Generalisation and Multiple Representation of Location-Based Social Media Data

Dirk Burghardt, Alexander Dunkel and Mathias Gröbe

Institute of Cartography, TU Dresden, 01062, Dresden, Germany; first_name.last_name@tu-dresden.de

Abstract: Generalisation and Multiple Representation encompass well established concepts and methods to enable the automated derivation of smaller scale maps. The paper explores the potential of applying them for the scale-dependent visual analysis and presentation of user-generated, spatial content in general and location-based social media data in particular. The original development of methods and models such as generalisation operators or digital landscape / cartographic models were driven by the requirements of National Mapping Agencies with focus on topographic map production. Here an attempt is made to rank generalisation operators according to their relevance for the derivation of general and categorised information out of the individual social media data and second to create readable maps without clutter.

Multiple representations of location-based social media data has the potential to support multi-scale visual analysis on different spatial and semantic granularities. Furthermore it could be interpreted as concept of representing various perceptions of different user groups. Two main strategies can be distinguished in order to generate the multi-scale views, on the one hand directly out of the data through internal derivation such as hierarchical clustering and on the other hand by external assignment of hierarchical structured reference units such as Quadtree or grid-based tessellations. Both strategies are illustrated with applications for the visual analysis of location-based social media data. Hierarchical clustering is used to generate tag maps out of meta data from georeferenced photos for purposes of landscape planning. The hierarchical data structure of Geohash is utilised for derivation of Micro Diagrams to analyse language tags from micro blogging content.

Keywords: generalisation, multiple representation, multi-scale view, location-based social media, user-generated spatial content, Volunteered Geographic Information, hierarchical clustering, Geohash, tag maps, micro diagrams

1. Introduction

Generalisation within traditional cartography pursues two objectives: abstraction and legibility. Abstraction aims at deriving general, relevant geographic concepts and relationships from individual geographic features – preserve or emphasize the essentials and simplify or omit the detail. Legibility supports the communication of cartographic information in a readable form. Both objectives are relevant for the analysis and visualisation of location-based social media data.

The way how location-based social media data is described and modelled varies greatly from the categorisation and management of administrative data. Nevertheless, methods applied in model generalisation are suitable to analyse location-based social media data at smaller scales. It supports the derivation and reasoning about higher order concept. Multi-scale views are generated to identify scale dependent pattern and relationships within locations-based social media data. Therefore methods of filtering, clustering and aggregation are applied. Closely related with the analysis of location-based social media data are questions of location privacy protection. Here abstraction can help to find suitable granularity levels and to achieve geo-indistinguishability.

Legibility requirements as drivers of cartographic generalisation are applicable for the geovisual presentation of social media data. In order to avoid overlays of too much information, to prevent clutter and to maintain minimum sizes of presented information automated methods of selection, displacement and typification can be applied. Furthermore methods from label placement can be adapted for the visualisation of tags and icons even within an interactive multi-scale web mapping environment.

The paper presents in Section 2 research from the literature about the generalisation and multiple representation of Volunteered Geographic Information and location-based social media data. Section 3 provides a ranking of generalisation operators according to their relevance for the analysis and presentation of lbs media content. In section 4 a distinction of two general strategies for the derivation of multiple representations is proposed to either internal derivation out of geodata or external assignment with hierarchical structured reference units. For every strategy a concrete approach is presented with application examples for the analysis of photo meta data (Flickr) and georeferenced microblogging content (Twitter).

2. Background

2.1 Research on the generalisation of VGI and location-based social media

Volunteered Geographic Information and location-based social media data can be subject of abstraction and generalisation to become appropriate for applications at various level of detail. Typically this kind of data were captured with a spatial granularity influenced by the available resolution provided through GPS receiver and Smartphones, which can be far to detailed for small scale analysis or overview presentation. Second VGI and location-based social media data from different platforms might be used together e.g. Wikidata and OpenStreetMap data or in combination with authoritative data which can pose challenges regarding geometric matching and semantic modelling (Sester et al., 2014). Third because of the often huge amount of individually generate data filtering and aggregation operations are beneficially to remove clutter and present information in a legible way (Touya et al. 2016).

The potential of VGI use in map production were discussed in Olteanu-Raimond et al. (2016). Opportunities for map revision, change detection and updating are conceivable, on the other side there are limitations because of issues related to data quality, coverage, sustainability and legal issues. A comparison of generalisation within the OpenStreetMap project compared to the generalisation of authoritative data is provided by Klammer and Burghardt (2014). Structural differences between official and crowd controlled map generalisation were analysed. The automated generalisation within OSM is part of the symbolisation and carried out on-the-fly during rendering of map tiles. Basic generalisation functionalities there are semantic filtering, simplification of linear geometries and label placement. A specific MRDB concept for managing different forms of representations as typically can be found within NMA´s is not applied within the OSM project.

Privacy issues have to be taken seriously when working with VGI and location-based social media. Abstraction and aggregation of data can contribute to the anonymity protection. A transformation of GPS trajectories based on spatial generalisation and k-anonymity was applied to hide locations by replacing exact positions in the trajectories through approximate centroids of areas (Andrienko et al., 2009; Monreale et al., 2010). The utilisation of geometry type change with transformation of cycling trajectories into privacy-perserving heat maps has been proposed by Oksanen et al. (2015). Chatzikokolakis et al. (2015) developed methods that achieve geo-indistinguishability by perturbating the actual location with random noise.

A comparison between map generalisation research and multimedia summarisation research was carried out by Touya (2015). This is particular relevant as multimedia summarisation have been also applied to provide visualised summaries of social events and corresponding microblogs (Bian et al. 2015). While Touya (2015) was mainly interested in deriving guidelines for map generalisation from the multimedia summarisation research, this paper will also look into the other direction to show that concepts and methods from generalisation and multiple representations might be beneficial for the visual analysis of VGI and location-based social media.

