

Chapter 10

Improvement of Task-Oriented Visual Interpretation of VGI Point Data



Martin Knura and Jochen Schiewe

Abstract VGI is often generated as point data representing points of interest (POIs) and semantic qualities (such as accident locations) or quantities (such as noise levels), which can lead to geometric and thematic clutter in visual presentations of regions with numerous VGI contributions. As a solution, cartography provides several point generalization operations that reduce the total number of points and therefore increase the readability of a map. However, these operations are applied rather general and could remove specific spatial pattern, possibly leading to false interpretations in tasks where these spatial patterns are of interest. In this chapter, we want to tackle this problem by defining task-oriented sets of map generalization constraints that help to maintain spatial pattern characteristics during the generalization process. Therefore, we conduct a study to analyze the user behavior while solving interpretation tasks and use the findings as constraints in the following point generalization process, which is implemented through agent-based modeling.

Keywords Point generalization · Constraints · Agent-based modeling

10.1 Introduction

As shown by the variety of different aspects and applications which are observed in this book, the volume and relevance of Volunteered Geographic Information (VGI) have immensely increased in recent years. In many cases, this VGI data is generated and visualized as point data, e.g., representing the location of a point of interest (POI), an event, or a data source. However, utilizing VGI data needs to take some specific characteristics into account in comparison to geospatial data acquired and processed in the “traditional” way. In particular, VGI “is produced by heteroge-

M. Knura (✉) · J. Schiewe
HafenCity University Hamburg, Hamburg, Germany
e-mail: martin.knura@hcu-hamburg.de; jochen.schiewe@hcu-hamburg.de

© The Author(s) 2024
D. Burghardt et al. (eds.), *Volunteered Geographic Information*,
https://doi.org/10.1007/978-3-031-35374-1_10

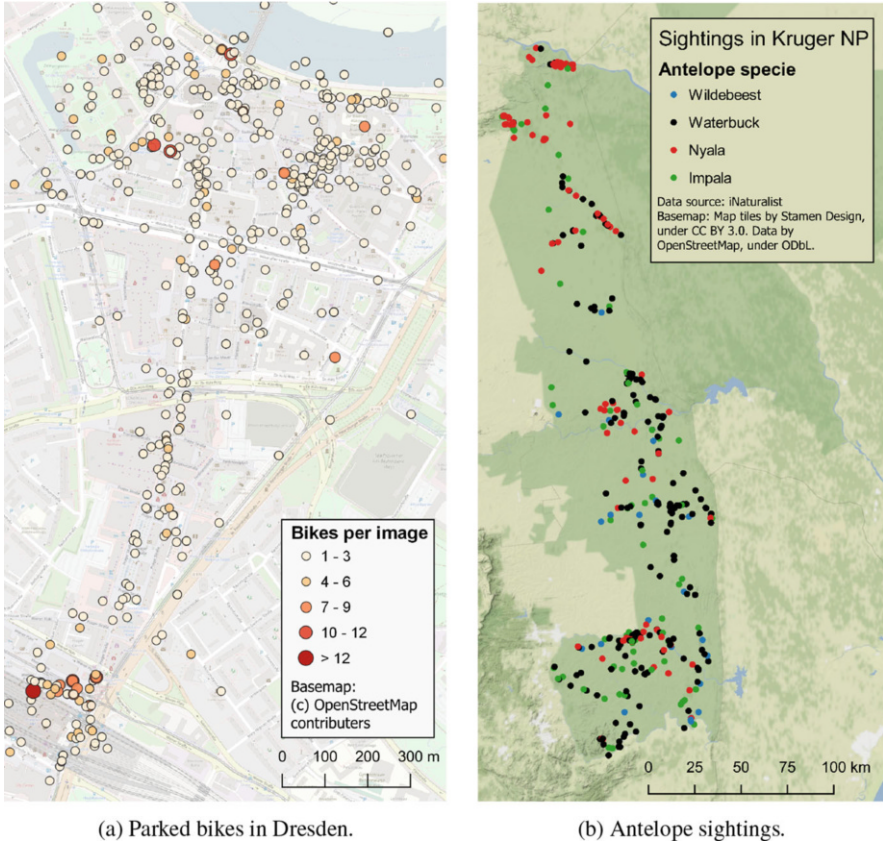


Fig. 10.1 Examples for VGI point data with point clutter. (a) Parked bikes in the city of Dresden, detected on Flickr images between 2004 and 2014 according to Knura et al. (2021). (b) Sightings of selected antelope species in Kruger National Park uploaded on the platform iNaturalist

neous contributors, using various technologies and tools, having different levels of details and precision, serving heterogeneous purposes, and a lack of gatekeepers” (Senaratne et al. 2017), leading to an enormous volume and heterogeneity within the data. All of these characteristics could harm the usability of the data, especially when it comes to the visual presentation and exploration of very dense and even overlapping point markers or symbols (see Fig. 10.1a and b), commonly known as geometric point clutter (Moacdieh and Sarter 2015).

As a solution to this clutter problem, cartography provides several point generalization operations such as selection, aggregation, or displacement, which rearrange or reduce the total number of points and therefore increase the readability of a map. However, these operations are applied rather general and could remove a specific spatial pattern, possibly leading to false interpretations in tasks where these spatial patterns are of interest.

The aim of the TOVIP project is to tackle this problem by defining a set of cartographic constraints—i.e., conditions a generalized map should satisfy—that preserve these spatial patterns throughout the whole generalization process. The first research question we want to answer in this chapter is therefore:

- What is the minimum set of constraints and constraint measures that should be used to evaluate interpretation tasks based on VGI point visualizations, such as pattern identification, pattern comparison, or relation seeking?

