TECHNISCHE UNIVERSITÄT DRESDEN

DFG Deutsche Forschungsgemeinschaft

Generalisation and Multiple Representation of Location-Based Social Media Data

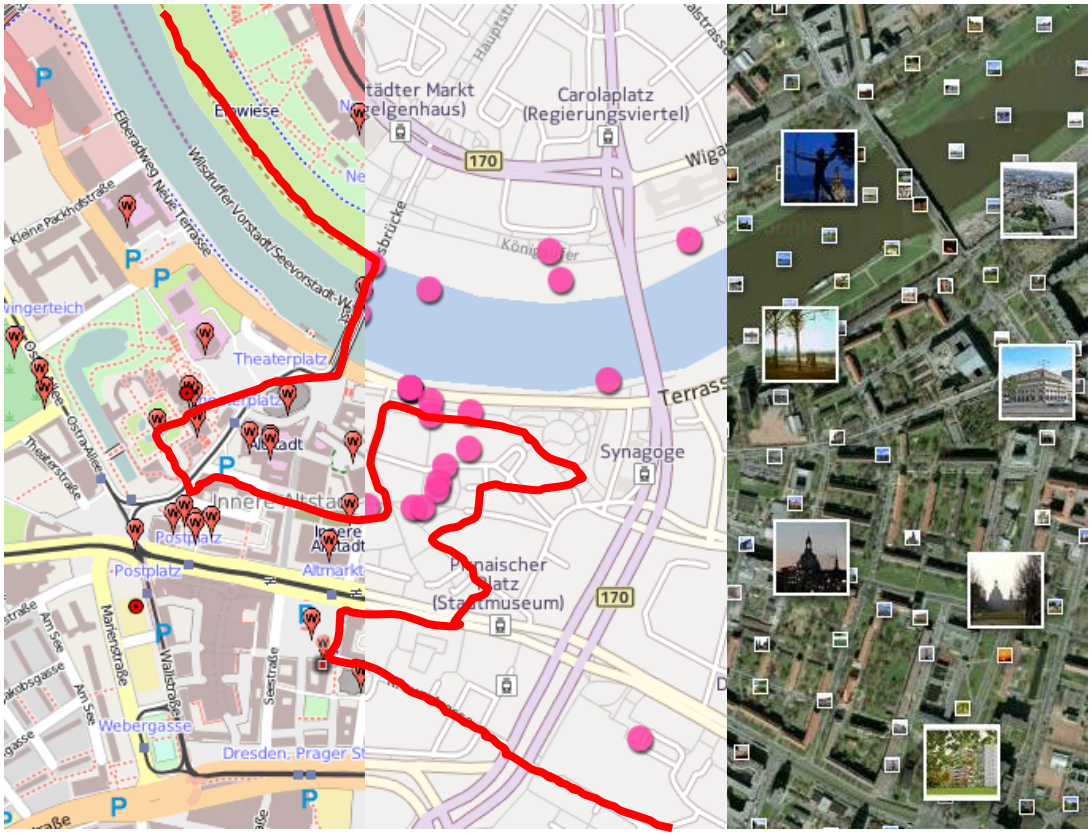
Dirk Burghardt, Alexander Dunkel and Mathias Gröbe,
Institute of Cartography

Outline

1. Motivation
   - VGI and spatial data from Location-Based Social Media
   - Potentials and challenges
2. Generalisation operators applied to LBSM
3. Derivation of Multiple Representations
4. Application examples
   - Tag Clouds Maps
   - Micro Diagram
5. Conclusions

Availability and retrieval of Volunteered Geographic Information

- broad range of volunteered geographic information (OSM-data, GPS-tracks, sensor data, Wikipedia, georeferenced photographs, social networks, microblogging, …)
- data sources are often very large, with high update rates (e.g. 500 Mill. Tweets per day)
- include not only factual but also subjective information → noise or signal
- spatial- /temporal reference is given either completely or partially



