to a centre of a cluster the larger membership grade it is assigned to the cluster. The iteration stops when the change in sum of the distance from one iteration to the next is less than ϵ or when n iterations has been done, where ϵ and n are set in advance.

As mentioned, the number of clusters has to be chosen by the user, in advance. In our case, the clusters should partly correspond to the building types as presented by Berghauer Pont and Haupt (2010) with the addition of N. The choice of the number of clusters is therefore equal to the number of building types presented by Berghauer Pont and Haupt (ibid., figure 21, p 191) where more clusters are added in case of a large variation in N within these building types. The amount of building types was six of which three show a high variation in N, resulting in nine clusters. Further, since we also have an idea about where a specific building type is located in the Spacemate scatter graph, its centres can be estimated which are then used as set starting points for the cluster centres (Table 1). This means that from the iteration process described above (1) is not needed and only (2) and (3) are iterated over.

In-data and software

In both the centroid-based clustering and the hierarchical clustering, we use GSI, L and N as in-data. We use these three variables

Typology names (Berghauer Pont and Haupt, 2010)		Starting points centroid-based clustering			
		FSI	L	GSI	N
1	low-rise point	0,17	1,6	0,11	0,011
2	low-rise strip	0,44	2,4	0,18	0,015
3	low-rise strip	0,47	2,3	0,21	0,027
6	low-rise block	0,78	2,3	0,35	0,027
7	mid-rise strip	0,92	4,8	0,19	0,011
8	mid-rise strip	0,90	4,8	0,19	0,020
4	mid-rise block	1,66	4,8	0,34	0,014
5	mid-rise block	1,70	4,5	0,38	0,022
9	high-rise point/strip	1,35	10,8	0,12	0,010

Table 1. Starting point for the centroid-based clustering for nine clusters (based on the results presented in Berghauer Pont and Haupt, 2010, figure 21, p. 191).

and not all five variables discussed earlier to ensure that the variables used for clustering have the lowest possible correlation, which is desirable because co-linearity is known to bias clustering results. Of the four built density variables FSI, GSI, OSR and L, the variables that correlate least are GSI and L. FSI is the product of GSI and L and OSR is derived from FSI and GSI using the formula $(1-GSI)/FSI$. N is independent from the other four density variables and can thus be used without a risk of co-linearity.

For the hierarchical clustering, IBM SPSS Statistics 22 is used and the results will be discussed using a dendrogram complemented with figure-ground maps of a selection of neighbourhoods. For the centroid-based clustering the Fuzzy c-means clustering method in Matlab is used. Besides the cluster membership of each neighbourhood, also the degree of belonging will be presented for the clustering using Fuzzy c-means clustering. In both cases, the resulting clustering will be shown in scatter graphs for the three planes of the three-dimensional Spacematrix showing FSI(GSI), FSI(N) and N(GSI).

Study area

The same 111 neighbourhoods (i.e. observations) as were used in the book 'Space, Density and Urban Form' (Berghauer Pont and Haupt, 2010, p. 125-168) are used except for the medieval areas in Barcelona that because of their extreme high values both in terms of FSI and GSI will anyhow form a separate cluster. These values are statistically speaking outliers and can disturb the clustering. The scale level of analysis is what is called the urban fabric which is similar to the scale and definitions used for a plan unit, as described by Conzen, and the tessuto, as used by Caniggia (Moudon, 1994). The urban fabric consists of a collection of islands, as well as the network that surrounds these islands. Boundaries are drawn in the middle of the streets. In circumstances where there is no street, the boundaries of the fabric are set by the lot boundaries.

Results: Urban fabric types Hierarchical clustering

The results of the hierarchical clustering are presented in a dendrogram (see figure 2) showing that, when the neighbourhoods are divided in only two clusters, the first (cluster I) represent high-rise building types and consists of three sub-clusters (6, 14 and 11) of which a few neighbourhoods are shown as figure-ground maps in figure 2. All other neighbourhoods end up in the second cluster (II). A further division in three clusters results in a subdivision of cluster II into two separate clusters where IIa represent the most dense building types of all neighbourhoods used in the analysis and is composed of two sub-cluster (2 and 8) that are shown as figure-ground maps in figure 2. Cluster IIb groups all other neighbourhoods and is still too crude to distinguish between neighbourhoods with different morphological characteristics such as villa-areas and areas with mid-rise slab buildings typical for the modernist international style propagated by CIAM (Berghauer Pont and Haupt, 2010). We continue increasing the amount of clusters until we arrived at a convincing set of clusters and found this to be 14 clusters. The dendrogram in figure 3 shows this hierarchy of grouping neighbourhoods which is so typical for this clustering method with the numbers representing the 14 clusters. Further, the figure-ground maps show convincingly that this statistical method is able to distinguish neighbourhoods with different morphological characteristics, both in terms of building type and N.

