Discretization of Urban Areas using POI-based Tesselation

Tobias Hagen, Jonas Hamann, Siavash Saki

Working Papers
Fachbereich Wirtschaft und Recht
Frankfurt University of Applied Sciences
www.frankfurt-university.de/fb3

Nr. **23**

Januar/January 2022
ISSN-Nr. 2702-5802
DOI: https://doi.org/10.48718/7jjr-1c66

Fachbereich 3
Wirtschaft und Recht | Business and Law

Wissen durch Praxis stärkt

Frankfurt University of Applied Sciences
Fachbereich 3: Wirtschaft und Recht
Nibelungenplatz 1
60318 Frankfurt am Main

Working Papers des Fachbereichs Wirtschaft und Recht der Frankfurt University of Applied Sciences dienen der Verbreitung von Forschungsergebnissen aus laufenden Arbeiten im Vorfeld einer späteren Publikation. Sie sollen den Ideenaustausch und die akademische Debatte befördern. Die Zugänglichmachung von Forschungsergebnissen in einem Fachbereichs Working Paper ist nicht gleichzusetzen mit deren endgültiger Veröffentlichung und steht der Publikation an anderem Ort und in anderer Form ausdrücklich nicht entgegen.

Working Papers, die vom Fachbereich Wirtschaft und Recht herausgegeben werden, geben die Ansichten des/der jeweiligen Autor*innen wieder und nicht die der gesamten Institution des Fachbereichs Wirtschaft und Recht oder der Frankfurt University of Applied Sciences.

Abstract English

Urban area tessellation is a crucial aspect in many spatial analyses. While regular tessellation methods, like square-grid or hexagon-grid, are suitable for addressing pure geometry problems, they cannot take the unique characteristics of different subareas into account. Irregular tessellation methods allow the border between the subareas to be defined more realistically based on the urban features like road network or POI data. This paper studies and compares five different tessellation methods: Squares, hexagons, adaptive squares, Voronoi diagrams, and city blocks. We explain how (open-source) POI data can be integrated into the tessellation process to build what we call "Local Geographic Units" (POI-based tiles). These units are flexible and adaptable to the structure of the studied area and underlying data and could improve the performance of further analyses. The results of the various tessellation methods are demonstrated for the city of Frankfurt am Main in Germany. A simple clustering of Local Geographic Units for the studied city indicates that city blocks perform better than the other methods in the city segmentation in terms of reflecting the structure of this city.

Abstract Deutsch

Die Tessellierungen urbaner Gebiete ist ein entscheidender Aspekt bei räumlichen Analysen. Regelmäßige Tessellierungen, wie die Unterteilung in Quadrate oder Hexagons, eignen sich zwar für Probleme rein geometrischer Natur, berücksichtigen aber die Charakteristika der enthaltenen kleineren geographischen Einheiten nicht. Unregelmäßige Tessellierungen ermöglichen eine realitätsnahe Unterteilung basierend auf städtischen Merkmalen, wie dem Straßennetz oder POI-Daten. In diesem Beitrag werden fünf verschiedene Tessellierungsmethoden vorgestellt und verglichen: Quadrate, Hexagons, adaptive Quadrate, Voronoi-Diagramme und City-Blocks. Die Integration von (Open-Source) POI-Daten in den Tessellierungsprozess führt zu sogenannten „Lokalen Geographischen Einheiten". Diese POI-basierten Einheiten sind flexibel und passen sich sowohl der Struktur des zu untersuchenden Gebiets, als auch der zugrundeliegenden Daten an und erlaube dadurch darauf aufbauende, detailliertere Analysen. Alle vorgestellten Tessellierungsmethoden werden an dem Beispiel Frankfurt am Main durchgeführt und präsentiert. Ein einfaches „Clustering" der Lokalen Geographischen Einheiten zeigt, dass City-Blocks die Struktur der Stadt besser abbilden können, als die anderen vorgestellten Methoden.

Discretization of Urban Areas using POI-based Tessellation

Tobias Hagen[1], Jonas Hamann[2] and Siavash Saki[3]

Frankfurt University of Applied Sciences

ReLUT – Research Lab for Urban Transport

Urban area tessellation is a crucial aspect in many spatial analyses. While regular tessellation methods, like square-grid or hexagon-grid, are suitable for addressing pure geometry problems, they cannot take the unique characteristics of different subareas into account. Irregular tessellation methods allow the border between the subareas to be defined more realistically based on the urban features like road network or POI data. This paper studies and compares five different tessellation methods: Squares, hexagons, adaptive squares, Voronoi diagrams, and city blocks. We explain how (open-source) POI data can be integrated into the tessellation process to build what we call "Local Geographic Units" (POI-based tiles). These units are flexible and adaptable to the structure of the studied area and underlying data and could improve the performance of further analyses. The results of the various tessellation methods are demonstrated for the city of Frankfurt am Main in Germany. A simple clustering of Local Geographic Units for the studied city indicates that city blocks perform better than the other methods in the city segmentation in terms of reflecting the structure of this city.

Keywords: urban analysis, tessellation, spatial tessellation, urban studies

[1] Prof. Tobias Hagen, director of the ReLUT – Research Lab for Urban Transport. Contact: thagen@fb3.fra-uas.de

[2] Jonas Hamann, Research Assistant at ReLUT. Contact: jonas.hamann@fb3.fra-uas.de

[3] Siavash Saki, Research Assistant at ReLUT. Contact: siavash.saki@fb3.fra-uas.de

1 Introduction

Cities and urban areas are complex systems (Healey, 2006; Portugali and Stolk, 2016). With advancing digitization and innovative mobility, the amount of data collected is growing rapidly. Smart cities and smart mobility are becoming more and more important for urban planners (Alvares et al., 2011). In Germany in particular, almost every decision relates to administratively defined city or district boundaries, which does not consider the actual structure of the city and the interdependency between city districts. A cross-district view taking the structure of the city into account is more suitable for many topics such as traffic planning. Therefore, it is necessary to discretely capture the structure of urban areas without using administrative boundaries (Briant et al., 2010). The process of discretization of space into subspaces with no overlaps and no gaps is called tessellation. Tessellation is essential in understanding geographical space and provides a framework for analyzing geospatial data (Gold, 2016). There are different tessellation methods, which are used (and can be combined) for different applications and analytics (Xing et al., 2020). White and Kiester (2008) argue that topology affects the outcome of geospatial models.

Many spatial tessellation methods for urban areas are based on discrete global grid systems (DGGS). These systems have already been well studied and extensively summarized in Sahr et al. (2003). The advantages and disadvantages of different methods are analyzed by Li and Stefanakis (2020). Tessellation methods are grouped into “regular”, “irregular”, and “semi-regular”. The square grid is one of the easiest methods to implement and is widely used. For example, Boontore (2011) uses square tessellation to analyze the degrees of evenness in the spread of urban development and to enable a comparison between subareas. Samsonov et al. (2015) use square grid to characterize urban canyons. A well-established implementation of square tessellation is Microsoft’s Bing Map[4] Tile system, which is also used in this paper. Another common regular tessellation is the hexagon grid. Hexagon tessellation is used in Asamer et al. (2016) to optimize charging station locations regarding the charging demand of electric taxis and to find optimum regions rather than exact positions. For analyzing public transport services, Biazzo et al. (2019) cover a city with a hexagonal grid to study the accessibility and inequality in different zones as well as lowering the computation times. For the analysis of soil contamination in urban environments, Griffith (2008) uses hexagons to take the impact of spatial autocorrelation into account and to find optimal numbers as well as positions of soil sample locations. Using e-scooter data and a hexagonal grid, McKenzie and Romm (2021) measure the similarity between regions in a city (and cross cities). Smucker et al.

[4] https://docs.microsoft.com/en-us/bingmaps/articles/bing-maps-tile-system

(2016) use hexagon tessellation to study landscape effects on stream water quality. Feick and Robertson (2015) use hexagons to capture spatial expressions of Vancouver by using geotagged photograph data (GTP) from Flickr. Uber takes hexagons as a basis for its applications and publicly provides an implementation of the hexagon tessellation for GIS applications, called the h3 grid system (Brodsky, 2018).

