LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
FAKULTÄT FÜR MATHEMATIK, INFORMATIK UND STATISTIK
INSTITUT FÜR INFORMATIK
LEHRSTUHL FÜR DATENBANKSYSTEME UND DATA MINING

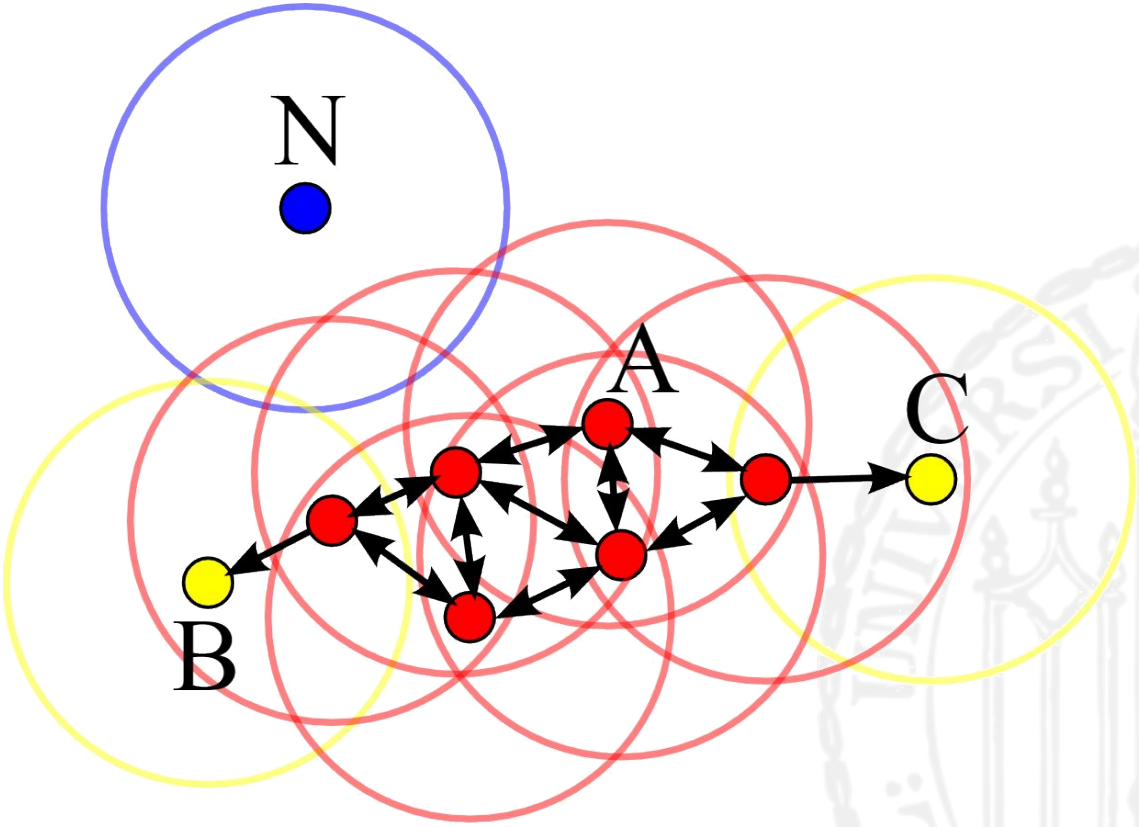
Chaindetection for DBSCAN

Janis Held

DBS

Wissenschaftlicher Betreuer:
Anna Beer
Verantwortlicher Professor:
Prof. Dr. Thomas Seidl

DBSCAN



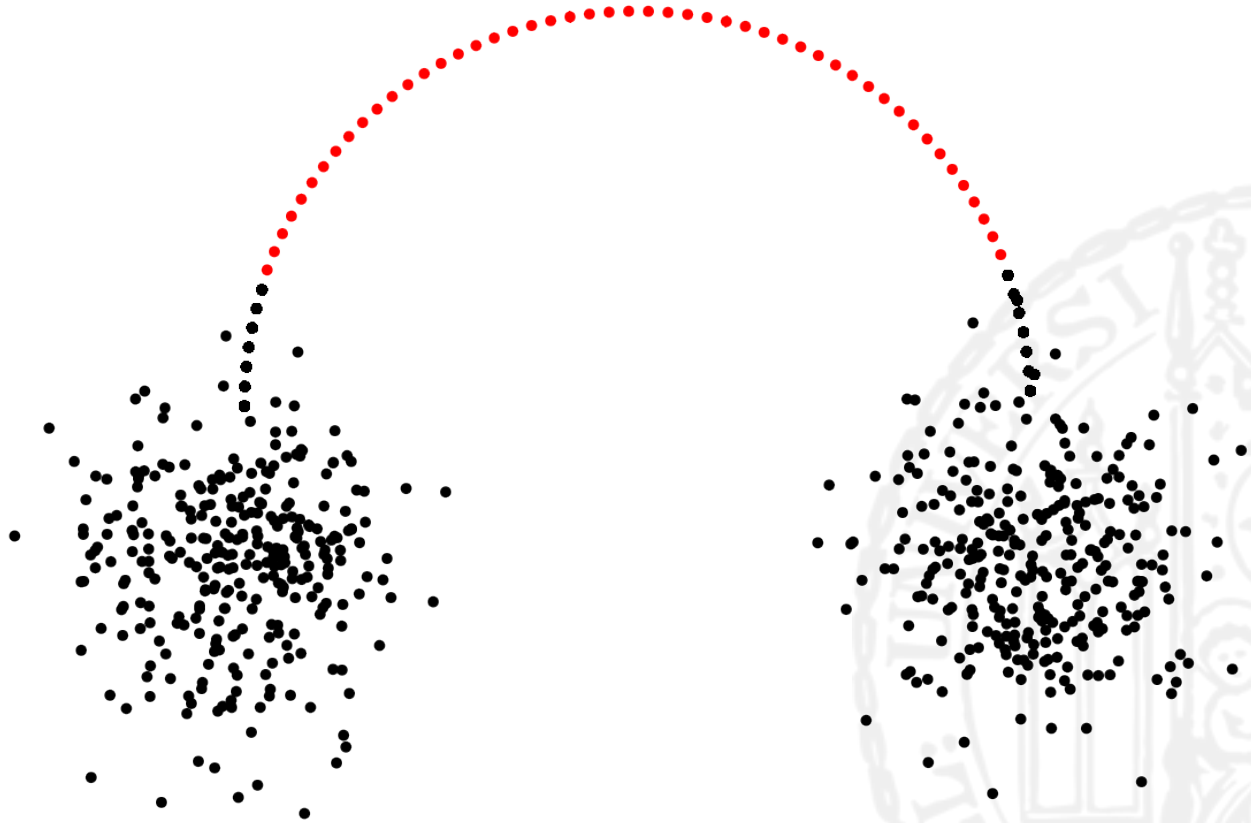
DBSCAN Cluster

Besseres Clustering

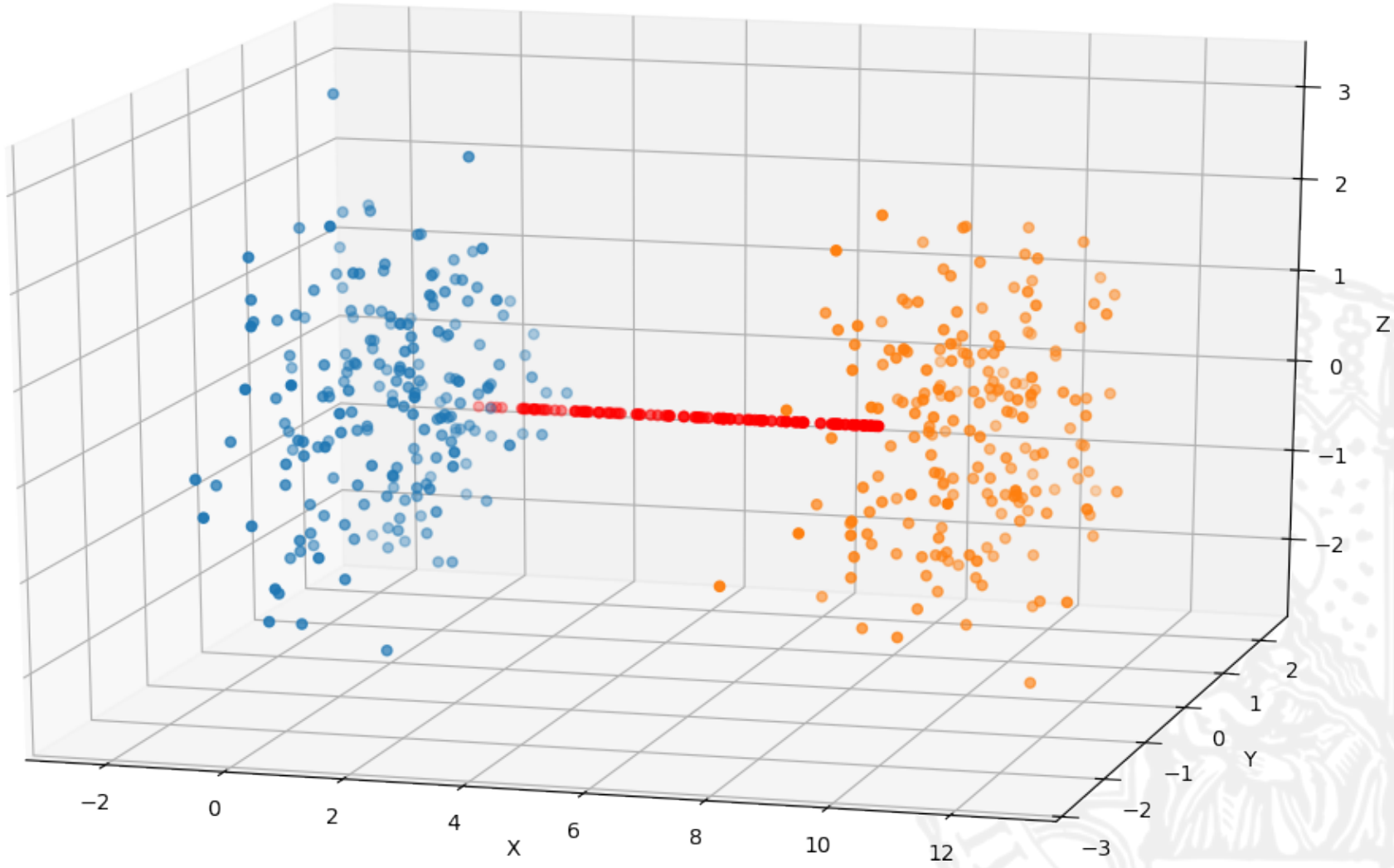
Was sind chains?