The division in 3 clusters grouped the neighbourhoods with similarities in building type (cluster IIa), but still a clear difference in street patterns and size of the islands (i.e. blocks). The subdivision in 14 clusters successfully distinguishes even this important characteristic: the size of the islands in cluster 2 are twice the size of the sizes of the islands in cluster 8 (150m respective 80m in square)³. The same can be seen for the high-rise building types where the subdivision in three clusters (6, 14 and 11) is due to the variations in the grain of the street layout or size of the islands with the smallest islands in cluster 11 and the largest in cluster 14.

The scatter graphs in figure 3 show the clustering in the three dimensional Spacematrix and the three planes of this graph: FSI(GSI), N(GSI) and FSI(N) respectively. In the FSI(GSI) scatter graph it is clear that the method successfully distinguishes the mid-rise block types (cluster 2 and 8) and high-rise types (cluster 6, 11 and 14). We can also clearly see that the differentiation in network density in the N(GSI) scatter graph which is the reason they end up in different clusters. The two types do not differ when it comes to their GSI values, but differ greatly when it comes to N. Recall the figure-ground maps of figure 2 discussed earlier. The same can be seen for the three high-rise clusters. Three other clusters (9, 10 and 12, in blue points in the graphs) are worth mentioning. These are all of the low-rise type, but differ greatly when the size of the islands (or N) are considered. Cluster 9 has very large islands of 240m in square and cluster 10 and 12 have islands of only 60m in square. Despite the similarities in L,GSI and FSI, they differ in N and therefore form separate clusters.

Centroid-based clustering

The centroid-based clustering with nine clusters and set starting points for the cluster centres resulted in a typology that shows similarities to the building types as discussed by Berghauser Pont and Haupt (2010). In figure 4 both the scatter graphs are shown and an overview of the degree of belonging. In cases where neighbourhoods belong less than 30% to their “main” cluster, the dots in the graphs are marked with a cross and in the table the median degree of belonging of each cluster is given as well as the percentage of neighbourhoods that fit the 30% threshold. Cluster 4 has the lowest levels: the median degree of belonging is only 34% and one third of the neighbourhoods have a level less than 30%. In comparison to the other clusters this is low which means that the neighbourhoods within this cluster are of ambiguous character, that is, they belong to more than one cluster. Berghauser Pont and Haupt (2010) discuss this as hybrid types and this is actually natural for any kind of

categorization; there will always be objects that are more or less positioned just in between. However, it can also be an indication of wrong starting setting, which is the amount of clusters and the chosen values for the centroids (i.e centres of the clusters). There can be reason to test clustering with more clusters where some of these in-between neighbourhoods may be defined as new types or other centroid values that could increase the overall degree of belonging. Further, when looking more closely at two of the clusters with low degree of belonging (cluster 4 and 5), we have to conclude that the centroid-based method with the starting settings, is not as successful as the hierarchical method in distinguishing variation in N. Although it is clear that cluster 4 captures the larger blocks, some of the neighbourhoods such as Eixample (see figure 4) with large blocks, ends up in cluster 5 (belonging 24%). Two other neighbourhoods with large blocks are Grachtengordel and Hackesche Höfe both belonging mostly to cluster 4 (31% resp. 28%) and could as well have ended up in cluster 5 as the belonging to that cluster is almost of similar dignity. One reading is that these neighbourhoods are hybrid types (in between cluster 4 and 5), but we argue that more accurate is to add a cluster from nine to ten. This discussion actually confirms one of the weak points of this clustering methods as we discussed earlier, which is the fact that the amount of clusters has to be defined in advance.

Conclusion and discussion

The results of the two methods demonstrate the strong and weak points of the respective methods where the process of clustering using the hierarchical clustering is easier to follow and results in convincing clusters of different morphological types. One might argue that we need rather many clusters (14 clusters), but that is not necessarily negative. The centroid-based method with nine clusters is especially powerful in identifying the in-between types, which is mostly thanks to the application of soft (i.e. fuzzy) clustering. What is further interesting with the use of this fuzzy method is that we

can speak of archetypes where the degree of belonging is high. The centroid-based-method is thus not only able to cluster neighborhoods, but also show hybrid or in-between types and archetypes. The hierarchical method, however, is more convincing in distinguishing neighbourhoods with similarities in building type, but differences in street layout and block size.