2.2 Research on multiple representation of VGI and location-based social media

The term "multiple representations" was used in conjunction with geographic databases to describe representations of the same geographic entity or phenomena, which were stored as different objects and linked (Sarjakoski, 2007). In the context of location-based social media researcher start to develop multi-scale approaches to analyse space, scale and meaning. Thereby, instead of "multiple representation", the term "multi-scale views" is commonly applied, which gives more attention to the varying patterns at different scales than to the linkage of individual objects.

A scale-structure identification method was proposed by Rattenbury et al. (2007) to extract place and event semantics from Flickr tags. The search for tag occurrences over space and time were carried out at multiple spatial and temporal scales to find the most appropriate one. Feik and Robertson (2016) analysed tag frequencies after aggregation of tags to hexagon tessellations with varying cell sizes. The cell size was chosen to be appropriate for analysis of local urban place, street scale, neighbourhoods and larger districts. They suggest that scale effects should be considered explicitly when analysing VGI and location-based social media data. Multi-scale event detection based on Twitter data were carried out by Dong et al. (2015) with utilisation of wavelets. They aim on modelling the relationships between the temporal and spatial scales to separate events of different scales in a meaningful way.

The derivation of multiple representation data bases (MRDB) was and is a major aim for National Mapping Agencies to support topographic map production and efficient updating at different scales. Methods of model generalisation were applied to derive different representation levels with varying degree of geometric and semantic abstraction within the MRDB. Underlying schemata were defined in advance following a top-down approach with unique classification for all geographic features. In contrast the categorisation of location-based social media data typically uses a bottom-up approach with tagging. Thereby a so called folksonomy evolves, which directly reflects on the user vocabulary. This system might also include category schemes that have the ability to describe content with different levels of spatial and semantic granularity. Meysam and Weibel (2016) proposed methods for the integration of folksonomies into the process of map generalisation. In OpenStreetMap tags are used to describe and structure geographic features. Methods to infer the level of detail (LoD) of VGI features are suggested by Touya and Reimer (2015). Furthermore methods presented for the detection of inconsistencies between features at different LoD as well as for the harmonisation of LoD heterogeneities (Touya and Brando, 2013).

Multiple representations were also derived with utilisation of hierarchical data structures. Methods for the real-time generalisation of point data using Quadtrees and space deforming algorithms were proposed by Bereuter and Weibel (2013, 2017). Vario-scale data structures were developed for the derivation of mixed scale representations (van Oosterom et al., 2014) and continuous, smooth zooming (van Oosterom and Meijers, 2014). The derivation of multiple representation with gradual content zoom were applied in the context of web mapping and mobile applications (Bereuter et al., 2013; Huang et al., 2016). Thereby content zoom gives access to more detailed information of thematic content independent from spatial zoom.

3. Generalisation operators in application for location-based social media data

Generalisation operators are traditionally used for the derivation of smaller-scale maps based on cartographic models from larger scale data (Mackaness et al., 2017). Thereby generalisation aims at the abstraction of spatial information from a high level of detail to a lower level. These operations can be applied for both digital landscape models and location-based information models. Within this section the relevance of generalisation operators will be summarised for the geovisual analysis of location-based social media data such as microblogging content, image tags or GPS-trajectories. Therefore a qualitative categorisation into very relevant, indirect relevant and less relevant is applied.

The **classification** operator applied in model generalisation is very relevant for the analysis and categorisation of location-based social media data. While digital landscape/cartographic models uses predefined categorisation schemata, the social media content is described with non-hierarchical keywords such as tags or hashtags. For purposes of analysis and application it is necessary to structure this information and associate them with a smaller number of themes or categories. Therefore automated methods are developed such as topic modelling, theme based clustering, correlation analysis, etc. (Wanner et al., 2014). Scale dependent analysis and geovisualisation of tags related to location-based social media will be a future research topic, especially as the abstraction of information enables presentation of overview and detail as well as the detection of patterns and relations at varying level of details (see also next section).

A second very relevant generalisation operator for analysis and visualisation of location-based social media is provided through **elimination** and **selection**. Typically elimination is realised through different types of filter operations. While in model generalisation the elimination operation focusses mainly on geometric criteria (minimal size, minimal length) for the removal of features, the social media content can be filtered according to spatial, temporal, semantic or social criteria. Examples of filtering social media content are the presentation of tags in a map which are characteristic for a specific region, theme, user group or time frame.

Aggregation and **typification** are very relevant generalisation operators in particular for the graphical presentation of social media content. The most often used methods are clustering approaches applied for the grouping of point features. Results might be visualised in a quantitative way. Therefore symbol size can be changed according to the number of aggregated social media content. An associated interaction can be utilized to give access to the underlying details, e.g. through a simple info-box or a "cluster spiderfier" (see example in Table 1). Aggregation operation can also be carried out also as part of model generalisation. Thus social media content or trajectory data can be aggregated to ensure privacy issues of individuals.

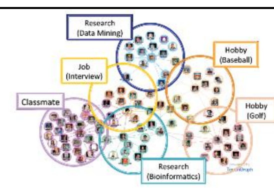
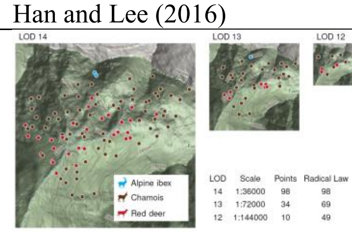
Combination operation are indirect relevant for the analyses of georeferenced social media data to identify regions out of several points where a specific tag or topic might be used. Possible methods are buffering and/or combined with calculation of convex hull. Furthermore density surfaces might be derived based on the quantitative analysis of point features (Hollenstein and Purves, 2010). Linear features respective movement trajectories can be constructed by combination of selective point features, e.g. uploaded pictures of a user within a given time frame.