Keine chain?

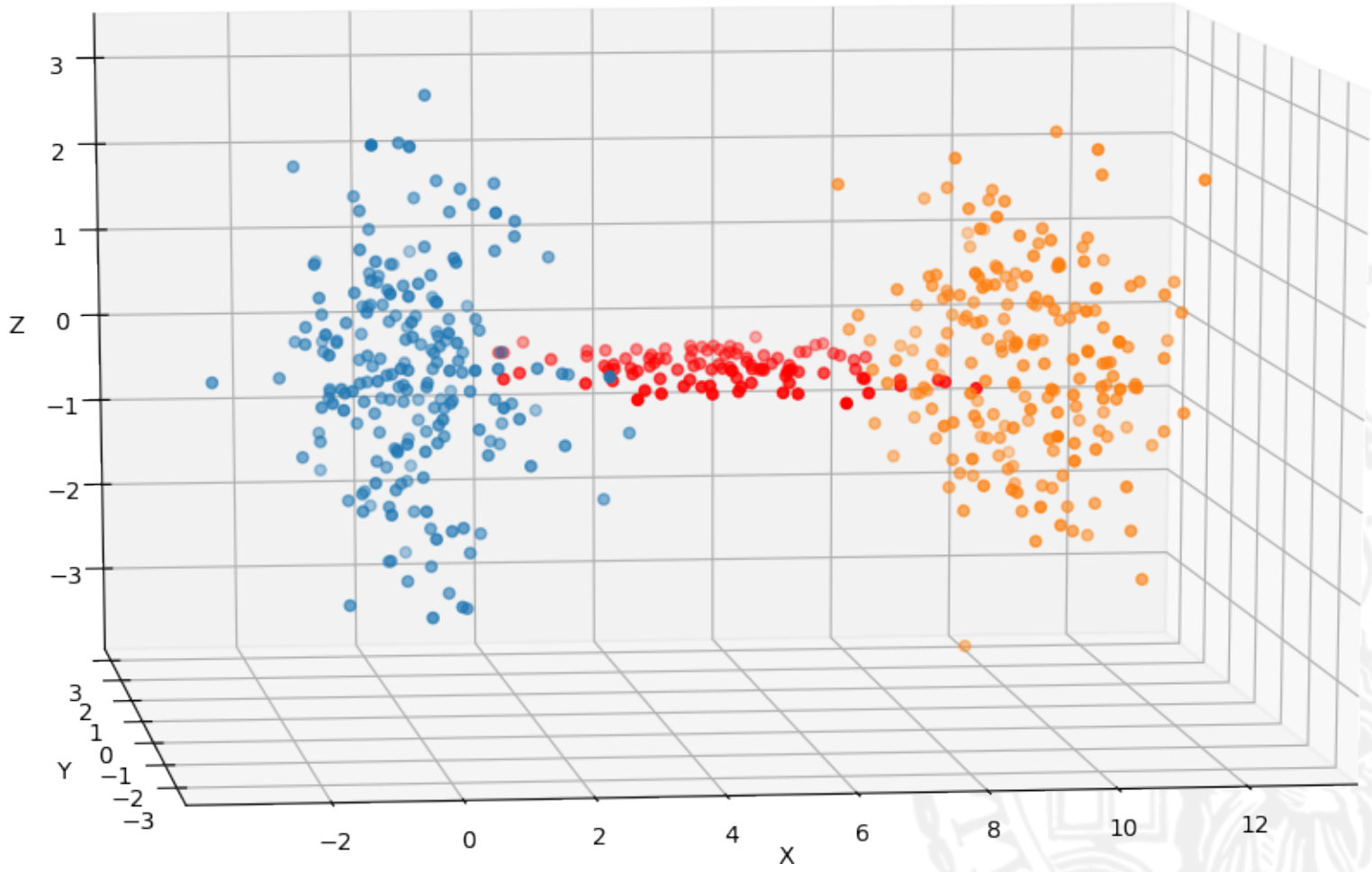
Ist das eine chain?