A qualitative comparison of the two clustering methods is shown in figure 5 using one figure-ground map as representative of each cluster. On the left side the nine centroid-based clusters are shown; on the right the corresponding hierarchical clusters. Three clusters using the centroid-based method are identical to the hierarchical clustering results (i.e. centroid based clusters 2, 7 and 9). Centroid cluster 8, though, does not fit the results of the hierarchical clustering at all and different hierarchical clusters fall within one and the same centroid cluster. Also centroid cluster 3 shows a large variety of hierarchical clusters. Here, however, the results partly overlap and the reason more hierarchical clusters are part of the centroid cluster might be the amount of clusters to start with, we started with nine for the centroid-based method and the hierarchical clustering resulted in 14 clusters. The same can be said about the centroid clusters 1, 4 and 5 where the centroid clusters partly overlap with the hierarchical clusters. For these three, we can see that the centroid-based method is not very successful in differentiating between street layouts, or in other words, block sizes. We can thus conclude that the hierarchical clustering is more convincing in developing an urban fabric typology where network density and block size is central. However, the hierarchical method cannot handle large dataset and it would therefore be interesting to test whether we can use the results of the hierarchical clustering and use the centre means of each cluster as input for a centroid-based clustering. In this

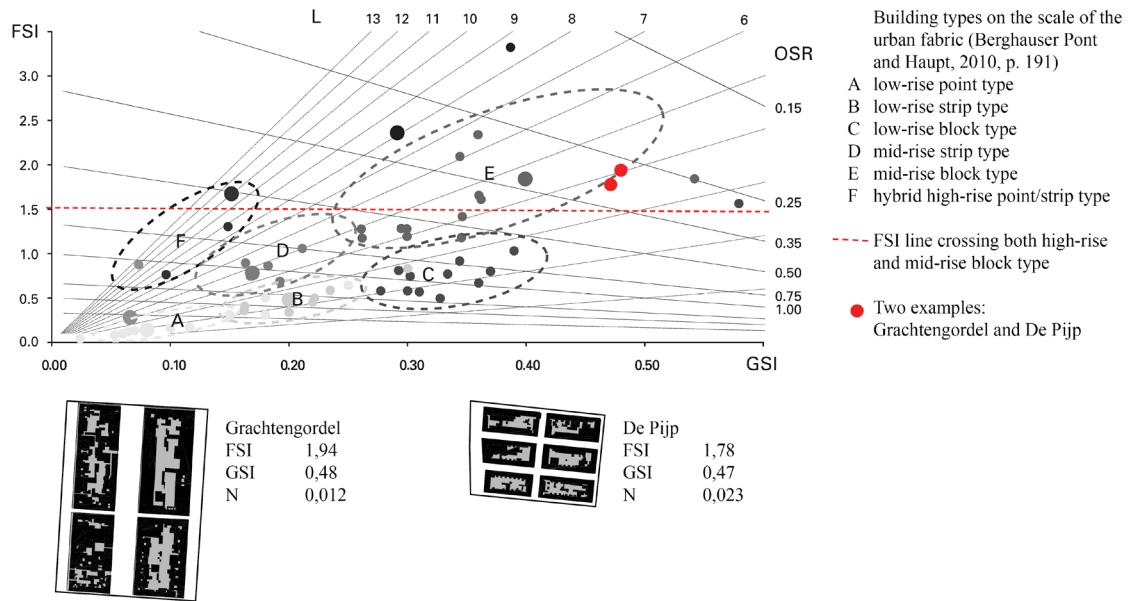


Figure 1. Building typology as presented in Berghauer Pont and Haupt (2010), figure 21, p. 191 and two neighbourhoods in Amsterdam (Grachtengordel and De Pijp) of similar building type, but differences in network density (N).

