

# Visual Analytics of Urban Environments using High-Resolution Geographic Data

Peter Bak<sup>1</sup>, Itzhak Omer<sup>2</sup> and Tobias Schreck<sup>3</sup>

<sup>1</sup> Visual Analytics Group, Dept. of Computer and Information Science., University of Konstanz, Germany

<sup>2</sup> Urban Space Analysis Laboratory, Dept. of Geography and Human Environment, Tel Aviv University, Israel

<sup>3</sup> Interactive Graphics Systems Group, Technische Universität Darmstadt, Germany

## Abstract

High-resolution urban data at house level are essential for understanding the relationship between objects of the urban built environment (e.g. streets, housing types, public resources and open spaces). However, it is rather difficult to analyze such data due to the huge amount of urban objects, their multidimensional character and the complex spatial relation between them. In this paper we propose a methodology for assessing the spatial relation between geo-referenced urban environmental variables, in order to identify typical or significant spatial configurations as well as to characterize their geographical distribution. Configuration in this sense refers to the unique combination of different urban environmental variables. We structure the analytic process by defining spatial configurations, multidimensional clustering of the individual configurations, and identifying emerging patterns of interesting configurations. This process is based on the tight combination of interactive visualization methods with automatic analysis techniques. We demonstrate the usefulness of the proposed methods and methodology in an application example on the relation between street network topology and distribution of land uses in a city.

## 1 Introduction

"All geographic data leaves its users, to some extent, uncertain about the nature of the real world" (Goodchild, 2005). This situation, which often stems from data resolution constraints and from their multidimensional character regarding space, time and objects, has affected socio-geographic research of the urban environment.

Until recently, urban environment research was limited to the use of large-scale aggregate data based on the level of administrative areas. The basic problem of using aggregated spatial data for geographical analysis stems from the fact that the distribution of objects within the areas is unknown. As a result, the aggregated data are not sufficient to capture the micro-scale situations, in which the main dimensions of urban environment – the built-up environment's properties, the individuals' socio-demographic properties and the individuals' behavior and perception – come together.

This situation has nonetheless changed recently due to improvements in GIS (Geographic Information System) technology and the construction of new geographic databases. Today it is possible to obtain geo-referenced high-resolution data (i.e. there is a link between the attribute of the object and its geographic location) on different types of urban objects in urban locations: (1) daily movement data, e.g., positioning data collected by global positioning system (GPS) and location based services (LBS) technology to surrogate daily movement data from mobile usage patterns (e.g. Ratti et al., 2006); (2) distribution of built-up environmental objects e.g., road networks, or housing, and (3) functional and socio-demographic objects, e.g., house-level socio-demographic and land-uses data.

The increasing availability of geographic data at high resolution and in good quality actually motivates the investigation of the relation between these different types of urban objects at the same geographic scale, i.e. the house level scale, of diminishing gaps between data that represent different types of urban objects. This ability

can be essential for identifying and understanding the variety of socio-demographic phenomena that incorporate different urban objects at different spatial dimensions. For example, it is possible to investigate how building types or the location of urban services are correlated with the spatial distribution of the populations' socio-demographic attributes. It is also possible to integrate high-resolution built-up and socio-demographic data with empirical data (i.e. data collected during interviews) to better understand the individuals' preference features – for instance, the choice of urban parks and commercial areas and their mode of transportation.

Nevertheless, use of high-resolution geographic data comes at a price. Despite their potential, it is rather difficult to use these data in research. Unlike the situation of aggregate data at the level of geographic areas, high resolution geo-referenced data at house level have no defined geographical boundaries and therefore, the main challenges of research are to identify patterns of interest based on huge amounts of urban objects with respect to their attributes and the complex spatial relation between them.

In this paper, we apply our methodology for assessing the spatial relation between high resolution geo-referenced urban land-uses (e.g. residence buildings, commercial open spaces as well as infrastructure objects such as street network) and the topological attributes of the urban street network, in order to identify typical or unique land-use spatial configurations and to characterize their geographical distribution. Configuration in this sense refers to the combination of the two geo-spatial attributes.

The proposed methodology uses a Visual Analytics framework, which combines interactive visual information displays with automatic data analysis techniques. The framework consists of first displaying geographic information, on which users can interactively segment the data into spatial configurations of interest. These are then clustered to reveal frequent or significant groups of configurations. In addition, appropriately designed cluster visualization is interactively linked with the original map display. Thus, the system is designed to support the user in understanding the joint properties of geospatial and multivariate data.

The structure of the paper is as follows: In Section 2, we provide a review of the related work from the geographic and visual analytics perspective. In Section 3, we introduce our methodology and an exemplary implementation. In Section 4, we apply our framework on a high-resolution real world dataset of an urban environment, and demonstrate the flexibility it offers to analyze the data for various interesting correlations. In Section 5, we discuss our approach, the obtained results, and outline future work in the area.

## 2 Related Work

### 2.1 Geographic Data Analysis for Urban Environments

Visual and analytical comparison between spatial distributions of objects and attributes within GIS framework is an essential tool for understanding and explaining geographic phenomena in urban areas. To resolve the problems entailed by high-resolution data (i.e. neighborhoods definition and cartographic constraints resulting from mapping high resolution data over large area) previous studies suggested local geo-statistics methods. One of the main methods for analyzing and presenting high-resolution data is to use *local indices of spatial association* (LISA). These indices are based on the comparison of the characteristic of a given spatially located object and its neighbors (Anselin, 1995; Benenson & Omer, 2003). Applying these measures helps the observer identify spatial variance and small clusters in the spatial distribution of socio-demographic attributes in urban areas (Talen & Anselin, 1998). A production of thematic maps at different aggregation scales enables also to identify the natural geographic scale for the investigated phenomena. Previous studies suggested such local geo-statistics measures and aggregations for analyzing and overcoming the cartographic constraints of high-resolution data for assessing accessibility to urban services (e.g. Talen & Anselin, 1998; Omer, 2005), residential segregation (e.g. Omer & Benenson, 2002; Wong, 2003) etc.

