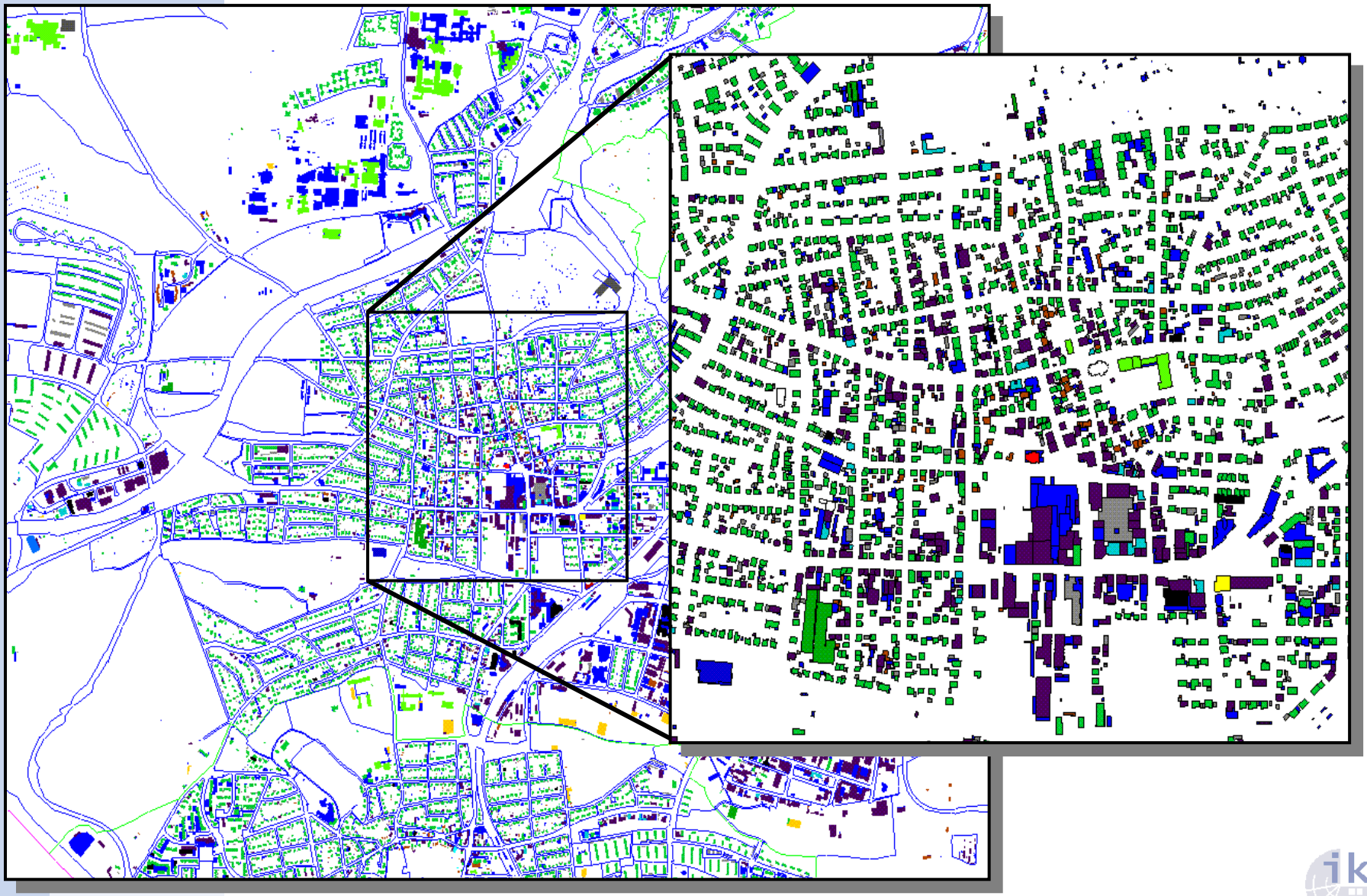
Hierarchical Graph Clustering to Find Groups of Objects

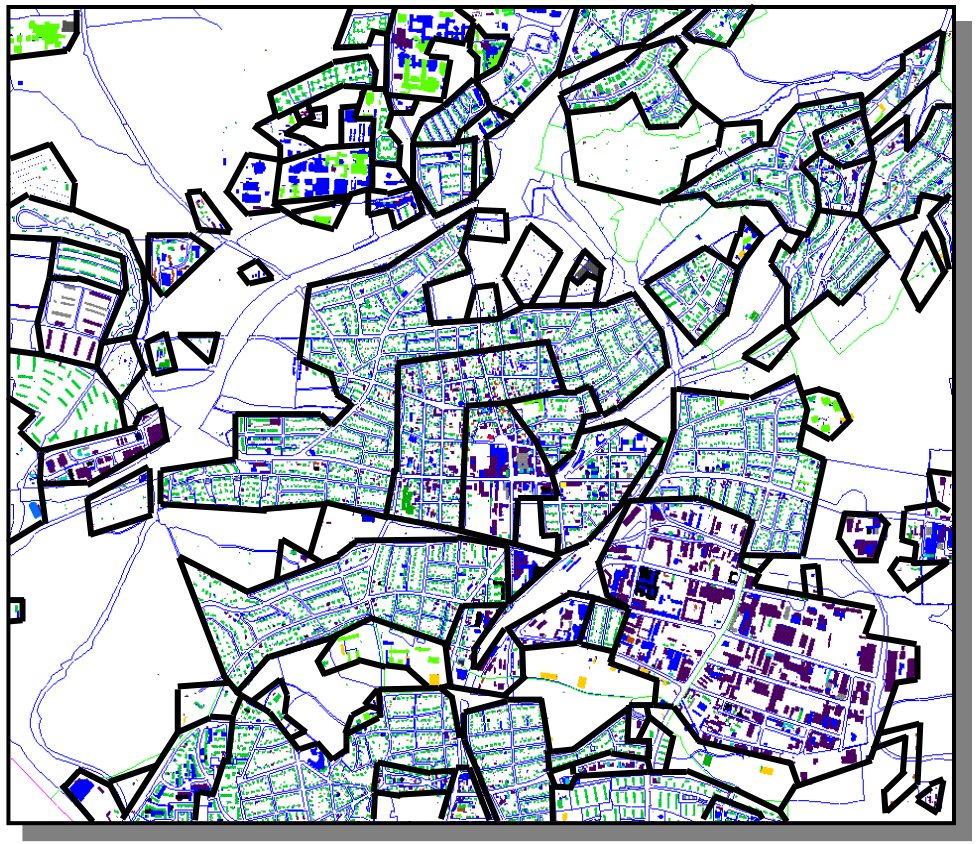
Karl-Heinrich Anders
University of Hannover