OSM-Daten | Sensor data / tracks | Images / microblogging text

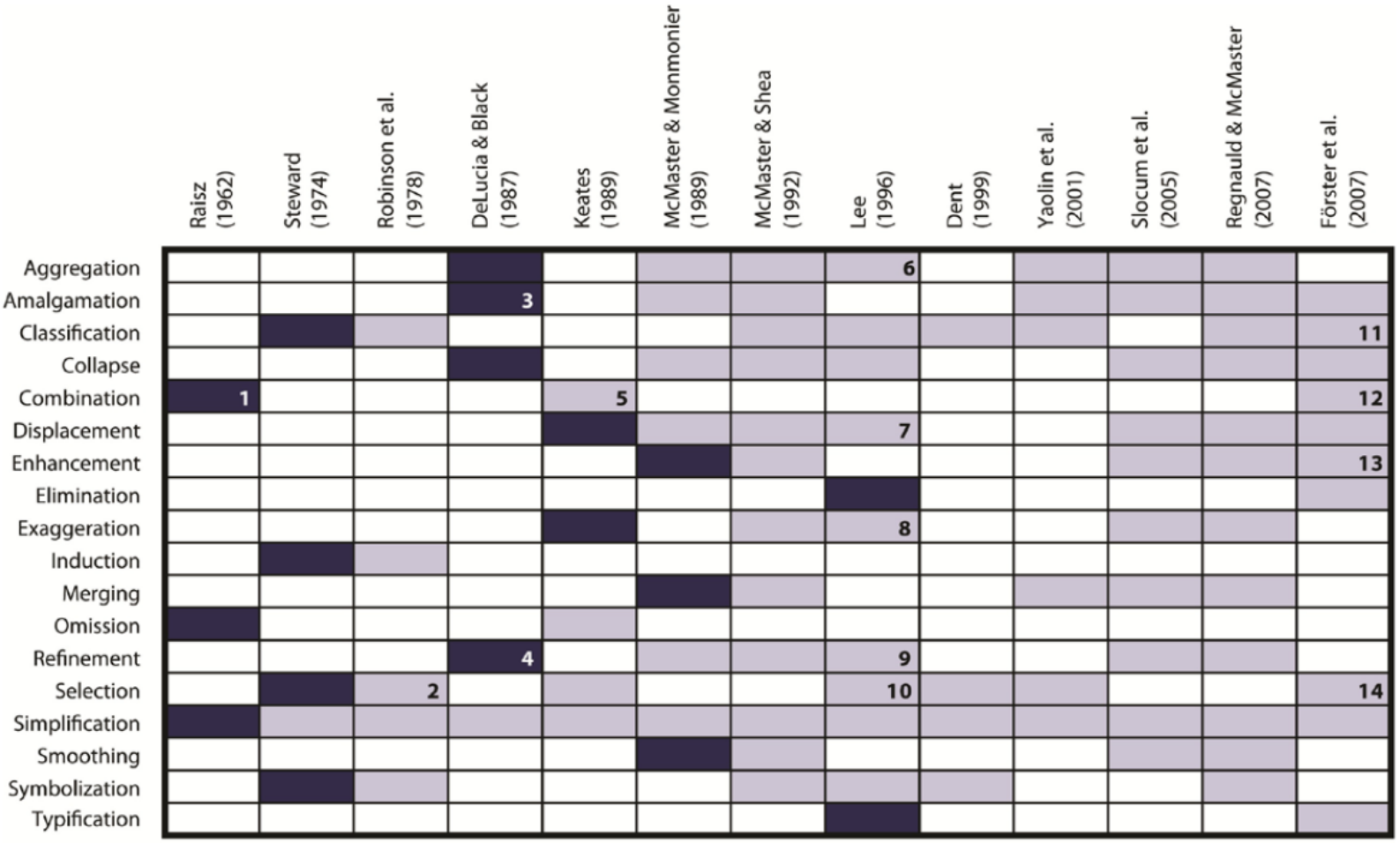
Term definition

- VGI -Volunteered Geographic Information (Goodchild, 2007)
  - introduced by Michael Goodchild (2007)
  - special case of user generated content (UGC) with direct or indirect spatial reference
  - concept „Humans as Sensors“ refers to users who uses mobile technology and low-cost sensors for various tasks
- type of user involvement
  - active data collection (e.g. citizen science projects)
  - passive data generation via location-enabled mobile devices

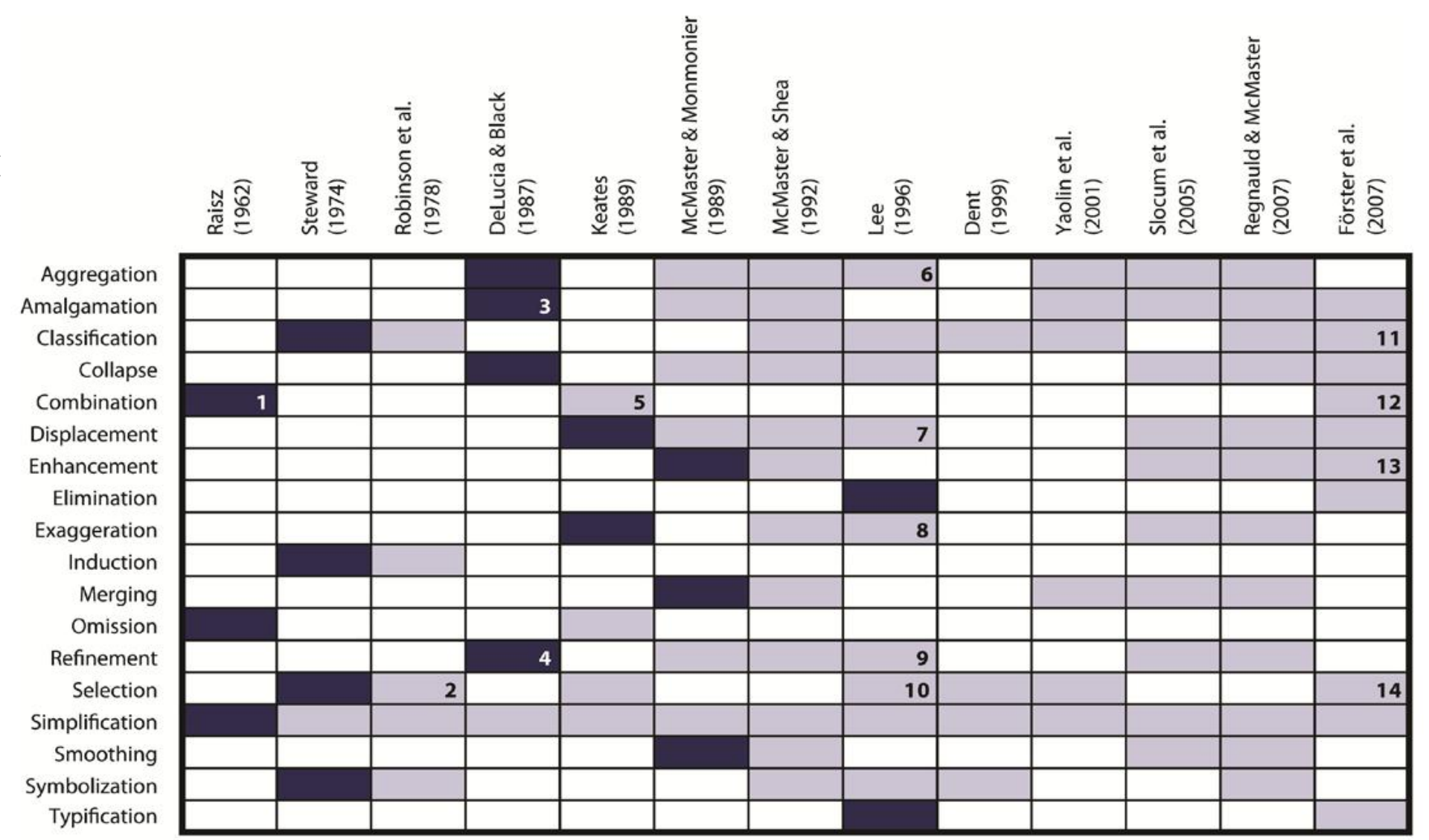
Outline

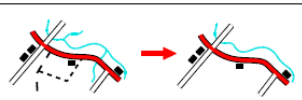
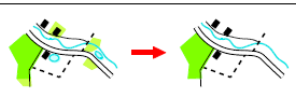
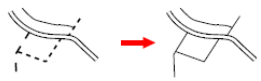
1. Motivation
   - VGI and spatial data from Location-Based Social Media
   - Potentials and challenges
2. Generalisation operators applied to LBSM
3. Derivation of Multiple Representations
4. Application examples
   - Tag Cloud Maps
   - Micro Diagram
5. Conclusions