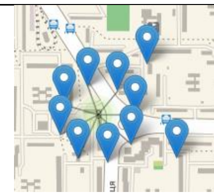
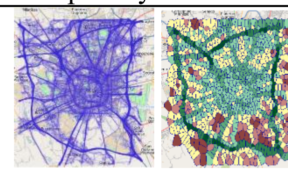
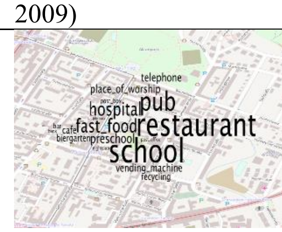
Displacement operation can be indirectly relevant for the geovisualisation of social media content through annotation or labelling. For example georeferenced word clouds require the (dis-)placement of labels without overlapping each other. A direct application of displacement operations to remove clutter or overlapping markers is only applicable if not too many objects are displayed.

Simplification is less relevant for georeferenced social media data which are provided typically with only one associated location. In case of trajectory data, e.g. hiking or cycling track the simplification operation can be carried out as a preprocessing operation.

Enhancement operation is less relevant for geometric transformation of social media data, but might be used to guide focus during visualisation, e.g. through usage of colour or other highlighting methods.

Table 1. Relevance of generalisation operators applied to location-based social media

Generalisation operator	**Relevance to location-based social media**	**Corresponding methods**	**Examples**
Classification / Reclassification	very relevant	(spatial) topic modelling, theme based clustering, correlation analysis	 Han and Lee (2016)
Elimination and Selection	very relevant	filtering according to spatial, temporal, semantic or social criteria	 Bereuter (2015)

Aggregation and Typification	very relevant	spatial or distance based clustering, aggregation of points or lines (trajectories), anonymisation through aggregation	 "cluster spiderfier" with markers spiderfy into a circle
Combine	indirect relevant	convex hull, buffering, combination of points or lines to area objects, e.g. for anonymisation	 trajectories points are combined and presented through choropleth map (Andrienko et al. 2009)
Displacement	indirect relevant	displacement of annotations and labels	 Label placement of georeferenced word clouds (Hahmann and Burghardt, 2011)
Simplification	less relevant	pre-processing of trajectories	
Enhancement	less relevant	guide focus during visualisation with highlighting methods	

4. Multiple representation of location-based social media data

There exist several reasons for the derivation of multiple representations from location-based social media data. First multiple representation enable an efficient visualisation of overview and detail accessible either interactively through continuous zoom or through parallel presentation with multiple linked views. Second the multiple representation support multi-scale analysis to see different, scale dependent pattern and relations between thematic content and geographic features. Third, if information is associated with regions, statistical units or regular grids and therefore aggregated the effects caused by the modifiable areal unit problem (MAUP) can be easily studied and misinterpretations be avoided.

A wider interpretation of multiple representation is the reproduction of different human perspectives and perceptions. As the creation of location-based social media data is very much influenced by subjective interpretations of humans, the multiple representations could select different views depending on age, gender, origin, thematic groups, etc. This allows enhancing the representativeness of the analysis for specific user groups. In the domain of cartography and geoscience the multiple representations are used mainly in connection with scale changes, nevertheless in social geography, landscape architectures, journalism and other application fields the wider interpretation would be beneficial.

The derivation of multiple representations of location-based social media data can be realised in two different ways either through an *internal derivation* directly out of the georeferenced content or with an *external assignment* of hierarchical structured reference units (see Fig. 1). The internal derivation can be achieved through utilisation of hierarchical clustering methods with consideration of spatial and semantic criteria. The external assignment aggregates the content within hierarchical structured reference units such as Quadtree, Geohash or Hexagons[1]. All of these tessellations do not consider context and thus can be applied for all kinds of location-based social media content.

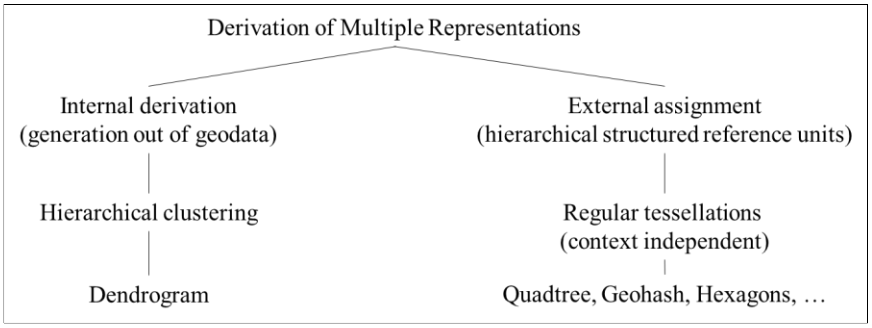


Fig. 1. Derivation of multiple representations for location-based social media through internal derivation of hierarchies or external assignment of hierarchical structured reference units.

A context dependent assignment of hierarchical structured reference units could be imagined, e.g. usage of functional units or statistical units. Therefore selective (thematic) analysis of location-based social media data is required as well as pre-knowledge about underlying geographic phenomena. In the case of statistical data, countries, federal states, municipalities and counties (e.g. NUTS 0 – 3 for the member states of the EU) might provide a suitable hierarchical structured reference frame. Another approach might be a scale dependent selection of contextual important point features (e.g. larger cities) followed by construction of a hierarchical Voronoi graph.

In the following two subsections methods of multiple representation for the analysis of location-based social media data will presented. First a grid-based derivation of micro diagrams will enable multiple representation of Tweets languages and second an object-based clustering will be utilised for the scale-dependent presentation of Flickr tag maps.