Different cartographic constraints often describe contradicting aspects with no optimal solution, as it is not possible during map generalization to maintain all information—i.e., fulfill all information preservation constraints—while keeping the map readable, i.e., fulfill all legibility constraints. Constraint-based generalization is therefore an optimization task, which tries to find a solution that satisfies as many constraints as good as possible, and has been implemented in recent years through multi-agent systems (Duchêne et al. 2018). We want to contribute to this research and define our second research question as:

- Is it possible to optimize the task-oriented generalization using an agent-based modeling approach?

The following chapter describes the workflow to answer the research questions as follows: in Sect. 10.2, we introduce the cartographic concept of constraint-based generalization, on which the TOVIP project is based upon. Section 10.3 summarizes the results of a user study, which analyzed the user behavior while working with spatial patterns in point data sets. In Sect. 10.4, we translate the findings of the previous section into measurable constraints that could be utilized in map generalization practice. In Sect. 10.5, we apply these constraints in an agent-based generalization model. That followed, we discuss our findings in Sect. 10.6 before concluding in Sect. 10.7.

10.2 Constraint-Based Map Generalization

Cartography provides a variety of different point generalization operations—and various combinations between them—to solve the aforementioned clutter problem. As an example, a *simplification* describes a straight reduction of source points based on geometric criteria (e.g., only points which have a minimum distance to their neighbors are preserved; (Slocum et al. 2009)). When semantic criteria are used, a *selection* operation could take place. For example, points can be selected based on respective information filtering methods (Huang and Gartner 2012) or scale-dependent (Gröbe and Burghardt 2021). *Aggregation* takes place when multiple points are replaced by a single aggregator marker. Most frequently, points are grouped through clustering with a respective initialization method (e.g., random, k-means, Voronoi-based; (Yan and Weibel 2008)), while alternatives, for example, use heat maps (Meier 2016), or geometry objects (Zahtila and Knura 2022) to aggregate

point data. A different solution to overcome clutter, but with the possibility to preserve the cardinality (i.e., the overall number of points) of the dataset, is to *displace* the points. During the displacement operation, an iterative workflow of overlap detection, relocation, and re-evaluation is executed (Mackaness and Purves 2001). Furthermore, if the preservation of cardinality and the original topology is of interest, a *spatial distortion* based on pixels (Keim et al. 2004) or point density (Bak et al. 2009) could help.

An operational system for point generalization must implement a workflow to trigger and orchestrate the individual point generalization operations described above. First implementations used a rule-based approach where a predefined set of well-defined and unambiguous rules guided the generalization process (Beard 1991). Each rule thereby states what has to be done in a process at a certain condition, so each condition was connected to a specific action (Harrie and Weibel 2007). The problem that occurred with this approach was that the enormous variety of spatial and non-spatial characteristics that exist in the world and therefore in maps led to a number of rules which were not possible to handle anymore. This leads to the constraint-based approach, which focused on the requirements that the final map should fulfil instead of providing a set of isolated generalization operations, leaving more flexibility within the generalization process on how to reach these results. According to Beard (1991), these constraints can be classified into aspects related to position, topology, shape, structural, functional, and legibility. Furthermore, it is necessary to introduce respective measures for these constraints, which are grouped by Mackaness and Ruas (2007) into either internal or external and either micro, meso, or macro.

If the constraints are defined in a complete and measurable way, there are different techniques available for implementation. Looking at optimizing single generalization methods, there is considerable work done, for example, regarding the displacement operation by applying least squares adjustment (Sester 2000), simulated annealing (Ware and Jones 1998), or snakes (Burghardt 2005). For more complex processes, agent-based modeling has shown great success in terms of applicability (Duchêne et al. 2018). In this approach, agents represent autonomous map objects trying to minimize a given cost function, which is based on the fulfillment of the constraint measures. As a result, the whole complexity of the generalization workflow is distributed to a set of relatively simple interacting agents.

The agent-based modeling approach is also used in the TOVIP project. Regarding the aim of TOVIP—defining a set of constraints that optimizes the generalization workflow designed for visual interpretation tasks where specific spatial patterns are of interest—it is necessary to consider two potentially contradictory aspects: On the final map, the aforementioned spatial patterns have to be visible (*preservation constraints*), while the map must still be readable by the users (*legibility constraints*). Describing constraints that preserve the relevant information during the generalization process is often done with object-specific measures, e.g., preserving the area of a polygon before and after generalization (Harrie and Weibel 2007). On the other hand, legibility constraints ensure the readability of the map, for example, by avoiding any spatial conflict—i.e., display clutter—and showing objects in

a suitable degree of detail according to the scale of the map. A respective list of analytical legibility measures, such as the number of vertices or the object line length, was developed by Stigmar and Harrie (2011). For the aim of the TOVIP project, it is now of interest to find the minimum set of preservation and legibility constraints that allow the interpretation of specific spatial pattern even after generalization.

10.3 User Behavior When Interpreting VGI Point Data

Developing constraints that support users while interpreting specific tasks implies profound knowledge of their behavior while doing so. Before defining constraints that support interpretation tasks, it is therefore necessary to analyze the behavior of users working with VGI point data sets. We conduct a user study where participants have to perform different interpretation tasks—like finding clusters within a dataset, comparing point densities, or finding areas with a specific point distribution—using a novel method that combines postal questionnaires, think-aloud interviews, and techniques from visual analytics. A more detailed overview on the technical aspects and the execution of the user study, including a detailed description of the analysis of the think-aloud interviews, is given by Knura and Schiewe (2021). In this chapter, we want to summarize the results of the study (see also Knura and Schiewe 2022), focusing on the impact of the user behavior on the definition of a minimum set of constraints as described above.