### 2.2 Analysis the Effect of Street Network Topology

Much evidence has been collected indicating that the topological characteristics of a street network have the potential to affect the spatial distribution of activities and land-uses in the city. Studies have found that these topological properties of individual streets are significantly correlated to the spatial distribution of retail and services (Porta et. al., 2006) and human movement rates (e.g. Hiller et. al., 1993; Jiang, 2007). However, we have no sufficient knowledge on these relations i.e. why certain topological properties are more appropriate than others for predicting human activities in the city.

### 2.3 Visual Analytics of High-Dimensional Data

Owing to our data transformation applied (see Section 3), this work relates also to the wider area of visual analytics in high-dimensional data sets. Work in this area is concerned with finding appropriate visual representations for data sets, which by their dimensionality exceed the number of direct visual variables available (e.g., position, color, and shape (Ware, 2004)). *Mapping* approaches such as Parallel Coordinates (Inselberg and Dimsdale, 1990), Iconic Displays (Everitt and Nicholls, 1975), Dimensional Stacking (LeBlanc et al, 1990), or Scatter Plot Matrices (Wilkinson et al, 2005) define certain mappings from multiple dimensions to visual variables and geometric arrangements that represent the properties of all dimensions simultaneously. *Interaction techniques* address the problem of high-dimensional data analysis by allowing efficient navigation through the space of low-dimensional projections. Examples for these kinds of approaches can be found in (Wilkinson et al, 2005, and Elmqvist et al, 2008). *Dimension reduction* approaches operate by first reducing the number of dimensions before visualization takes place. Prominent dimensionality reduction techniques include Principal Components Analysis (Jolliffe, 2002) or Multidimensional Scaling (Cox and Cox, 2001). The general aim of dimensionality reduction is to capture as much information as possible in a limited number of dimensions. However, the reduced dimensions often are a linear or non-linear combination of all input dimensions, and therefore not straightforward to interpret by the user.

### 2.4 The Self-Organizing Map Approach in the Geographic Context

The SOM algorithm (Kohonen, 2001) is a technique which combines dimensionality and data reduction, and implicitly yields a mapping of data elements to position. The algorithm is especially suited for visualization of its output (Vesanto, 1999). It has been successfully applied to many data analysis problems including in geotemporal data (Guo et al, 2006), textual data (Honkela et al, 1997), and financial data (Deboeck and Kohonen, 1998). The SOM approach has also been leveraged in Geospatial data analysis before. In (Spielmann, 2008) an interactive system linking a standard land-covering map with a SOM-based clustering of demographic records was introduced. It allowed the user to select demographic records and simultaneously display them on the SOM (allowing inspection of multivariate properties) and on the map (allowing inspection of spatial distribution). The system was shown to be useful for joint analysis of multivariate and geospatial data properties. In (Bacao, 2005), the basic SOM algorithm was extended by a learning constraint that considers for each data record also its geospatial coordinates. This algorithm produces clusters, which also partially reflect the geospatial position of data records in the SOM network, a property which SOM usually does not consider. Our approach relates to both works (Spielmann, 2008; Bacao, 2005) in that we apply the SOM algorithm in a joint geospatial and multivariate data analysis. In extension to (Spielmann, 2008), we provide an improved SOM display, which allows perception of the distribution of multivariate records already on the SOM display, reducing the need for additional detail views to understand the multivariate data properties.

## 3 Analytic Methodology

Analysis of high resolution urban environments requires methods to define units of investigation (for example, buildings, streets, etc), and allows analysts to interactively extract configurations and patterns of interest. Here we propose a methodology that enables analysts to conduct their investigations exactly for this purpose based on a modular pipeline, in which information visualization and guided analysis constitute the core part. Consequently, the proposed methodology can be viewed as a modular framework, in which techniques can be chosen by their appropriateness for the research tasks and data.

The pipeline of analysis is triggered by analytic questions and suitable data sources. Here the first step is to find a visual mapping that allows analysts to view and understand the distribution of the data, which is a highly complex task since high resolution data in the geographic domain is large and heterogeneous. The second step requires analysts to define and segment units of investigation, in which configurations and pattern of interest may be possible. However, a high level of refinement through iterations is required, once the results are obtained. These steps are followed by automatic analysis techniques that can extract frequent patterns and significant configuration that reflect the analysts' expectations. The results of the automatic techniques have to be made accessible in an interactive visual way, in order to allow reasoning and refinement of previous decisions that are required to sharpen and finalize the results. As such, the pipeline and methodological framework are highly iterative and combine visual and automatic analysis techniques. The following diagram (Figure 1) provides a high level overview of the described pipeline.