Although these regular approaches are well suited for the analysis of a variety of research topics, as the review of previous research shows, they do not adapt to the structure of cities. A data-driven irregular tessellation approach to overcome this problem is Voronoi diagrams (Aurenhammer, 1991). Generators are used to create polygons based on data. These polygons subdivide the space without gaps or overlaps. Generators can be footprints (Fleischmann et al., 2020), street network data (He et al., 2017), or even (clustered) points of interest (POI). There is a wide range of studies that use Voronoi diagrams, for example, as in Xinqi et al. (2010) for optimizing regional urban system planning, as in Sadahiro (2002) for analyzing the spatial structure of an administrative system, and as in Bełej and Figurska (2020) for geospatial analysis of real estate prices. By analyzing Twitter activity, Frias-Martinez and Frias-Martinez (2014) argue that Voronoi diagrams are useful to partition the area while preserving the topological characteristics of geolocated tweets as well as respecting the actual shape of the studied area. Another common usage of Voronoi diagrams is the analysis of phone-call data when transceiver (cell) towers are used as generators (Yuan and Raubal, 2012). Another approach to tessellating urban areas is to create city blocks using road networks. Using a raster-based model Yuan et al. (2012) show how street segments can be used to create these polygons. The created city blocks with this method are used in Zheng et al. (2011) for analyzing GPS trajectories of taxis. Graser (2017) proposes a new approach for creating city blocks based on Voronoi diagrams. It uses street intersections as generators and builds polygons centered around street intersections rather than closed areas between the street segments.

The absence of a uniform definition for local tiles that can represent a spatial area with additional information leads to the necessity to introduce a new concept, which we call *Local Geographic Units* (LGUs). Tiles are the product of tessellations in general (Ammann et al., 1992) in the form of polygons representing only spatial data. To include additionally non-spatial attributes such as POI, we propose the concept of LGUs, which are a special case of tiles. LGUs combine spatial and non-spatial data within a data-driven approach that adapts to the structure of cities. Therefore, even "dynamic LGUs" are possible since they can be continuously updated regarding their shape and attribute according to the changes in cities. LGUs can be a better basis for the analysis of land-use, socio-demographic trends or traffic patterns, as they do not adhere to administrative boundaries, but apply to the whole continuous area. The need for a general unit for spatial and non-spatial data to capture urban areas' complex structures, which can be used in any kind of urban area, city or even rural areas, motivates the POI-based tessellation methods applied in this paper. It studies and compares five approaches to generate LGUs: two so-called *regular tessellations* (squares, hexagons) and an extension we call *adaptive square tessellation*, and two *irregular tessellations* (Voronoi diagrams and city blocks).

The paper is structured as follows. Section 2 provides basic definitions and introduces the general concept of LGUs. In section 3, the data sources are described. Different methods to generate LGUs are proposed in section 4. We use the city of Frankfurt am Main in Germany as case study to demonstrate different concepts of tessellations. The results of the different tessellation methods are presented in section 5. As an example of application, section 6 shows the performance of different LGUs as the basis for clustering urban areas. Finally, section 7 concludes.

2 Basic Definitions and Concepts

A tessellation of a d-dimensional (Euclidean) space, $\mathbb{R}^{\mathbf{d}}$, can be defined as a set of d-dimensional regions which cover the whole $\mathbb{R}^{\mathbf{d}}$ without overlaps or gaps. A formal definition can be found in Okabe et al. (2000). Within the context of urban areas tessellation, we work with two-dimensional spatial data. A two-dimensional tessellation is also called *planar tessellation* (Okabe et al., 2000).

There are various tessellation methods to discretize the Earth's surface. These methods are based on one of the two following approaches: The first approach is to "cut" the globe into a two-dimensional plane and discretize the two-dimensional space afterwards. The second approach is based on *discrete global grid systems (DGGS)*, which uses a platonic solid and lays a corresponding grid onto the surface. DGGS are spatial references that use a hierarchical tessellation of cells to partition and address the entire Earth's surface (Peterson, 2016). A HDGGS (hierarchical DGGS) consists of increasingly finer resolution grids; i.e., the grids in the series have a monotonically increasing number of cells (Sahr et al., 2003).

We propose the concept of *Local Geographic Units* (LGUs) as a special case of tessellation tiles that contain both spatial attribute (coordinates) and non-spatial attributes (e.g., population, area size, number of restaurants). They discretize the spatial space and allow further analysis using their non-spatial attribute(s). A formal definition of a LGU is given below.

Let $D \subset \mathbb{R}^{\mathrm{d}}$ be a finite space, $\mathrm{T} = \{\tau_1, \ldots, \tau_n\}$ be a tessellation of D, and $M = \{z_1, \ldots, z_n\}$ be a set of non-spatial attributes. A **Local Geographic Unit** $l_i = (\tau\_i, z\_i\ ), where\ \tau_i \in \mathrm{T}\ and\ z_i \in M$ is called:

regular, when the tiles are congruent, which is $\tau_i \cong \tau_j, \forall i, j = 0, \ldots, n,$

semi-regular, when the tiles are similar, which is $\tau_i \sim \tau_j, \forall i, j = 0, \ldots, n,$

irregular, otherwise.

The regularity of LGUs only refers to the spatial attribute (geometry) of the LGUs and differs only in the shape and size[5] of the LGUs. While regular LGUs all have the same shape and size (apart from minor errors due to their local projections that can be neglected), semi-regular LGUs are only similar in shape, and differ in size. Irregular LGUs have completely different shapes and sizes. In this paper, regular, irregular and semi-regular tessellations are considered and applied to analyze the structure of cities and urban areas.

With tessellation in general and LGUs in particular, the issues of finding the appropriate method and settings (parameters) for each method arise. Too many LGUs (i.e., very small LGUs in relation to the space) can result in too many LGUs with zero non-spatial attributes, which would imply poor results of the subsequent data analytics methods. On the other hand, too few LGUs (i.e., very large LGUs in relation to the space) can contain many different non-spatial data, resulting in losing variance between the LGUs, increasing variance within the LGUs, and making subsequent analyses difficult or even meaningless. Finding "good" sizes and shapes for LGUs is somehow a trade-off between too much and too little information per LGU, which could be reflected by an appropriate metric. To the best of our knowledge, such a metric to compare tessellation methods and different settings has not yet been developed. In order to get some insights into the difference in the performance of the approaches, the resulting distribution of the number of POI per LGU and the results of simple clustering of LGUs are compared in Section 6.

3 Data Sources

The largest open-source geospatial database is OpenStreetMap (OSM) (OpenStreetMap Wiki contributors). In the past decade, the quality of OSM data was analyzed in several studies (Haklay, 2010; Jekel, 2012). Compared to other potential sources, like Facebook

[5] Size in the sense of area (e.g., m^2).

or Foursquare, the POI (Points of Interest) locations are much more accurate in OSM. Hochmair et al. (2018) find that POI data from Facebook or Foursquare show higher mean offsets from their true location compared to the OSM Data. The fact that OSM data is open-source and has a large amount of data makes it especially useful for approaches that should be generalizable to other geographical areas. A further advantage of the OSM platform is the simplicity of collecting data from its database. For this paper, all data preparations and analyses were done in Python. Details can be found in the supplementary materials.

All the applied methods for generating LGUs should be transferable to any city or rural area. To guarantee this, OSM is the only data source used.

4 Methods

4.1 Squares

Being easy to implement, square tessellation is widely used. By creating squares, the most relevant question is which projection should be used. Depending on the projection, the earth's surface could be represented in ellipsoid, rectangle or other shapes.

We apply Microsoft Bing Maps Tile System[6], which is an implementation of square tessellation based on Mercator projection. This projection distorts the size and shape of areas and gets problematic near the poles; however, it can be neglected since there are no large urban areas to analyze in these areas (Winter and Goel, 2021). It preserves the shape of relatively small objects, which is essential when analyzing urban areas. To guarantee this, one has to choose a sufficiently detailed square size, depending on the size of the city.

4.2 Adaptive Square Tessellation

We extend square tessellation to an adaptive version. The idea of adaptive refinement comes from solving complex differential equations. In numerical analysis, this is called adaptive mesh refinement (AMR). The underlying grid[7] is refined at certain regions (e.g., turbulences or singularities) to ensure a higher accuracy (Gerya, 2019). In this paper, regions to be refined are LGUs that exceed a threshold, e.g., LGUs that contain "too much" information. These LGUs are subdivided into four smaller squares. "Too much" could mean that a given threshold of the number of POI per LGU is exceeded or there are too many different POI types included in an LGU. Then this LGU is split up until

[7] Commonly used girds are triangulations or square-elements.

it fulfills the given threshold. Possible thresholds are, for example, the 90% Quantile of the number of POI per LGU, the mean number of POI per LGU in the first step, or even an arbitrary number depending on the application. The result is a tessellation in which densely populated areas are more finely divided. This method is somehow a combination of irregularity and regularity since the LGUs have different sizes, but they are still similar[8] according to shape.