Generalisation operators applied to LBSM

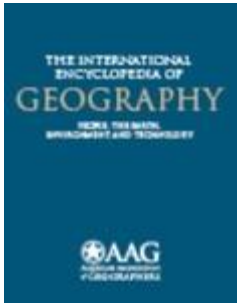


Quelle: Roth et al. 2011
A typology of operators for maintaining legible map designs at multiple scales


Raisz (1962)
Steward (1974)
Robinson et al. (1978)
DeLucia & Black (1987)
Keates (1989)
McMaster & Monmonier (1989)
McMaster & Shea (1992)
Lee (1996)
Dent (1999)
Yaolin et al. (2001)
Slocum et al. (2005)
Regnauld & McMaster (2007)
Förster et al. (2007)
Aggregation
Amalgamation
Classification
Collapse
Combination
Displacement
Enhancement
Elimination
Exaggeration
Induction
Merging
Omission
Refinement
Selection
Simplification
Smoothing
Symbolization
Typification
1 - used to describe any manner of combining items
2 - considered a pre-processing step
3 - called agglomeration
4 - meant as typification
5 - meant as aggregation
6 - meant as both aggregation and merging
7 - called conflict resolution
8 - meant as a change in size
9 - meant as smoothing
10 - called preselection
11 - first to call classification 'reclassification'
12 - meant as aggregation
13 - includes smoothing, exagerration, and a size change, but not enhancement
14 - called class selection, used to include selection and refinement
First Appearance
Appeared Previously

Operator / Description	Example with linear objects	Example with area objects
Class Selection Model ↔ Carto Select the classes of features and attributes that the model/map should contain.		
Reclassification Model ↔ Carto Changes the class membership of the feature and reduces the number of modeled attributes.		
Combine Model ↔ Carto Regrouping a set of feature into a more abstract feature, often of higher dimension.		
Collapse Model ↔ Carto Reduction in the geometric dimension. Point features are often represented by icons.		
Simplification Model ↔ Carto Eliminates the unimportant details, while the general characteristics are preserved.		
Elimination Model ↔ Carto When congestion occurs, less important and short or small objects are eliminated.		
Aggregation Model ↔ Carto Common boundaries, small gaps between neighbour objects are eliminated.		
Enhancement Model ↔ Carto Enhancement is used to exaggerate parts of an object or enlarge the whole object.		
Displacement Model ↔ Carto Important objects remain on their locations. Unimportant objects are moved away.		
Typification Model ↔ Carto Reduces the number of objects, while their distribution and patterns are preserved.		

10 Generalisation Operator

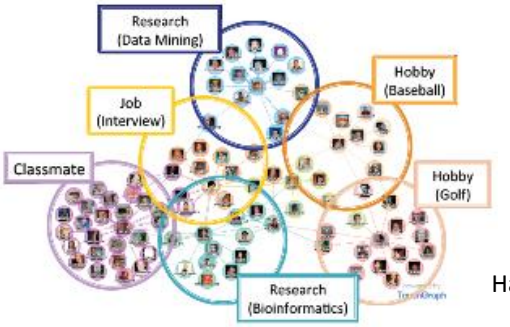


Mackaness, W. A., Burghardt, D. and Duchene, C. (2017). **Map Generalization.** *The International Encyclopedia of Geography*: People, the Earth, Environment, and Technology. John Wiley & Sons.

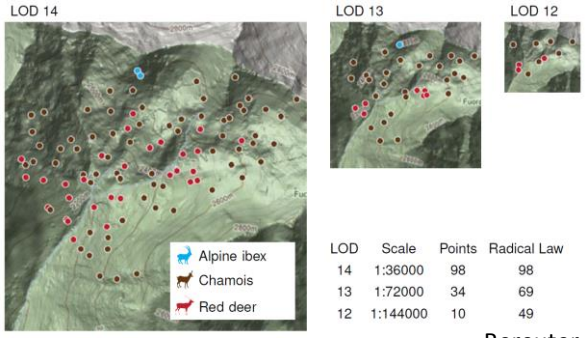
http://onlinelibrary.wiley.com/doi/10.1002/9781118786352.wbieg1015/abstract

Relevance of generalisation operators applied to LBSM

Generalisation operator	Relevance to LBSM	Corresponding methods
Classification / Reclassification	very relevant	(spatial) topic modelling, theme based clustering, correlation analysis
Elimination and Selection	very relevant	filtering according to spatial, temporal, semantic or social criteria
Aggregation and Typification	very relevant	spatial or distance based clustering, aggregation of points or lines (trajectories), anonymisation through aggregation

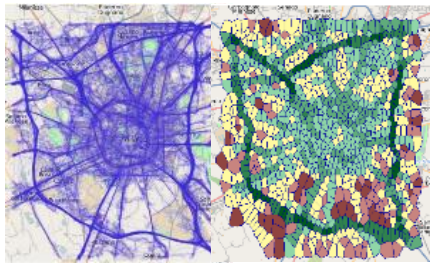


Han and Lee (2016)



Bereuter (2015)

"cluster spiderfier"



trajectories points are combined and presented through choropleth map (Andrienko et al., 2009)