Motivation: Finding Object Groups in Spatial Data

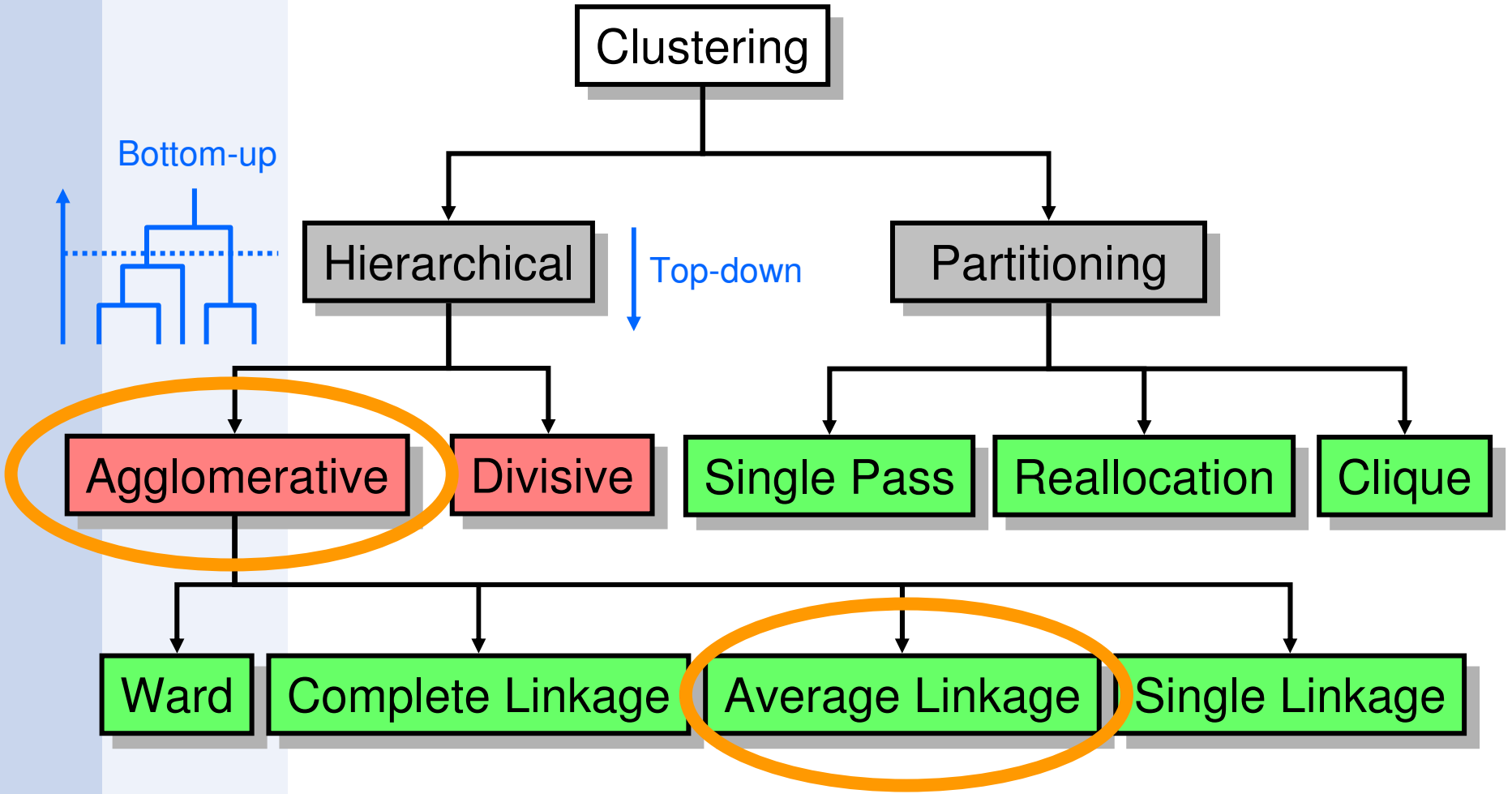


Motivation: Finding Object Groups in Spatial Data

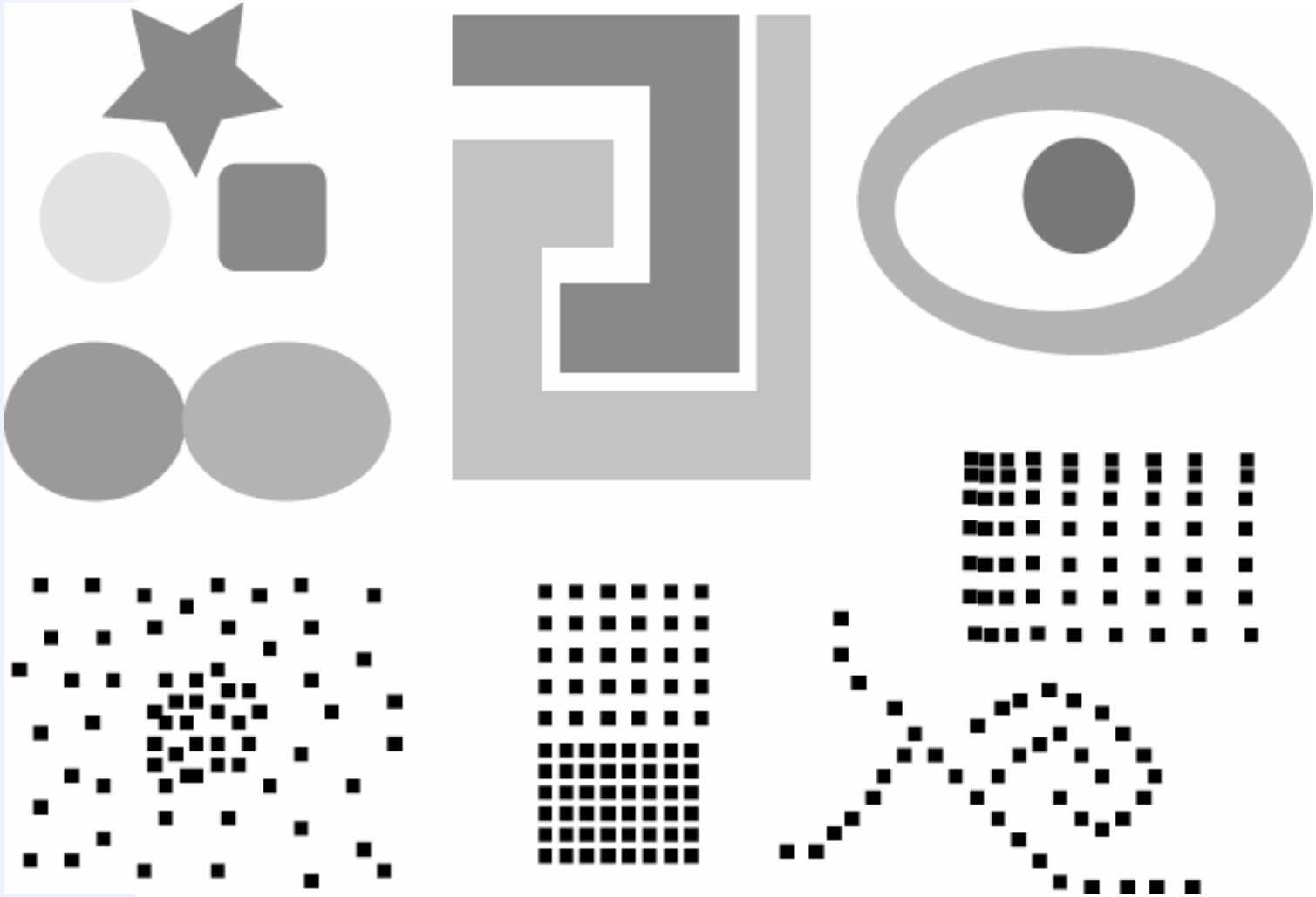
- Equal Objects
- Equal Spatial Distribution
- Equal Distribution of Mixed Object Types