4.3 Hexagons

Unlike squares, it is not possible to divide one parent hexagon into multiple smaller hexagons without over-lapping or gaps. However, the main advantage of this approach is that the distances of each hexagon center point with the center points of its neighbors is equal[9]. A more detailed explanation can be found in Sahr (2014). Uber is an example that uses this approach. With its h3 grid system, Uber combines the hierarchical subdivisions with a hexagonal grid. The Uber h3 grid system has 16 different resolutions (0 to 15). The most detailed resolution 15 has an average edge length of 0.5m. We apply this implementation provided by Uber.

4.4 Voronoi Diagrams

After presenting two of the main regular tessellations and a semi-regular tessellation, we describe and apply one widely used irregular tessellation, the Voronoi diagram. Generalized Voronoi diagrams are created by generators in the plane, like points, lines or polygons, by assigning each location to the nearest generator; nearest in the sense of the used metric, e.g., Euclidean or Manhattan. This method differs from regular tessellations as it is defined on finite subsets. The *Ordinary Voronoi Diagram* (OVD) uses points to define a tessellation. The formal definition is of Voronoi diagrams is provided by Okabe et al. (2000).

The generators for Voronoi diagrams depend on the use case. We use POI in order to capture the structure of a city. If the number of POI is small enough relative to the (city) area, POI can directly be used as generators. However, in most cases, the input dataset has a vast number of POI. To tackle the problem of the high number of POI, first of all, the POI have to be clustered. For this purpose, different clustering algorithms are used to group POI, based on their spatial proximity and density. After grouping POI, a generator is derived from each POI cluster. We used cluster centroid (center of mass) to create the generators from the clusters.

[8] Similarity in the mathematical sense, that is, one can be obtained from the other by scaling uniformly.

[9] In contrast, a square-based DGGS has two different distances, one with the neighboring square, sharing an edge and one with the vertex. In the case of a triangular-based DGGS, there are even three different distances.

The resulting LGUs are of irregular shape because the size and shape depend on the POI and the clustering algorithm to generate those. Here, k-means and HDBSCAN[10] are used to cluster the POI before defining the generators. Figure 1 shows a layered map with the individual steps for generating Voronoi polygons using HDBSCAN Algorithms as an example. The procedure is the same for k-means.

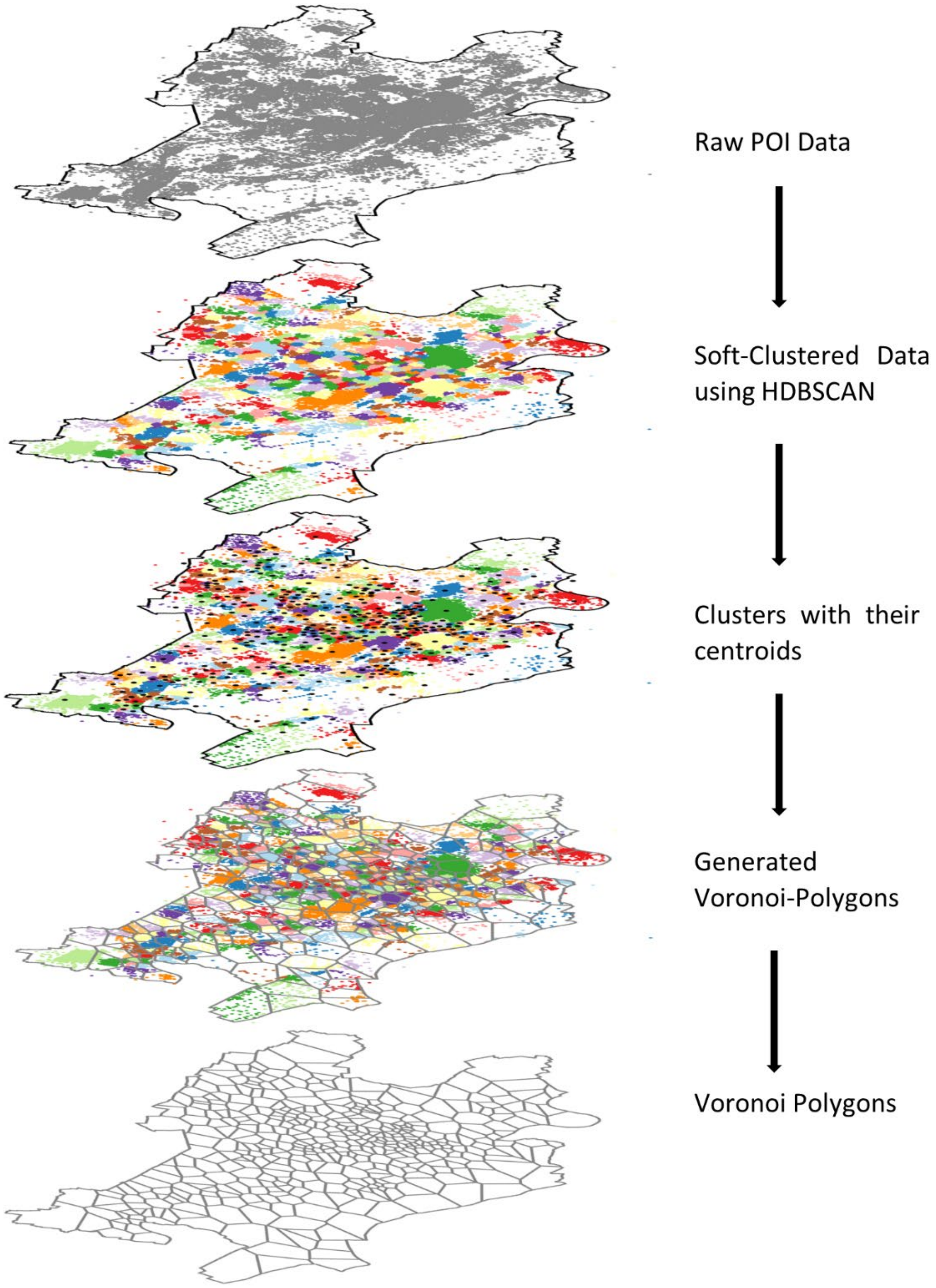

Figure 1: Procedure for generating Voronoi polygons using HDBSCAN for POI.

[10] HDBSCAN is a modification of DBSCAN (Ester et al. (1996)) clustering developed and introduced by McInnes et al. (2017).

4.5 City Blocks

City blocks are based on the road network[11] (or other line segments) of the area. In urban morphology, a city block (street block) refers to the smallest closed areas surrounded by road segments, which is used for building construction (Oliveira, 2016).

One of the two main problems with creating city blocks using this method is that not every part of an urban area is part of a city block. Therefore, there may be gaps between the city blocks. The second problem is that city blocks can be too small, e.g., a traffic island. Traffic islands are identified as city blocks since they are bounded by street segments. To overcome the first problem, the initially created city blocks are merged and subtracted from the whole boundary polygon of the urban area (or the planar space). Then, the resulting rest (multi-)polygon is cut into polygons. This guarantees that the whole area is filled without gaps. The problem of too small polygons can be addressed in multiple ways. One could define a threshold for minimum area size, e.g., $10{,}000m^2$, or use clustering algorithms to group small polygons.

To generate city blocks for an area, the following steps are done:

1. Define a boundary polygon and collect street network data (e.g., using OSMnx)
2. Use the polygonize function and the street network to generate the initial polygons
3. Merge these polygons temporarily resulting in a multi-polygon
4. Use the multi-polygon to identify the gaps and a define a “Rest-Multi-Polygon”
5. Polygonize the “Rest-Multi-Polygon”
6. Merge small contiguous polygons using clustering algorithms to generate the LGUs. Here we use hierarchical clustering to identify the merging polygons. For this reason, the number of LGUs is equal to the number of clusters which can be arbitrarily set.

The process is visualized in the following Figure 2.

[11] Road networks include pedestrian and cycle roads.