10.3.1 Task-Solving Strategies

We analyzed the strategies of the participants by dividing the overall task-solving process into three sequential actions: (1) finding a start position, (2) obtaining information, and (3) decision-making. Apart from a task where the participants have to find a similar pattern compared to a given reference, the point density of a cluster—as a combination of proximity and cardinality of points—was the most important factor when selecting a starting position on the map, followed by the point color. For the process of obtaining information, point density was again the most important factor, as more dense clusters were described and analyzed earlier and more often. Moreover, density was the main evaluation measure in comparison tasks and during decision-making. Although we had different categories of synoptic interpretation tasks, which—in contrast to elementary tasks—include pattern identification, pattern comparison, and relation seeking (see Andrienko and Andrienko 2006), the task-solving strategies did not differ significantly between different kinds of tasks. As a first result of the study, we state that point density has the biggest impact on the task-solving behavior of the participants and has to be addressed in the first place when defining constraints.

10.3.2 Influence of Point Data Cardinality and Background Map

A key factor that could have an impact on the user behavior during visual interpretation is the map complexity. There are numerous definitions and concepts of map complexity in cartography (see Touya et al. 2016). As most of them distinct between the intellectual complexity, which relates to the cognitive process of map reading, and the graphical complexity, which relates to the visual perception of individual map objects, we vary the maps for some of the tasks with respect to these two categories. To learn more about the influence of the intellectual complexity on user behavior, we varied the data cardinality—i.e., the number of points—for two of the tasks. Although we recognized some minor differences in the behavior between the user groups, the overall task execution strategy remains unchanged with a higher data cardinality. For analyzing the impact of the graphical complexity, we varied the background map source between Google Maps, Bing Maps, and Stamen Terrain. This time, we identified both an implicit and explicit influence from the background map. Implicitly, because participants frequently identified clusters which were visually supported by the background map and explicitly because they refer to the characteristics of the background map when explaining their strategies. But again, and despite the influence of the background map on the reasoning, the overall task-solving strategies described in the section above remain unchanged between different levels of graphical complexity.

10.3.3 Implications for Constraints Supporting Interpretation Tasks

Following the results of our study, there are two main aspects that have to be considered while defining constraints for map generalization. First, it is of major importance to preserve the original pattern proportions during the generalization process. Our study revealed that the point density had the biggest impact on the task-solving process, and participants discussed both interrelations between clusters with different density, as well as between different classes of points within the same cluster. Information preservation constraints regarding the point density should therefore:

- Retain the proportion of points between areas with different densities
- Preserve the ranking of densities between different areas
- Preserve proportions between classes while maintaining at least one point per class
- Preserve Gestalt law rules regarding similarity and proximity of clusters

The second aspect to consider is the use of cartographic techniques to guide the interpretation of points. The use of specific colors to draw attention is common in

cartography, and this can be applied to other map objects with the aim of lowering the graphical complexity. Respective constraints could ensure to:

- Use cartographic style elements where pattern preservation is difficult to manage
- Optimize the guiding effect of the background map (e.g., preservation of other map objects in close proximity to point clusters)

These constraints could be categorized as both preserving information and legibility, and they address not only the location and visibility of the point symbols but also their style and the surrounding map areas.

10.4 Defining Constraints and Measures for Spatial Pattern Interpretation

The previous section revealed the importance of preserving point densities during generalization. In this section, we collect a list of different approaches—both from the literature and own experiments—to define constraints and respective measures, which can help to preserve point densities and spatial pattern and test them on exemplary point distributions. The aim is thereby to find a minimum set of constraints that fit best to the list of requirements described above. We thereby focus on information preservation constraints regarding the point distributions. Constraints related to cartographic techniques are a key aspect of our future work.

10.4.1 Measures Describing Spatial Pattern and Densities

When defining measures for spatial pattern and densities, we follow the categorization of Mackaness and Ruas (2007), who distinguish between macro-measures that deal with all point objects of interest, micro-measures that deal with individual characteristics of objects (i.e., points), and meso-measures that deal with the specific properties of different groups of objects (i.e., point clusters). The authors also distinguish between internal and external measures, which states if a measure can be calculated based on a single dataset (internal) or is a relation between two datasets (e.g., before and after a generalization operation; external).

10.4.1.1 Macro-Measures

Macro-measures are able to describe the entirety of information and respective characteristics in a single value. One of the most basic macro-measures is the radical law (Töpfer and Pillewizer 1966), which estimates how many features should be

maintained at a smaller scale in the generalization process. It is defined as:

$$n_D = n_S \sqrt{\frac{m_S}{m_D}}, \quad (10.1)$$

where n is the number of objects of the derived ($_D$) resp. source ($_S$) map, and m is the scale denominator. In the context of this project, it is worth noting that the calculation should be based on a readable map, i.e. without any point clutter. If this measure is calculated from a map with point clutters, the calculated value should be interpreted as the **maximum number of objects** on the derived map. Even more basic is the measure that describes the **amount of information** N_i as the number of all map objects (Harrie and Stigmar 2010), calculated as:

$$N_i = \sum_{i=1}^n \sum_{j=1}^{m_i} O_{ij}. \quad (10.2)$$

For objects other than points, this measure can be expanded with the number of object points, calculating the overall measure as the sum of all object points of all map objects.

Beside measures that deal with the amount of information in general, global measures can also describe a specific characteristic of the dataset or the map. In the same work, Harrie and Stigmar (2010) defined an index to characterize the **spatial distribution of points** I_{SDP} based on Voronoi regions. The index is calculated as:

$$I_{SDP} = \frac{\sum_{i=1}^{NP} P_{SDP,i} \log P_{SDP,i}}{\log \frac{1}{NP}}, \quad (10.3)$$

where $P_{SDP,i}$ is the relative size of the Voronoi region for a point i and NP is the number of points. I_{SDP} converges to 1 the more even the sizes of the Voronoi regions are. Zhang et al. (2009) use the **Voronoi region size as the variable of interest in Moran's I** to discern if point distributions are clustered, dispersed, or random:

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum_{i=1}^N (x_i - \bar{x})^2}, \quad (10.4)$$

where N is the number of spatial units indexed by i and j , x is the size of the Voronoi region A_V , \bar{x} is the mean of x , w_{ij} is a matrix of spatial weights with zeroes on the diagonal (i.e., $w_{ii} = 0$), and W is the sum of all w_{ij} .