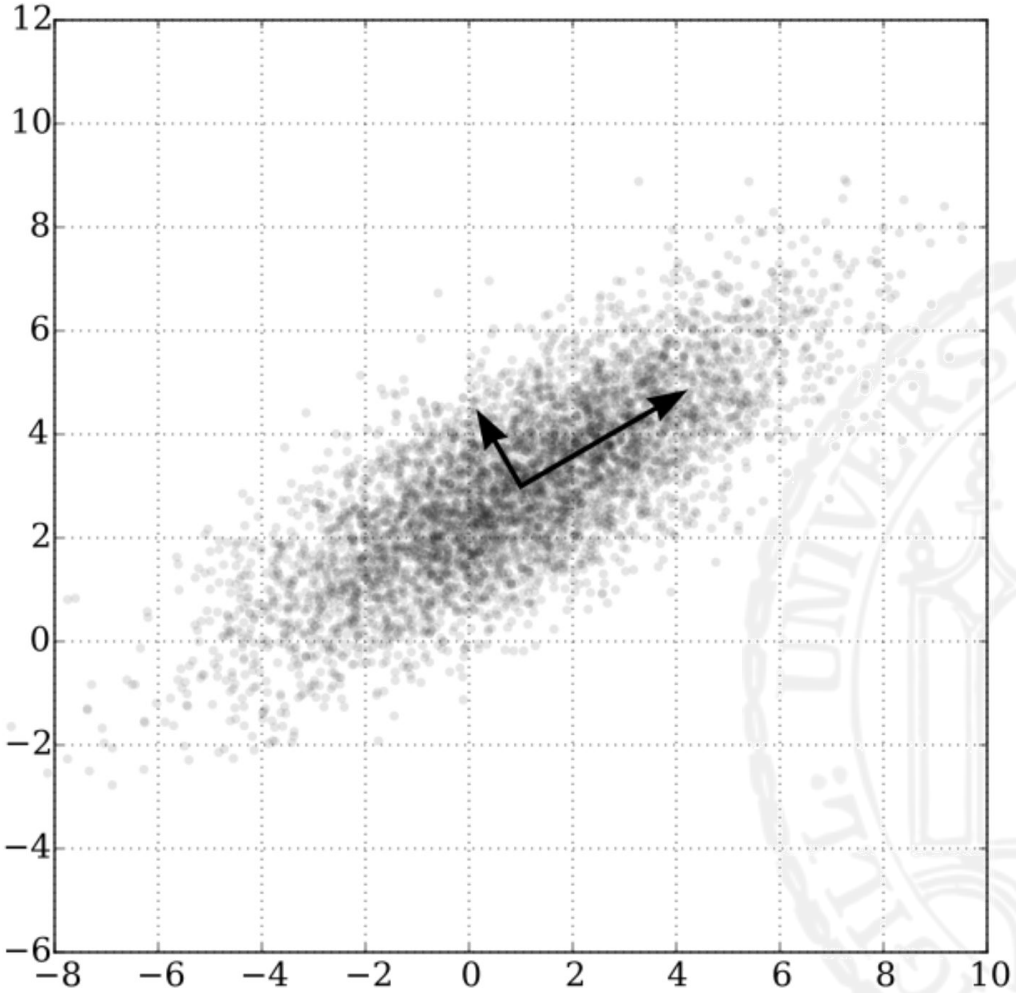
Eine 1D chain im 3D Raum



Eine 2D chain im 3D Raum

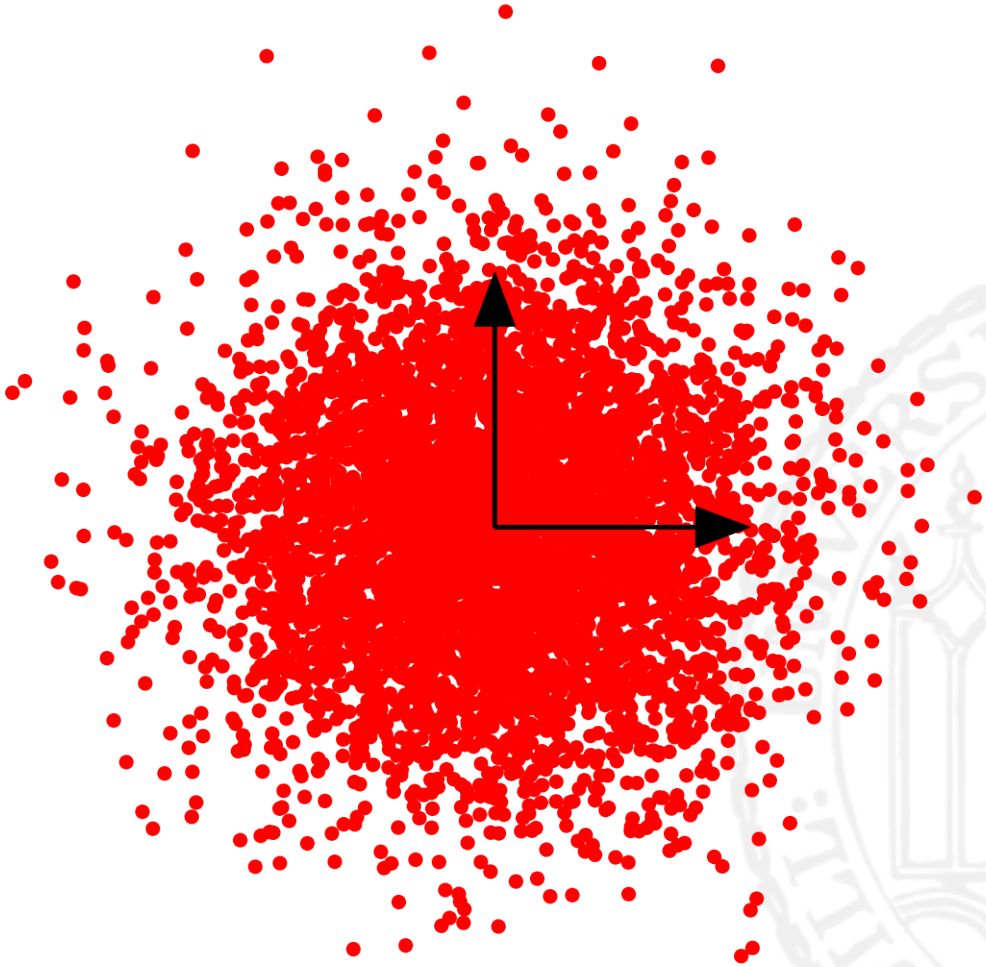


Die Idee: Normierte Eigenwerte der Kovarianzmatrix



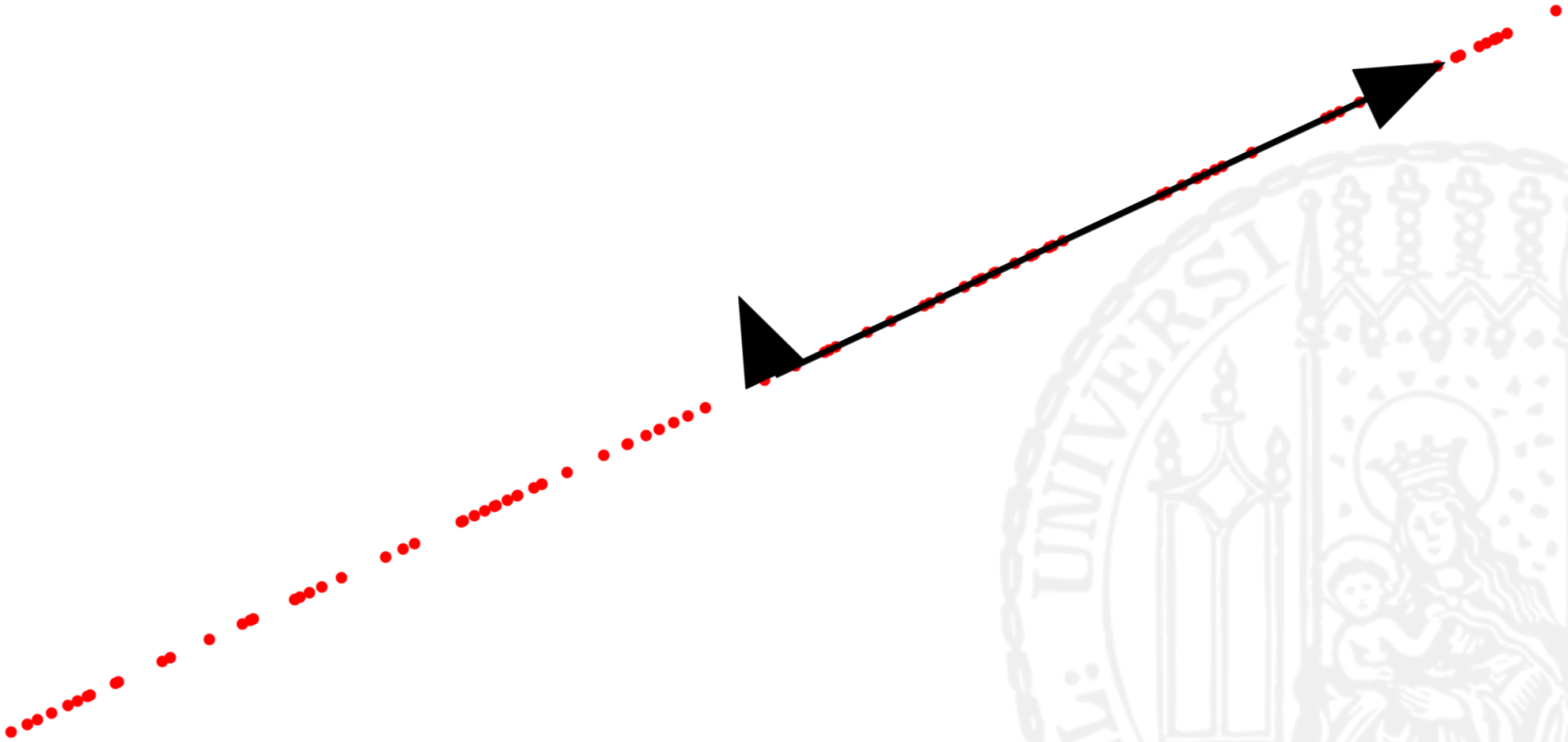
Eigenwerte der Kovarianzmatrix

Alle Eigenwerte gleich



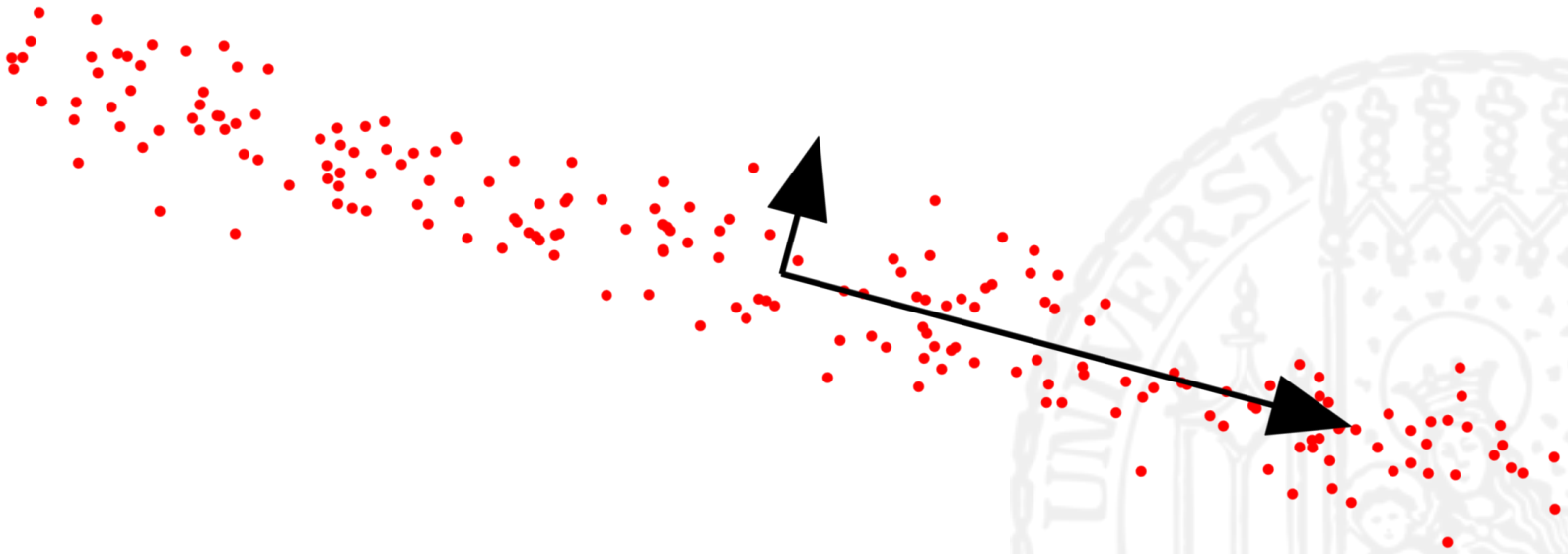
Eigenwerte der Kovarianzmatrix

Nur ein Eigenwert ungleich 0



Eigenwerte der Kovarianzmatrix

Ein Eigenwert nahe 0

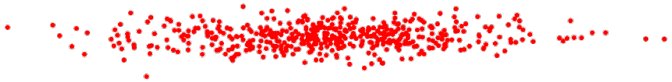


Anwendung

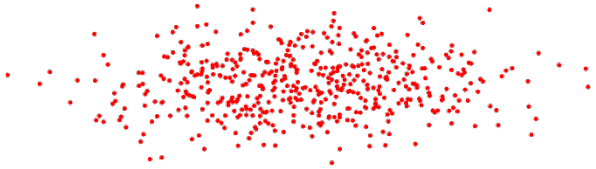
- Parameter:
  chainDim aus *{0,...,d}*,
  allowedVariation aus *[0,1[*
- Betrachte *epsilon* Umgebung eines Punktes und berechne normierte Eigenwerte der Kovarianzmatrix.
- Berechne normierten Fehler aus den *d* - *chainDim* kleinsten normierten Eigenwerten.
- Punkt ist Kandidat wenn der normierte Fehler kleiner gleich *allowedVariation* ist.