Clustering Methods



Cluster Shapes

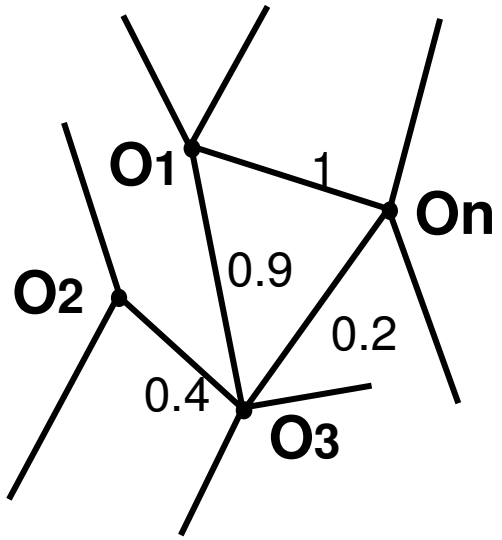


Graphbased Clustering I

- Powerful methods for clustering in difficult problems, [Jaromczyk und Toussaint, 1992].
  - Any cluster shape (convex, non convex).
  - Best agreement with human performance.
- Simple basic idea
  - Remove / Insert edges from / To a graph by a given criteria.
  - The resulting forest is the clustering
- Example
  - Shared Near Neighbours Method [Jarvis, 1973]

Graphbased Clustering II

$$A(O_1,\ldots,O_n) = \begin{pmatrix} 1 & 0 & 0.9 & . & . & . & 1 \\ 0 & 1 & 0.4 & . & . & . & 0 \\ 0.9 & 0.4 & 1 & . & . & . & 0.2 \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ 1 & 0 & 0.2 & . & . & . & 1 \end{pmatrix}$$



Distance matrix

Part of the dual graph

Which Distance Measure?

Which Graph?

Proximity Graphs

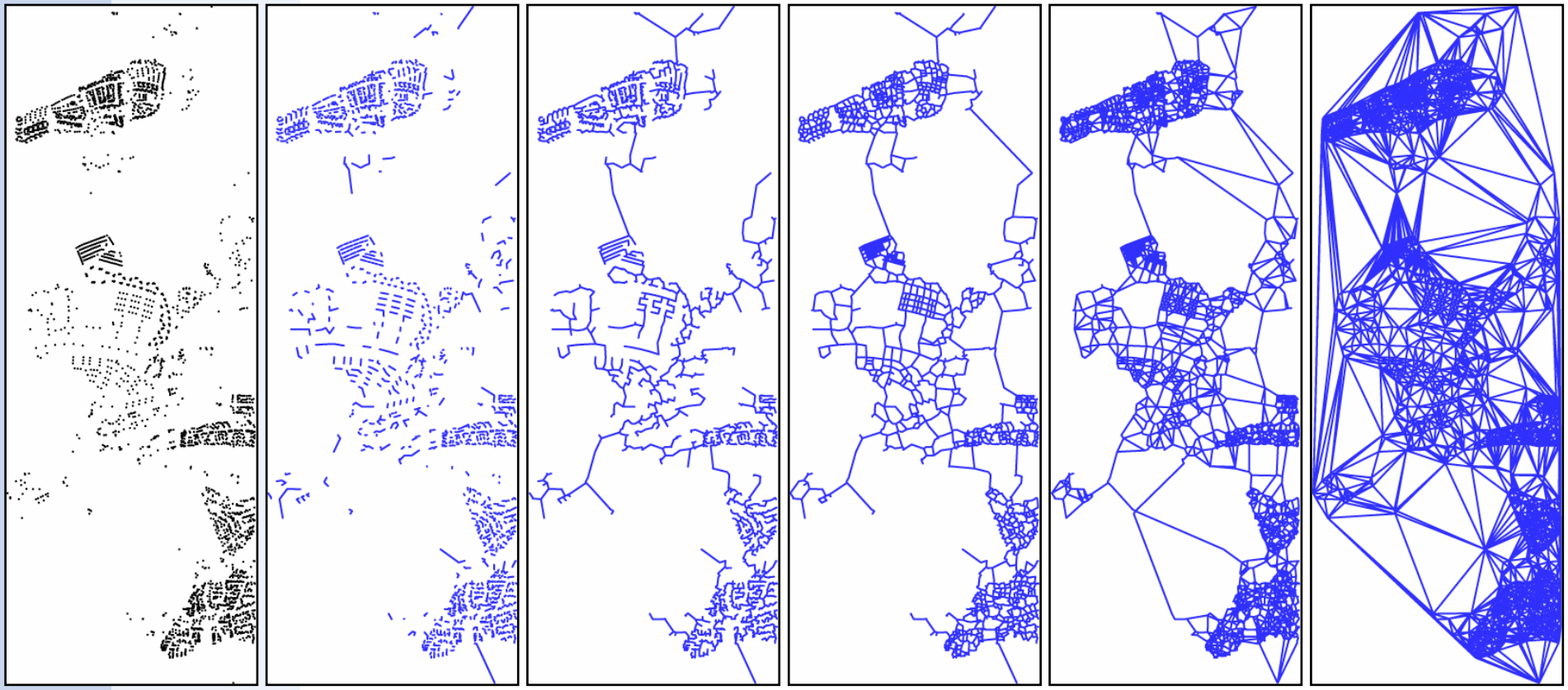
- Nearest Neighbourhood Graph (NNG)
- K-Nearest Neighbourhood Graph (k-NNG)
- Minimum Spanning Tree (MST)
- Relative Neighbourhood Graph (RNG)
- Gabriel Graph (GG)
- Delaunay Triangulation (DT)
- Sphere of Influence Graph (SIG)

In any L_p metric, for a fixed set V and $\beta \in [1,2]$ the following hierarchy is valid: $NNG \subseteq MST \subseteq RNG \subseteq G_\beta \subseteq GG \subseteq DT$.