10.4.1.2 Micro-Measures

Micro-measures describe characteristics of individual objects and therefore can take the local neighborhood into account. Analog to the macro-measures before, calculating the size of the Voronoi region of a point A_V is used as a fundamental metric to describe local density. Based on this, Zhang et al. (2008) calculate the *object-oriented density* OD as:

$$OD = \frac{1}{A_V}. \quad (10.5)$$

A higher object-oriented density implies a smaller Voronoi region and therefore a higher point density in the local neighborhood. Vice versa, a small object-oriented density indicates a bigger Voronoi area and a more dispersed distribution around that point.

Besides density measures, qualitative and quantitative information about the points in close proximity are also of interest when point generalization operations like selection are used. Therefore, Delauney triangulations are often used to identify “natural” neighbors in point distributions (Sadahiro 1997). Applying this tessellation to a point data set provides a *list of neighbors* for each point, and micro-measures like the *number of natural neighbors*, the *mean neighbor distance*, and the existence of *local extreme values* can be calculated (see Fig. 10.2). Delauney triangulation also helps to define clusters, so the *cluster affiliation* can also be defined in this way.

All the measures introduced in this section are internal because they can be calculated solely based on one dataset. However, it is possible to compare the measures of an individual point to measures of the same point during the

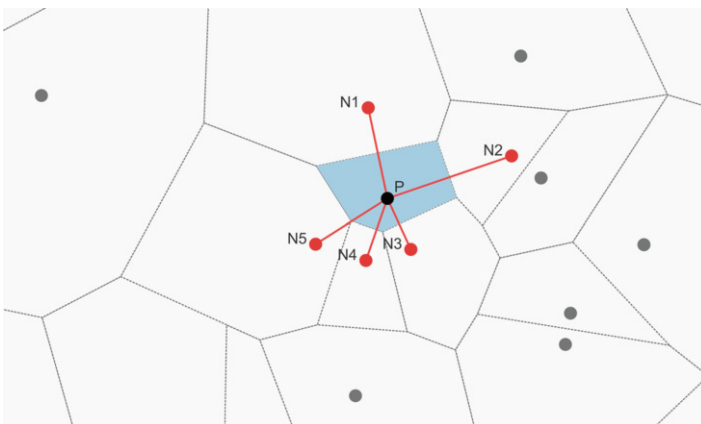


Fig. 10.2 Example of micro-measures for point P. Natural neighbors N1 to N5 in red, other points in gray. Voronoi region A_V for point P in light blue

generalization process, and note the amount of change as an additional external measure. In the same way, the *distance to the origin location* of a point is of interest when displacement operations take place during the generalization.

10.4.1.3 Meso-Measures

Compared to macro- and micro-measures, meso-measures are not bound to a predefined number or list of points they describe. The first step to calculating meso-measures is therefore to define which points are members of the group of interest. In the TOVIP project, we focus on spatial pattern, and so the definition of clusters is relevant for further processing. Clustering can be made—among many other techniques—by cluster algorithms such as k-means and HDBscan, based on Delauney triangulation (Sadahiro 1997), or by using grids (Yan et al. 2021). These clusters can then be described by meso-measures such as the *number of group members*, the *existence of different point categories*, and the *mean distance* between members or between members and the group centroid. Comparing the measures of the respective clusters, it is also possible to define *cluster rankings*. Furthermore, according to the findings of the user study presented in Sect. 10.3, measures regarding the shape and the orientation of the clusters can be of interest. Common methods to represent *the shapes of point clusters* are convex hulls or alpha shapes (Edelsbrunner et al. 1983). The *orientation of a cluster* can be described by the minimum rotated rectangle, a technique which is usually used for building orientation (Duchêne et al. 2003). Furthermore, all macro-measures defined in Sect. 10.4.1.1 can also be applied on clusters with a defined border.

10.4.2 Deriving a Minimum Set of Constraints

Based on the list of different measures, we test the suitability of the measures to control the different aspects which help to fulfil the information preservation constraints we developed in Sect. 10.3. We thereby subdivide the constraints and respective measures into three groups:

1. Measures describing the overall distribution of points and the density ranking between different areas of the map
2. Measures preserving pattern-specific characteristics like hot spots, extreme values, cluster density, etc.
3. Measures describing Gestalt law rules

Furthermore, we compare the measures and their performance on different point distributions to identify redundancies, and we examine the robustness on point cardinality, which is essential when applied in map generalization operations. We create a series of experimental point distributions with 100, 200, 500, and 1000 points and different characteristics: a regular and a random distribution, distributions

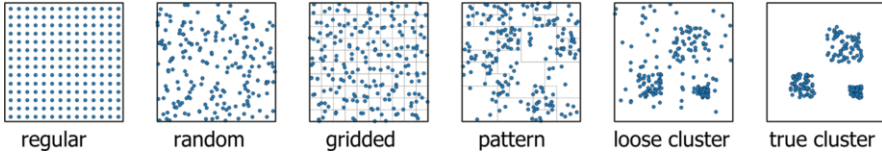


Fig. 10.3 Point distributions with different characteristics, using the example of 200 points. Gray borders in the background of the gridded and pattern distributions indicate the areas in which the same number of points were randomly distributed

where we predefined regular (gridded distribution) and irregular areas (pattern distribution) to control the density, and distributions with loose and clear clusters (see Fig. 10.3). Furthermore, we tested the behavior of micro-measures on points in different VGI datasets to evaluate their utility.