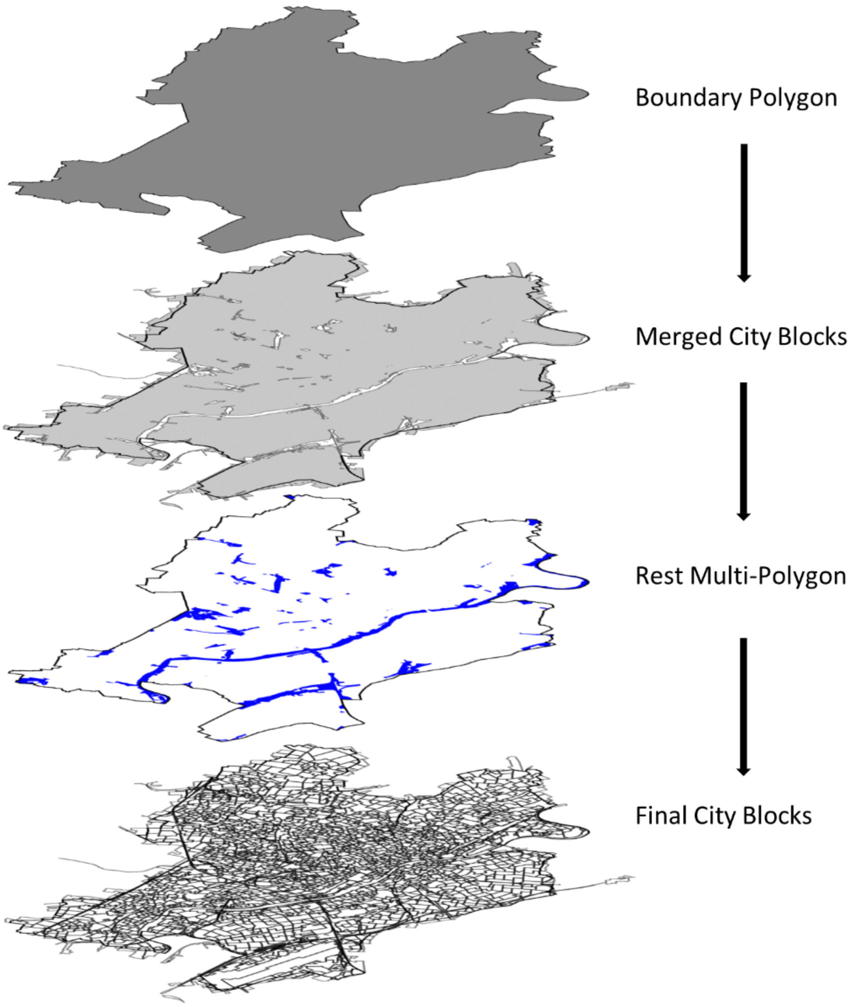

Figure 2: Procedure for creating city blocks.

5 Results

In this section, we demonstrate the resulting tessellations based on the different LGU approaches. The city of Frankfurt am Main in Germany is chosen as an example. Frankfurt has many different geographical attributes that may be difficult to tessellate, especially using the city block method. Besides rivers and parks, it includes the largest airport in Germany in its boundary polygon. An additional challenge arises from the fact that the airport includes not only runways but also many POI such as stores and hotels.

5.1 Regular and semi-regular tessellations

When using Bing Maps Tile System to create a square grid, a resolution value to define square sizes has to be specified. We choose resolutions 16 and 17. The results of the square tessellation are shown in Figure 3. The appropriate resolution depends on the use case. For example, when considering all the available POI of a city in OSM, the resolution should not be too detailed (e.g., resolution 18) and not too rough (e.g., resolution 14). As discussed in Section 6, an indication for a “good” resolution could be the POI count per LGU (here per square) and the associated distribution.

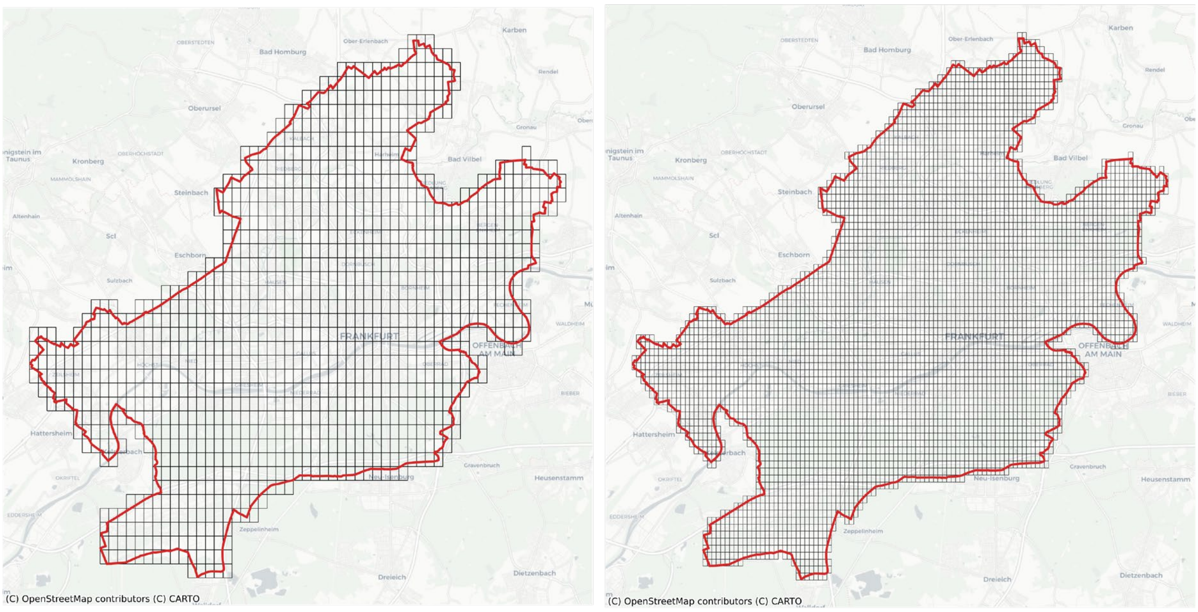

Figure 3: Squares as LGUs for Frankfurt am Main. Left and right images show square tessellation using Bing Maps Tile System in resolutions 16 (1,786 squares) and 17 (6,811 squares), respectively.

Figure 4 shows tessellation using adaptive squares. The problem of choosing a “good” resolution using this approach turns out to be secondary here, but the choice of the threshold is more relevant. The method starts with a low resolution concerning the number of POI (e.g., a resolution of 14 for all of the available POI of a city). Then the implemented algorithm subdivides the whole area until the discussed threshold is not exceeded anymore. Here, we used the POI data for Frankfurt, containing approximately 190,000 POI. Adaptive squares for Frankfurt are demonstrated in Figure 4 using two different settings. On the left image, the starting resolution is 14 and the threshold is arbitrarily set to 1000. Squares, shown on the right image, start at resolution 15. The threshold is the median number of POI per LGU at the starting level 15, i.e., 187.5. As expected, the city center is tessellated with a higher resolution than more rural areas. Already based on these LGUs, a rough structure is visible.

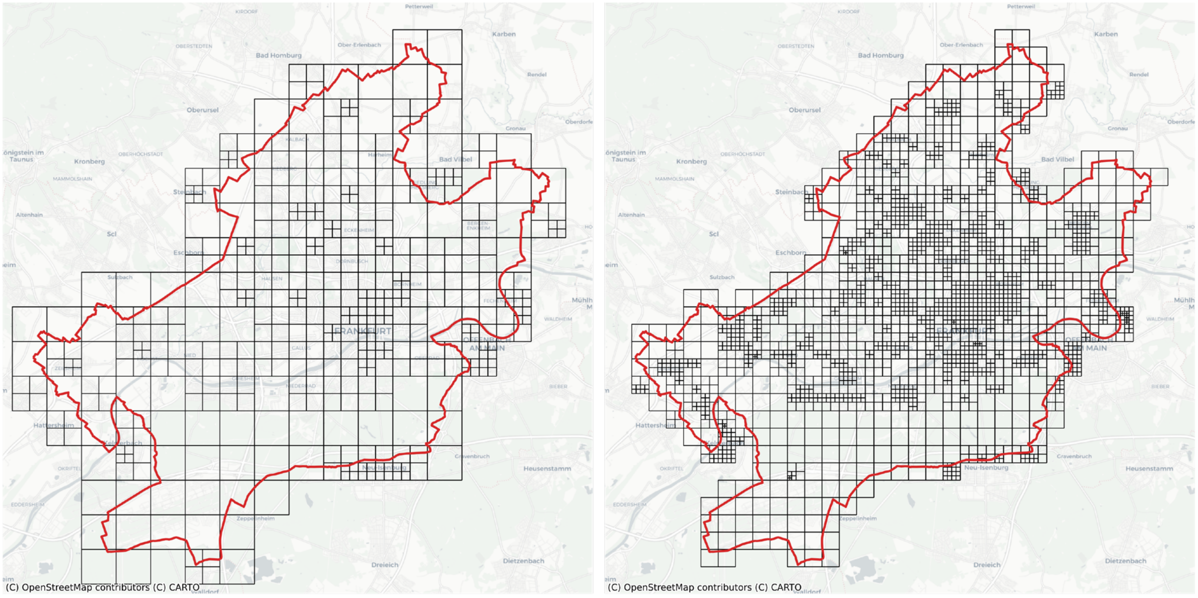

Figure 4: Adaptive squares as LGUs for Frankfurt am Main. 494 squares on the left image from resolution 14 to 16, 2300 squares on the right image from resolution 15 to 18.