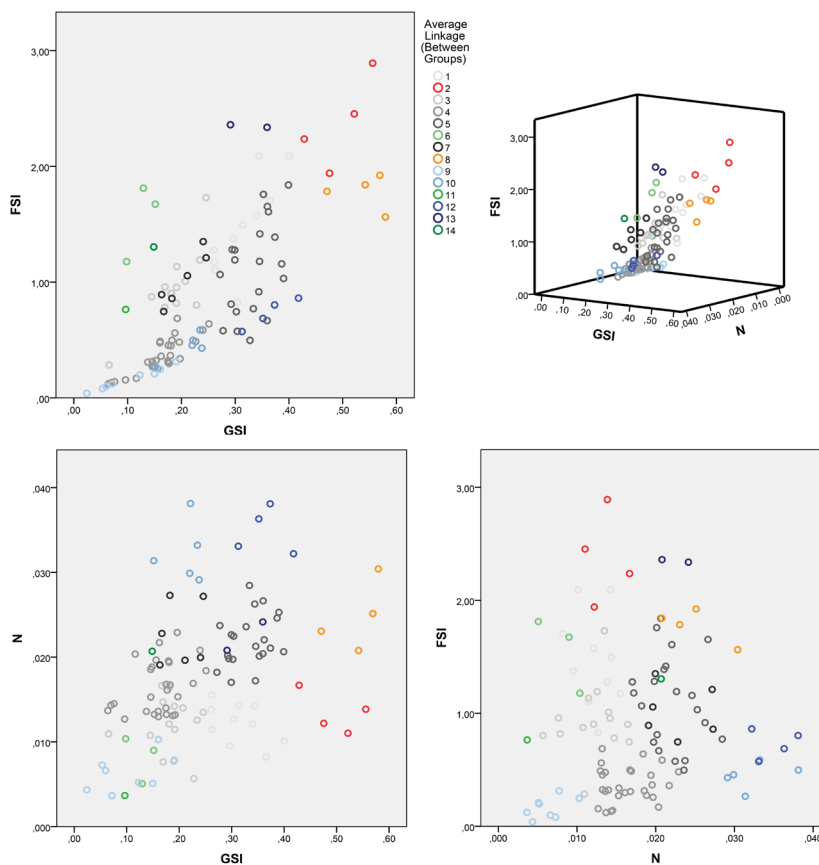


Figure 3. Result of the hierarchical clustering presented in scatter graphs: FSI(GSI), FSI(N), N(GSI) and the three-dimensional scatter graph.