10.4.2.1 Preserve the Overall Distribution of Points and the Density Ranking Between Areas

The first subgroup of measures combines the first two constraints on information preservation in Sect. 10.3.3 and can be controlled through a combination of macro- and meso-measures. We tested both the Voronoi-based Moran’s I and the spatial distribution of points with our series of different artificial point distributions. Table 10.1 shows the calculated measures and the standard deviation over the different point cardinalities. We can see that the spatial point distribution measure is to a certain degree stable toward the point cardinality and is smaller when the points

Table 10.1 Results for point distribution measures. Values with (*) signs indicate that there was a small deviation to the given number of points because of distribution characteristics (e.g., 196 instead of 200 points for the regular distributions)

Point distribution measure						
Number of points	regular	random	gridded	pattern	l. cluster	t. cluster
100	1.0000	0.9673	0.9924	0.9493*	0.8624	0.7623*
200	0.9994*	0.9718	0.9904	0.9405*	0.8603	0.7481*
500	0.9987*	0.9789	0.9847	0.9394*	0.8664	0.7079*
1000	0.9991*	0.9805	0.9832	0.9391*	0.8704	0.6658*
<i>std</i>	<i>0.0005</i>	<i>0.0061</i>	<i>0.0044</i>	<i>0.0048</i>	<i>0.0044</i>	<i>0.0434</i>
Voronoi-based Moran’s I						
Number of points	regular	random	gridded	pattern	l. cluster	t. cluster
100	0.4683	0.1707	0.0032	0.1819*	0.5665	0.2733*
200	0.4372*	0.1919	-0.0196	0.2126*	0.623	0.4653*
500	0.4690*	0.1575	0.1995	0.2988*	0.6378	0.4990*
1000	0.4817*	0.2907	0.1809	0.3841*	0.2891	0.5187*
<i>std</i>	<i>0.0189</i>	<i>0.0603</i>	<i>0.1151</i>	<i>0.0911</i>	<i>0.1629</i>	<i>0.1126</i>

are more clustered. Moran's I on Voronoi regions is more sensitive to variations of point cardinality, as it uses the area size as the variable of interest—whose value decreases with more points. As a result, we decided to use the point distribution measure to control the overall distribution and the distribution within clusters. That also includes the fact that the measures for amount of information and cluster size are variables within the calculation of the spatial point distribution on the macro- and meso-level. In contrast, the cluster ranking measure has no overlap with other measures and is therefore inevitable.

10.4.2.2 Preserve Pattern-Specific Characteristics

Pattern-specific measures are a crucial part of the goals of the TOVIP project. If local extreme values and the existence of different point categories are of interest in an interpretation task, it is mandatory to preserve these points and therefore control them with related measures. As this measure requires a Delauney triangulation to define the neighborhood, respective measures that are based on this can be performed with low additional effort, even if not compulsory. As an example, the mean distance to neighbors can be calculated this way. As an alternative, the distance to all points within the predefined cluster can be used to decide which points are overlapping and thus should be a controlling measure. If a displacement operation is implemented, the distance to the origin location of a point can be of interest. For the other pattern-specific measures, we did not find a unique behavior in which we see an additional utility for our model.

10.4.2.3 Preserve Gestalt Law Rules

The maximum number of points can be utilized as a target value for the generalization process, although it is not mandatory if all legibility measures are satisfied. Measures related to the shape and orientation of clusters are utilizing common techniques from the field of geospatial analysis, such as calculating the minimum bounding rectangle, the convex hull, or the alpha shape of a point set. We compared the different approaches on different data sets and decided to use the convex hull to describe the shape, as it needs no additional parameter compared to the alpha shape and is more detailed than the rectangular bounding box. If the orientation is of interest, the longer side of the minimum bounding rectangle can be utilized.

Table 10.2 shows the selected measures which we initially implemented in the agent-based model, together with additional measures that could be relevant for certain tasks and were also recognized. Nevertheless, because most of the measures are defined in code blocks outside the actual agent-based model, it is possible to adopt measures from other scale levels during model optimization.

Table 10.2 Subgroups of measures and selected measures in column “Set”. (X) indicates the measure is not mandatory in certain applications

	Macro	Meso	Micro	Set
Overall distribution of points/cluster rankings				
<i>amount of information</i>	X			
<i>spatial distribution of points</i>	X	X		X
<i>Moran’s I on Voronoi regions</i>	X	X		
<i>cluster density ranking</i>	X			X
<i>cluster size</i>		X		
Pattern-specific characteristics				
<i>local extreme values</i>			X	X
<i>point category preservation</i>		X		X
<i>number of natural neighbors</i>		X	X	
<i>mean distance to cluster members</i>		X	X	X
<i>mean neighbor distance</i>			X	
<i>object-oriented density</i>			X	
<i>distance to the origin location</i>			X	(X)
Gestalt Law				
<i>maximum number of points</i>	X			(X)
<i>shape of a cluster</i>		X		X
<i>orientation of a cluster</i>		X		(X)

10.5 Application Using Agent-Based Modeling

The set of constraints and respective measures developed in the previous section is a key component for the implementation of a map generalization process, which preserves spatial patterns. We apply the constraint-based approach using agent-based modeling, which is a powerful method for controlling complex processes (Harrie and Weibel 2007). As the model is currently in the final phase of development, this section will focus on the architecture and parametrization we implemented: First, we introduce the software framework we use and explain the different components within the model. In the second part, we describe the integration of global map specifications and the translation of measures to a satisfaction scale, which helps the agents to better evaluate their fulfillment of constraints. Evaluation of the model results will be part of our future work.