Similar to the square grid, to create a hexagon grid, a suitable hexagon size must be initially specified. The hexagon grid in the following examples and figures are created using Uber's h3 implementation. The size of hexagons can again be modified changing the resolution value. Our recommendation is to use a relatively detailed resolution; otherwise gaps may appear at the boundary polygon. Larger hexagons create larger gaps. As mentioned early, a good indication for a "good" resolution (hexagon size) could be the number of POI per LGU (see Section 6). Figure 5 shows the tessellation of Frankfurt with hexagons at levels 8 and 9.

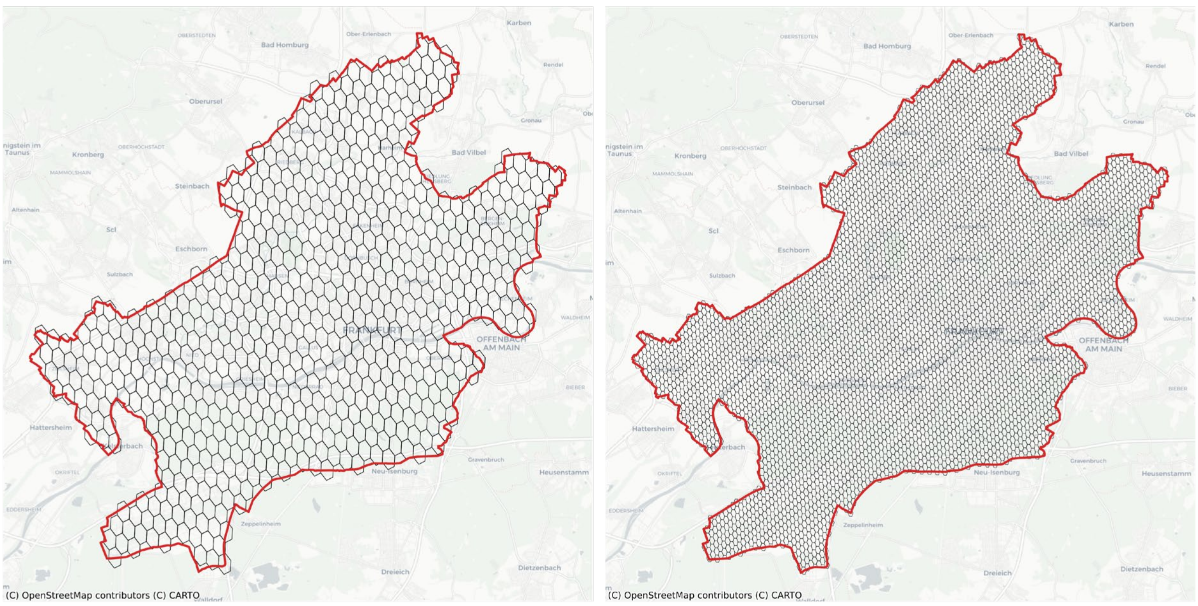

Figure 5: Hexagons as LGUs for Frankfurt am Main. Left and right images show hexagon tessellation using uber h3 in resolutions 8 (366 hexagons) and 9 (2,590 hexagons), respectively.

5.2 Irregular tessellations

Figure 6 shows the final city blocks for Frankfurt (see Section 4.5). For a plausible comparison with the results of the other approaches, it is intended to have roughly the same number of city blocks as the number of hexagons and squares generated by the previously described approaches. According to Figure 6, city blocks seem to adapt well to the structure of the city. The city center is more likely to have a denser road network than urban areas, which results in more city blocks. Smaller LGUs can be seen in the center and busy areas, while the surrounding areas have larger blocks. The airport (southeast) is the biggest city block, followed by the city block representing the river Main. It is a further indication of the good performance of the city blocks approach that the airport and river main are identified as a coherent area. The same is true for parks and forests.

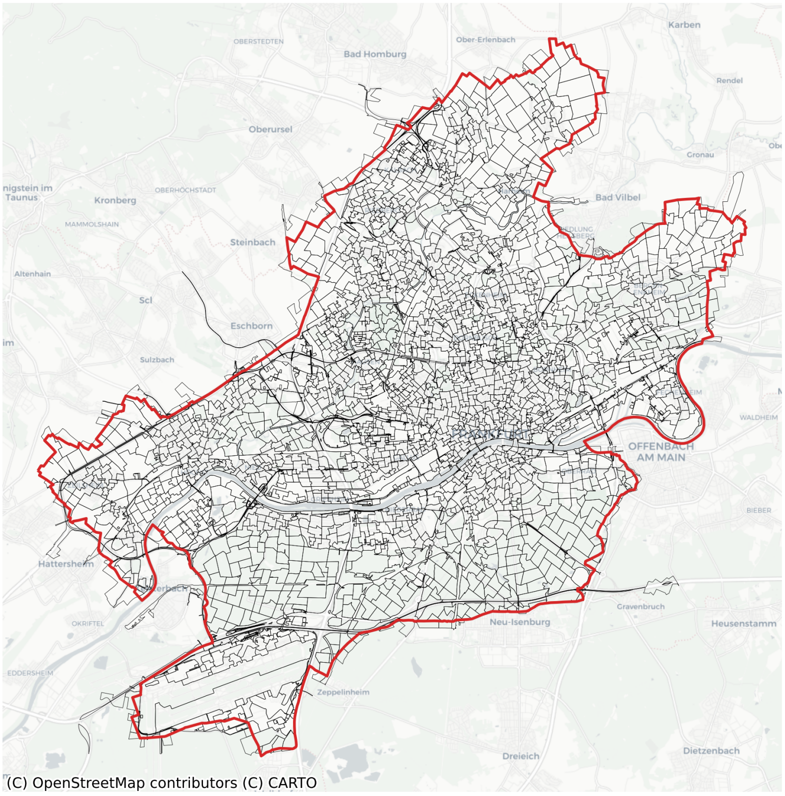

Figure 6: 2,357 city blocks as LGUs for Frankfurt am Main.

Figure 7 shows the area of Frankfurt, subdivided into LGUs with the Voronoi diagrams based on two different preprocessing methods. Because of their large number, POI cannot be used directly as generators and need to be clustered. In the left and right examples in Figure 7, k-means and HDBSCAN are used respectively to cluster the POI. The centroids of the clusters are then used as generators in the Voronoi algorithm. Similar to the city blocks, the city center has smaller LGUs than the surrounding area. Each clustering method in the preprocessing step has its own advantages and disadvantages for the application here. For example, HDBSCAN leads to smaller LGUs where the density of POI is higher, e.g., in the city center. This is useful when a finer separation of the area in the city center is desired. Using k-means, all the POI of a cluster end up in the same Voronoi polygon.