4.1 Multiple representation of Micro Diagrams with utilisation of Geohash

The multiple representation of location-based social media data can be derived by assigning the data to a hierarchical spatial data structure. In this example we utilise the Geohash as a geocoding system which supports the hierarchical representation of point data through an indexed grid structure (Fox et al. 2013). The Geohash represents a coordinate pair as a string, whereby the length of the string is proportional to the spatial resolution. It is often used as index in spatial databases. The Geohash enables real-time geovisualisation with varying granularities showing either distributions of individual points or aggregated, increasingly abstract grid-like visualisations (Fig. 2). Through interactive control it is possible to choose spatial zoom and grid-width independently. This way the user can first select the spatial zoom level followed by an interactive variation of the thematic granularity through *content zoom*. Flexible change between spatial and content zoom supports the user in analysing scale dependent pattern and distributions.

[1] Hexagon tessellations are chosen for multiscale spatial analysis because of compactness and simple recursive hierarchical nesting (Feik and Robertson, 2016)

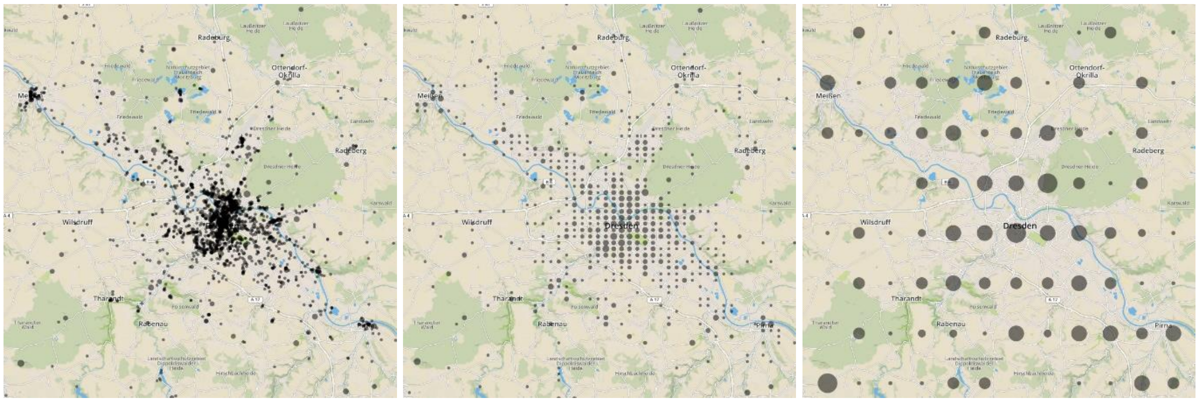


Fig. 2. Grid-based real-time geovisualisation based on Geohash with increasing number of characters (left 7, centre 6, right 5). The content zoom enables the quantitative presentation of tweet distributions through either individual points (left) or with varying degree of aggregation and abstraction (middle and right).

The example in Fig. 2 shows a quantitative representation of georeferenced Tweets captured within one week. While the aggregated, more abstract representations (Fig. 2 middle and right) enable better estimation of tweet numbers per grid-cell the higher granular visualisation emphasises the local distribution patterns (Fig. 2 left).

A multivariate quantitative analysis can be carried with Micro Diagrams (Gröbe and Burghardt, 2016). Thereby a quantitative visualisation of several attributes is realised within every grid-cell. As point related values have to be aggregated, both the size of the grid cell and the zoom scale influence the resulting distributions. The interactive control enables visual analysis of aggregation effects, especially in sub regions with inhomogeneously distributed data. This way MAUP (modifiable areal unit problem) becomes less a problem, as changing areal units through spatial and content zoom does provide alternative views on the underlying location-based social media data. The example in Fig. 3 analyses the language tags associated with tweets of one week. Fig. 3 left shows the volume of tweets and the language distribution within Europe. Fig. 3 middle represents tweet numbers and associated languages for Germany while Fig. 3 right zooms to Berlin as the region of interest. It becomes obvious that for the higher zoom level (e.g. around Berlin) new sub patterns can be identified.

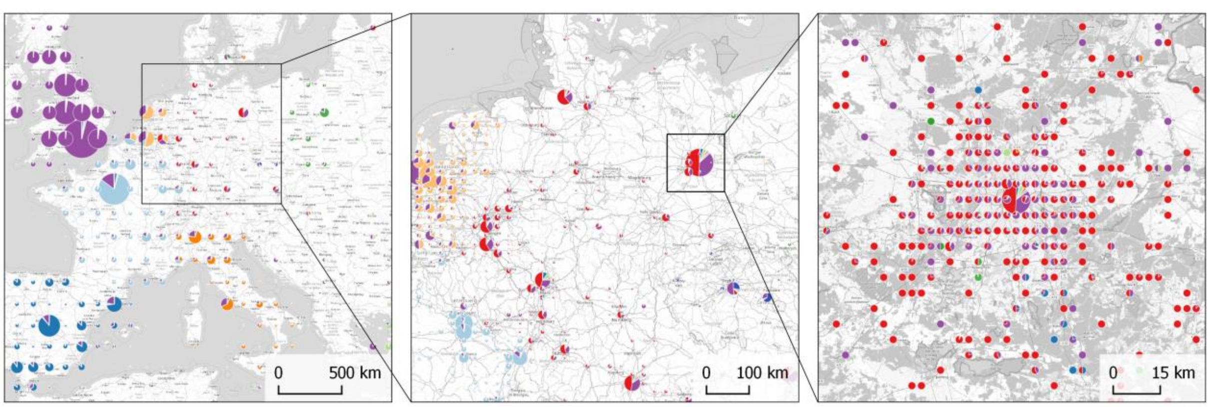


Fig. 3. Multi-scale visualisation with Micro Diagrams for quantitative representation of language tags from tweets.