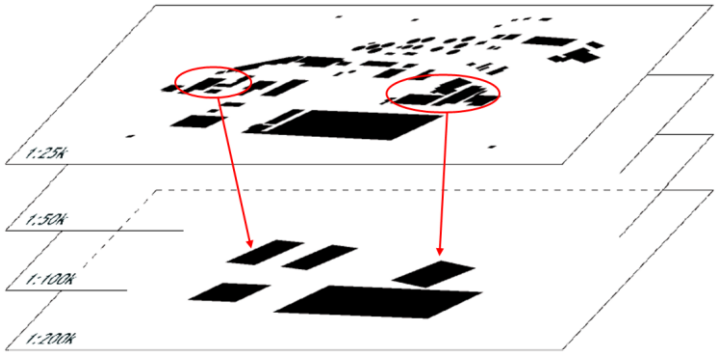
label placement, georeferenced word clouds (Hahmann and Burghardt, 2009)

Outline

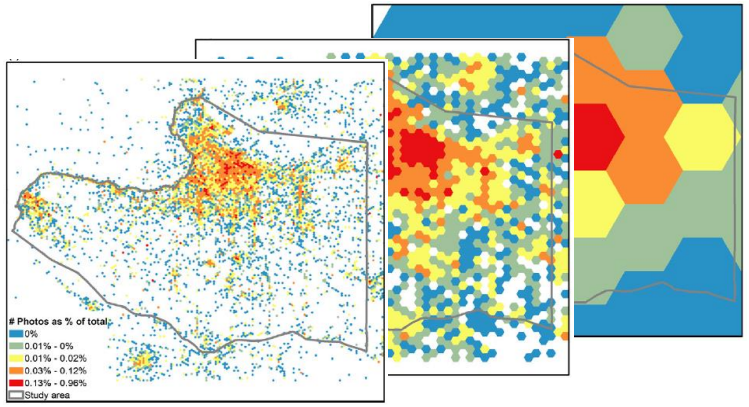
1. Motivation
   - VGI and spatial data from Location-Based Social Media
   - Potentials and challenges
2. Generalisation operators applied to LBSM
3. Derivation of Multiple Representations
4. Application examples
   - Tag Cloud Maps
   - Micro Diagram
5. Conclusions

Terminology – "Multiple representation" vs. "Multi-scale views"

- the term "multiple representations"
  - is used in the context of topographic map production for scale dependent storage of object geometries of the same entity (Sarjakoski, 2007)
  - explicit linkage between representations enable update propagation, consistency checks and support of continuous zooming
- related to the analysis of LBSM data the slightly different term "multi-scale view" is applied
  - gives more attention to the varying patterns at different scales than to the linkage of individual objects



Linkage within MRDB (Cecconi, 2003)



Multi-scale approach applied to geotagged photographs (Feik and Robertson, 2015)

Multiple representation of location-based social media data

Reasons for derivation

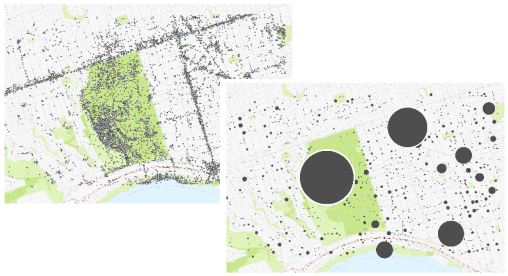
- enable visualisation of overview and detail
  - accessible either interactively through continuous zoom or through parallel presentation with multiple linked views
- support multi-scale analysis
  - to identify scale dependent pattern and relations between thematic content and geographic features
  - avoid modifiable areal unit problem (MAUP) as different aggregations can be study at various scales

Derivation of Multiple Representations

Internal derivation
(generation out of geodata)

Hierarchal clustering
(Dendrogram)

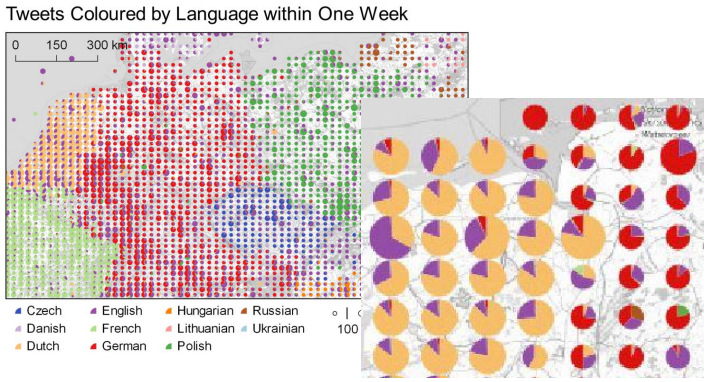
Tag Cloud Maps



External assignment
(hierarchical structured reference units)