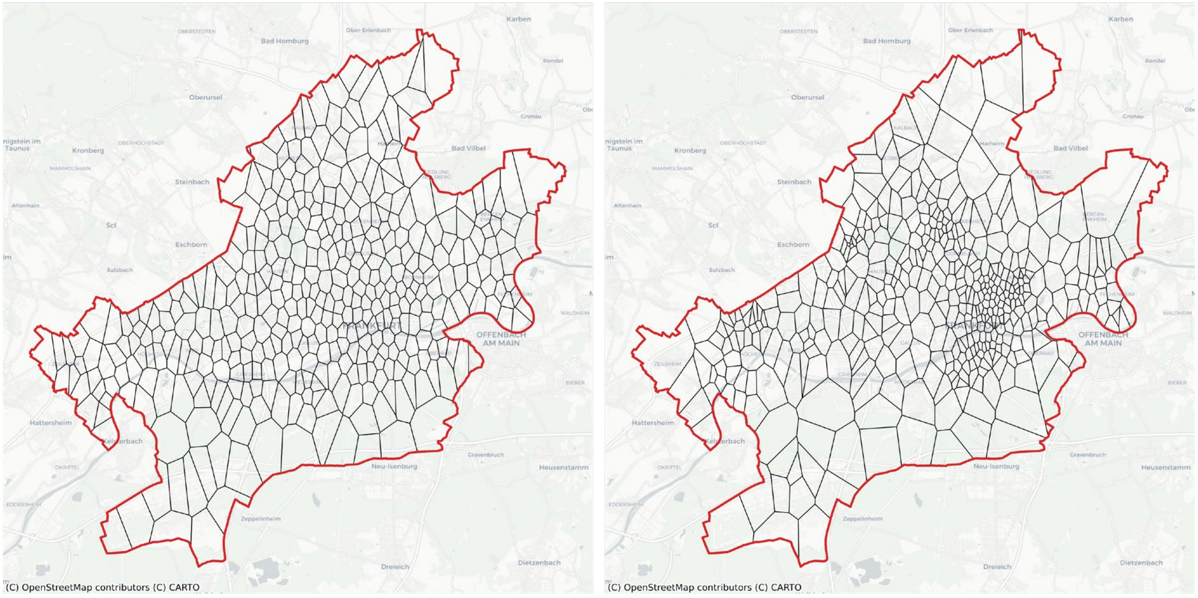

Figure 7: Voronoi polygons as LGUs for Frankfurt am Main. k-means (444 LGUs) and HDBSCAN (400 LGUs) are used in the preprocessing step in the left and right images, respectively.

6 Discussion

To demonstrate the performance of the approaches, a simple clustering algorithm is used to cluster the LGUs of Frankfurt based on its POI data. This is one possible application for LGU: to identify similar areas (clusters) within the cities based on POI. The POI data is restricted to five categories (out of 29 OSM primary categories): amenities, shops, offices, buildings and public transport stations. These five categories contain a considerable share of POI in Frankfurt (approximately 140,000 POI accounting for about 54% of the entire tagged POI in Frankfurt).

Figure 8 shows the distribution of the number of POI per LGU for the different approaches. City blocks, hexagons, and squares (all do not consider POI data) result in more LGUs with zero POIs. The reason is that a large part of Frankfurt is vegetation and does not contain any POI in these five categories. This results in a significant share of LGUs with zero POI when the polygons are uniformly distributed over the city, i.e., squares and hexagons, and also when polygons are created using the road network, i.e., city blocks. In contrast, POI are more systematically distributed in LGUs by the methods which consider these POI data, i.e., Voronoi diagram and adaptive squares. This is an intuitive result, as these methods tend to build smaller polygons where the POI are dense (e.g., near city center) and larger polygons where the POI are sparse (e.g., the vegetation part). However, adaptive squares still have a noticeable number of LGUs with zero POI. To reduce zeros when using this method, a rougher initial resolution can be chosen.

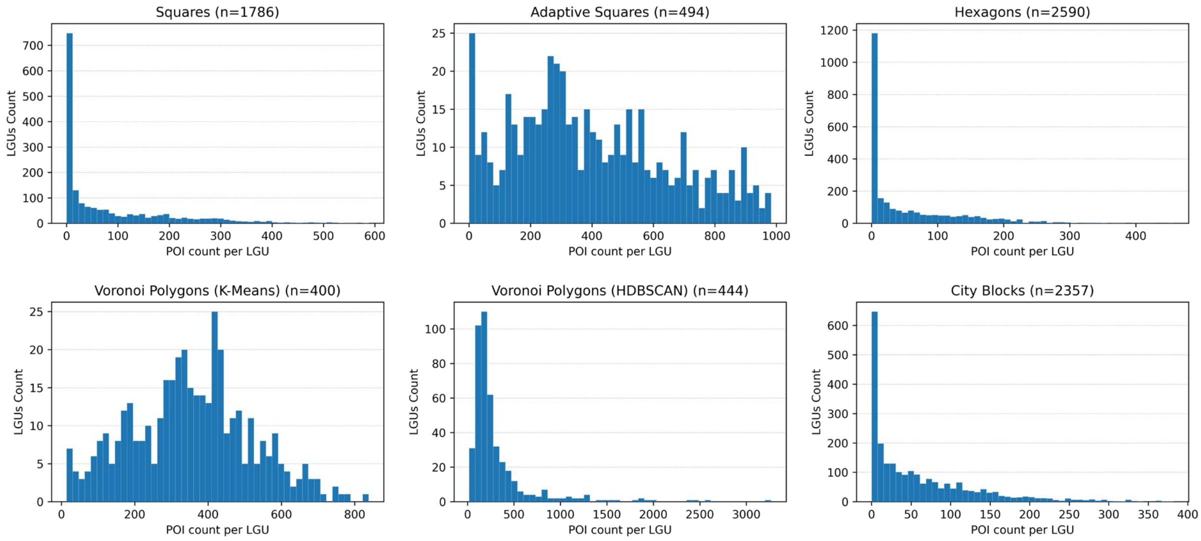

Figure 8: Histograms of the LGUs distribution regarding their POI count for the 6 investigated cases.

Using k-means, LGUs created by different tessellation methods are clustered into four groups. As a preprocessing step to the clustering methods, the number of POI in each LGU is normalized by the maximum number of POI per LGU. For city blocks, a further step is done, i.e., normalization by the LGU area, since the area varies drastically between the city blocks. For example, there are LGUs with enormous areas like the river Main or the airport, and also small LGUs like specific living blocks. This further normalization step is not necessary for other methods.

Figure 9 shows the results of clustering methods based on the different methods for generating LGUs, with the first row summing up the regular and semi-regular approaches and the second row showing the irregular approaches. For the Voronoi-based LGUs, two clustering methods (k-means and HDBSCAN) are presented to visualize the differences (see Section 4.4). The centroids of the clustered POI data are used as generators.

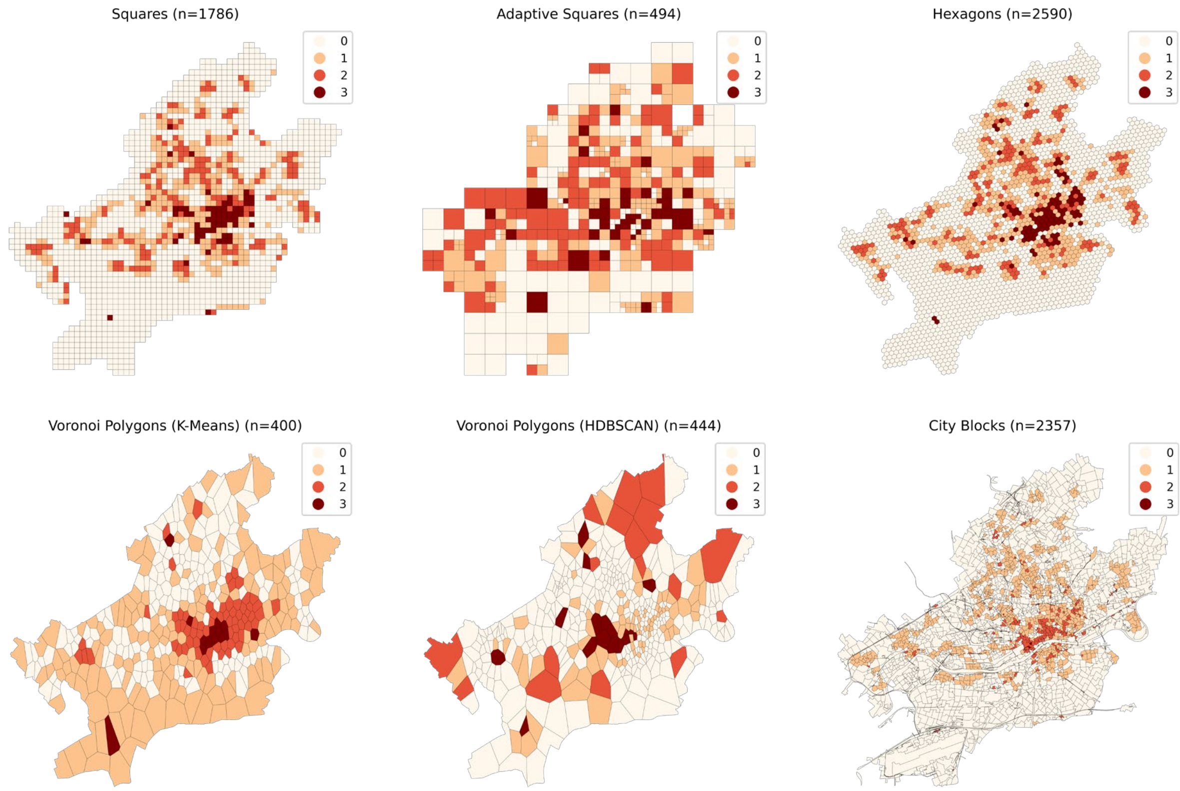

Figure 9: Clustering Results for Frankfurt am Main using different tessellation methods.