10.5.1 *Software and Components*

We implement our agent-based model¹ using the Mesa framework (Kazil et al. 2020). Mesa is an open-source framework for creating agent-based models written in Python. It includes four core components (Model, Agent, Schedule and Space), along with additional components for analysis and visualization. Thereby, the *Model* class is the core class for creating the environment of the model using the *Space* class, initializing the agents which are objects of the *Agent* class, and orchestrating the running model through the *Scheduler* class. Applied to the process of map generalization, our model has a map area which is implemented through a continuous space—providing a high flexibility for different map scales—and contains map agents which represent objects that generalize themselves by performing generalization operations, according to the perception of their current state and their fulfillment of given constraint measures. Besides micro-agents, which represent the individual points, our model also contains meso-agents, which are generated within the model initialization and control the pattern preservation.

(Map) agents and the implementation of their decision-making process are the most complex part of an agent-based model. Duchêne et al. (2018) decomposed the “brain” of map agent into three main components: capacities, mental representation, and procedural knowledge. We followed this approach and used these components in our model (see Fig. 10.4). The capacities of our agents include the ability to perceive their surrounding space, to evaluate themselves, and to perform generalization operations. The updating process of the first two capacities is thereby provided by the *Model* class, which performs several spatial analysis operations on the totality of map objects after each simulation step and transmits the calculated measures back to the individual micro- and meso-agents. The mental representation of the agents compares their current state with the goals they are aiming at—i.e., the fulfillment of map constraints—and calculates their satisfaction. It also memorizes all previous actions the agents took and the respective outcome of it. Finally, the procedural knowledge component is the decision-making unit of the agents. Based on the agent’s constraint satisfaction and the knowledge of the past steps, it decides which operation the agent should execute in the next step.

Besides the core functions for agent-based modeling, Mesa offers functionalities for data analysis and model visualization. The *DataCollector* class of Mesa is able to record, store, and export all relevant data of the agents for further analysis. It allows us to control the mechanisms of the model, as well as tuning the decision-making process of the agents. Via the visualization components, Mesa also provides a browser-based visualization of the running model, but until now, we haven’t implemented a respective function in our model yet. Instead, we set up and run the model via Jupyter Notebook and present the generalized map in an interactive browser map.

¹ Our source code is online: <https://gitlab.com/g2lab/tovip>.

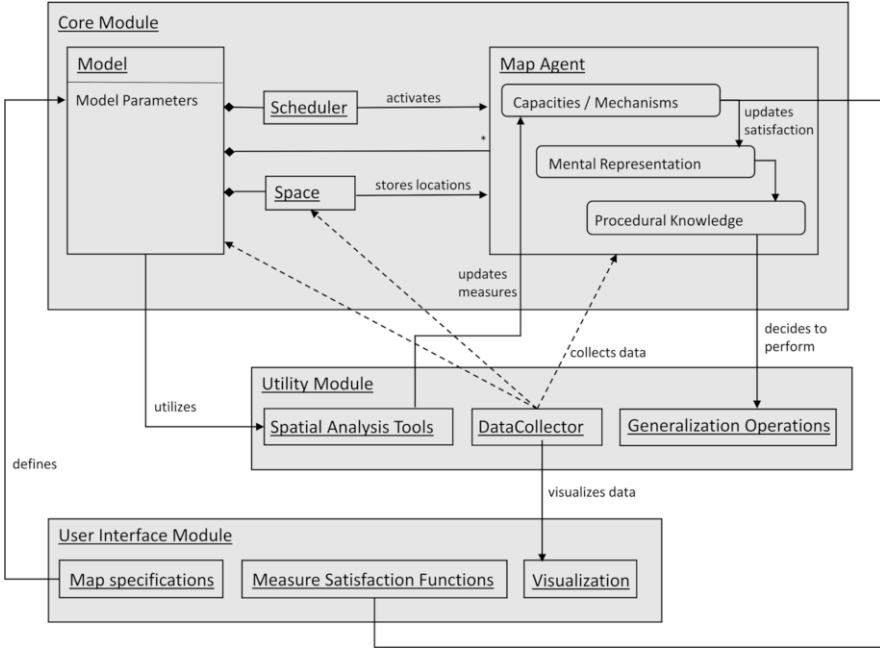


Fig. 10.4 Components of our model for agent-based map generalization

10.5.2 Global Map Specifications and Measure Satisfaction

Besides the specific constraints we defined in order to support interpretation tasks, there are also global map specifications and characteristics which are important for the process of generalization in general and for the point generalization in particular and which have to be defined in advance. For example, it is necessary to know the scale of the source map and if this map satisfies all legibility constraints regarding the point symbols (i.e., the source map has no point clutter and is readable). It is also required to define the target scale of the map and the (pixel) size of the point symbols. Moreover, it is of interest if the point data set contains different classes and, if it does, the respective scale of measurement. While these global map specifications are determined in most of the use cases for point generalization (e.g., the target map scale via predefined zoom levels), they can also be changed in the model setup.

Furthermore, the “brain” of the map agent requires determining a predefined behavior regarding the task of translating a list of measures into a value representing the satisfaction of an agent at its current status. The common workflow for this task consists of two steps (Touya 2012): First, the measures get translated into a Likert-like satisfaction scale, which ranges from 1 (“unacceptable”) to 8 (“perfect”). Each measure thereby has its own method for translation, which has to be defined in

advance. In the second step, a global satisfaction value is derived from the individual values, utilizing principles from Social Welfare Orderings (SWO). Again, the specific SWO method has to be chosen in advance and triggers different strategies the agents will follow. For example, using a SWO that emphasizes low values with a higher weight, the agents will try to minimize the number of low values and focus less on maximizing other values, while a utilitarian SWO fosters strategies where agents maximize the sum of all values.

Taillandier and Gaffuri (2012) proposed an approach to help the user with parameterization using a human-machine dialogue. We utilize this approach for our model and offer a guided user interface where parameter adjustments get visualized via samples on a map. It allows the user to adjust the satisfaction scales by modifying the class dividers, which are predefined as a function of respective global map attributes such as scale and point cardinality.