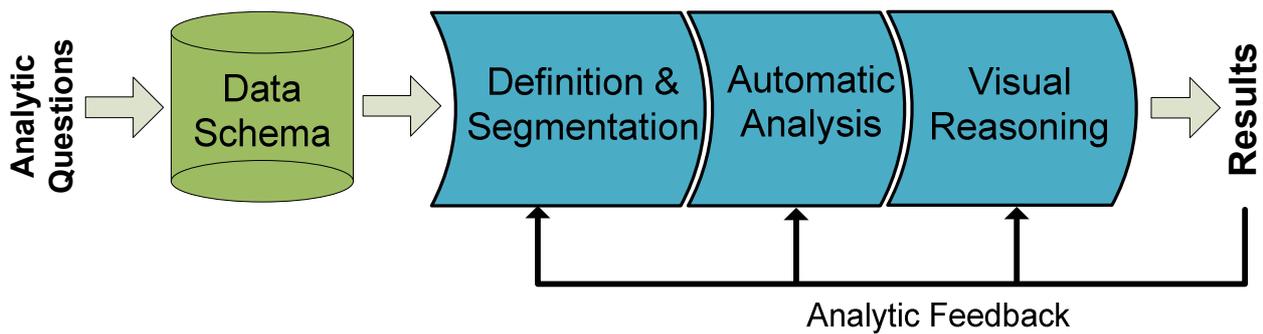


Figure 1. Visual Analytic Pipeline showing the interleaving stages of the analytic process.

In the following, we describe a concrete instantiation of the presented methodology by describing the key analytic question of interest and the database schemes available in a selected domain. We then present the details of the applied analysis procedures, visual mappings, and user interactions supported by our system.

### 3.1 Analytic Questions

Within the developed methodological framework, we aim to investigate the relation between the topological structure of urban street network and the spatial distribution of urban land-use in the city of Raanana, Israel. We investigate whether the spatial distribution of topological properties in a given street network correlates with the spatial distribution of land-uses. Due to the multidimensional nature of the proposed approach it can be used to identify typical land-uses spatial configurations and to investigate how they are influenced by the topological properties of urban street network. This means to investigate the effect of topological structure, not only on the spatial distribution of one land-use, but also on the formation of land-use spatial configurations, with respect to their geographical location. Accordingly, the analytical questions in this research are:

- Are there typical and significant correlations between the topological properties of streets and land-uses' spatial configurations?
- What are the geographic patterns of identified typical land-use configurations?
- Which topological properties are more significant than others for formation of land-use spatial configurations?

### 3.2 Data Schema of Concern

To illustrate the potential of the proposed methodology, we conducted a detailed investigation of one city's land-use and street network. The data obtained for analysis was the spatial distribution of the land-use and the topology of the street network in the city of Raanana, Israel. We used two kinds of geographic data sets of Raanana. A street network data set (a total of 324 streets), and a land-use data set, at the level of individual buildings (a total of 8664 buildings). The source of the data is the 2002 Infrastructure Database of the Israeli Central Bureau of Statistics ([www.cbs.gov.il](http://www.cbs.gov.il)), which are organized within a GIS framework.

On the land-use side, the data sets specify in real-world coordinates the presence of urban infrastructure elements. These include public-service installations such as education facilities, recreational areas, medical attention facilities, or recreational areas. They also include industrial area data, telecommunication installations and so on. By their nature, the infrastructure elements have a spatial extension and are therefore encoded by polygonal descriptions of the covered area.

The topology of urban streets takes individual streets as nodes (vertices) and street intersections as edges of a connectivity graph. The graph forms a basis for structural analysis using the centrality measures (Jiang & Claramunt, 2004; Jiang & Omer, 2007) initially developed for the description of social networks (Freeman, 1979). A graph  $G(V,E)$  is defined as a pair of a finite set of vertices  $V = \{v_1, v_2, \dots, v_n\}$  and a finite set of edges  $E = \{v_i, v_j\}$ . Three centrality measures – degree, closeness, and betweenness – are used to describe the status of individual streets, in terms of which streets intersect with other streets. Degree indicates how many

other streets are connected directly to a particular street, a characteristic that reflects the level of a street's integration with its neighboring streets. In a graph, the degree is the number of nodes that link a given node. Formally, the degree centrality for a given street (node)  $v_i$  is defined by:

$$C_D(V_i) = \sum_{k=1}^n r(v_i, v_k)$$

$$r_{ik} = \begin{cases} 1: \text{if } V_i, V_j \in E \\ 0: \text{otherwise} \end{cases}$$

Where  $n$  is the total number of streets (nodes) within a street network (vertices of the graph  $G$ ).

Closeness indicates how close a street is to other streets by computing the shortest distances between every street node to every other street node, a feature that reflects how well a street is integrated within the network. Formally, the closeness measure is defined by:

$$C_c(V_i) = \frac{n-1}{\sum_{k=1}^n d(V_i, V_k)}$$

where  $d$  is the shortest (topological) distance between two given streets ( $v_i, v_k$ ) in the street network (graph).

Betweenness centrality indicates the extent to which a street is located between pairs of streets; as such, it directly reflects the intermediate location of the specific street in the entire street network. Accordingly, we define the betweenness centrality as follows:

$$C_B(V_i) = \sum_j \sum_k \frac{P_{jik}}{P_{jk}}$$

where  $P_{jk}$  denotes the number of shortest paths from  $j$  to  $k$  and  $P_{jik}$  is the number of shortest paths from  $j$  to  $k$  that pass through street  $i$ , so  $C_B$  is the proportion of shortest paths from  $i$  to  $j$  that pass through  $k$ .

Figure 2 shows the topology and land-use of the urban environment. Topology is mapped to color using a heat map having red colors for high and yellow colors for low centrality values. The land-use is mapped to a diverging color schema for six categories (educational institutions, public services, culture, commerce, industrial buildings and parks).