The scale dependent geovisualisation with Micro Diagrams has the advantage of providing fast changes between overview and detail (I), enables analysis of distribution patterns at different scales (II) and offers appropriate visualisation options to avoid MAUP (III).

4.2 Multi-scale view of tag maps based on hierarchical clustering

A multi-scale view can be derived directly out of the location-based social media data through utilisation of hierarchical clustering methods. The distance thresholds applied during clustering influence the multi-scale presentation of the generated tag maps. Tag maps can be used to visualise associated meta-information of location-based social media content with help of georeferenced word clouds. The derivation method can be described as a two step procedure, with a first step for visualising photo location density and second step for map labelling based on tag occurrences (Dunkel, 2015).

Fig. 4 shows results of single linkage clustering applied to locations of Flickr images with scale dependent distance thresholds. The radius of the circles corresponds to the number of aggregated images. The resultant pattern are quite different for different clustering thresholds. While for the larger scales (smaller distance thresholds) popular photo locations along linear elements such as streets can be identified, for the smaller scales (larger distance thresholds) other aggregated locations such as the recreational park become visually dominant.

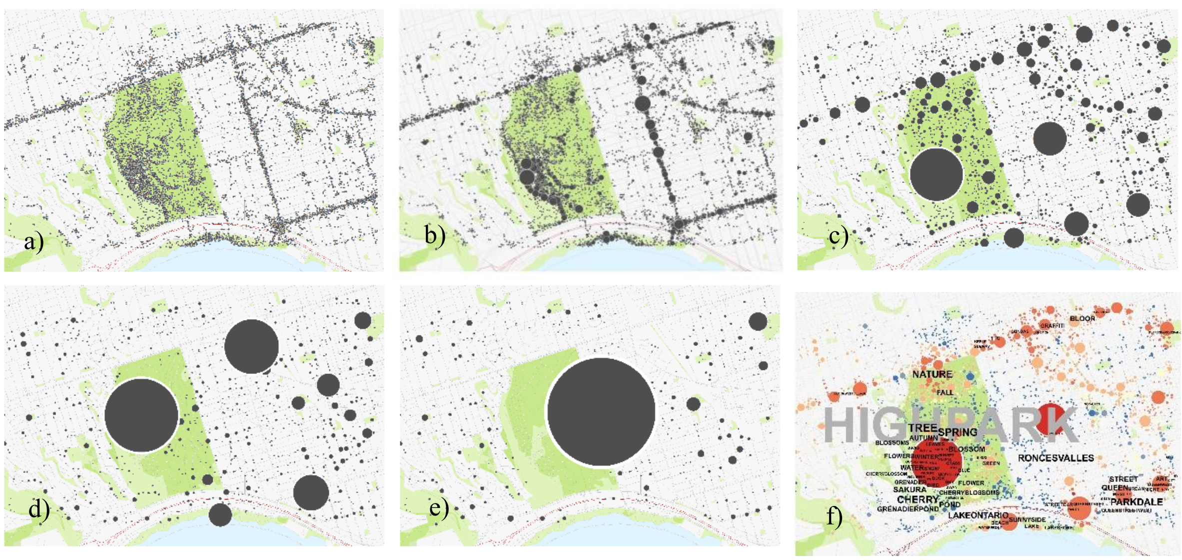


Fig. 4. Locations of Flickr images within "High Park" (Toronto) are shown in the figure a), followed by results of a single-linkage clustering with threshold distances of 10 m, 30 m, 50 m and 70 m. The radius of the circle is proportional to the number of aggregated images. Figure f) shows the resulting tag map generated with a clustering distance of 30 m.

The image in Fig. 4f) shows the final tag map generated for a clustering distance of 30 m, which enables local scale analysis of differences in attribution meaning. Red and blue colour is used to differentiate between hot and cold spots, indicating places where either more or less pictures were taken compared to the overall area of analysis. The size of the labels represents the number of times a tag was used. The tags with highest counts were placed with light grey colour as background labels in order to avoid offsetting many smaller label, which are often of equal interest to the investigator.