The two regular tessellation methods, square and hexagon, provide almost identical results. This is reasonable as their polygon sizes, polygon numbers, and POI distributions are similar too (Figure 8). Note that this example may not reveal all the differences between these two tessellation methods. These methods could still differ in other analyses. For example, when the contiguity of the polygons is crucial, e.g., when spatial autocorrelation is considered. The adaptive squares method produces rather similar results with significantly fewer clustering units (LGUs). This results in a lower computation time (while clustering) in comparison to squares and hexagons. The key insight is that near the city center where the variation of POI is higher, the polygons are small and the results are as detailed as regular squares, but in other parts where there are not many POI (e.g., the part of the city in the south with vegetation) polygons are larger, which reduces computation time.

Voronoi diagram seems to be well suited for finding the city center as a cluster of contiguous LGUs. A difference of Voronoi results with other methods is that the LGUs and clusters are more spatially separated. The clusters contain contiguous LGUs (especially by k-means Voronoi), unlike squares and hexagons that have clusters with spatially mixed LGUs.

At first sight, the city blocks method leads to the best results. The borders between clusters, which are defined by road segments, seem more realistic. Part of the airport is assigned to the cluster 3 by all six methods. It is a single square in case of the regular squares methods, a relatively big square in case of the adaptive squares method, two hexagons in case of the hexagon method, and random polygons in case of the Voronoi methods. However, in case of the city blocks methods, exactly the building in which the airport shops are located is identified. This is an example for a very accurate discretization. In contrast to the other methods, which have four visible clusters, cluster 3 in the city block result is not clear at first sight. An interesting finding is indicated by a deeper look into the results: The cluster 3 in the city block method contains only a couple of LGUs; the building in the airport where the shops are located and the central train station where also its shops are located.

In addition, the other clusters based on city blocks methods are easy to interpret. Cluster 0 shows the area predominantly covered with vegetation, cluster 1 represents the residential areas and cluster 2 has the characteristics of the city center. The sparse LGUs belonging to cluster 2, which can be found everywhere in the city, represent the local centers of different city districts. A possible reason that the clustering shows a better performance for LGUs based on the city blocks method could be that it is the only method that uses both POI data and road network data for creating and preprocessing the LGUs.

In summary, all these approaches can work well depending on the use case. The POI per LGU distribution can help to evaluate the performance of different LGU approaches and the chosen settings. POI-based tessellations are suitable when it is desired to have control over the number of POI per LGUs and the distribution of POI have a strong impact on the outcome.

7 Conclusion

Regular tessellations (squares and hexagons) can be easily implemented and require little computational effort. They are suitable for basic applications or when uniform/congruent LGUs are needed. Example use cases are rainfall (Goovaerts, 2000) or pollution (Janssen et al., 2008) in urban areas. These methods could also be used for spatial interpolation since most interpolation methods use mesh grids. A square-based tessellation can be seen as a kind of mesh grid (Goovaerts, 2000; Janssen et al., 2008). The studied POI-based methods are more sophisticated and complex to implement. They have the advantage of being adaptable to urban areas by considering non-spatial attributes. Adaptive Squares can also be used for the above-mentioned applications while allowing more complex spatial analysis and lowering the computation time. For spatial studies, POI-based irregular tessellations have shown to be suitable approaches in this paper. They also offer more control over the number of LGUs and the distribution of the number of POI per LGU. In particular, the Voronoi polygons and the city block method can generate an arbitrary number of LGUs by choosing the number of clusters in the clustering step. The distribution of POI per LGU can be adjusted by using different clustering methods before creating Voronoi polygons. Possible use cases are spatial clustering and the analysis of traffic patterns or socio-demographic structures. These methods are more complex to implement, but could provide more relevant and accurate results for specific analyses. The results of clustering the LGUs based on the methods taken the POI data into account indicate that city blocks, in particular, perform better in city segmentation in terms of reflecting the structure of the city.

All presented approaches could be generalized to (almost) any urban area since only open-source data is used. In addition, they are scalable to discretize metropolitan regions or even countries. Note, however, that a distortion of the LGUs due to the projection is inevitable in the case of discretization of large areas.

Future work should focus on the following points: First of all, the generalizability of the studied tessellation methods, e.g., to rural areas with spare POI data, should be evaluated. Furthermore, implementing other POI-based tessellation methods or extensions of the proposed ones can be valuable. Higher-order Voronoi polygons, weighted Voronoi polygons, other approaches creating city blocks, and other thresholds for adaptive square-based LGUs are possible extensions. The implementation of city blocks could be based on inhabitants instead of road networks. For example, another definition of a city block could be a polygon where a specific number of people live. Moreover, LGUs based on complex spatial clustering algorithms like DBRS (Wang and Hamilton, 2003) could be investigated. Finally, theoretical and data-driven approaches to find the optimal number of LGUs should be investigated. However, finding a general approach could be challenging as these parameters always depend on the research question, the area, and the data.

Funding

This work is funded by the state of Hesse and HOLM funding under the “Innovations in Logistics and Mobility” measure of the Hessian Ministry of Economics, Energy, Transport and Housing. [HA Project No.: 1017/21-19]

References

Alvares F, Cleary F, Daras P, et al. (2011) *The Future Internet: Future Internet Assembly 2012: From Promises to Reality*. Springer.

Ammann R, Grünbaum B and Shephard GC (1992) Aperiodic tiles. *Discrete & Computational Geometry* 8(1): 1–25.

Asamer J, Reinthaler M, Ruthmair M, et al. (2016) Optimizing charging station locations for urban taxi providers. *Transportation Research Part A: Policy and Practice* 85: 233–246.

Aurenhammer F (1991) Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)* 23(3): 345–405.

Bełej M and Figurska M (2020) 3D Modeling of Discontinuity in the Spatial Distribution of Apartment Prices Using Voronoi Diagrams. *Remote Sensing* 12(2): 229.

Biazzo I, Monechi B and Loreto V (2019) General scores for accessibility and inequality measures in urban areas. *Royal Society open science* 6(8).

Boontore A (2011) Stability tests of urban physical form indicators: The case of European cities. *Procedia-Social and Behavioral Sciences* 21: 336–345.

Briant A, Combes P-P and Lafourcade M (2010) Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics* 67(3): 287–302.

Brodsky I (2018) H3: Uber's Hexagonal Hierarchical Spatial Index. Available at: https://eng.uber.com/h3/.

Ester M, Kriegel H-P, Sander J, et al. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD*.

Feick R and Robertson C (2015) A multi-scale approach to exploring urban places in geotagged photographs. *Computers, Environment and Urban Systems* 53: 96–109.

Fleischmann M, Feliciotti A, Romice O, et al. (2020) Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale. *Computers, Environment and Urban Systems* 80: 101441.

Frias-Martinez V and Frias-Martinez E (2014) Spectral clustering for sensing urban land use using Twitter activity. *Engineering Applications of Artificial Intelligence* 35: 237–245.

Gerya T (2019) *Introduction to Numerical Geodynamic Modelling*. Cambridge University Press.

Gold C (2016) Tessellations in GIS: Part I—putting it all together. *Geo-spatial Information Science* 19(1): 9–25.

Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* 228(1-2): 113–129.

Graser A (2017) Tessellating Urban Space based on Street Intersections and Barriers to Movement. *GI_Forum* 1: 114–125.

Griffith DA (2008) Geographic sampling of urban soils for contaminant mapping: how many samples and from where. *Environmental geochemistry and health* 30(6): 495–509.

Haklay M (2010) How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and planning B: Planning and design* 37(4): 682–703.

He Y, Ai T, Yu W, et al. (2017) A linear tessellation model to identify spatial pattern in urban street networks. *International Journal of Geographical Information Science* 31(8): 1541–1561.

Healey P (2006) *Urban complexity and spatial strategies: Towards a relational planning for our times*. Routledge.