Normierte Fehler für Punktemengen mit *ChainDim = 1.*

Normierter Fehler ~ 0,0002



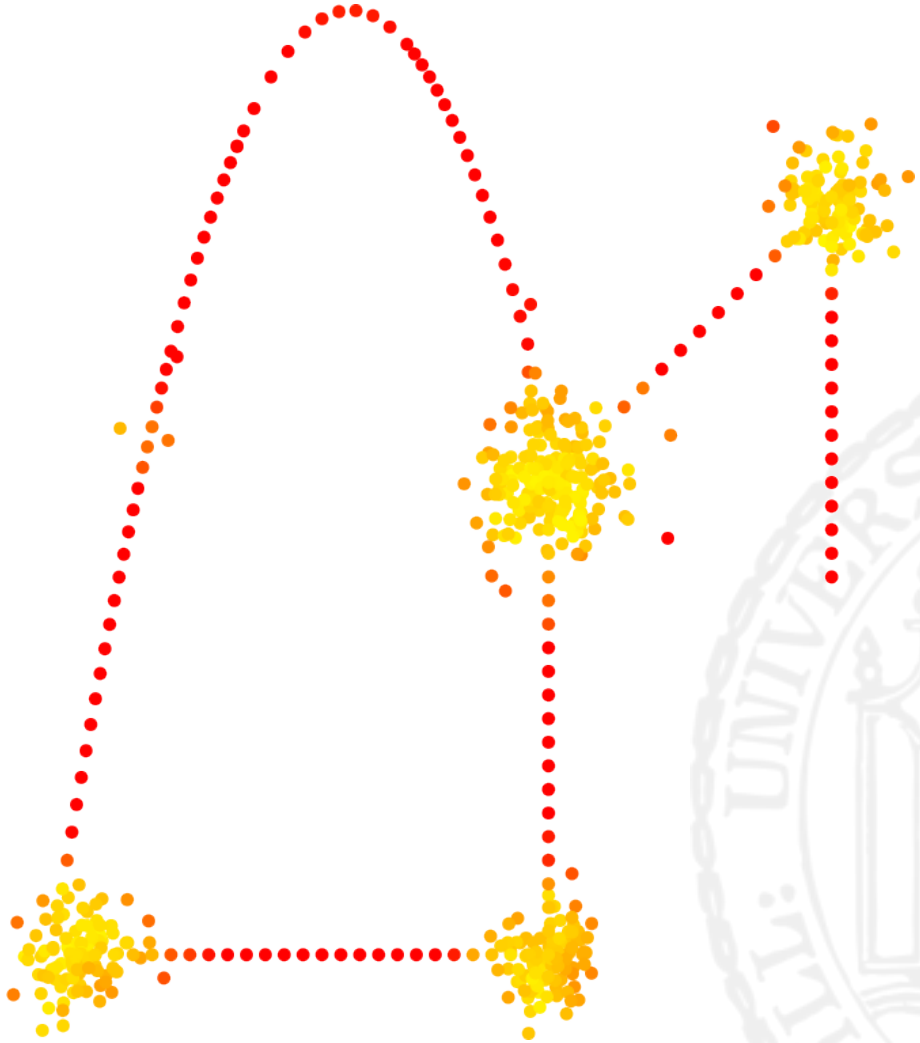
Normierter Fehler ~ 0,023



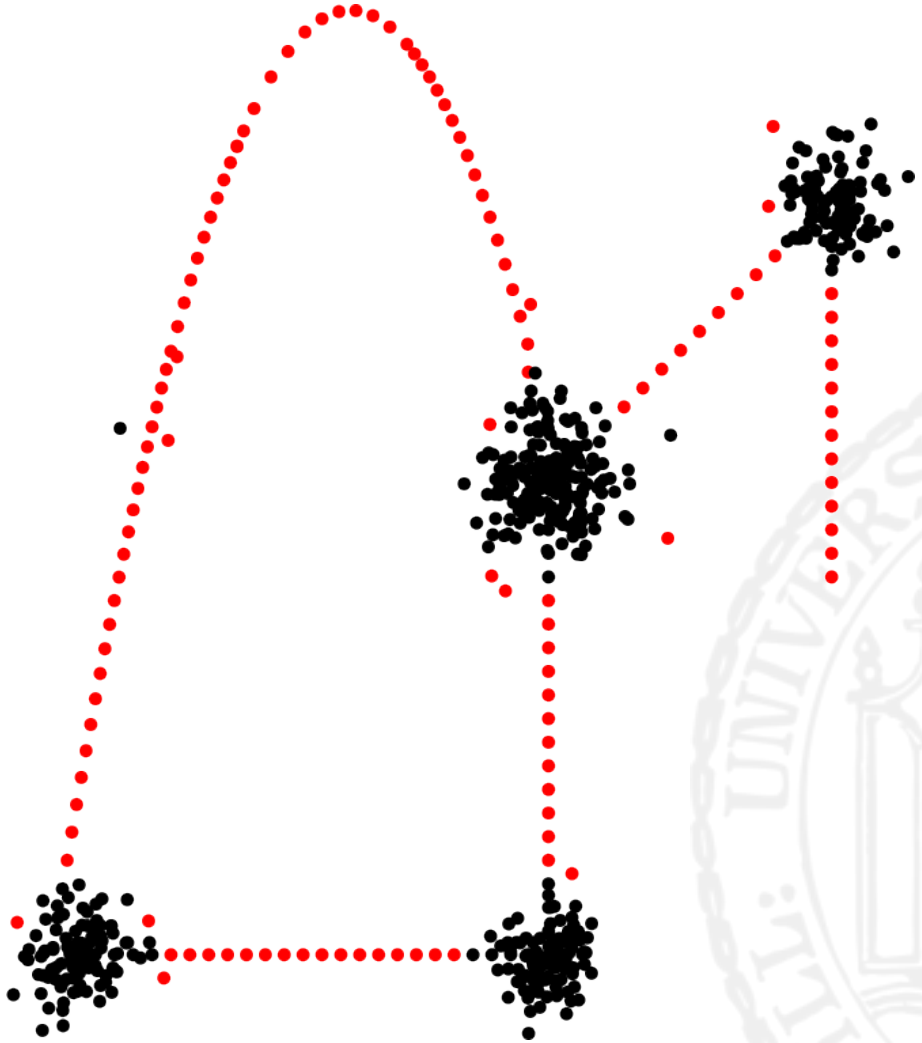
Normierter Fehler ~ 0,1563

Normierter Fehler ~ 0,9997

Beispiel: Normierte Fehler der Epsilon-Umgebungen

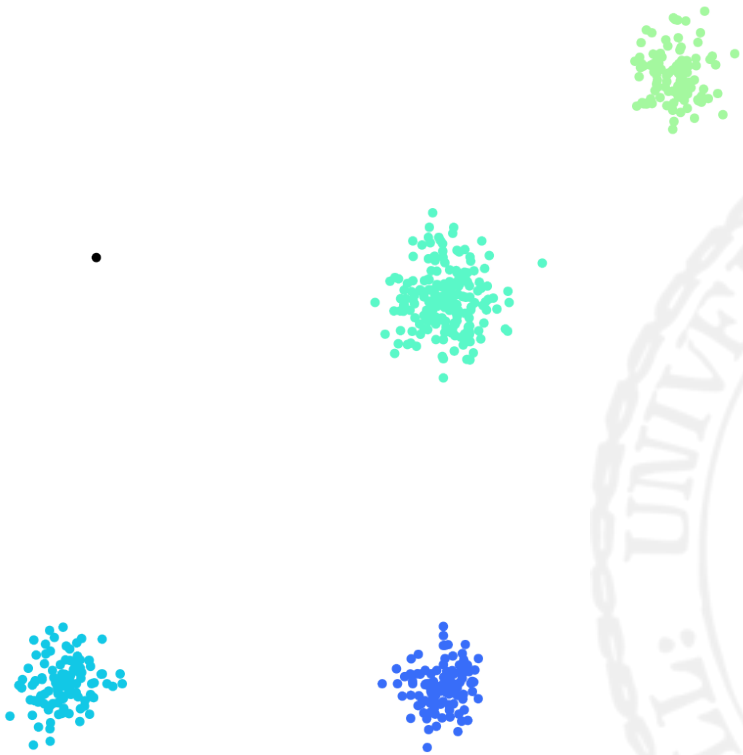


Kandidaten mit *allowedVariation* 0.2



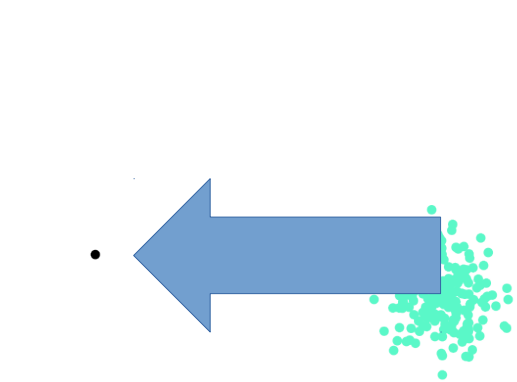
Verfeinerung der Kandidatenmenge

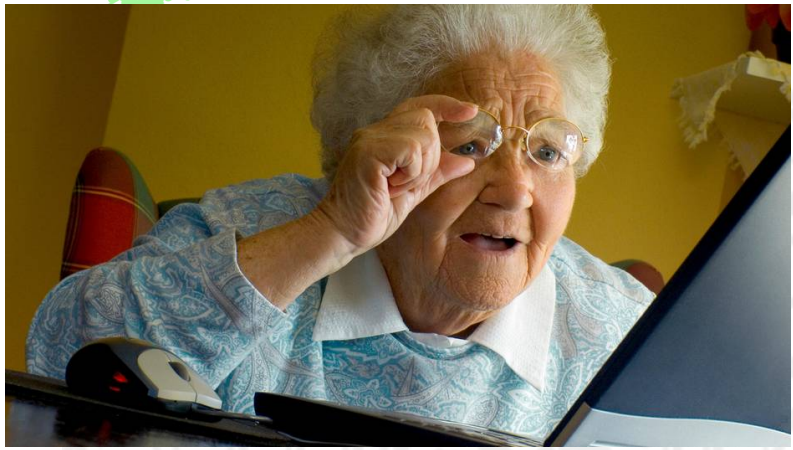
Clustering der Nicht-Kandidaten
Outlier werden zur Kandidatenmenge hinzugefügt



Verfeinerung der Kandidatenmenge

Cluster der Nicht-Kandidaten
Outlier sind werden zur Kandidatenmenge hinzugefügt





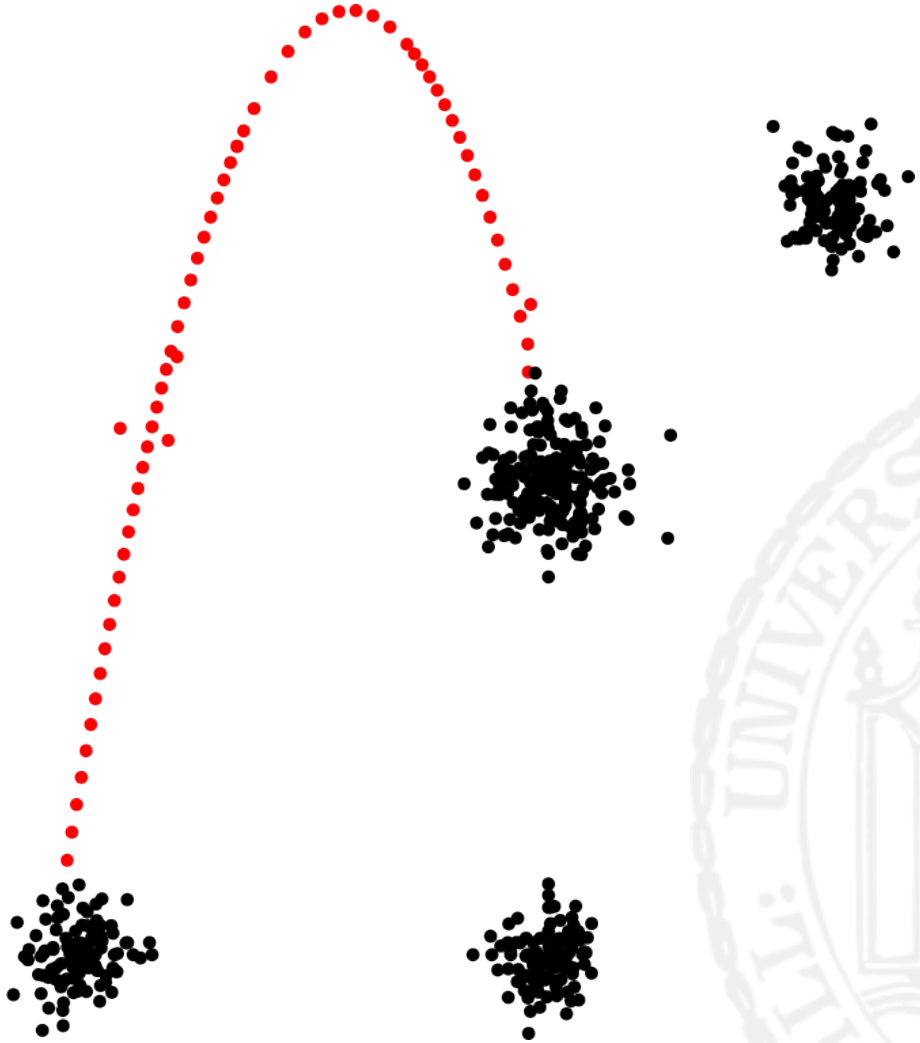
Chaindetection for DBSCAN
Bild: https://www.stern.de/neon/vorankommen/mobilitaet/memes--in-unserem-neon-online-knigge-geht-es-dieses-mal-um-memes-8397568.html