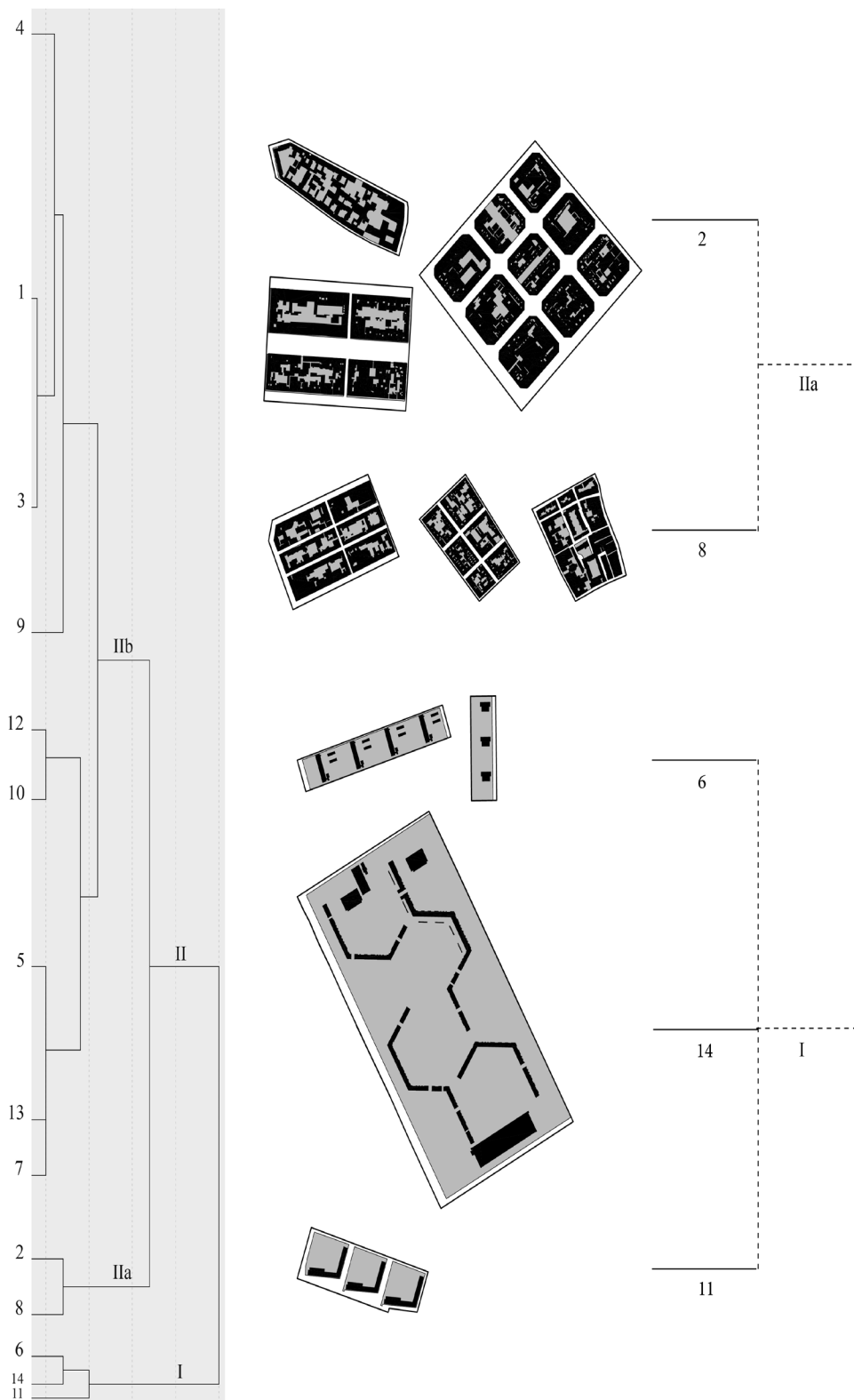


Figure 2. Result of the hierarchical clustering presented in a dendrogram and selection of figure-ground maps to illustrate the hierarchical clustering process.