Hochmair HH, Juhász L and Cvetojevic S (2018) Data Quality of Points of Interest in Selected Mapping and Social Media Platforms. *LBS 2018: 14th International Conference on Location Based Services*: 293–313.

Janssen S, Dumont G, Fierens F, et al. (2008) Spatial interpolation of air pollution measurements using CORINE land cover data. *Atmospheric Environment* 42(20): 4884–4903.

Jekel T (ed) (2012) *GI_Forum 2012: Geovisualisation, society and learning: Conference proceedings*. Berlin, Offenbach: Wichmann.

Li M and Stefanakis E (2020) Geospatial Operations of Discrete Global Grid Systems—a Comparison with Traditional GIS. *Journal of Geovisualization and Spatial Analysis* 4(2).

McInnes L, Healy J and Astels S (2017) hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software* 2(11): 205.

McKenzie G and Romm D (2021) Measuring urban regional similarity through mobility signatures. *Computers, Environment and Urban Systems* 89: 101684.

Okabe A, Boots B, Sugihara K, et al. (2000) *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Oliveira V (2016) *Urban Morphology: an introduction to the study of the physical form of cities*. Springer.

OpenStreetMap Wiki contributors Stats - OpenStreetMap Wiki. Available at: https://wiki.openstreetmap.org/w/index.php?title=Stats&oldid=2172694 (accessed 29 November 2021).

Peterson PR (2016) Discrete Global Grid Systems. *International Encyclopedia of Geography: People, the Earth, Environment and Technology*: 1–10.

Portugali J and Stolk E (2016) Complexity, Cognition, Urban Planning and Design. *Post-Proceedings of the 2nd Delft International Conference. Switzerland: Springer*. DOI: 10.1007/978-3-319-32653-5.

Sadahiro Y (2002) An Exploratory Method for Analyzing a Spatial Tessellation in Relation to a Set of other Spatial Tessellations. *Environment and Planning A: Economy and Space* 34(6): 1037–1058.

Sahr K (2014) *Central Place Indexing: Optimal Location Representation for Digital Earth*.

Sahr K, White D and Kimerling AJ (2003) Geodesic Discrete Global Grid Systems. *Cartography and Geographic Information Science* 30(2): 121–134.

Samsonov TE, Konstantinov PI and Varentsov MI (2015) Object-oriented approach to urban canyon analysis and its applications in meteorological modeling. *Urban Climate* 13: 122–139.

Smucker NJ, Kuhn A, Charpentier MA, et al. (2016) Quantifying Urban Watershed Stressor Gradients and Evaluating How Different Land Cover Datasets Affect Stream Management. *Environmental management* 57(3): 683–695.

Wang X and Hamilton HJ (2003) DBRS: A Density-Based Spatial Clustering Method with Random Sampling. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*: 563–575.

White D and Kiester AR (2008) Topology matters: Network topology affects outcomes from community ecology neutral models. *Computers, Environment and Urban Systems* 32(2): 165–171.

Winter S and Goel S (2021) *Smart Parking in Fast-Growing Cities. Challenges and Solutions*. TU Wien Academic Press.

Xing J, Sieber R and Roche S (2020) Rethinking Spatial Tessellation in an Era of the Smart City. *Annals of the American Association of Geographers* 110(2): 399–407.

Xinqi Z, Shuqing W, Meichen F, et al. (2010) Urban Cluster Layout Based on Voronoi Diagram. *Advanced Techniques in Computing Sciences and Software Engineering*: 13–18.

Yuan NJ, Zheng Y and Xie X (2012) Segmentation of Urban Areas Using Road Networks. *Microsoft Technical Report*(MSR-TR-2012-65).

Yuan Y and Raubal M (2012) Extracting Dynamic Urban Mobility Patterns from Mobile Phone Data. *International conference on geographic information science*: 354–367.

Zheng Y, Liu Y, Yuan J, et al. (2011) Urban Computing with Taxicabs. *Proceedings of the 13th international conference on Ubiquitous computing*: 89–98.

Bisher erschienene Working Papers des Fachbereich 3:

Nr. 22 Balioamoune, Mina; Bausony, Mohamed A.K.; Lutz, Stefan; K.A. Mohamed, Ehab (2022): International Ownership and SMEs in Middle Eastern and African Economies. DOI: https://doi.org/10.48718/c5sy-3h14

Nr. 21 Graf, Erika; Franz, Catharina; Rugbarth, Matthias; Schmidt, Leonard (2021): Wie beeinflussen die Namen von Fleischersatzprodukten die Akzeptanz der Verbraucher? DOI: https://doi.org/10.48718.smsh-p640

Nr. 20 Hagen, Tobias; Saki, Siavash; Scheel-Kopeinig, Sabine (2021): start2park – Determining, Explaining and Predicting Cruising for Parking.

Nr. 19 Graf, Erika (2021): Banken auf dem Holzweg? Eine empirische Untersuchung der Bewertung von Kreditkarten aus Holz.

Nr. 18 Ziegler, Yvonne; Uli, Vincenzo Uli; Kramer, Astrid; Tatari, Mahmoud Tatari (2021): Development of an innovative halal logistics concept for the air cargo supply chain.

Nr. 17 Hagen, Tobias; Scheel-Kopeinig, Sabine (2020): Would Customers be willing to use an alternative (chargeable) delivery concept for the last mile?

Nr. 16 Giegler, Nicolas; Schneider, Swen (2020): Leadership und Digitalisierung.

Nr. 15 Graml, Regine; Hagen, Tobias; Ziegler, Yvonne; Khachatryan, Kristine; Astrida Herman, Ricky (2020): Lesbische Frauen in der Arbeitswelt – The L-Word in Business.

Nr. 14 Voigt, Martina; Ruppert, Andrea (2018): Follow-up –Studie: Gendertypische Verhandlungskompetenz und ihre Auswirkungen auf Gehalts-und Aufstiegsverhandlungen.

Nr. 13 Celebi, Kaan; Hönig, Michaela (2018): Dynamic Macroeconomic Effects on the German Stock Market before and after the Financial Crisis.

Nr. 12 Lutz, Stefan (2018): R&D, IP, and firm profits in the North American automotive supplier industry.

Nr. 11 Voigt, Martina; Ruppert, Andrea (2016): Durchsetzungsorientiert, hart oder sachorientiert, kooperativ verhandeln –welche Einschätzungen und Erwartungen haben Studierende?

Nr. 10 Hagen, Tobias (2016): Econometric Evaluation of a Placement Coaching Program for Recipients of Disability Insurance Benefits in Switzerland.

Nr. 9 Weissenrieder, Caprice Oona; Spura, Anastassja (2015): Akzeptanz von Führungs-kräften- Analyse wahrgenommener Verhaltensweisen von Frauen und Männern in Führungspositionen.

Nr. 8 Graf, Erika (2015): Raising Sustainability Awareness and Understanding in Higher Education.

Nr. 7 Weissenrieder, Caprice Oona; Graml, Regine; Hagen, Tobias; Ziegler, Yvonne (2015): Explorative Untersuchung der Unternehmenskultur auf die Karrierechancen von Frauen.

Nr. 6 Ruppert, Andrea; Voigt, Martina (2014): Verhandlungsstrategien und Verhandlungs-taktiken in Gehaltsverhandlungen.

Nr. 5 Jungmittag, Andre (2014): Combination of Forecasts across Estimation Windows: An Application to Air Travel Demand.

Nr. 4 Jungmittag, Andre (2014): Der Trade-off zwischen deutschen Direktinvestitionen und Exporten: Wie wichtig sind die verschiedenen Dimensionen der Distanz?

Nr. 3 Hagen, Tobias; Waldeck, Stefanie (2014): Using Panel Econometric Methods to Estimate the Effect of Milk Consumption on the Mortality Rate of Prostate and Ovarian Cancer.

Nr. 2 Hagen, Tobias (2014): Impact of National Financial Regulation on Macroeconomic and Fiscal Performance After the 2007 Financial Shock Econometric Analyses Based on Cross-Country Data.

Nr. 1 Rieck, Christian; Bendig, Helena; Hünnemeyer, Julius; Nitzsche, Lisa (2012): Diversität im Aufsichtsrat. Studie über die Zusammensetzung deutscher Aufsichtsräte.

Alle Publikationen des Fachbereich 3 Wirtschaft und Recht finden Sie unter:

https://www.frankfurt-university.de/de/hochschule/fachbereich-3-wirtschaft-und-recht/forschung-und-transfer/publikationen/