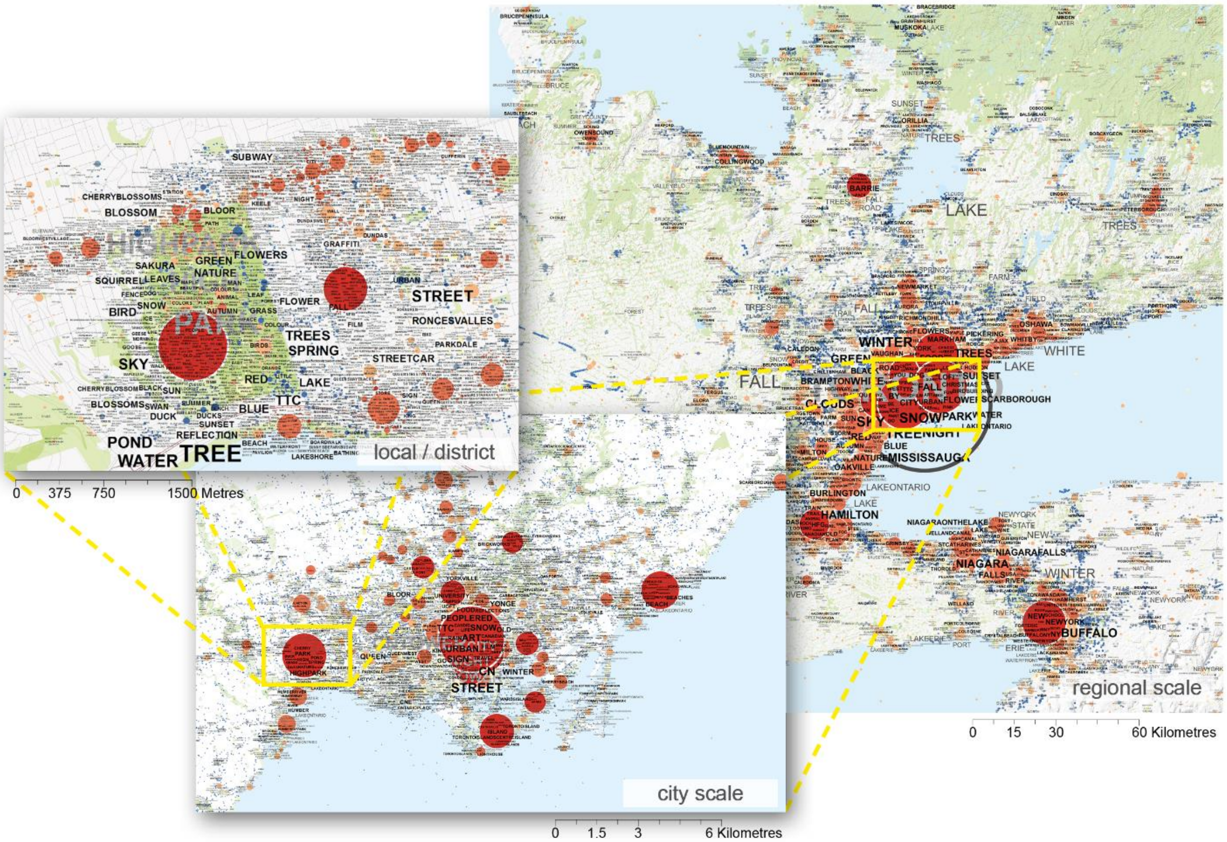


Fig. 5. Multi-scale views of Flickr tag maps showing photo location densities and most frequent occurrence of tags around Toronto for the local/ district level, the city scale and the regional scale.

The multi-scale view in Fig. 5 has been derived from aggregating Flickr data with varying distance thresholds reaching from 30 metre (local / district scale) to 50 metres (city scale) to 100 metres for the regional scale. A phenomenon that is typical for the presented multi-scale clustering of tags is that on larger scales relatively simple and general terms emerge, whereas on smaller scales, more and more references to local characteristics become visible (Dunkel, 2016). In the example set of maps above, this is visible for references to High Park, a popular and highly frequented park in Toronto. On the local scale, the tags ‘highpark’, ‘high’ and ‘park’ are most frequently used by photographers visiting the area and placed in background to avoid offsetting smaller labels that reference local characteristics of the park itself. On the city scale, references to High Park are still visible but appear next to many similarly important characteristics of the city of Toronto (such as its popular beaches or districts, Queen, Bloor, Yorkville and Yonge). Finally, on the regional scale, references to High Park are only indirectly available on the map, based on more general terms such as ‘park’, ‘tree’, or ‘flower’, which symbolise categories of characteristics that are present in many places of the city and, thus, collectively perceived and referenced. In landscape planning, for example, these maps can be used to explore characteristics of places and areas. This provides an opportunity to quantify aspects of the perceived landscape character, and thus may help in more accurately assessing a project’s consequences. To the investigator, the hierarchical visualisation of tag maps for different scales exposes relationships and presents a base for interactively probing further questions.

4.3 Multiple representation with reproduction of different human perspectives

Treating different people's viewpoints as a form of multiple representation is a rather contemporary extension of the traditional definition. But subjectivity is a key characteristic that users bring to bear on creating VGI, such as Flickr photo data, which therefore needs specific consideration in analysis. For example, when taking and tagging photographs, there exist different degrees of consensus for choosing relevant subjects between different users and groups of users for different areas, and these differences are reflected in patterns of the resulting data. On the one hand, this provides the opportunity to visualise perceived values from observers and what the environment actually means to the people who interact with it. Assessing such aspects was formerly reserved to the domain of surveys and other forms of active participation. On the other hand, when working with VGI, it is usually unknown who is actually represented by the data (Brown and Whitall, 2014). Each area offers data from a unique composition of users. One approach is to classify users based on their origin, which was presented by several studies in the context of Flickr (Wood et al. 2013; Popescu and Grefenstette, 2010).

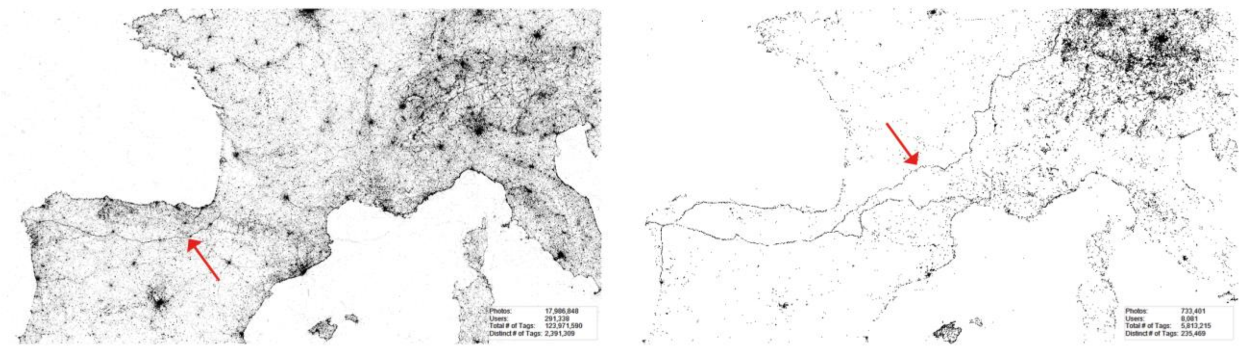


Fig. 6. Flickr photo locations in Europe (a) for all photographers (left map) and (b) photographers with origin set to Germany (right map). Arrows mark the routes of the Camino de Santiago emerging with different relative intensities in each map.

