Community Detection in a Large Collaboration Network of Nanotechnology Researchers

Scott Cunningham

Jan Kwakkel

TU Delft

Why study academic collaboration?

- Collaboration represents knowledge transfer
- Knowledge transfer is key in innovation

- Collaboration is determined by various types of proximity/distance
  - Geographical
  - Sectorial
  - Social
  - Epistemic
- Previous work has highlighted the importance of geographic proximity
  - Regional versus national versus supranational innovation systems

Community detection in networks

- Many networks divide naturally into communities
- Communities are sets of nodes between which many edges exist, but that have few edges to nodes outside the set
- Community detection
  - Graph partitioning
  - Hierarchical clustering
  - Modularity

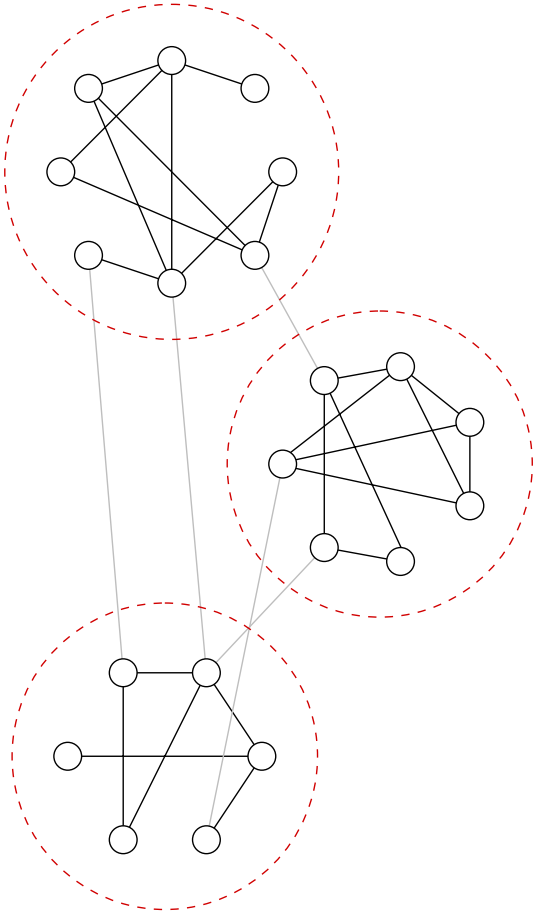


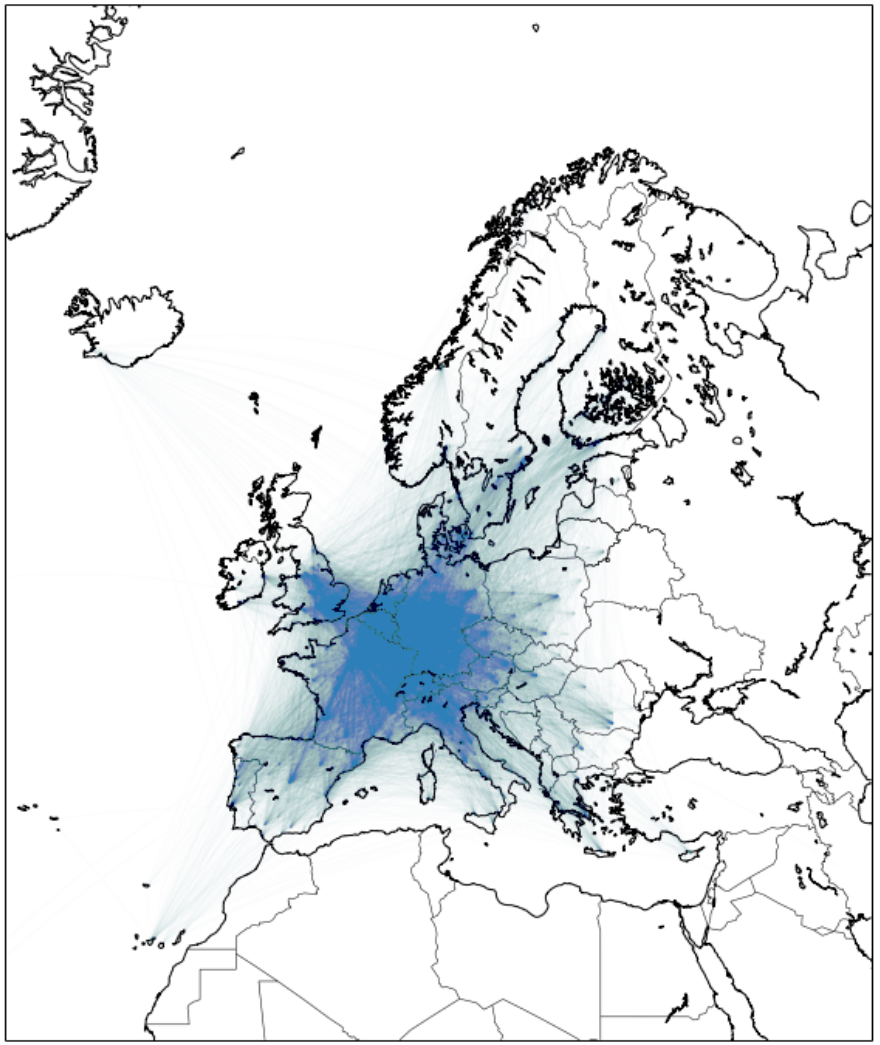
Figure taken from Newman & Girvan (2003)