→ **Natural Generalized Neighbourhood Model**

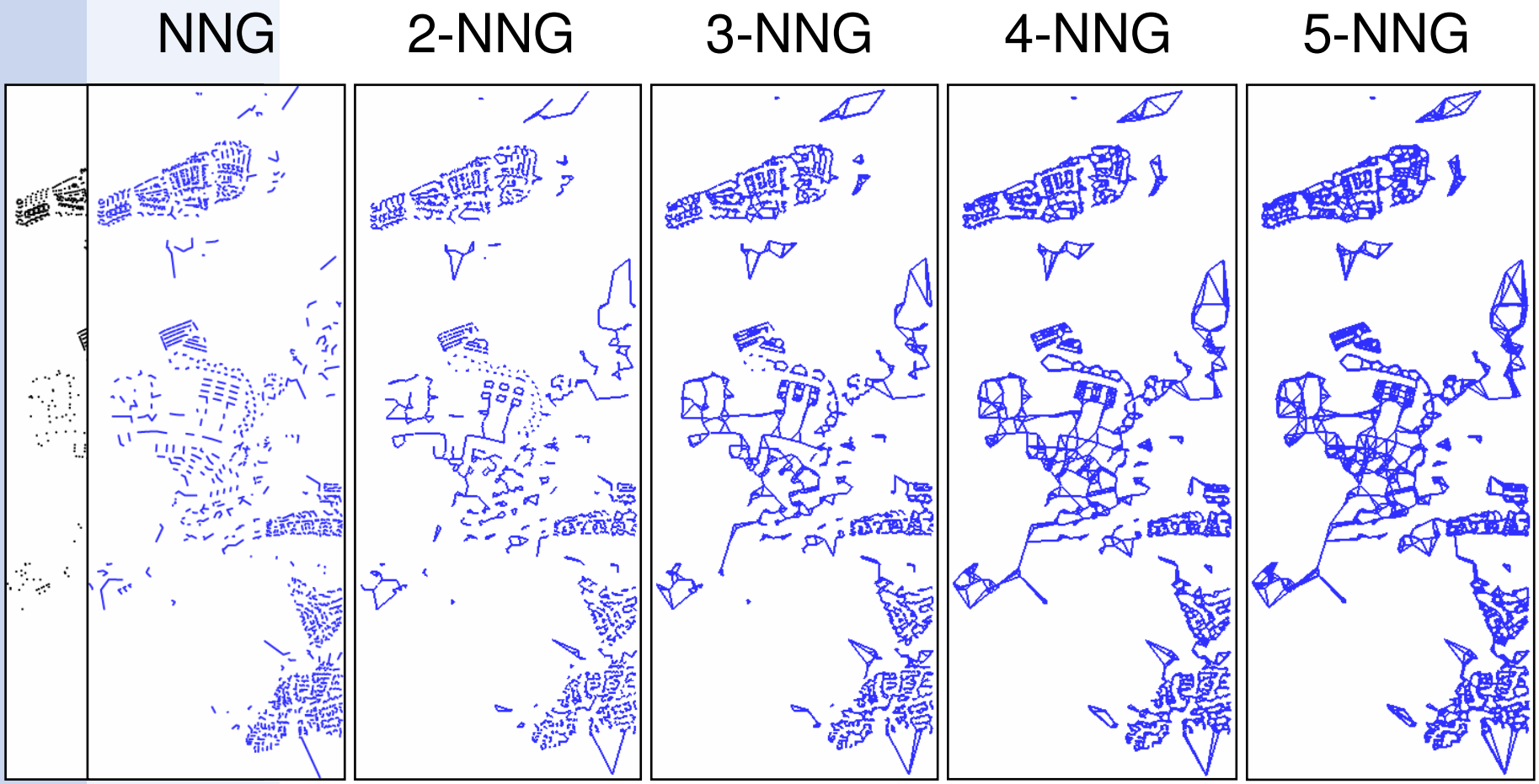
Neighbourhood Hierarchy

Local → Global

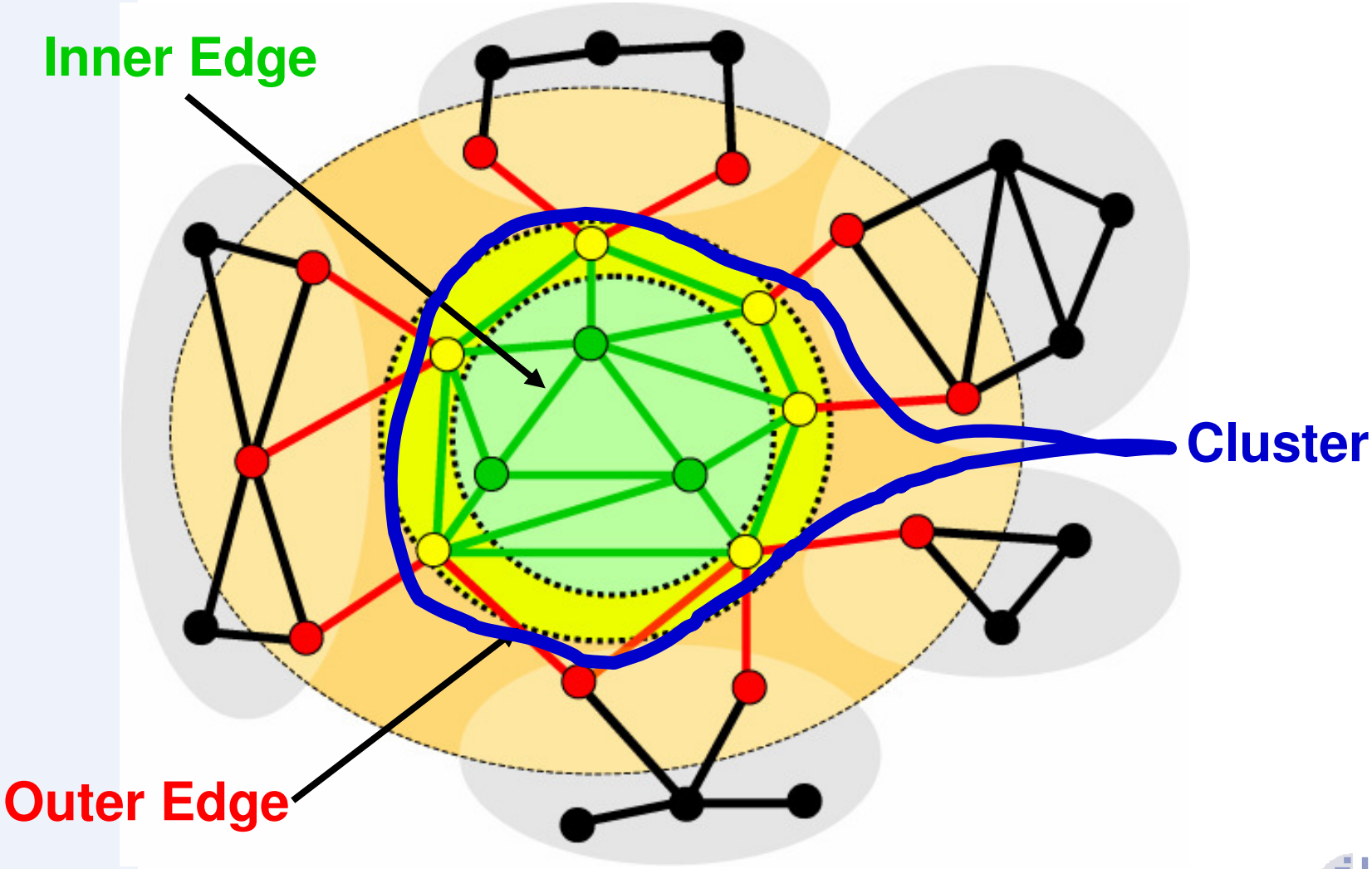


Point set NNG MST RNG GG DT

K-Nearest Neighbourhood Hierarchy



Cluster Definition

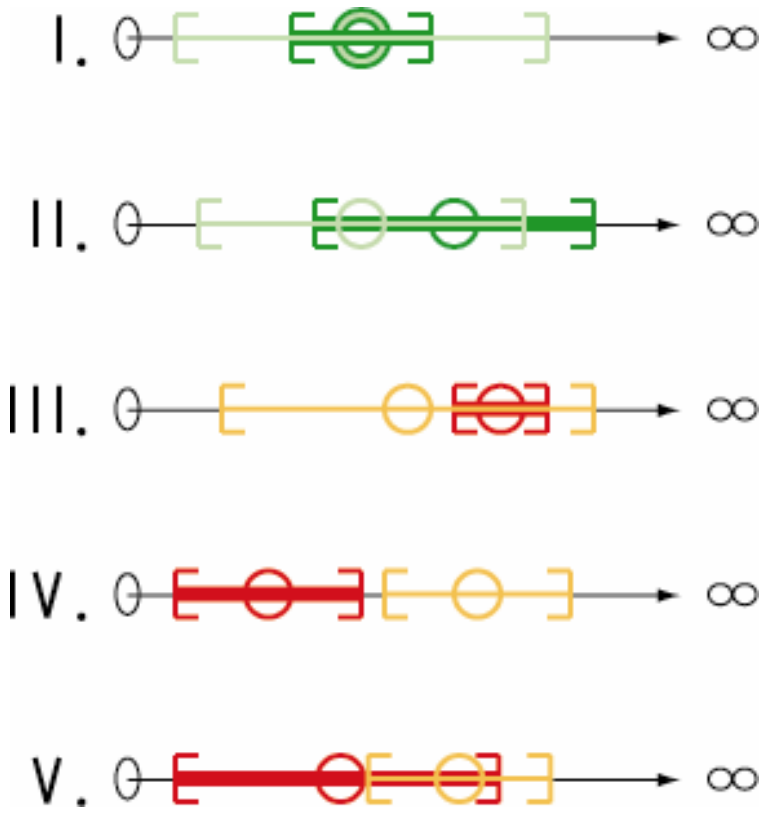


Clustering Rules

- Higher Density / Higher Priority
- Lower Variance / Higher Priority