10.6 Discussion

The previous sections described the workflow of defining a set of constraints and respective measures based on the findings from a user study and the implementation in an agent-based model. This already answers our research questions to a certain degree. In this section, we want to further discuss implications that occur with the results of the previous sections.

In Sect. 10.4.2, we define a set of constraints containing measures that control the spatial distribution of points, the ranking between clusters, the shapes of clusters, and the distances between the points of a cluster. Furthermore, the preservation of local extreme values and all point categories should be added if their existence is of interest in the interpretation task. Taking the different measure scales of two of the constraints into account, there are only six to eight measures, which can be used to control the generalization process. Still, this requires at least six different predefined parameter adjustments to translate measures into satisfaction values. The complexity of the parametrization process has been identified as one of the main drawbacks of the agent-based approach (Duchêne et al. 2018), and this is also the case in our model. As an example, defining a function to evaluate the macro-measure of point distribution is little intuitive, as it needs to define class dividers for narrow-value ranges, which differences are hard to visualize. However, six parameters and a user-friendly way to adjust them are still feasible in our opinion.

Manual parameter adjustment is one reason which makes it difficult to transfer our approach of point generalization to other applications. The second reason is the time-consuming calculation of measures that rely on rather complex geospatial operations. On-the-fly point generalization (Jabeur et al. 2006) is therefore not possible with our approach, and the computing time depends heavily on the number of points to generalize. A solution to this problem could be the integration of novel learning techniques (Touya et al. 2019). If a model can learn how to generalize

points while preserving the right information, it could predict a generalized point set on-the-fly.

10.7 Summary and Outlook

The generation of VGI data in general, and of points in particular, has shown an immense increase in recent years. As one of the main properties of VGI is its enormous volume and heterogeneity within the data, it leads to dense clutters when it is presented on maps. The cartographic solution to this problem is point generalization: rearranging or reducing the number of points. If this is applied rather general, specific spatial pattern could be eliminated—which is a major problem when these patterns were of high importance and subject of interest to the user. This chapter presents a workflow to resolve this problem by defining a set of constraint that can be used to control the generalization workflow. We developed the list of constraints based on a user study and applied them by implementing an agent-based model for point generalization.

VGI point data is often produced in multiple scale levels and over longer periods of time. In our future work, we want to factor this and develop our model further by adding functionalities for *multi-scale views*, which requires consideration of scale transitions, and *multi-temporal representations*, where the cognitive workload related to animations must also be considered. Furthermore, we plan to improve the usability of our agent-based model by developing a more intuitive user interface, which would allow more users to apply the findings of our model to their objectives. A third task we plan to work on in the near future is the integration of other cartographic techniques such as point color, point symbolization, and others in our generalization model. The overall plan is thereby to stepwise add functionalities for broader applications of map generalization into our system.

Acknowledgments This research was supported by the German Research Foundation DFG within Priority Research Program 1894 *Volunteered Geographic Information: Interpretation, Visualization and Social Computing* (VGIscience, TOVIP, SCHI 1008/11-1).

References

- Andrienko N, Andrienko G (2006) Exploratory analysis of spatial and temporal data. Springer, New York. <https://doi.org/10.1007/3-540-31190-4>
- Bak P, Schaefer M, Stoffel A, Keim DA, Omer I (2009) Density equalizing distortion of large geographic point sets. *Cartogr Geogr Inf Sci* 36(3):237–250. <https://doi.org/10.1559/152304009788988288>
- Beard K (1991) Constraints on rule formation. *Map generalization: making rules for knowledge representation*, pp 121–135