Modularity

- Fraction of intra-community edges versus fraction of inter-community edges
  - Score of 0 means that the network is essentially random and has no community structure
  - scores close to 1 imply very strong community structure
  - scores above 0.3 are robust evidence for community structure
- Algorithms based on modularity produce hierarchical results: communities of communities
- We use the algorithm proposed by Blondel et al (2008)
  - Bottom up merging based on local optima of modularity score
  - Applicable to weighted directed graphs

European collaboration in nanoscience

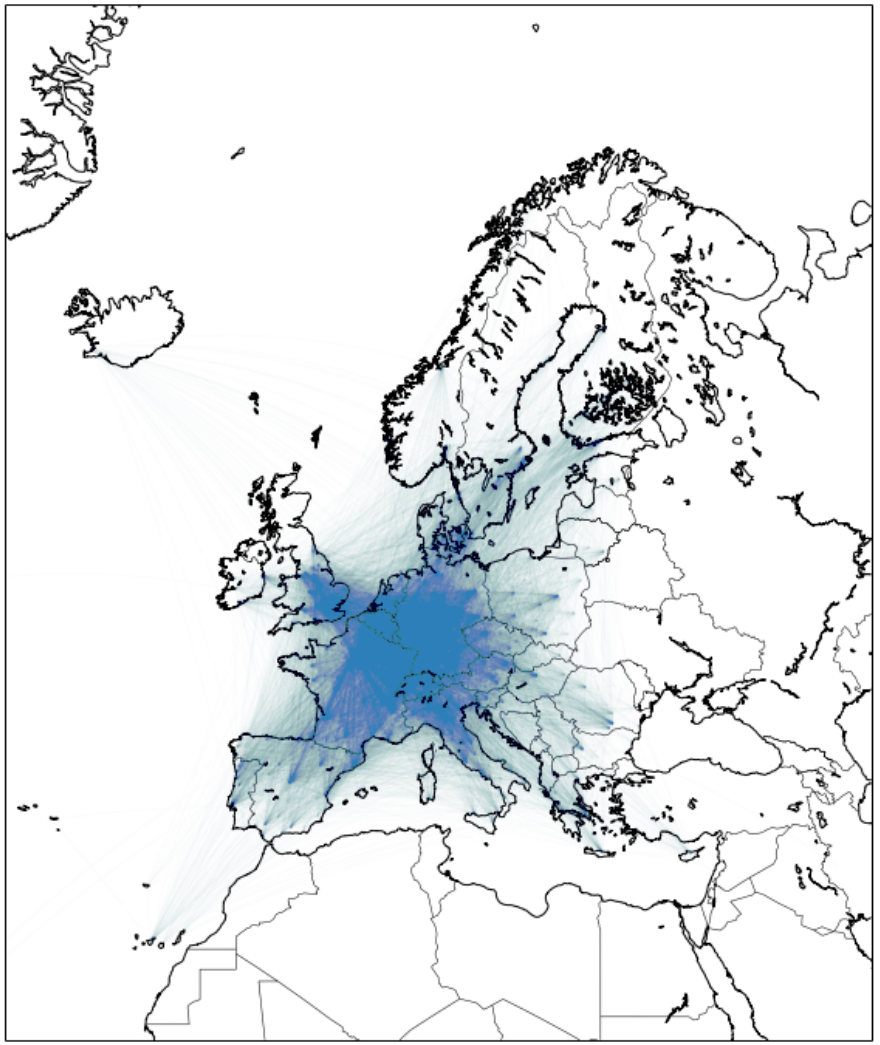
- Porter nano query
- Limited to Europe
- Collaboration aggregated at city level
  - Multiple authored papers are fractionated
  - Cleaning of addresses
- Google used for geolocating cities
- Supervised merging of nodes with same latitude and longitude