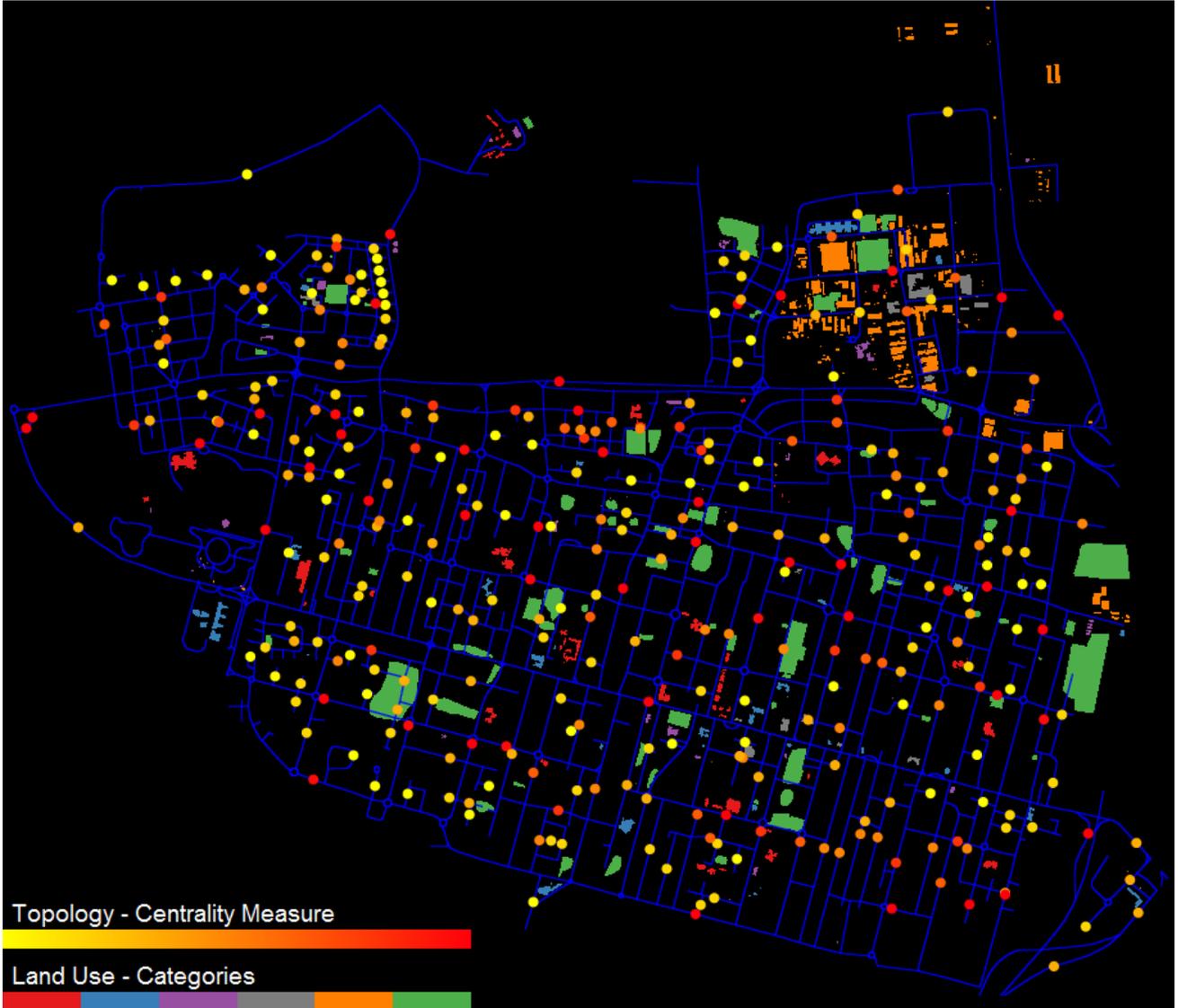


Figure 2. Map representation of the considered urban environment of the city of Raanana: Topology (degree centrality) on the points and land-use on the polygons of the image. Values of topology are mapped to a diverging colomap going from yellow (low values) to red (high values). Land-use types are mapped to a discrete color schema for educational institutions, public services, culture, commerce, industrial buildings and parks.

### 3.3 Segmentation and Definition

The data sets described are large and complex, requiring appropriate data segmentation to facilitate the analysis and visual representation. To facilitate the investigation, analysts have to define the combination of urban environmental variables of interest into spatial configurations. In order to find appropriate configuration units, we conduct a four stage process: Firstly, we structure the elements of our investigation by their spatial location and neighboring elements. In the particular case, the elements of our investigation are named streets described by the topological value. Secondly, we create a structure, in which each named street is described by its own topological value and those of the connected ones. Thirdly, we describe each street by the presence of infrastructural elements in its neighborhood. The overall description of each named street is then obtained as a (high-dimensional) vector of association frequencies for each infrastructure element type. Finally, we partition into configuration units consisting of created multidimensional feature vector for each named street. In the particular case, the partitioning referred only to represent each street individually. However, any other level of resolution is practicable for this step. A schematic representation of this process is shown in Figure 3.