For generating the maps shown in Figure 6, a relatively simple classification was made by geocoding the location that Flickr users publish on their profile. The left map shows locations for all photographers, whereas the right map shows only photo locations from users with a specified origin located in Germany. While some of the provided origins may be inaccurate, it is obvious that the right map reflects behaviour patterns related to a specific subgroup of photographers (i.e. 'German photographers'). Consequently, without further filtering, several routes of the Camino de Santiago are directly and dominantly visible on the right map (Fig. 6), whereas on the left map, only the shared, last part of the route leading through Spain is visible for photographers of any origin.

Conversely, these user-behaviour related characteristics of data also pose challenges to research dealing with VGI because data is neither created by a homogeneous set of rules nor by a consistent group of trained experts. For example, in the study from Popescu and Grefenstette (2010), contribution patterns from 5000 users from 6 different countries were analysed. The researchers found that Americans uploaded the largest number of images, followed by British contributors, while in turn Spanish users restricted themselves to the smallest size of vocabulary when tagging pictures. Ignoring these differences in contribution patterns may negatively affect results or lead to visualisations biased towards a specific group or particular very active user or group users, which hinders drawing accurate conclusions. Many techniques already exist to reduce these undesirable biasing effects in VGI for creating balanced visualisations for available perspectives, but careful consideration is necessary during all steps of processing data, interpreting results and drawing conclusions.

5. Conclusions

The consideration of scale is essential for the analysis and visualisation of location-based social media. Interactive multiple representations provide overview and detail from spatial point of view and enable the analysis of general and specific semantic descriptions. The size of the study area and the level of detail to be examined are mutually dependent. On a small scale, one can find a sub-selection of tags used to describe the place at larger scales, but also different, new tags can occur prominent. Thus patterns and relationships change at different scales. The derivation of smaller scale representations to analyse location-based social media data can be achieved through automated generalisation, whereby the most relevant operations are the classification of data, the filtering and selection as well as the aggregation and typification. Displacement and label placement operations are relevant to prevent clutter and enable the use of annotations as shown with example of tag maps.

Two derivation strategies can be distinguished for the generation of multiple representations: I) the internal derivation out of the location-based data through hierarchical clustering and II) the external assignment of hierarchical structured reference units. Both strategies were illustrated with examples to analyse location-based social media. The internal derivation with hierarchical clustering was applied to derive tag maps out of meta information associated with photos (from Flickr platform) for purposes of landscape planning. The external assignment was realised with a Geohash to visualise language tags (from Twitter platform) through Micro Diagrams at various scale. The examples were chosen to illustrate that interpretation of patterns are influenced by the selection of cluster distances and Geohash level with corresponding grid width. Thereby multiple representations enable visual interaction to avoid miss interpretations, because of selection of scale, spatial granularity or MAUP effects.

The concept of multiple representation can be applied to the different perceptions of topics or places for various user groups. First challenge thereby is the identification of user groups regarding different characteristics such as familiarity with place, attitude to certain themes or nationalities as shown in the examples. In other words, different human perspectives can be investigated, visualised and made explicit. This way the subjectivity of user-generated spatial content is not seen as a biasing effect but as a source for analysing differences in collective values and meaning.

New challenges regarding generalisation and multiple representation of location-based social media data arise in the case of real-time analysis and visualisation of streaming data. There methods of classification, filtering and clustering have to be extended for the processing of incremental updates. The usage of various sources of location-based social media data require research on similarity measurement, data matching and integration. Furthermore the investigation of interrelations between spatial and semantic hierarchies is promising for the derivation of place-based descriptions and purposes of landscape planning.

Acknowledgements

This work was supported by the German Research Foundation as part of the priority programme "Volunteered Geographic Information: Interpretation, Visualisation and Social Computing" (VGIscience, SPP 1894).

References

Andrienko G., Andrienko, N., Giannotti F., Monreale A., Pedreschi D. (2009). Movement data anonymity through generalization. In Proceedings of the 2nd SIGSPATIAL ACM GIS 2009 International Workshop on Security and Privacy in GIS and LBS, SPRINGL '09.

Bereuter P., Weibel R. (2017, submitted). Variable-scale maps in real-time generalisation using a quadtree data structure and space deforming algorithms. ICC 2017 International Cartographic Association. Washington, USA.

Bereuter, P., Weibel, R., Burghardt, D. (2013). Content zooming and information exploration for mobile maps. International Journal of Geomatics and Spatial Analysis, Revue internationale de géomatique, 23(3-4), 295-321.

Bereuter, P., Weibel, R. (2013). Real-time Generalization of Point Data in Mobile and Web Mapping Using Quadtrees. Cartographic and Geographic Information Science, 40(4), 271-281.

Bian, J., Yang, Y., Zhang, H., Chua, T. S. (2015). Multimedia summarization for social events in microblog stream. IEEE Transactions on Multimedia, 17(2), 216-228.

Brown, G., Kelly, M. and Whitall, D. (2014). Which “public”? Sampling effects in public participation GIS (PPGIS) and volunteered geographic information (VGI) systems for public lands management. Journal of Environmental Planning and Management, 57(2), 190–214. http://doi.org/10.1080/09640568.2012.741045

Chatzikokolakis, K., Palamidessi, C., Stronati, M. (2015). Geo-indistinguishability: A Principled Approach to Location Privacy. Proceedings of the 11th International Conference on Distributed Computing and Internet Technology, Lecture Notes in Computer Science, Springer, pp. 49-72.