- Density Compatibility
- Distance Compatibility
- Variance Compatibility

- Median based Outlier removal



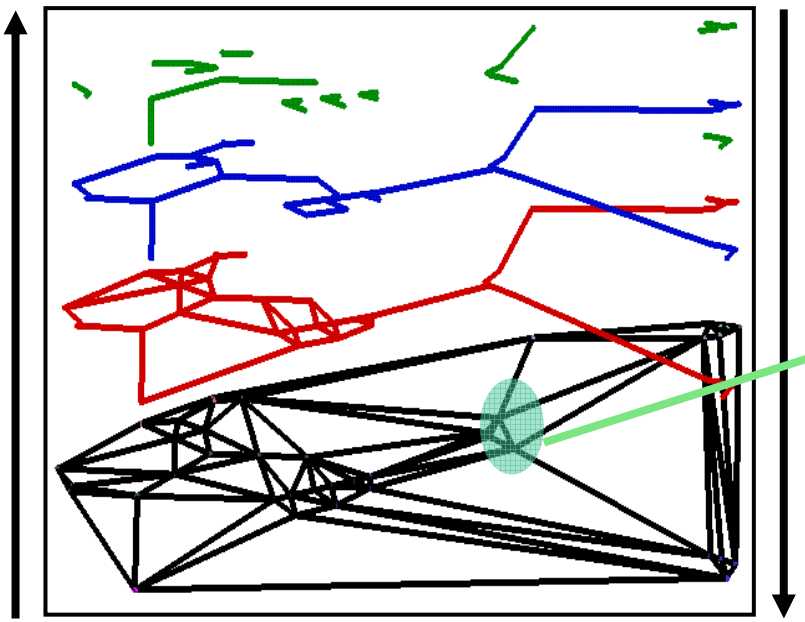
Algorithm

NNG

RNG

GG

DT



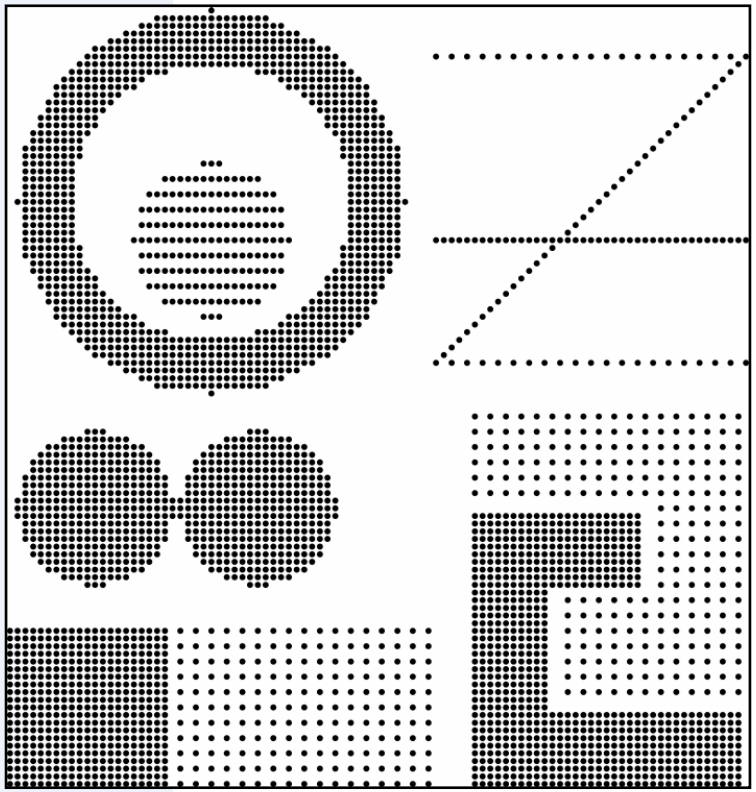
Outer edges

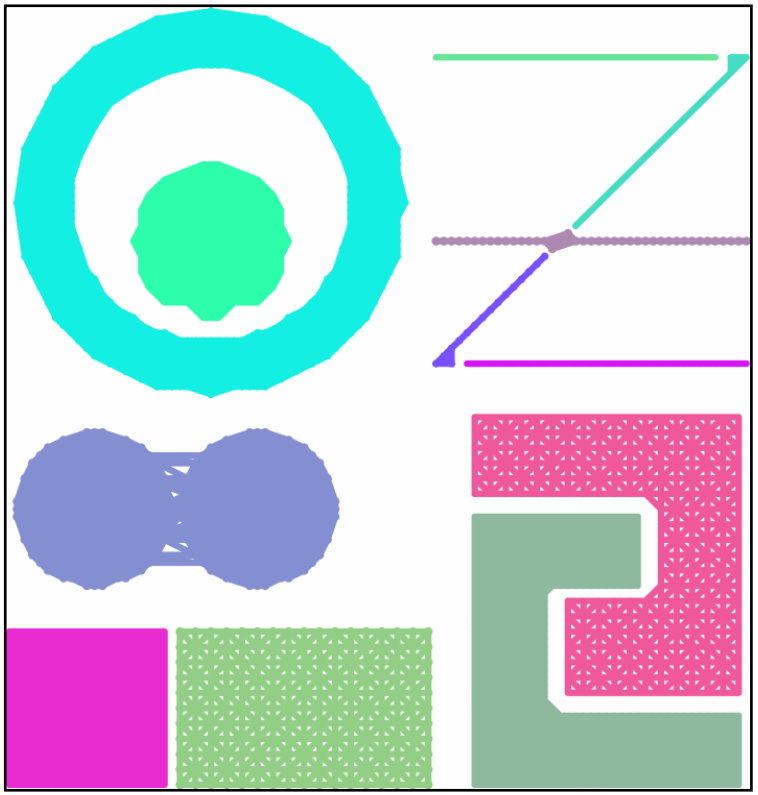
Inner edges

Computation of the graphs

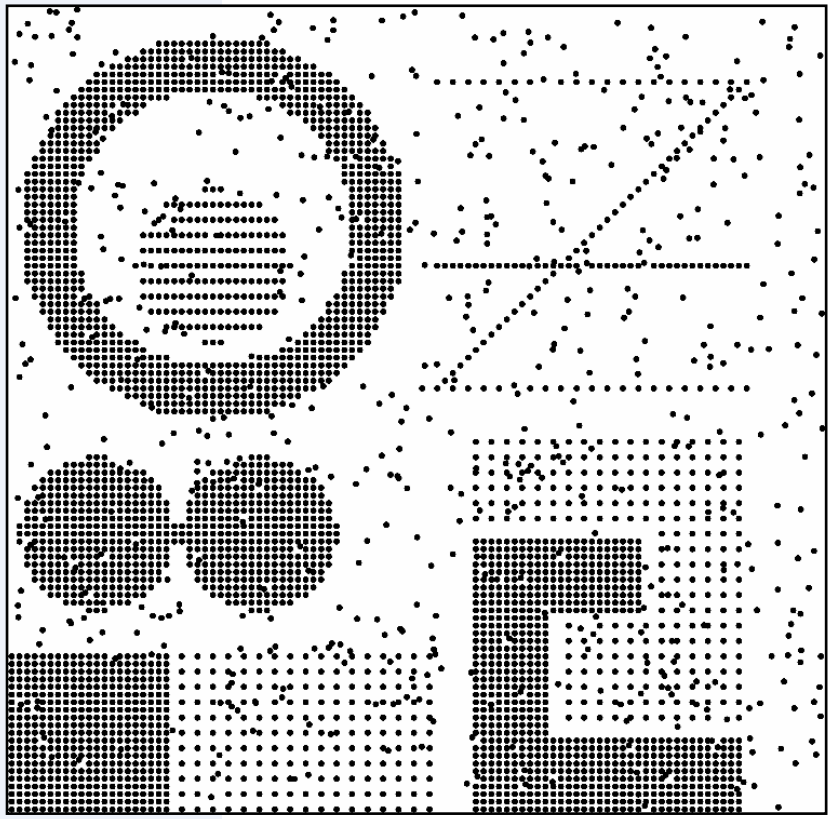
Clustering

Example: Artificial Point Set