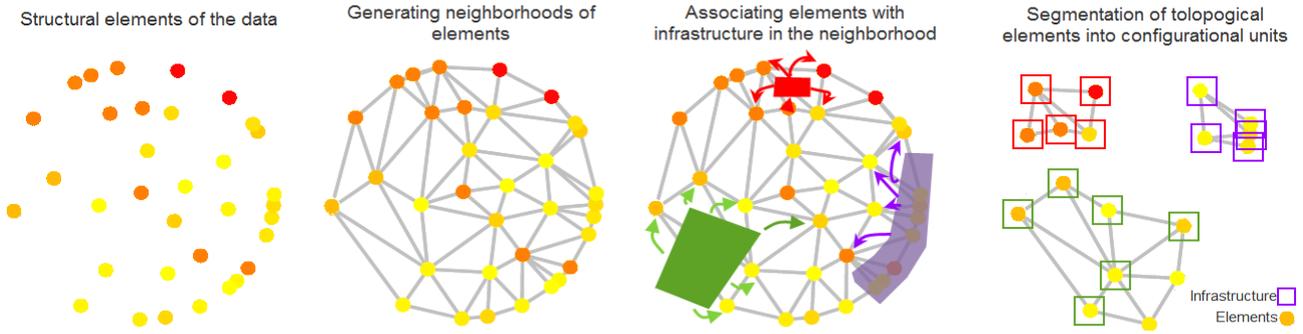


Figure 3. Stepwise generation of neighborhoods of relating elements and associating these with the neighboring infrastructural elements. As a result, each element is described by a multidimensional vector consisting of all its neighbors and surrounding infrastructural elements.

### 3.4 Analysis of Configuration Units

Having applied the above mentioned preprocessing, we obtain a large number of street descriptors, which represent the local spatial pattern of land-uses and infrastructure of a given city. In order to perform a correlation analysis between these local patterns and a selected overall/global target variable, like the topological structure of street network, we first conduct a cluster analysis of all named street descriptions.

We chose to use the SOM algorithm (Kohonen, 2001) for cluster analysis. It is a combined vector quantization and projection algorithm. It produces a network of reference (prototype) vectors from a set of input data vectors by means of a competitive learning process. During the learning process, which takes place after an appropriate initialization of reference vectors has been performed, input data is sequentially presented to the network. Then, the currently best matching (in the nearest neighbor sense) cluster prototype is determined, and this prototype together with a neighborhood of prototypes is then adjusted toward the presented input (cf. Figure 4). As a function of time, during learning the degree of adjustment of prototype vectors is reduced, and stable results are obtained. The SOM reference vectors represent clusters in the input data set and are typically modeled on a 2D grid. Practically, one important property observed on the SOM analysis output is that the arrangement of prototype vectors approximates resembles the topology of data vectors in input space.

We perform SOM cluster analysis based on the vector descriptions of the infrastructure descriptions associated on average with each neighborhood (see Section 4). For setting of SOM parameters, we rely on rule-of-thumb settings typically recommended (Kohonen et al, 1996).

As a result, we obtain a network of clusters describing prototypical distributions of infrastructure elements over named streets in our data sets.

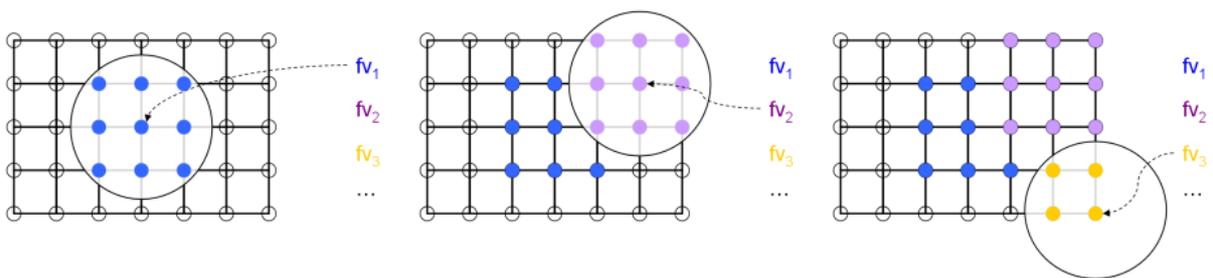


Figure 4. During the SOM learning process, sample input vectors are iteratively presented to the map, which is gradually adjusted to the presented input. The process yields set of cluster prototype vectors which are arranged on a regular grid that approximately represents the topology of the input data (Kohonen, 2001).

### 3.5 Visualization of Configuration Types

The SOM analysis yields an intermediate cluster result, which we visualize together with the quantitative information of the selected street network topology measure. We visualize land-use characteristics of each group (cluster) of named street patterns as a radial Parallel Coordinate Plot (Van Long, 2009). The basic idea is to map each land-use dimension to one axis emanating radially from the origin in an equally-spaced angular direction. Following the parallel coordinate approach, we connect the coordinate positions of each dimension

in the SOM prototype vector by straight, bold lines. A high-dimensional glyph results in form of a radar-like chart. On this chart, we also overlay the set of street samples represented by the SOM prototype vector, by means of opacity bands (Fua, 1999). The street network topology value which is to be correlated with the land use prototypes is mapped to the background color of the group diagrams. To this end, we again use the yellow-red color-map introduced in Figure 2. The final analytic view is constructed by drawing radar charts for each group of named streets yielded by the SOM analysis, using the grid structure of the SOM. The display allows to visually assessing several data aspects. The distribution of land-use over the different groups can be assessed by comparing the shape of the diagrams. The land-use properties can be correlated with the network topology properties by means of the background coloring. The opacity bands allow to assess the crispness of the groups in terms of the spread of group member attributes around the prototypes. Figure 5 illustrates the construction of one group glyph.