Verfeinerung der Kandidatenmenge

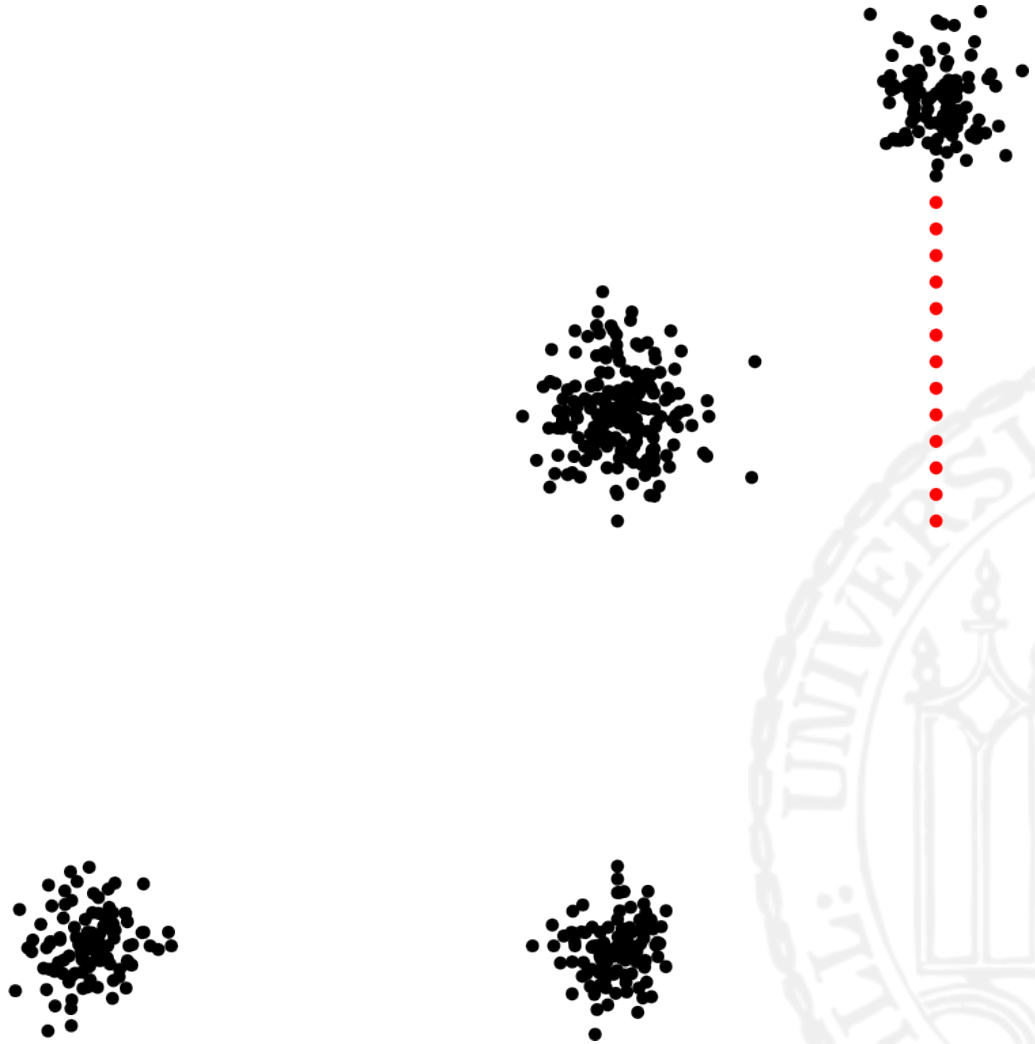
Cluster der Kandidaten

Outlier werden von der Kandidatenmenge entfernt

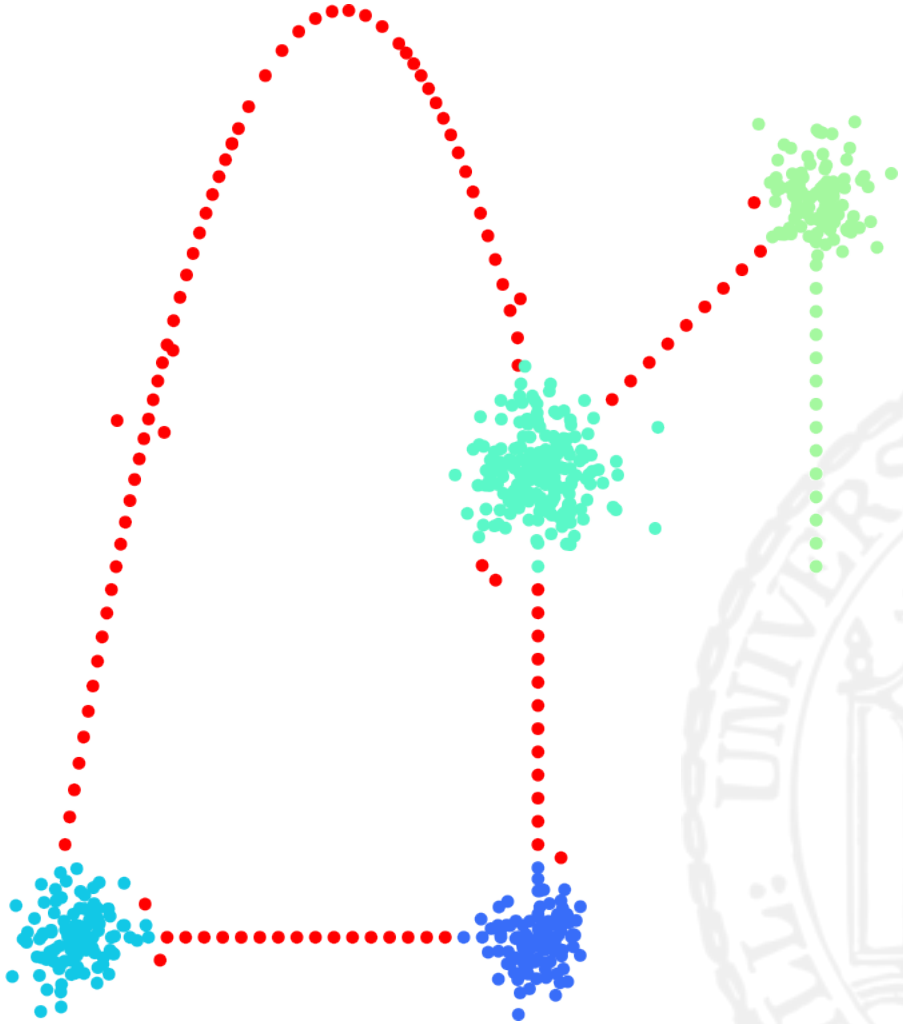
Verifizierung der Chaincandidates



Verifizierung der Chaincandidates

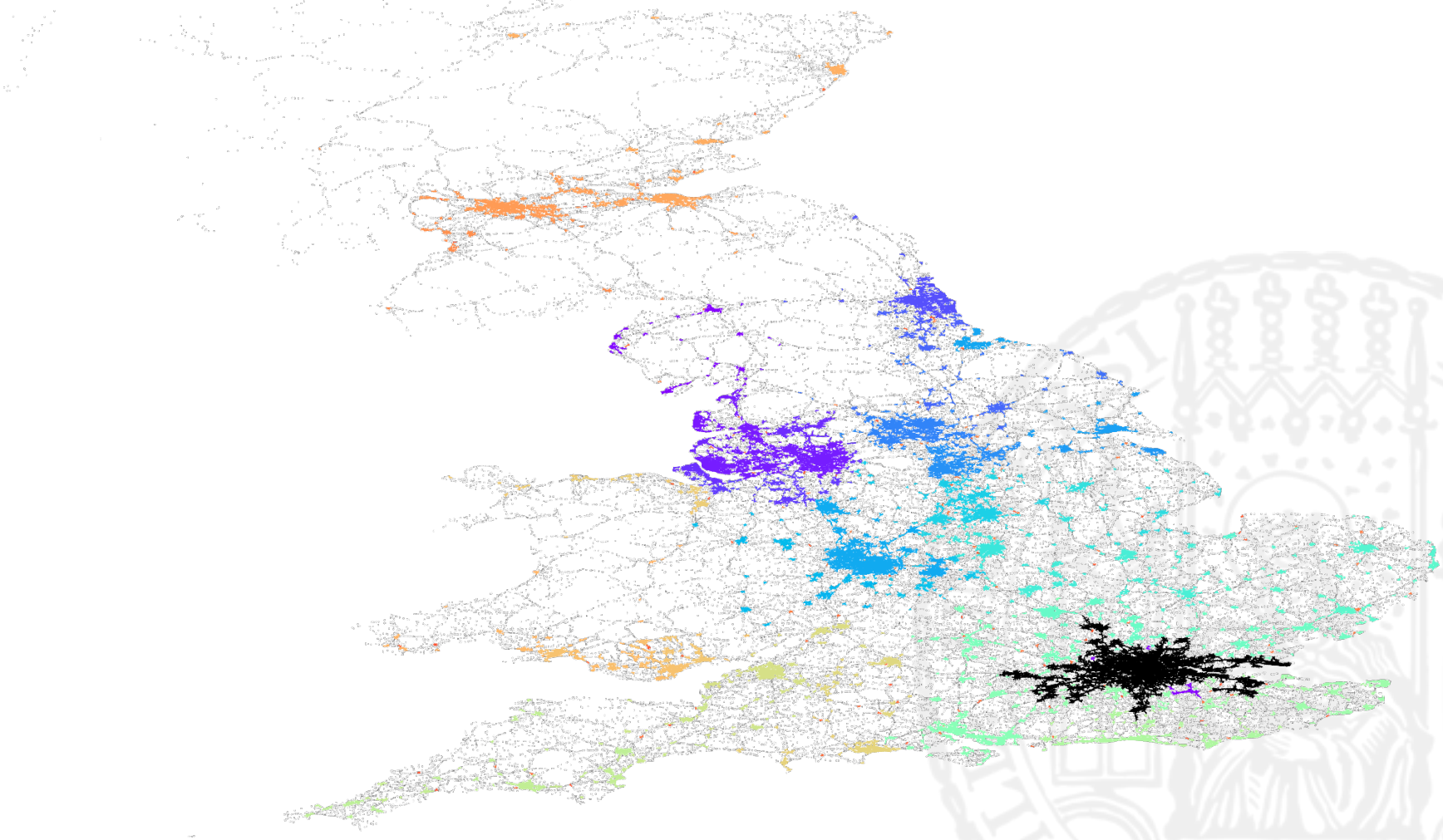


Clustering mit Chaindetection