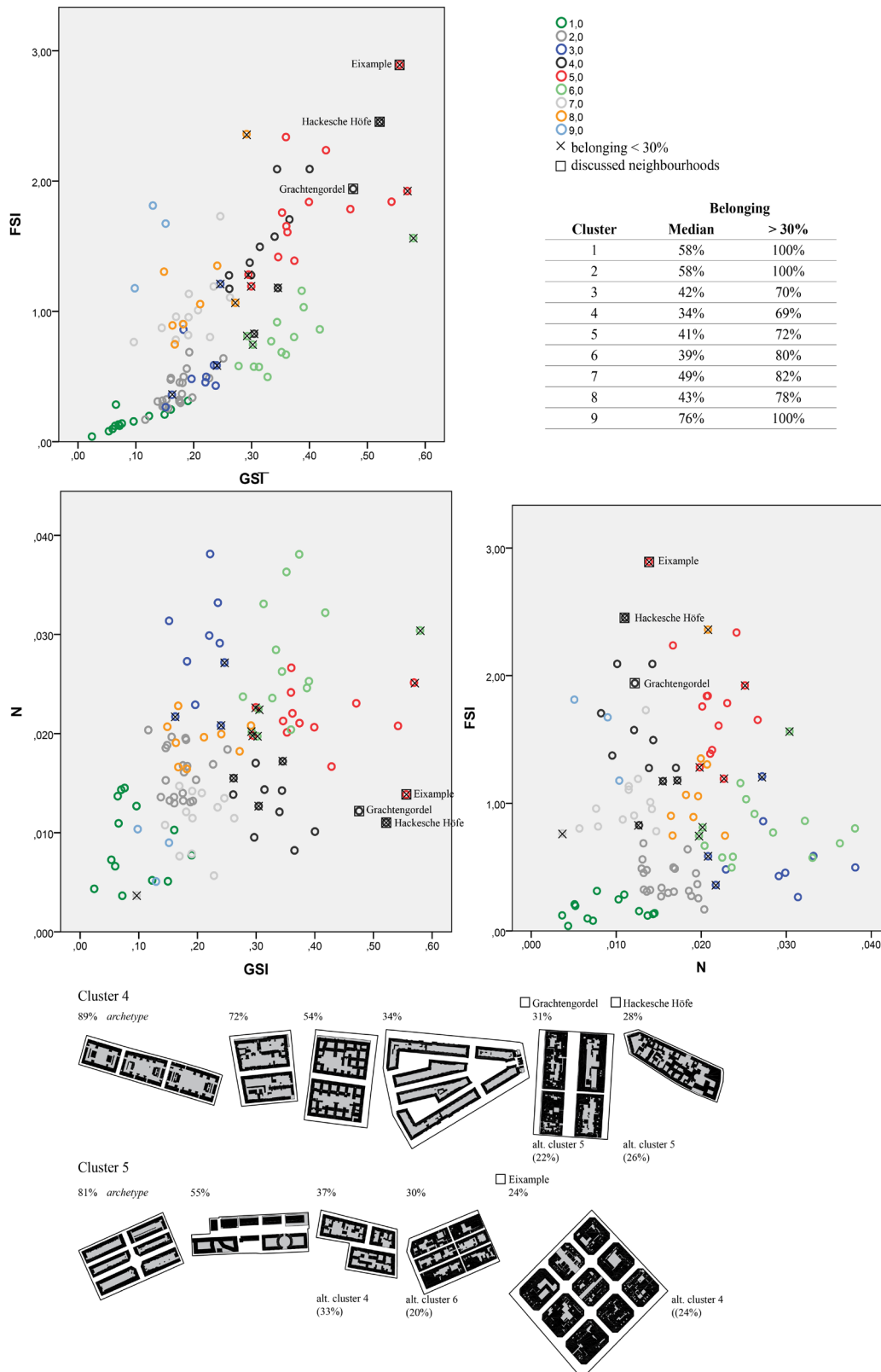


Figure 4. Result of the centroid-based clustering presented in scatter graphs: FSI(GSI), FSI(N) and N(GSI) and table with the degree of belonging of the neighbourhoods to the clusters and figure-ground maps of cluster 4 and 5.