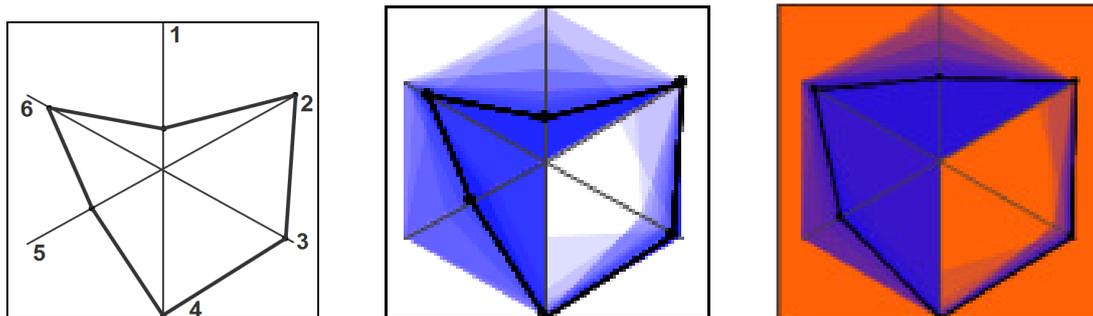


Figure 5. We show the properties of land-use types occurring in a group (cluster) of configuration types by means of a radar-plot (left image). The six radial axes refer the six land-use types (as described previously). Samples of the cluster are overlaid by opacity bands, indicating the distribution of represented sample data points (second image from left). The street topology measure is mapped to the background color of the image (third image from left) using the same color-map as for the data themselves.

### 3.6 Interaction Facilities

In order to facilitate user interaction, we implemented three major interaction techniques at different stages of the analytic pipeline (as shown the methodology section (Section 3) in Figure 1). In the data mapping stage users are able to select the relevant variables for their analysis. Users have to determine the “independent variable” of their analysis, which was the land-use in the current example. Users also have to select the “dependent variable” of the analysis, which was the topology of street networks in the presented example.

The main concern of the analysts is defining the neighborhoods of the elements, in the current example we used a Delaunay triangulation, which can obviously anytime be replaced by any other structural analysis technique or spatial clustering method. Choosing the right methods for this task is crucial and requires domain knowledge and optimal parameterization of the methods.

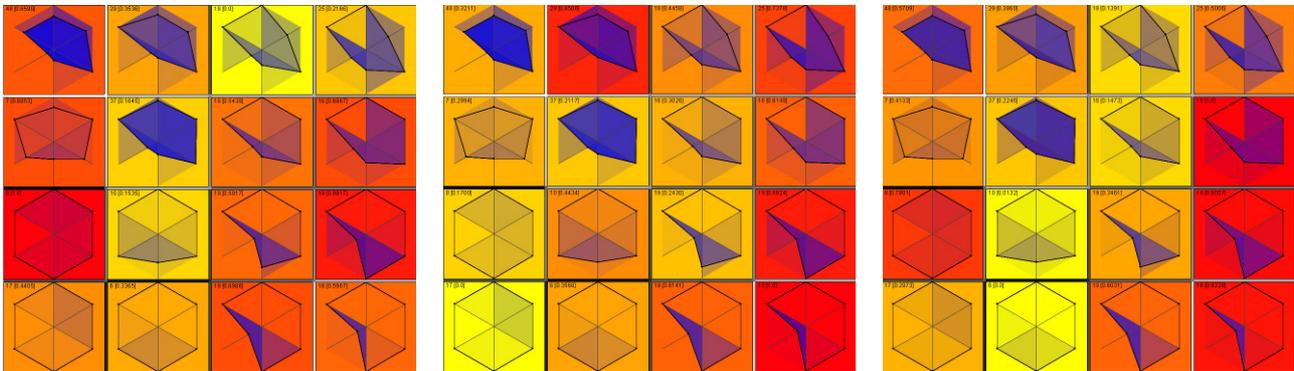
The number of automatic pattern extraction techniques is practically unlimited, if we take the combination of their parameters also into account. In the current example we successfully showed that Self-Organizing Maps are effective and useful for the current task and data. However, this method can be exchanged for different algorithms. Finding the best parameter settings is a highly iterative task, since only rule-of-thumb suggestions exist that require constant refinement. In order to facilitate this interaction, we provide a mouse-over function for the generated SOM-clusters which show the location of the cluster members in a geographic map. Consequently, the analysts can investigate the spatial features and distribution of the created clusters in addition to the distribution of the selected variables in the SOM-cluster itself. As a result, refinement of the properties of the clustering algorithm can be used to obtain smaller/larger units and higher/lower levels of distinction between cluster centers.

One of the required properties of information displays is that users can alter the color maps and their scaling in all visualizations. This feature is implemented at every stage of the analytic pipeline, in which visualization is involved. Currently we implemented a continuous heat-map color-map (from yellow to red) for the topological variable, and a discrete color map for the land-use representation. Users can also apply non-linear (square-root and logarithmic) scaling to the continuous color-mapping, in case this is required to compensate for skewed data distributions.

## 4 Results

In order to show the usefulness of the proposed methodology's instantiation, we present here its potential to identify typical and significant land-use spatial configurations, to locate land-use spatial configurations of interest and to characterize their geographical distribution. As described above, the investigation was conducted based on the relation between the spatial distribution of land-use and the topology of street network in the city of Raanana. The first action in implementing the methodology is a creation of clusters for different spatial land-use configurations using self-organizing maps. The resulting clusters are then colored with the three centrality measures of the street topology, as shown in Figure 6. Therefore, the resulting representations have the same cluster configurations, but different coloring for closeness, betweenness, and degree centrality measures. This possibility enables us to compare between these topological measures in term of their relation with each of the identified land-use configurations. In addition, the number of occurrences for each cluster is indicated in the upper left corner together with the average centrality measure in brackets.

Such presentation opens the possibility for identification of frequent and typical land-use spatial configurations with respect to the relation between land-use and street topology.



**Figure 6.** The correlation of Closeness (left), Betweenness (middle) and Degree (right) values with the spatial configurations of land-uses in Raanana.