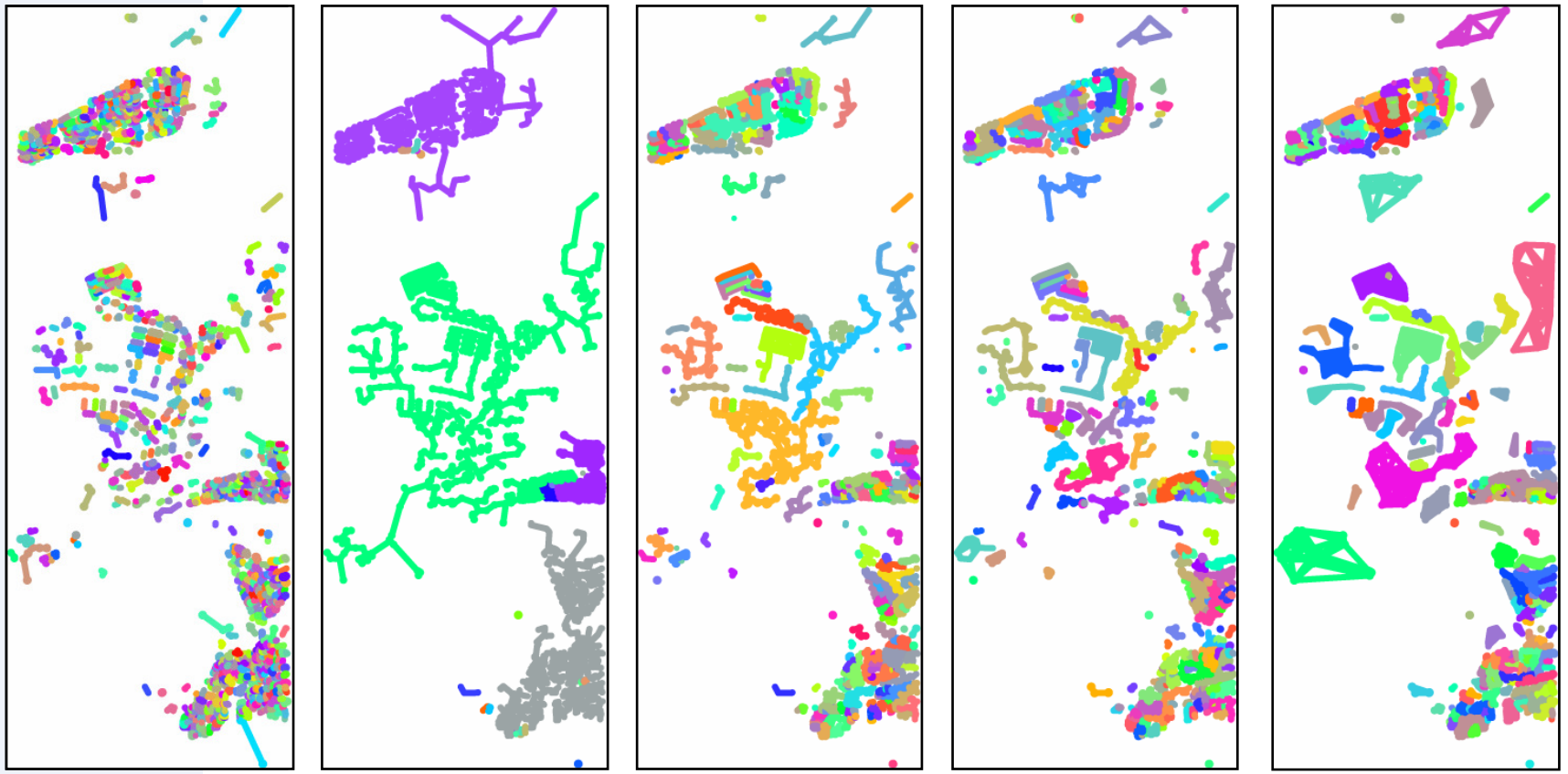




Example: Point Set with Noise

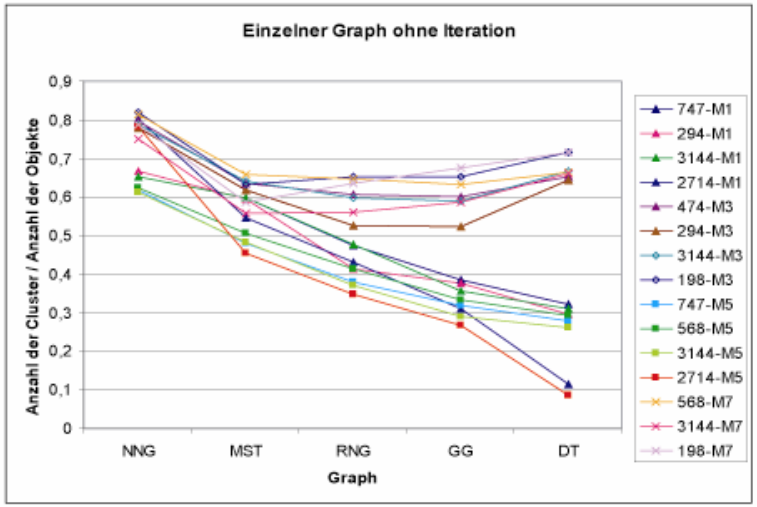


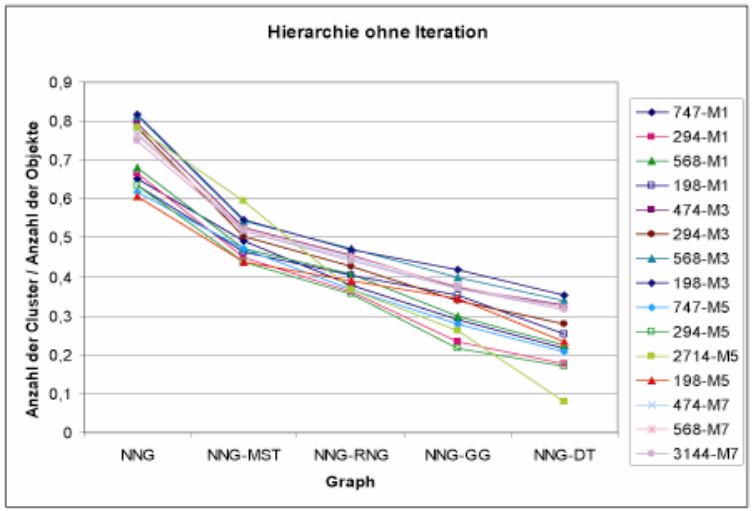
Example: Settlement Structure

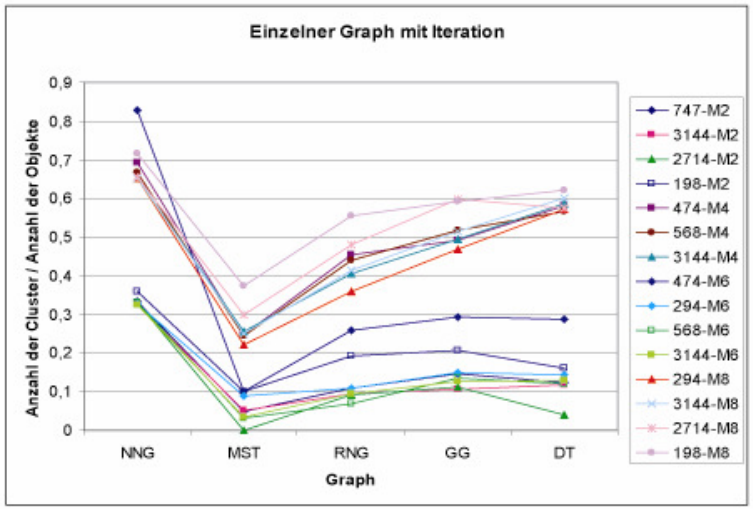


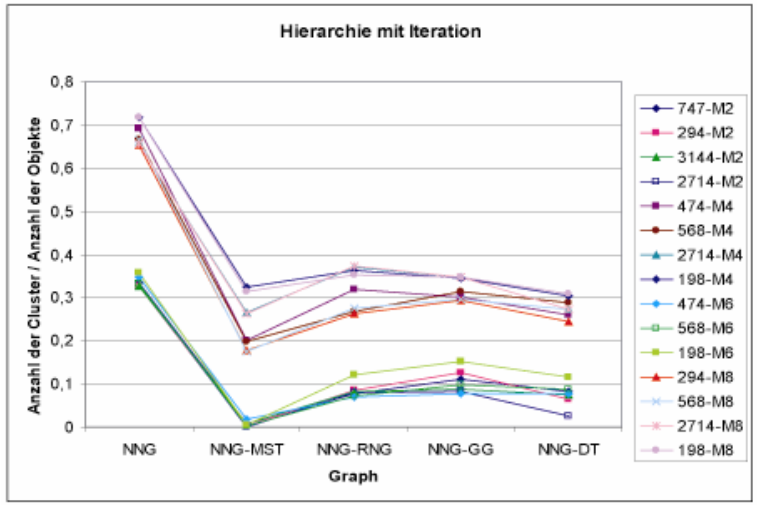
NNG NNG-MST NNG-RNG NNG-GG NNG-DT

Number of Detected Clusters

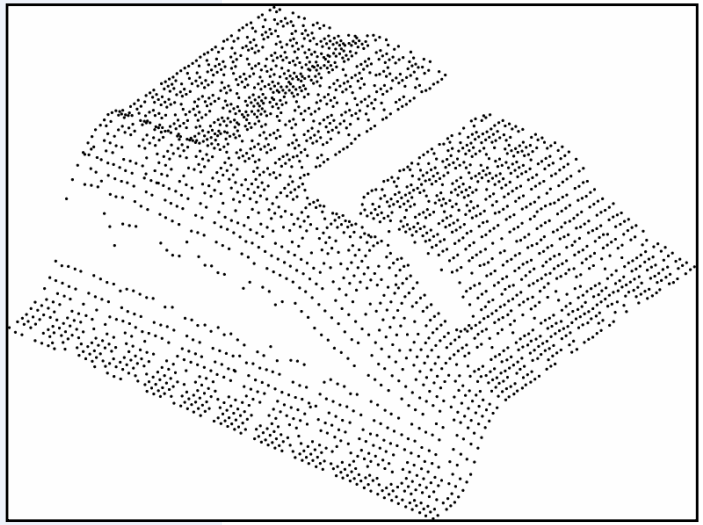


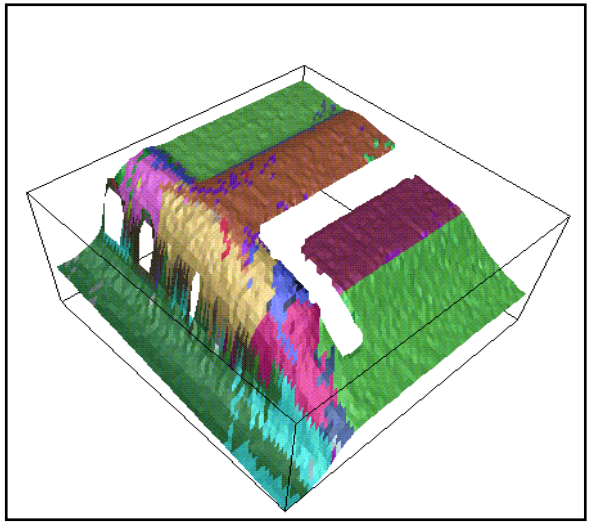