Hierarchical tessellation
with Grids, Quadtree, Hexagons, Geohash

Micro-Diagramme

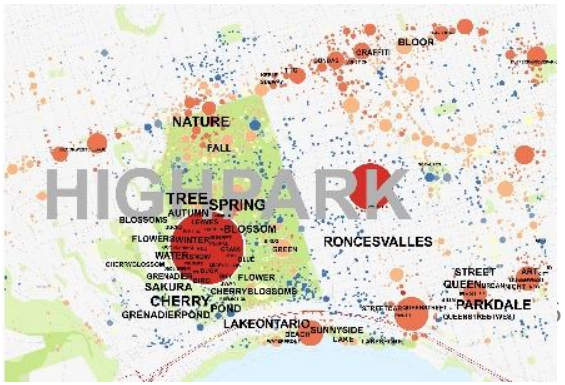


Outline

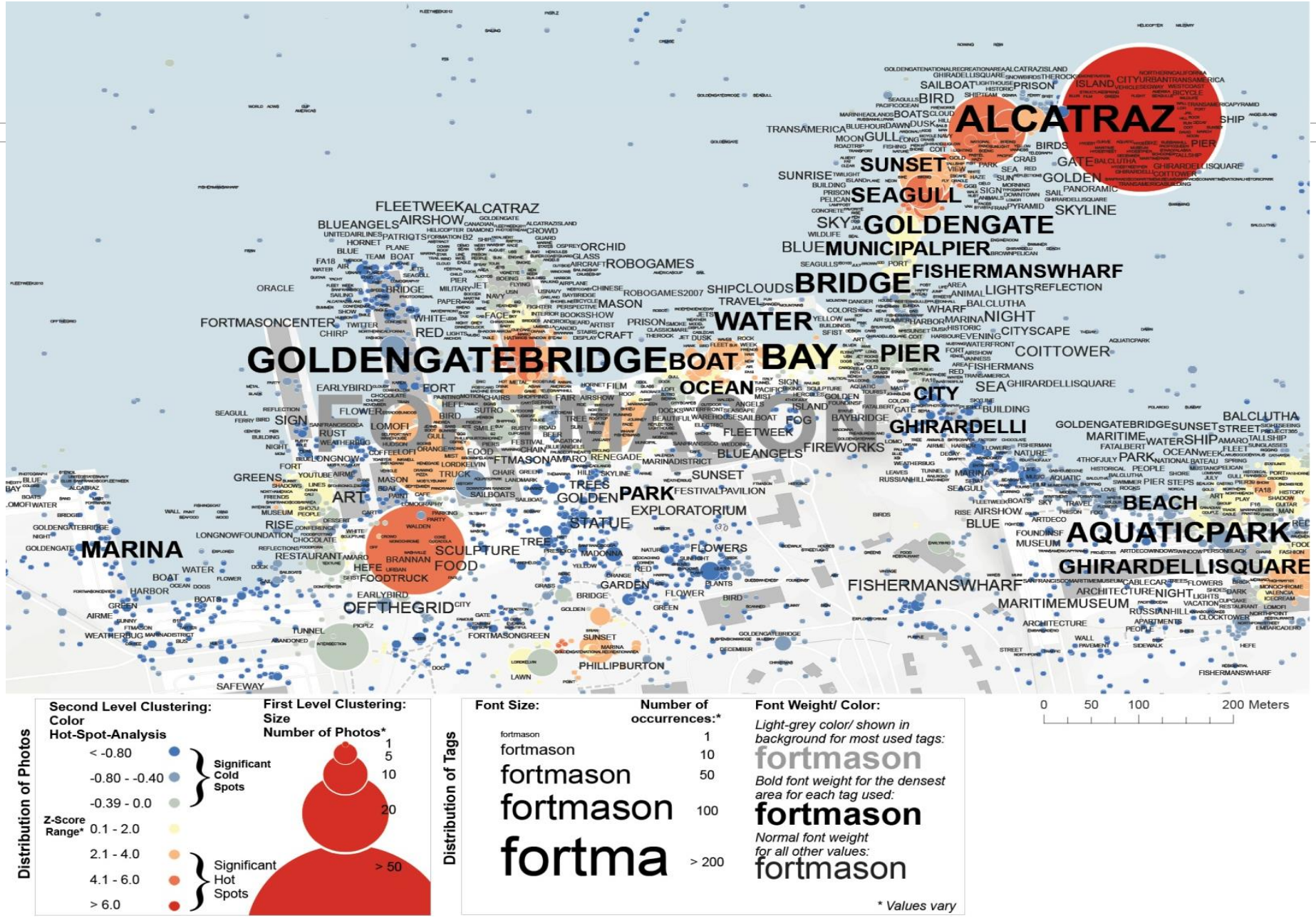
1. Motivation
   - VGI and spatial data from Location-Based Social Media
   - Potentials and challenges
2. Generalisation operators applied to LBSM
3. Derivation of Multiple Representations
4. Application examples
   - Tag Cloud Maps
   - Micro Diagram
5. Conclusions

Tag Maps based on georeferenced word clouds

- Input: point locations of georeferenced photos from sharing application (e.g. Flickr, Panoramio) with various attributes/tags
- aim on derivation of „tag maps“ for landscape and urban planning (PhD thesis Dunkel, 2016)
  - Aggregation of photo locations based on hierarchical clustering for
  - Visualisation of the most common tags



PhD thesis: A. Dunkel (2016)
Assessing the perceived environment through crowdsourced spatial photo content for application to the fields of landscape and urban planning.



Second Level Clustering:
Color
Hot-Spot-Analysis

Distribution of Photos

Z-Score Range*	
< -0.80	Significant Cold Spots
-0.80 - -0.40	
-0.39 - 0.0	
0.1 - 2.0	
2.1 - 4.0	Significant Hot Spots
4.1 - 6.0	
> 6.0	

First Level Clustering:
Size
Number of Photos*

1
5
10
20
> 50

Distribution of Tags

Font Size:	Number of occurrences:*
fortmason	1
fortmason	10
fortmason	50
fortmason	100
fortma	> 200

Font Weight/ Color:

Light-grey color/ shown in background for most used tags:
fortmason

Bold font weight for the densest area for each tag used:
fortmason

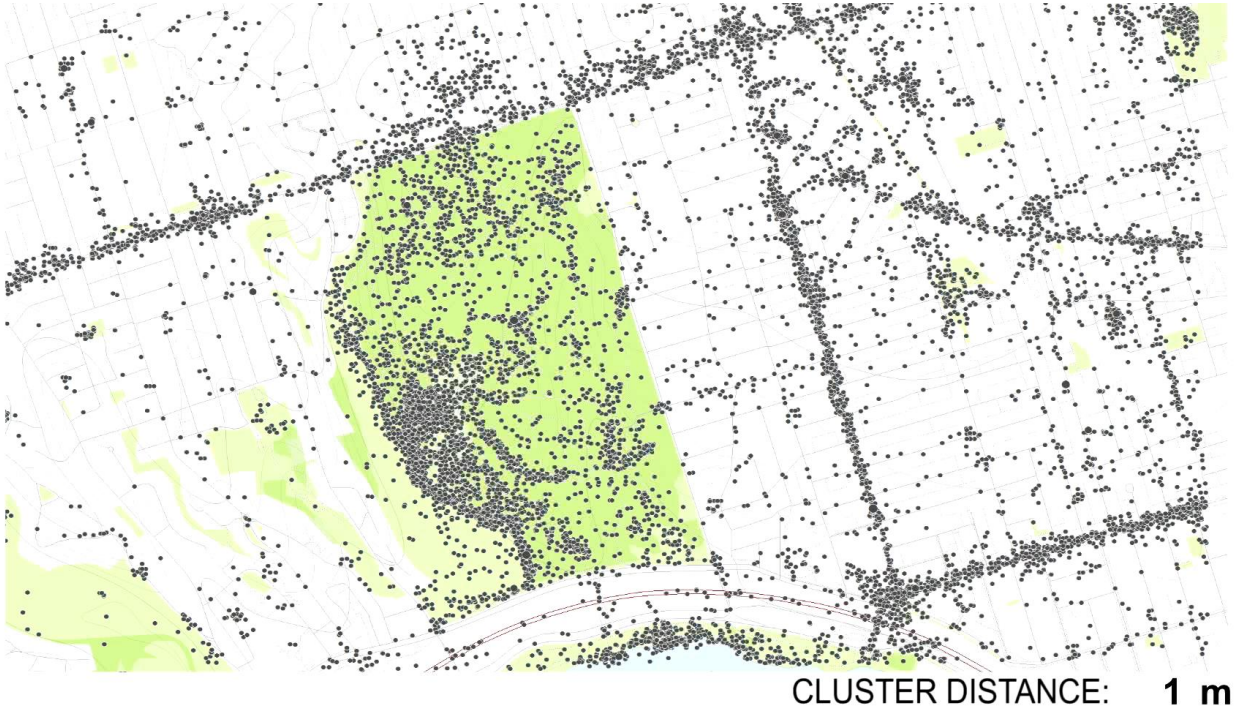
Normal font weight for all other values:
fortmason

* *Values vary*

Source: Dr.-Ing. Alexander Dunkel, Institute of Cartography, TU Dresden

Hierarchical clustering

Variation of cluster distance →
Derivation of multiple representation for various scales

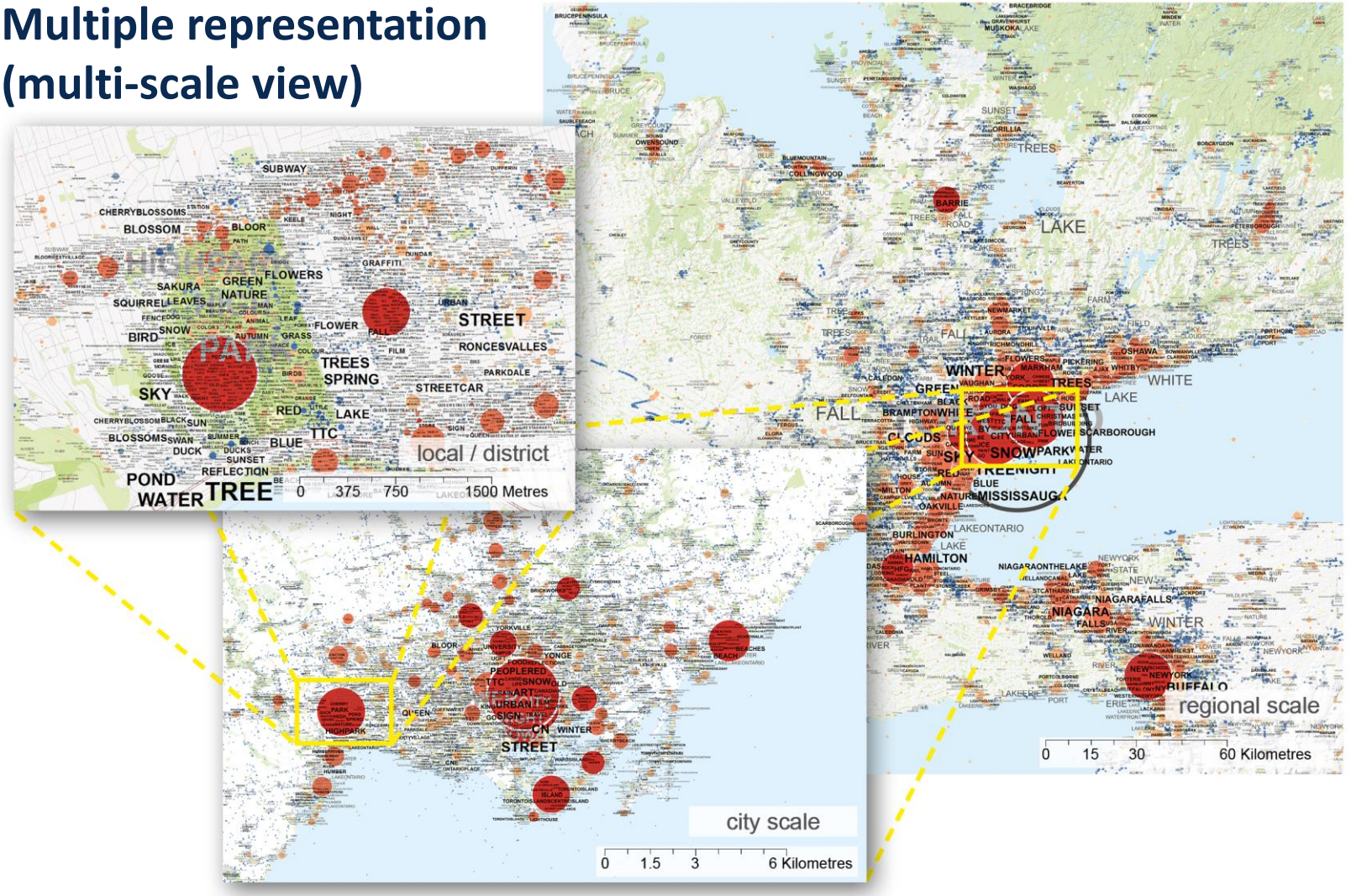


Source: Dr.-Ing. Alexander Dunkel, Institute of Cartography, TU Dresden

Various patterns generated by hierarchical clustering

Source: Dr.-Ing. Alexander Dunkel, Institute of Cartography, TU Dresden

Multiple representation (multi-scale view)