Dong, X., Mavroeidis, D., Calabrese, F., Frossard, P. (2015). Multiscale event detection in social media. Data Min Knowl Disc, 29, 1374-1405.

Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. Landscape and Urban Planning 142: 173-186.

Dunkel, A. (2016). Assessing the perceived environment through crowdsourced spatial photo content for application to the fields of landscape and urban planning. PhD thesis, TU Dresden. http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-207927

Feick, R., Robertson, C (2016). A multi-scale approach to exploring urban places in geotagged photographs. Comput. Environ. Urban Syst. 53, 96–109.

Fox, A., Eichelberger, C., Hughes, J., Lyon, S. (2013). Spatio-temporal indexing in non-relational distributed databases. Big Data, 2013 IEEE International Conference on. IEEE, pp. 291-299.

Hahmann, S., Burghardt, D. (2011). Maple – a Web Map Service for Verbal Visualisation using Tag Clouds Generated from Map Feature Frequencies. Lecture Notes in Geoinformation and Cartography, Advances in Cartography and GIScience, Vol. 1, Springer, pp. 3-12.

Hollenstein, L., Purves, R. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. Journal of Spatial Information Science: Iss. 1, 21-48.

Huang, L., Meijers, M., Suba, R., van Oosterom, P. J. M. (2016). Engineering web maps with gradual content zoom based on streaming vector data. ISPRS Journal of Photogrammetry and Remote Sensing, 114(April), 274-293. DOI: 10.1016/j.isprsjprs.2015.11.011

Klammer, R., Burghardt, D. (2014). Generalisation within the OpenStreetMap Project Compared to the Generalisation of Authoritative Data (Chapter 5, Case Study II). In Burghardt, D., Duchene, C. and Mackaness, W. (eds.). Abstracting Geographic Inforamtion in a Data Rich World. Lecture Notes in Geoinformation and Cartography, Springer, pp. 139-145.

Mackaness, W. A., Burghardt, D., Duchêne, C. (2017). Map Generalization. The International Encyclopedia of Geography: People, the Earth, Environment, and Technology. 1-16. The Association of American Geographers. DOI: 10.1002/9781118786352.wbieg1015

Meysam, A., Weibel, R. (2016). Integration of folksonomies into the process of map generalization. In: 19th ICA Workshop on Generalisation and Multiple Representation, Helsinki (Finnland).

Monreale, A., Andrienko, G., Andrienko, N., Giannotti, F., Pedreschi, D., Rinzivillo, S., Wrobel, S. (2010). Movement data anonymity through generalization. Trans. Data Privacy, 3, pp. 91–121.

Oksanen, J., Bergman, C., Sainio, J., Westerholm, J. (2015). Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data, JTRG, 48, pp. 135–144, http://dx.doi.org/10.1016/j.jtrangeo.2015.09.001

Olteanu-Raimond, A.-M., Hart, G., Touya, G., Kellenberger, T., Foody, G., Demetriou, D. (2016). The scale of VGI in map production: A perspective of European National Mapping Agencies, Transactions in GIS, doi:10.1111/tgis.12189

Popescu, A. and Grefenstette, G. (2010). Mining user home location and gender from flickr tags. ICWSM 2010: International Conference on Weblogs and Social Media.

Rattenbury, T., Good, N., Naaman, M. (2007). Towards automatic extraction of event and place semantics from Flickr tags. In: ACM SIGIR conference on research and development on information retrieval, Amsterdam

Sarjakoski, L.T. (2007). Conceptual Models of Generalisation and Multiple Representation. In: Ruas, A., Mackaness, W. A. and Sarjakoski, L.T. (eds.). Generalisation of Geographic Information: Cartographic Modelling and Applications, Series of International Cartographic Association, Elsevier, pp. 11-35.

Sester, M., Arsanjani, J. J., Klammer, R., Burghardt, D. and Haunert, J.-H. (2014). Integrating and Generalising Volunteered Geographic Information. In Burghardt, D., Duchene, C. and Mackaness, W. (eds.). Abstracting Geographic Inforamtion in a Data Rich World. Lecture Notes in Geoinformation and Cartography, Springer, pp. 119-155.

Touya, G., S. Christophe and C. Hoarau (2016). Clutter and Map Legibility in Automated Cartography: A Research Agenda, Cartographica: The International Journal for Geographic Information and Geo-visualization, 51(4), 198-207.

Touya, G, Reimer, A. (2015). Inferring the Scale of OpenStreetMap Features. In Arsanjani, J. J., Zipf, A., Mooney, P., Helbich, M. (eds.). OpenStreetMap in GIScience. Lecture Notes in Geoinformation and Cartography, Springer, pp. 81-99.

Touya, G. (2015) Lessons Learned From Research on Multimedia Summarization. Proceedings 18th ICA Workshop on Generalisation and Multiple Representation, Rio de Janeiro, Brazil.

Touya, G., Brando, C. (2013). Detecting Level-of-Detail Inconsistencies in Volunteered Geographic Information Data Sets, Cartographica: The International Journal for Geographic Information and Geo-visualization, 48(2), 134-143, doi:10.3138/carto.48.2.1836

van Oosterom, P. J. M., Meijers, M. (2014). Vario-scale data structures supporting smooth zoom and progressive transfer of 2D and 3D data. International Journal of Geographical Information Science, 28(3), 455–478.

van Oosterom, P. J. M., Meijers, M., Stoter, J., Suba, R. (2014). Data Structures for Continuous Generalisation: tGAP and SSC. In Burghardt, D., Duchene, C. and Mackaness, W. (eds.). Abstracting Geographic Inforamtion in a Data Rich World. Lecture Notes in Geoinformation and Cartography, Springer, pp. 83-117.

Wood, S. A., Guerry, A. D., Silver, J. M. and Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. Scientific Reports, 3.