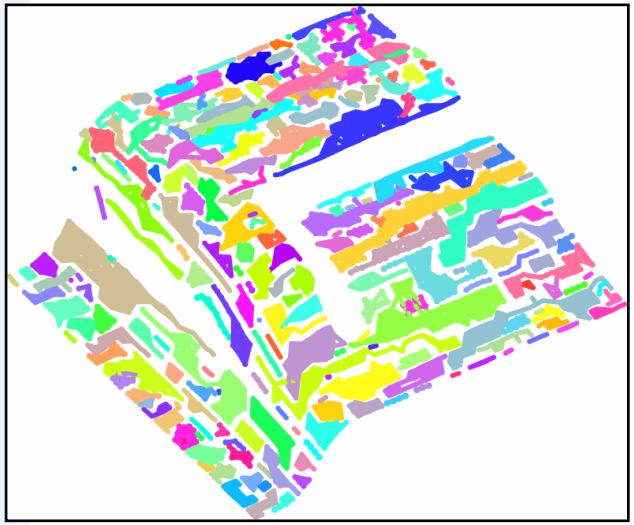
Example: 3D Range Data

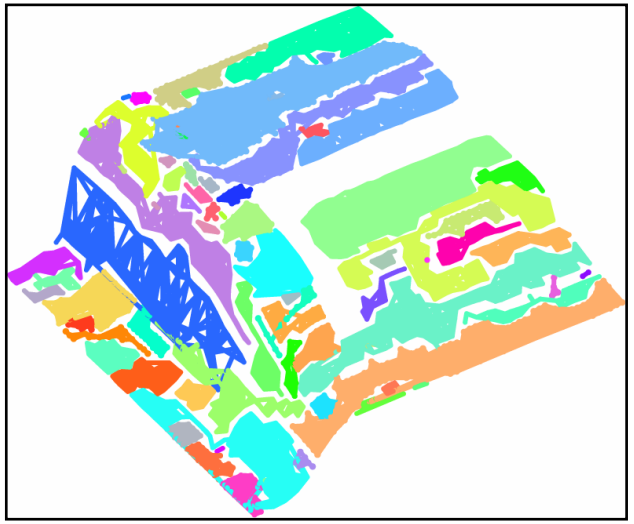




Special Curvature Based Seg-mentation

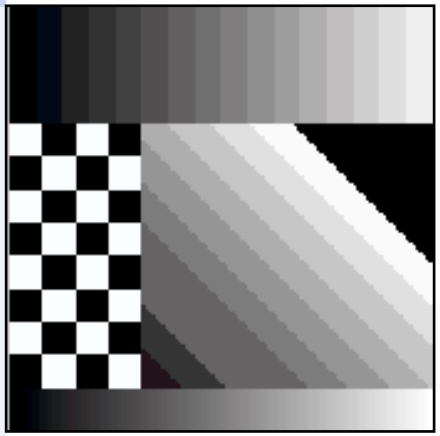
No Iteration

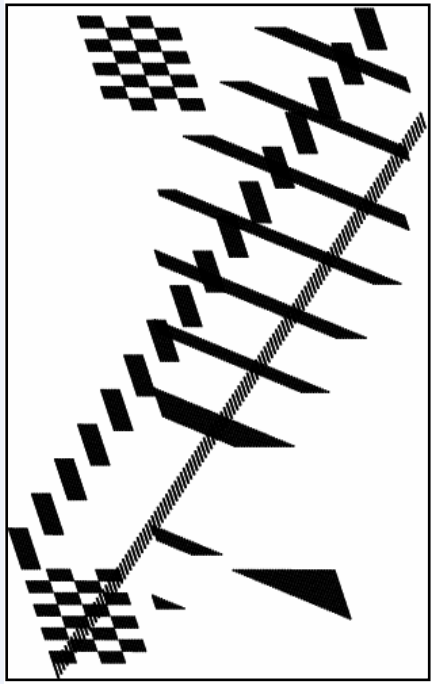


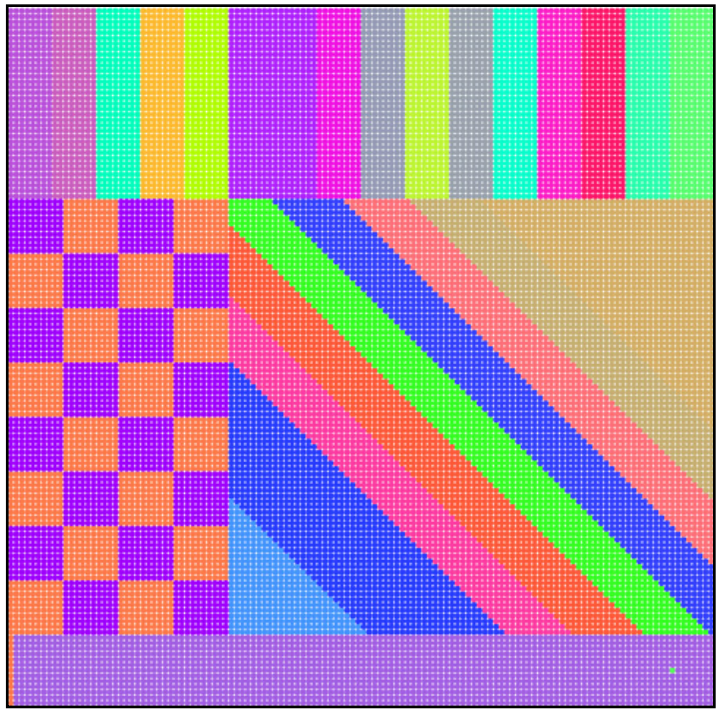


Iteration

Example: Greyscale Image







Conclusion

- Proximity graphs are well suited to find spatial object cluster.
- They provide a natural hierarchical neighbourhood (similarity) model.
- A more detailed approach should use a connected component analysis.
- The Delaunay Triangulation is only usefull for 2 or 3 dimensional feature spaces.
- The k-Nearest Neighbour Hierarchy should be used for high dimensional feature spaces .
  - Disadvantage: Introduction of the parameter k