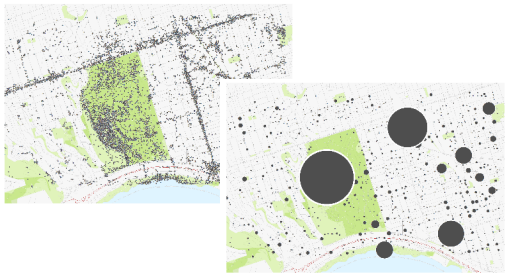


Derivation of Multiple Representations

Internal derivation
(generation out of geodata)

Hierarchal clustering
(Dendrogram)

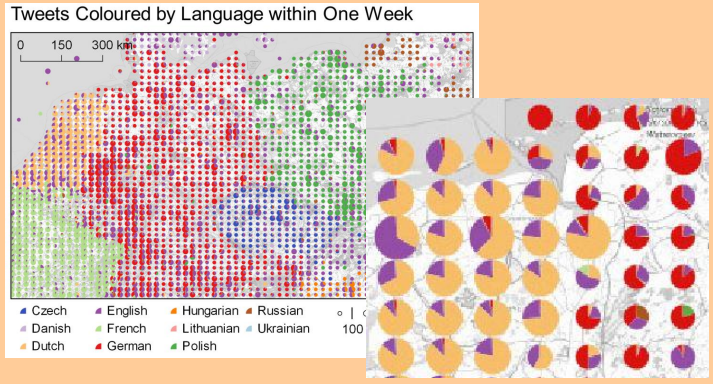
Tag Cloud Maps



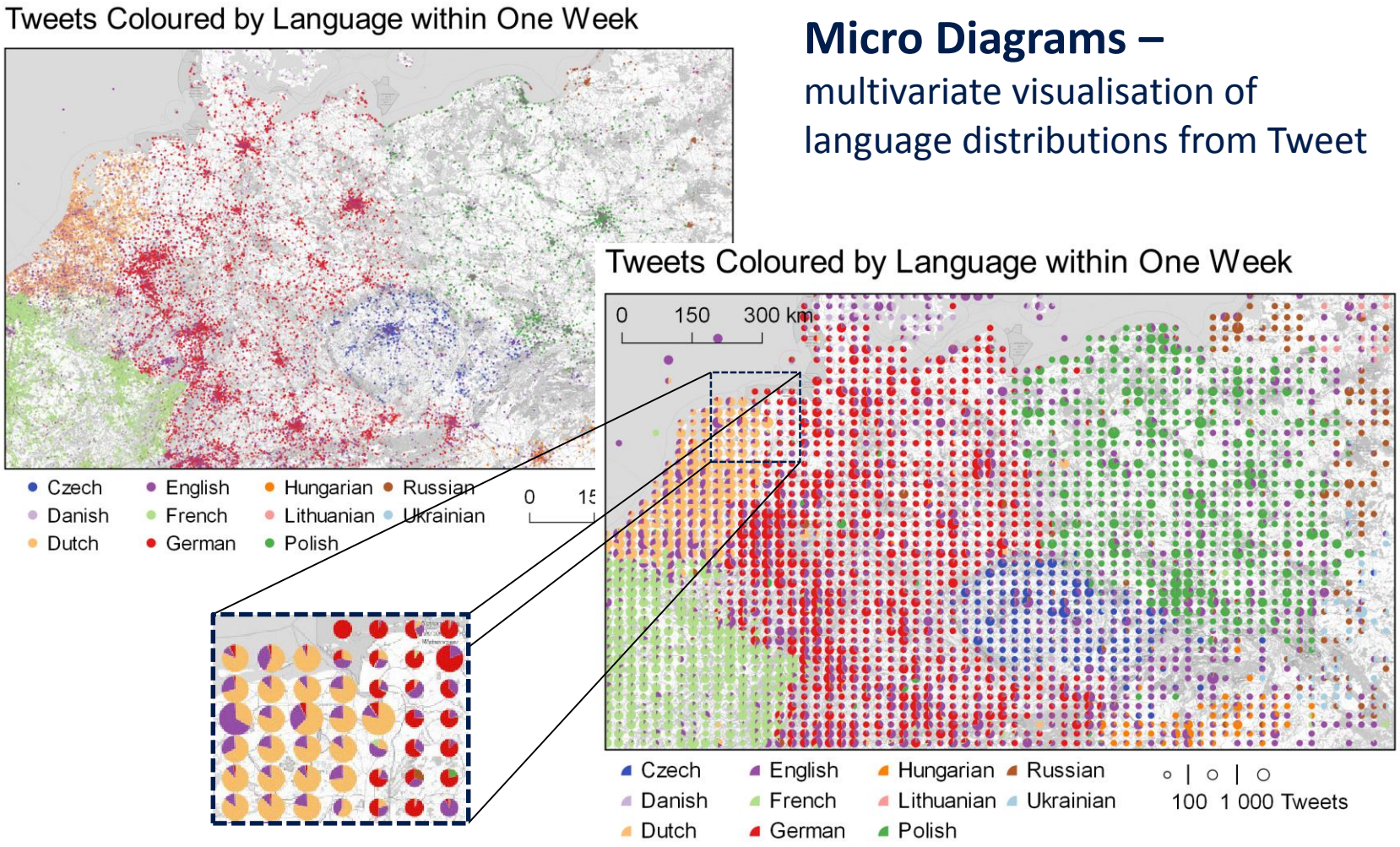
External assignment
(hierarchical structured reference units)