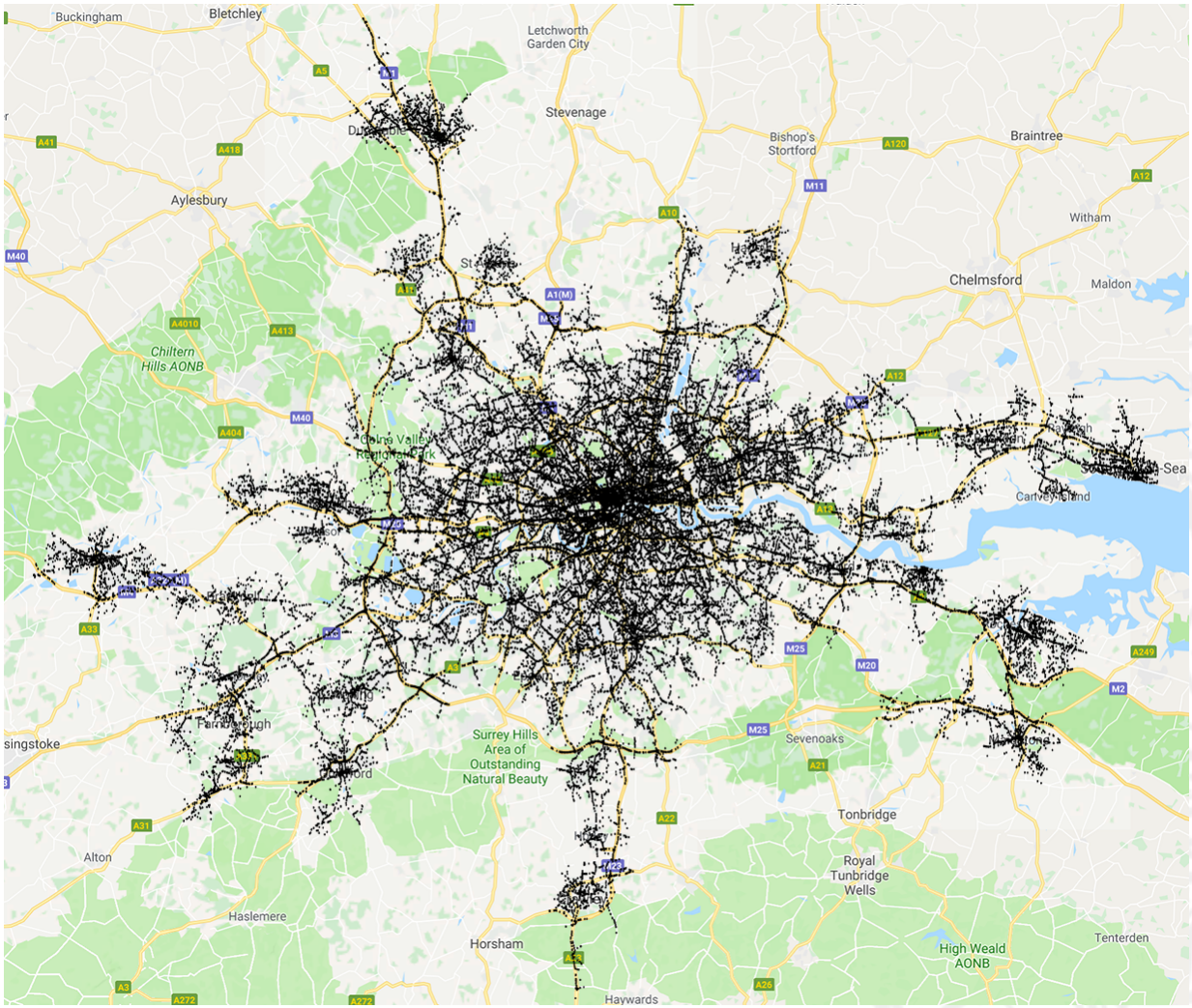


Verkehrsunfälle in England

DBSCAN mit eps = 0.01 und minPts = 15

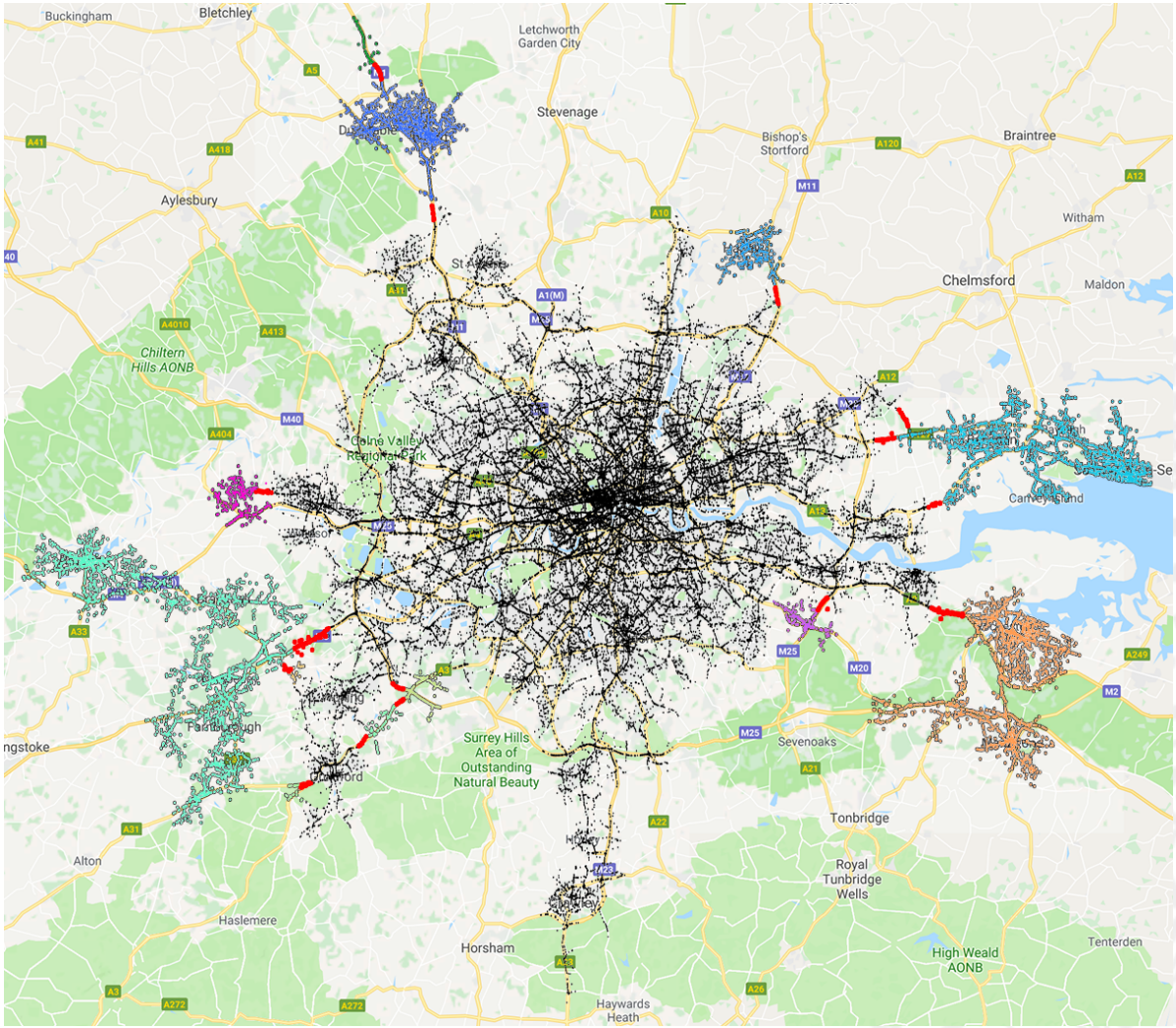


Londoncluster



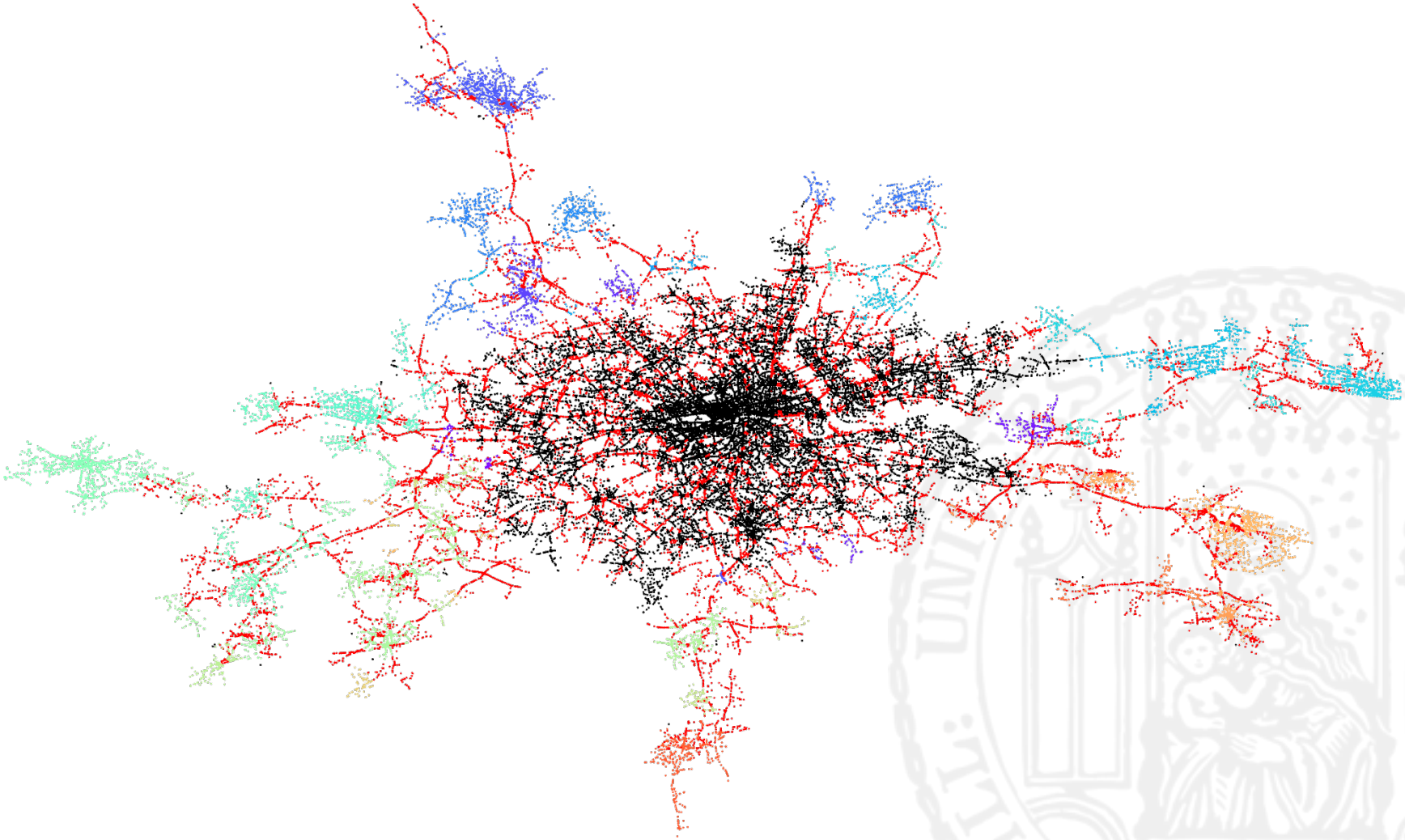
Londonclustering mit Chaindetection

Chaindection mit chainDim = 1 und allowedVariation = 0.2



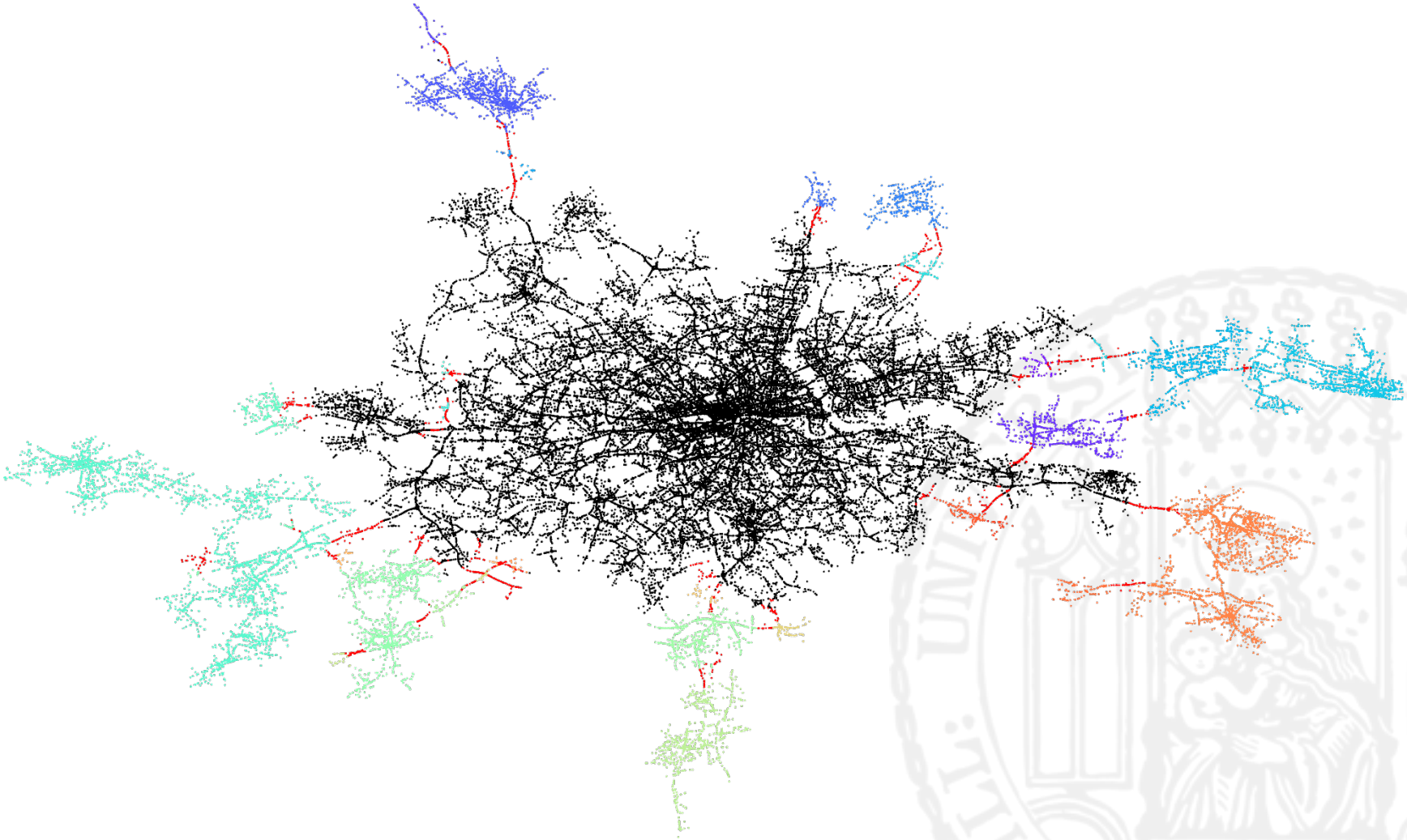
Londonclustering mit Chaindetection