- Burghardt D (2005) Controlled line smoothing by snakes. *GeoInformatica* 9(3):237–252. <https://doi.org/10.1007/s10707-005-1283-3>
- Duchêne C, Touya G, Taillandier P, Gaffuri J, Ruas A, Renard J (2018) Multi-agents systems for cartographic generalization: feedback from past and on-going research. Research report, IGN (Institut National de l'Information Géographique et Forestière), LaSTIG, équipe COGIT. <https://hal.archives-ouvertes.fr/hal-01682131>
- Duchêne C, Bard S, Barillot X, Ruas A, Trévisan J, Holzapfel F (2003) Quantitative and qualitative description of building orientation. In: ICA Generalisation Workshop, April 2003
- Edelsbrunner H, Kirkpatrick D, Seidel R (1983) On the shape of a set of points in the plane. *IEEE Trans Inf Theory* 29(4):551–559. <https://doi.org/10.1109/TIT.1983.1056714>
- Gröbe M, Burghardt D (2021) Scale-dependent point selection methods for web maps. *KN J Cartogr Geogr Inf* 71(3):143–154. <https://doi.org/10.1007/s42489-021-00079-y>
- Harrie L, Stigmar H (2010) An evaluation of measures for quantifying map information. *ISPRS J Photogramm Remote Sens* 65(3):266–274. <https://doi.org/10.1016/j.isprsjprs.2009.05.004>. Theme issue “Visualization and exploration of geospatial data”
- Harrie L, Weibel R (2007) Modelling the overall process of generalisation. In: Mackaness WA, Ruas A, Sarjakoski LT (eds) *Generalisation of geographic information*. International Cartographic Association, Elsevier Science B.V., Amsterdam, pp 67–87. <https://doi.org/10.1016/B978-008045374-3/50006-5>
- Huang H, Gartner G (2012) A technical survey on decluttering of icons in online map-based mashups. *Online Maps with APIs and WebServices*
- Jabeur N, Boulekrouche B, Moulin B (2006) Using multiagent systems to improve real-time map generation. In: Lamontagne L, Marchand M (eds) *Advances in artificial intelligence*. Springer, Berlin, Heidelberg, pp 37–48
- Kazil J, Masad D, Crooks A (2020) Utilizing python for agent-based modeling: The mesa framework. In: Thomson R, Bisgin H, Dancy C, Hyder A, Hussain M (eds) *Social, cultural, and behavioral modeling*. Springer International Publishing, Cham, pp 308–317
- Keim DA, Panse C, Sips M, North SC (2004) Pixel based visual data mining of geo-spatial data. *Comput Graph* 28(3):327–344. <https://doi.org/10.1016/j.cag.2004.03.022>
- Knura M, Schiewe J (2021) Map evaluation under covid-19 restrictions: a new visual approach based on think aloud interviews. *Proc ICA* 4:60. <https://doi.org/10.5194/ica-proc-4-60-2021>
- Knura M, Schiewe J (2022) Analysis of user behaviour while interpreting spatial patterns in point data sets. *KN J Cartogr Geogr Inf* 72(3):229–242. <https://doi.org/10.1007/s42489-022-00111-9>
- Knura M, Kluger F, Zahtila M, Schiewe J, Rosenhahn B, Burghardt D (2021) Using object detection on social media images for urban bicycle infrastructure planning: A case study of dresden. *ISPRS Int J Geo-Inf* 10(11). <https://doi.org/10.3390/ijgi10110733>
- Mackaness WA, Purves RS (2001) Automated displacement for large numbers of discrete map objects. *Algorithmica* 30(2):302–311. <https://doi.org/10.1007/s00453-001-0007-9>
- Mackaness WA, Ruas A (2007) Evaluation in the map generalisation process. In: Mackaness WA, Ruas A, Sarjakoski LT (eds) *Generalisation of geographic information*. International Cartographic Association, Elsevier Science B.V., Amsterdam, pp 89–111. <https://doi.org/10.1016/B978-008045374-3/50007-7>
- Meier S (2016) The marker cluster: A critical analysis and a new approach to a common web-based cartographic interface pattern. *Int J Agric Environ Inf Syst* 7:28–43. <https://doi.org/10.4018/IJAIEIS.2016010102>
- Moacdieh N, Sarter N (2015) Display clutter: A review of definitions and measurement techniques. *Hum Factors* 57(1):61–100. <https://doi.org/10.1177/0018720814541145>. pMID: 25790571
- Sadahiro Y (1997) Cluster perception in the distribution of point objects. *Cartogr Int J Geogr Inf Geovis* 34(1):49–62. <https://doi.org/10.3138/Y308-2422-8615-1233>
- Senaratne H, Mobasheri A, Ali AL, Capineri C, Haklay MM (2017) A review of volunteered geographic information quality assessment methods. *Int J Geogr Inf Sci* 31(1):139–167. <https://doi.org/10.1080/13658816.2016.1189556>
- Sester M (2000) Generalization based on least squares adjustment. *GeoInformatica* 6:233–261

- Slocum T, McMaster R, Kessler F, Howard H (2009) Thematic cartography and geovisualization. Prentice Hall series in geographic information science. Pearson Prentice Hall, Hoboken
- Stigmar H, Harrie L (2011) Evaluation of analytical measures of map legibility. *Cartogr J* 48(1):41–53. <https://doi.org/10.1179/1743277410Y.0000000002>
- Taillandier P, Gaffuri J (2012) Designing generalisation evaluation function through human-machine dialogue. *CoRR abs/1204.4332*. <http://arxiv.org/abs/1204.4332>
- Touya G (2012) Social welfare to assess the global legibility of a generalized map. In: *Geographic information science*. Springer Berlin, Heidelberg, pp 198–211. https://doi.org/10.1007/978-3-642-33024-7_15
- Touya G, Hoarau C, Christophe S (2016) Clutter and map legibility in automated cartography: A research agenda. *Cartogr Int J Geogr Inf Geovis* 51(4):198–207. <https://doi.org/10.3138/cart.51.4.3132>
- Touya G, Zhang X, Lokhat I (2019) Is deep learning the new agent for map generalization? *Int J Cartogr* 5(2-3):142–157. <https://doi.org/10.1080/23729333.2019.1613071>
- Töpfer F, Pillewizer W (1966) The principles of selection. *Cartogr J* 3(1):10–16. <https://doi.org/10.1179/caj.1966.3.1.10>
- Ware JM, Jones CB (1998) Conflict reduction in map generalization using iterative improvement. *GeoInformatica* 2(4):383–407. <https://doi.org/10.1023/A:1009713606524>
- Yan H, Weibel R (2008) An algorithm for point cluster generalization based on the voronoi diagram. *Comput Geosci* 34(8):939–954. <https://doi.org/10.1016/j.cageo.2007.07.008>
- Yan X, Chen H, Huang H, Liu Q, Yang M (2021) Building typification in map generalization using affinity propagation clustering. *ISPRS Int J Geo-Inf* 10(11):732. <https://doi.org/10.3390/ijgi10110732>
- Zahtila M, Knura M (2022) Visualizing point density on geometry objects: Application in an urban area using social media vgi. *KN J Cartogr Geogr Inf* 72(3):187–200. <https://doi.org/10.1007/s42489-022-00113-7>
- Zhang X, Ai T, Stoter J (2008) The evaluation of spatial distribution density in map generalization. In: *ISPRS 2008: Proceedings of the XXI congress: Silk road for information from imagery: the International Society for Photogrammetry and Remote Sensing*, 3–11 July, Beijing, China. Comm. II, WG II/2. International Society for Photogrammetry and Remote Sensing (ISPRS), Beijing, pp 181–187. https://www.isprs.org/proceedings/XXXVII/congress/2_pdf/2_WG-II-2/03.pdf
- Zhang X, Ai T, Stoter JE (2009) A voronoi-like model of spatial autocorrelation for characterizing spatial patterns in vector data. In: *2009 Sixth International Symposium on Voronoi Diagrams*, pp 118–126. <https://doi.org/10.1109/ISVD.2009.19>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