Hierarchical tessellation
with Grids, Quadtree, Hexagons, Geohash

Micro-Diagramme



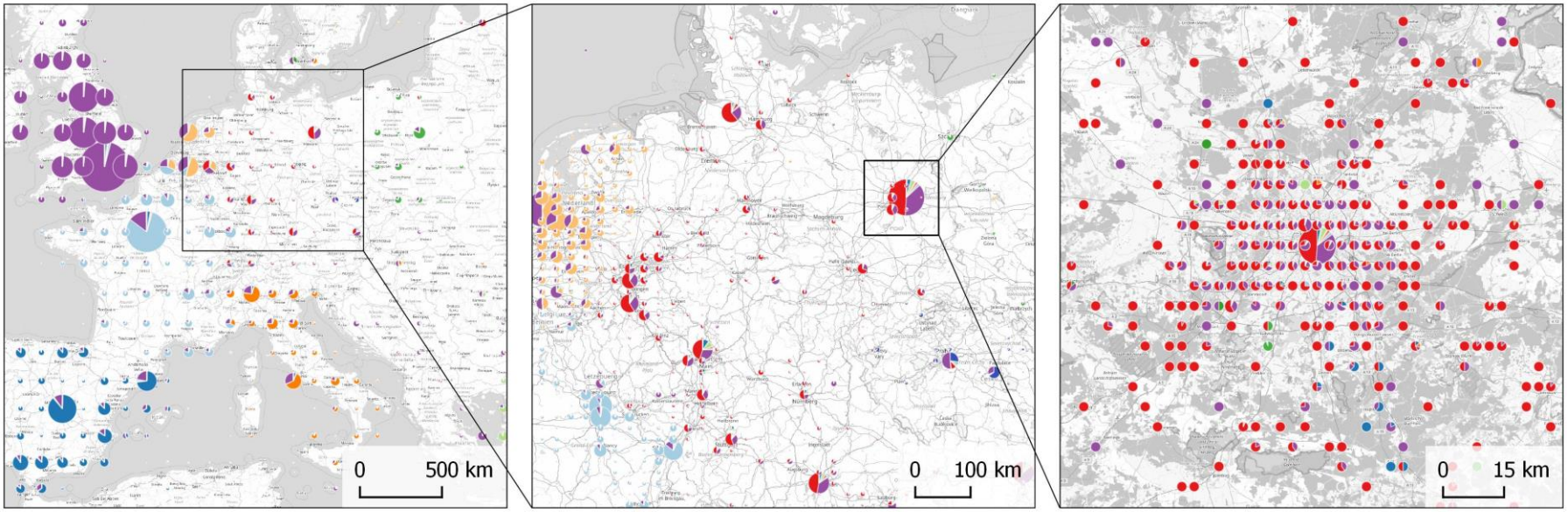
Multiple Representation through external assignment



Micro Diagrams –
multivariate visualisation of language distributions from Tweet

Gröbe, M. and Burghardt, D. (2017). Micro Diagrams: A Multi-Scale Approach for Mapping Large Categorised Point Datasets. Agile Conference, Wageningen, The Netherlands.

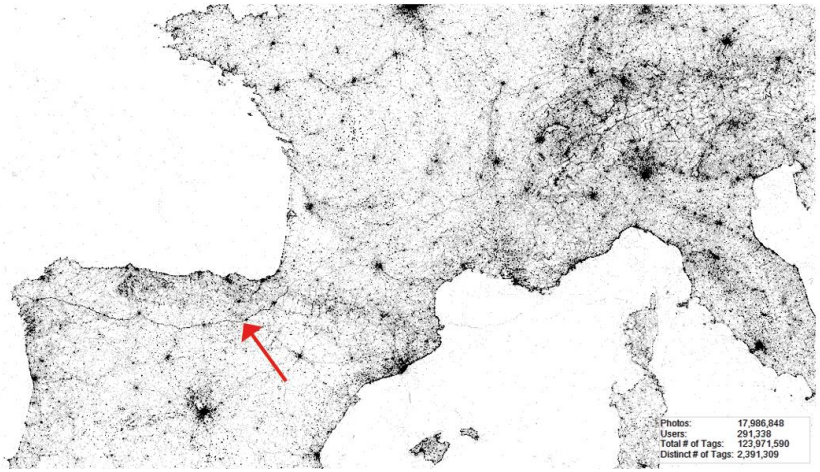
Derivation of multiple representation

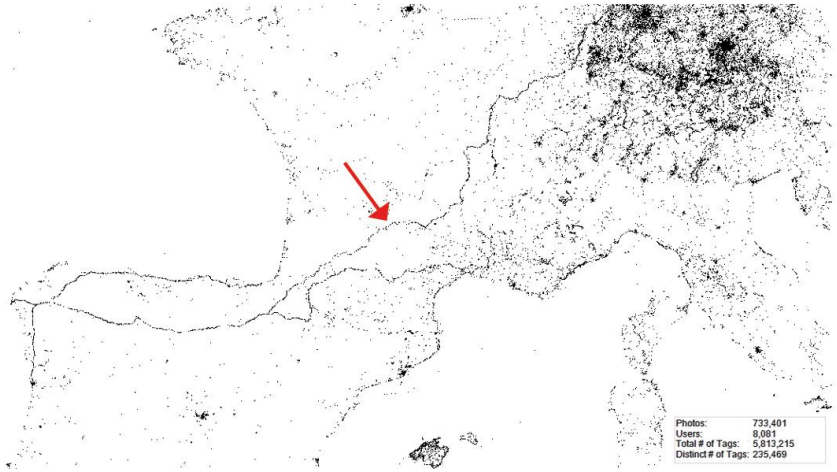


Source: M.Sc. Mathias Gröbe, Institute of Cartography, TU Dresden

Multiple representation with reproduction of different human perspectives

- a broader definition of “multiple representation” could be used to reflect on the varying viewpoints of people (subjective perception)
  - requires specification of different user groups, e.g. regarding age or home town (taken from user profile)
  - example: routes of Camino de Santiago perceived by German photographers





Flickr photo locations in Europe (a) for all photographers (left map) and (b) photographers with origin set to Germany (right map)

Conclusion

- consideration of scale is essential for the analysis and visualisation of location-based social media
  - the derivation of smaller scale representations of LBSM data can be achieved through automated generalisation (most relevant operations are classification, filtering, selection as well as aggregation, typification)
- interactive multiple representations provide overview and detail from spatial and semantic point of view → patterns and relationships change at different scales
- 2 derivation strategies for the generation of multiple representations:
  I) the internal derivation through hierarchical clustering → "Tag Maps"
  II) the external assignment of hierarchical structured reference units → "Micro Diagrams
- LBSM data provide new challenges related to abstraction and derivation of multiscale views