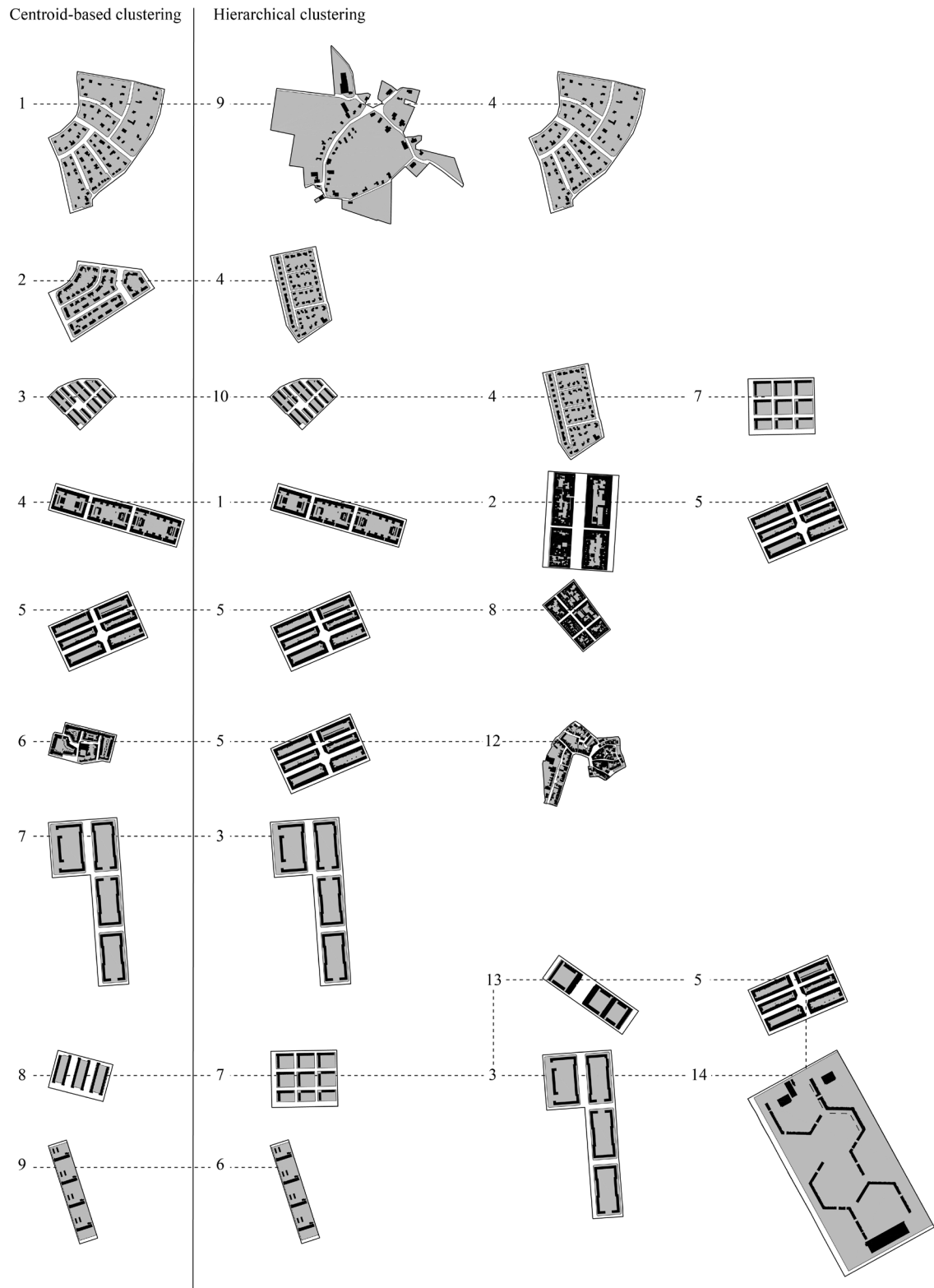


Figure 5. Comparison of the two clustering methods (hierarchical clustering and centroid-based clustering) using a selection of figure-ground maps.

manner, the strong sides of both methods will be combined. Further study has to confirm this as well as the success of applying this on the scale of whole cities.

Notes

1. A comparable term for FSI, first used in New York City's Zoning ordinance in 1940, is the Floor to Area Ratio, FAR (Noble, et al., 1993).
2. The ancient Romans called their urban blocks *insulae*, or island, reflecting the topological containment of buildings and land parcels (lots) within a continuum of public space primarily constituted by the system of public streets (Panerai, 2002, p. 11–14).
3. Median N for cluster 2 is 0,013 and for cluster 8 it is 0,024 and block size is calculated as $2/N$ (see for formulas Berghauser Pont and Haupt, 2010, p. 110).

References

- Berghauser Pont, M., Stavroulaki, G., Gil, J., Marcus, L., Serra, M., Hausleitner, B., Olsson, J., Abshirini, E., Dhanani, A. (2017). 'Quantitative comparison of cities: Distribution of street and building types based on density and centrality measures'. Proceedings XI space syntax conference, Lisbon.
- Berghauser Pont, M. and Haupt, P. (2010), *Spacematrix. Space, density and urban form* (Rotterdam: NAI Publishers).
- Berghauser Pont, M., and Marcus, L. (2014). 'Innovations in measuring density. From area and location density to accessible and perceived density', *Nordic Journal of Architectural Research*. Issue2, pp.11-30
- Conzen, M.R.G. (1960). *Alnwick, Northumberland: A Study in Town-Plan Analysis*. London: Philip (in series: The Institute of British Geographers, publication 27).
- Heeling, J., H. Meyer and J. Westrik, (2002). *Het ontwerp van de stadsplattegrond*. Amsterdam: SUN.
- Marshall, S. (2005). *Streets & Patterns*. Oxon: Spon Press.
- Martin, L. and March L. (1972). *Urban Space and Structures*. Cambridge: Cambridge University Press.
- Moudon, A.V., 1994. Getting to know the built landscape: Typomorphology. In K. Franck and L.H:L. Schneekloth, eds. 1994. *Ordering space. Types in architecture and design*. New York: Routledge. pp 289–311.
- Panerai, P. (2002), 'De schaal van het bouwblok', in: S. Komossa et al., *Atlas van het Hollandse bouwblok*. Bussum: Uitgeverij THOTH.
- Sola-Morales, M. de (1978). 'Towards a Definition: Analysis of Urban Growth in the Nineteenth Century', *Lotus*, 19 June 1978, 28-36.
- Steadman, P. (2013). 'Density and built form: integrating «Spacemate» with the work of Martin and March'. *Environment and Planning B: Planning and Design*, 40, pp. 341–358.
- Whitehand, J. (2001). 'British urban morphology: the Conzenian tradition'. *Urban Morphology*. 5(2), pp. 103-109.
- Wilmink, F.W. and Uytterschaut, H.T. (1984). 'Cluster analysis, history, theory and applications', in G. N. van Vark and W. W. Howells (eds.), *Multivariate Statistical Methods in Physical Anthropology*, 135-175. D. Reidel Publishing Company.