In general, the spatial distribution of education (1), culture & leisure (3) and parks and open spaces (6) is relatively high in all clusters and has limited variation between the clusters. Medium spatial distribution of the public services (2) is visible with high variability of the centrality measures. High variations in the spatial distribution of commerce (4) and industry (5) are visible, which is assumed to be influenced by the centrality measures of street network topology.

Closeness reveals a positive correlation with the availability of commerce, as shown in columns 3 and 4 of Figure 6 (left). Low closeness values show low availability, and high closeness values a high availability of commerce. This means that the difference between these configurations located on the 'edges of closeness centrality' – the most accessible places versus inaccessible places – is in the availability of commerce. Interestingly, the lowest closeness value (Column 3 – Row 1) has low availability of commerce and also of public services and industry. The highest closeness values (Columns 3-4 – Rows 3-4) show high availability of commerce and public spaces with low availability of industry.

When comparing these findings with the levels of betweenness and degree, three interesting configurations can be extracted as perceivable in Figure 6:

*Configuration 1:* High availability of commerce, with low availability of industry (Columns 3 and 4 – Rows 3 and 4) having high values of closeness, showing also high values for betweenness and degree. Such a configuration is expected to be located in the center of the city. In the case of Raanana, it is located along the main street (Weitzman Street) of the city. This configuration is shown in the left image in Figure 7.

*Configuration 2:* High availability of commerce and industry (Columns 1 and 2 – Rows 3 and 4) having high and medium closeness values and low or medium betweenness and degree values. This constellation describes the industrial area of Raanana in the north-east corner of the city. The described pattern is typical for industrial areas, which are accessible, but are located in a well separated district of the city. This configuration is shown in the middle image in Figure 7.

*Configuration 3:* High availability of educational institutions, culture and leisure and parks and open spaces (Columns 1 and 2 – Row 1) with tendencies to lower closeness values and higher betweenness values. Such configurations are mostly characterized residential areas with high socio-economic standards. This finding seems reasonable since such a combination of topological properties means to live in residence places which are close to other parts of the city but which are not served for movement or transit between other parts of the city. A geographical mapping of this configuration of interest (e.g., for spatial equity, socio-spatial planning policy) shows clearly that it has an expected peripheral pattern. This configuration is shown in the right image in Figure 7.

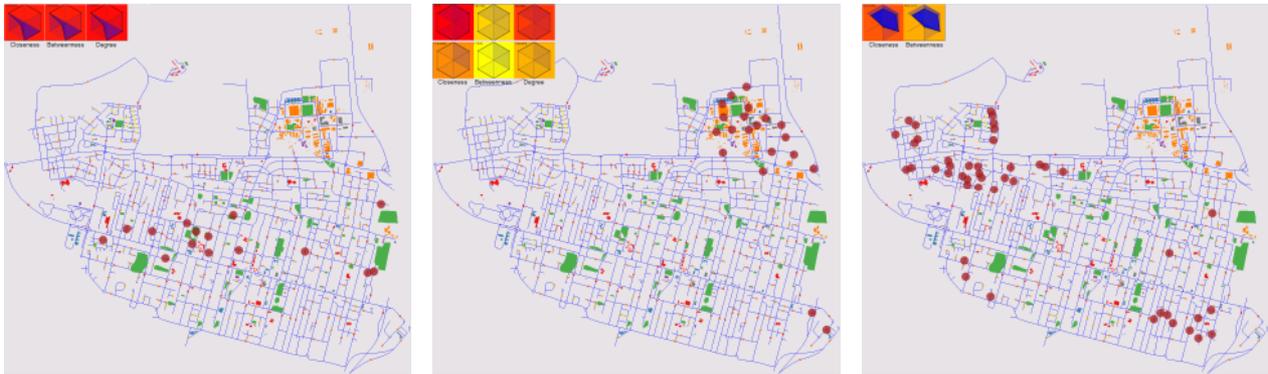


Figure 7. Geographic location of different spatial configurations: High commerce area (left), industrial area (middle) and residential areas (right). The SOM-clusters showing the different configurations are shown in the left upper corner. The named streets included in the clusters are highlighted on the map in red color.

These examples illustrate how locating the identified typical configurations on the geographic map are helpful in defining empiric findings. The explanation may concern previous knowledge on the development history and planning policy of Raanana as well as theoretical models that are suggested to elucidate on the spatial structure of land-uses in the city. Thus, the methodology enables us to identify typical and significant correlations between the topological properties of streets, or their combinations, and land-use spatial configurations, and further to explore their geographic patterns. This framework also helps to determine how the topological properties of a given street and its interrelation affect the functional content of its surroundings, i.e., which topological properties are significant for the formation of land-use spatial configurations?

## 5 Discussion and Conclusion

We described an analytic framework to assess urban spatial configurations. The methodology is applied on local land-use spatial configurations, and the local (i.e. degree centrality) and global (i.e. closeness centrality and betweenness centrality) of street networks' topologic properties in the city of Raanana. The methods used for analysis are based on SOM-clustering to group similar configurations, and on geographic views, which support analysts to iteratively extract interesting configuration patterns. These views, the more abstract cluster visualization, and the more concrete geographic map are highly interactive and strongly coordinated to each other. The contribution of the suggested methodological framework is clear: traditional local spatial analysis methods (i.e. local geo-statistics measures and aggregation at different scales) for analyzing and presenting high-resolution geographic data are typically limited to geographical presentation of one attribute only and blur the results by aggregation. Against this background, the proposed methodology has potential to shed light on the relation between multiple structural and geographical dimensions of an urban environment by keeping the individual objects as the level of investigation. Thus, the methodology enables investigation of local spatial relations between huge amounts of individual buildings with respect to their local and global attributes without loss of data as a result of aggregation. It is also possible to apply this methodology at different geographic scales and to explore the 'natural' geographic scale for the investigated phenomena. For now, we suppose that the main practical application of the proposed methodology is a an examination of a variety of urban spatial forms in order to reveal their unique spatial configuration with respect to functional and social composition of different city types. Such application has the potential to improve our knowledge on the relationship

between urban forms and the formation of land use spatial distribution in cities and can be used to support urban spatial policy.