Chaindection mit chainDim = 1 und allowedVariation = 0.6



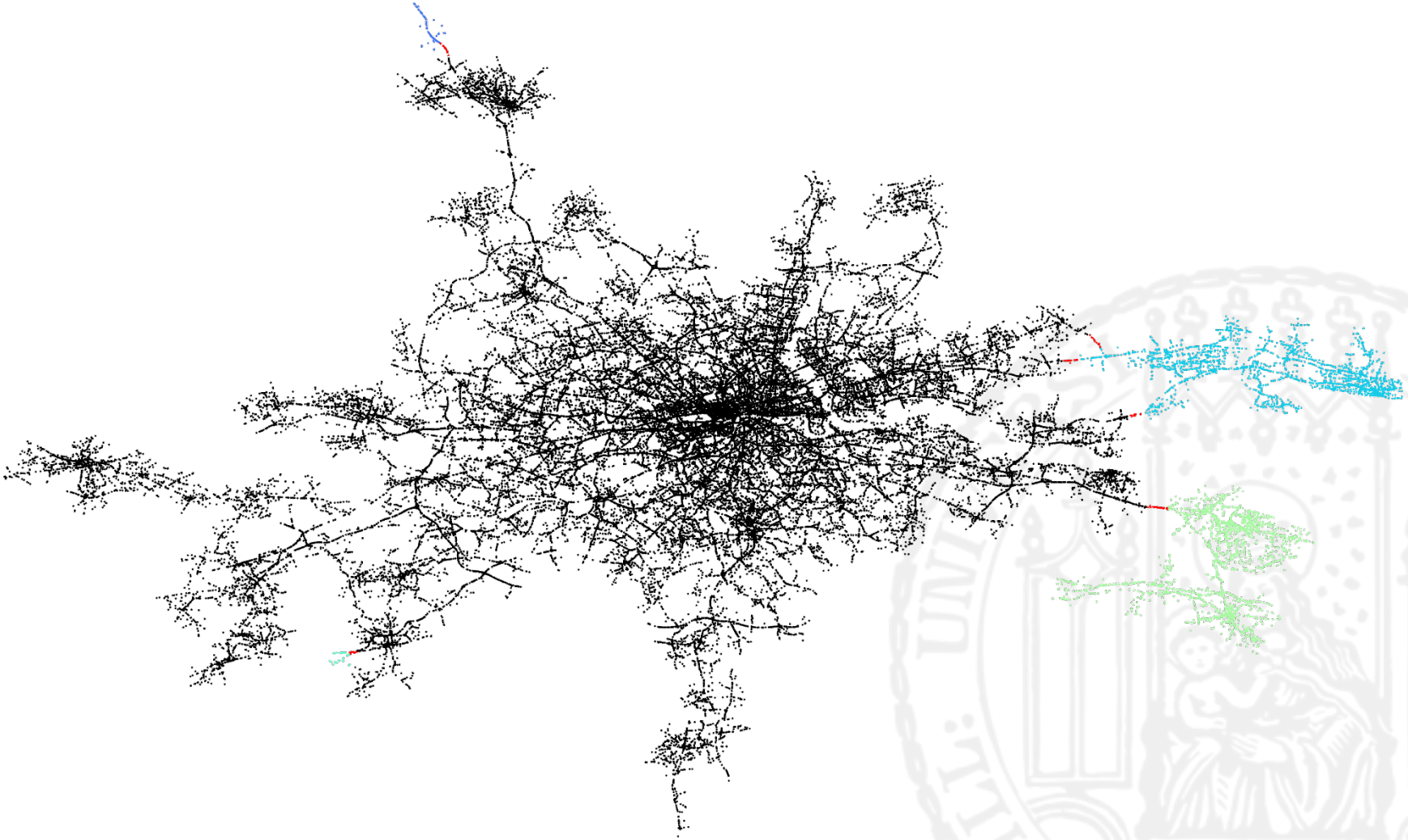
Londonclustering mit Chaindetection

Chaindection mit chainDim = 1 und allowedVariation = 0.3



Londonclustering mit Chaindetection

Chaindection mit chainDim = 1 und allowedVariation = 0.1



Vielen Dank für die Aufmerksamkeit

Fragen?