Future work will explore additional multivariate visualization options, e.g., other glyph based approaches and graph-based representations. Currently, we support the visual correlation analysis between multidimensional variables in an abstract manner (Section 3.5), and in a separate geographic visualization of selected clusters (Section 4). Joint representation of these views is a challenge, which we like to address in the future. The method proposed on (Bacao, 2005) seems an interesting starting point to this end. On the algorithmic part, many options are open to implement additional algorithms (multi dimensional scaling, principal component analysis, etc.), which might be appropriate for different analytic questions in this context. Applying our approach on other domains is certainly a long-term perspective.

## References

- Anselin, L. (1995) Local Indicators of Spatial Association – LISA, *Geographical Analysis* 27(2), 93–115.
- Bacao, F., Lobo, V., Painho, M. (2005). The Self-Organizing Map, the Geo-SOM, and Relevant Variants for Geo-Sciences. *Computers & Geosciences*, 31, 155–163.
- Benenson, I., & Omer, I. (2003). High-Resolution Census Data: A Simple Way to Make Them Useful. *Data Science Journal* 2 (26), 117-127.
- Cox, M. & Cox, M. (2001). Multidimensional Scaling. *Chapman and Hall*.
- Deboeck, G., Kohonen, T. (1998). Visual Explorations in Finance With Self-Organizing Maps. *Springer*.
- Elmqvist, N., Dragicevic, P., Fekete, J.-D (2008). Rolling the Dice: Multidimensional Visual Exploration Using Scatter-Plot Matrix Navigation. *IEEE Transactions on Visualization and Computer Graphics*, 14, 1141–1148.
- Everitt, B. S. & Nicholls, P. (1975). Visual Techniques for Representing Multivariate Data. *The Statistician*, 24(1), 37-49.
- Fua, Y.-H., Ward, M., Rundensteiner, E. (1999). Hierarchical Parallel Coordinates for Exploration of Large Datasets. *Proceedings of IEEE Conference on Visualization (VIS)*, 43-50.
- Guo, D., Chen, J., MacEachren, A. M., Liao, K (2006). A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP). *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1461–1474.
- Hillier, B., Penn A., Hanson J., Grajewski T. and Xu J. (1993) Natural Movement: Configuration and Attraction in Urban Pedestrian Movement, *Environment and Planning B*, 20, pp. 29-66.
- Honkela, T., Kaski, S., Lagus, K., Kohonen, T. (1997). WEBSOM—Selforganizing Maps of Document Collections. *Proceedings Workshop on Self-Organizing Maps*, 310–315.
- Inselberg, A. & Dimsdale, B. (1990). Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry. *Proceedings Conference on Visualization*, 361–378.
- Jiang, B. (2007). A Topological Pattern of Urban Street Networks: Universality and Peculiarity. *Physica A*, 384, 647-655
- Jiang, B. & Harrie L. (2004) Selection of Streets from a Network Using Self-Organizing Maps, *Transactions in GIS*, 8(3): 335–350
- Jolliffe, I. (2002). Principal Components Analysis. *Springer*, 3rd edition.
- Kohonen, T. (2001). Self-Organizing Maps. *Springer*, 3rd edition.
- Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J. (1996). SomPak: The Self-Organizing Map Program Package. *Helsinki University of Technology Technical Report*.
- LeBlanc, J., Ward, M.O., Wittels, N (1990). Exploring N-dimensional Databases. *Proceedings Conference on Visualization*, 230 - 237.
- Omer, I., & Benenson I., (2002). Investigating Fine-Scale Residential Segregation by Means of Local Spatial Statistics. *Geography Research Forum* 22, 41-60.
- Porta, P. Crucitti, P. & Latora V. (2006). The network analysis of urban streets: A primal approach. *Environment and Planning B: Planning and Design* 2006, 33, 705–725
- Ratti C, Pulselli RM, Williams S, Frenchman D (2006) Mobile Landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design* 33: 727-748
- Spielman, S. & Thill, J.-C. (2008). Social Area Analysis, Data Mining, and GIS. *Computers, Environment and Urban Systems*, 32, 110–122.
- Talen, E. (2003). Neighborhoods as Service Providers: a Methodology for Evaluating Pedestrian Access. *Environment and Planning B: Planning and Design*, 30, 181- 200.
- Talen, E., & Anselin, L. (1998). Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds. *Environment and Planning A*, 30, 595–613.

- Van Long, T. & Linsen, L. (2009). MultiClusterTree: Interactive Visual Exploration of Hierarchical Clusters in Multidimensional Multivariate Data. *Proceedings of Eurographics / IEEE Symposium on Visualization*.
- Vesanto, J. (1999). SOM-based Data Visualization Methods. *Intelligent Data Analysis*, 3(2):111–126.
- Ware, C. (2004). Information Visualization: Perception for Design. *Morgan Kaufmann*, 2nd Ed.
- Wilkinson, L., Anand, A., Grossman, R. (2005). Graph-Theoretic Scagnostics. *Proceedings of the IEEE Symposium on Information Visualization*.
- Wong D.W.S, (2003) Spatial Decomposition of Segregation Indices: A framework toward measuring segregation at multiple levels, *Geographical Analysis*, 35 (3), 179-184.