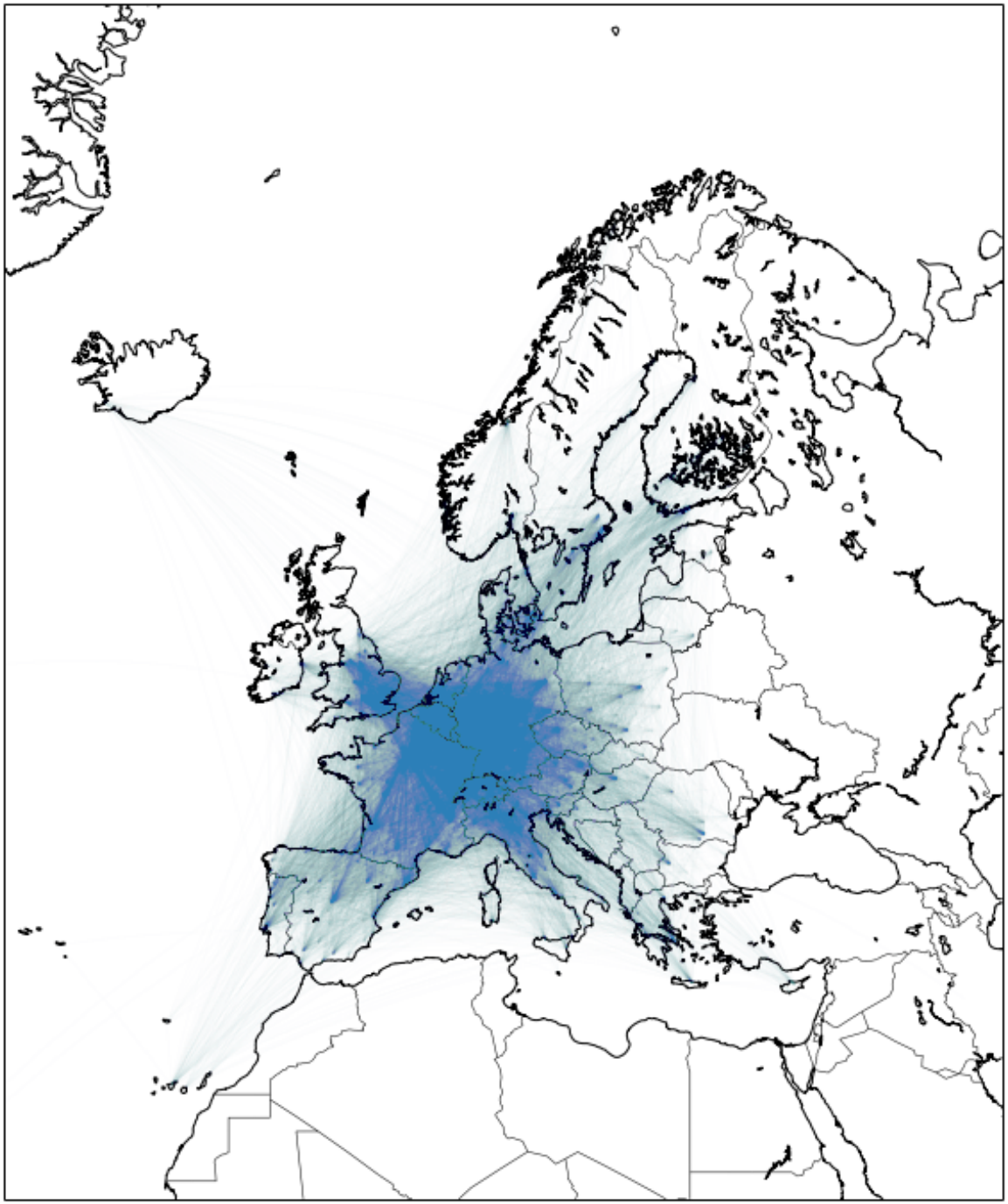


European collaboration in nanoscience

- 2716 cities, ranked ordered by volume of publication
- 33 cities with missing location information
  - Wrongly parsed addresses
  - Misspelled names
- 35623 edges
- 87 edges associated with 33 cities missing location information

- IPython notebook and data are available for full reproducibility



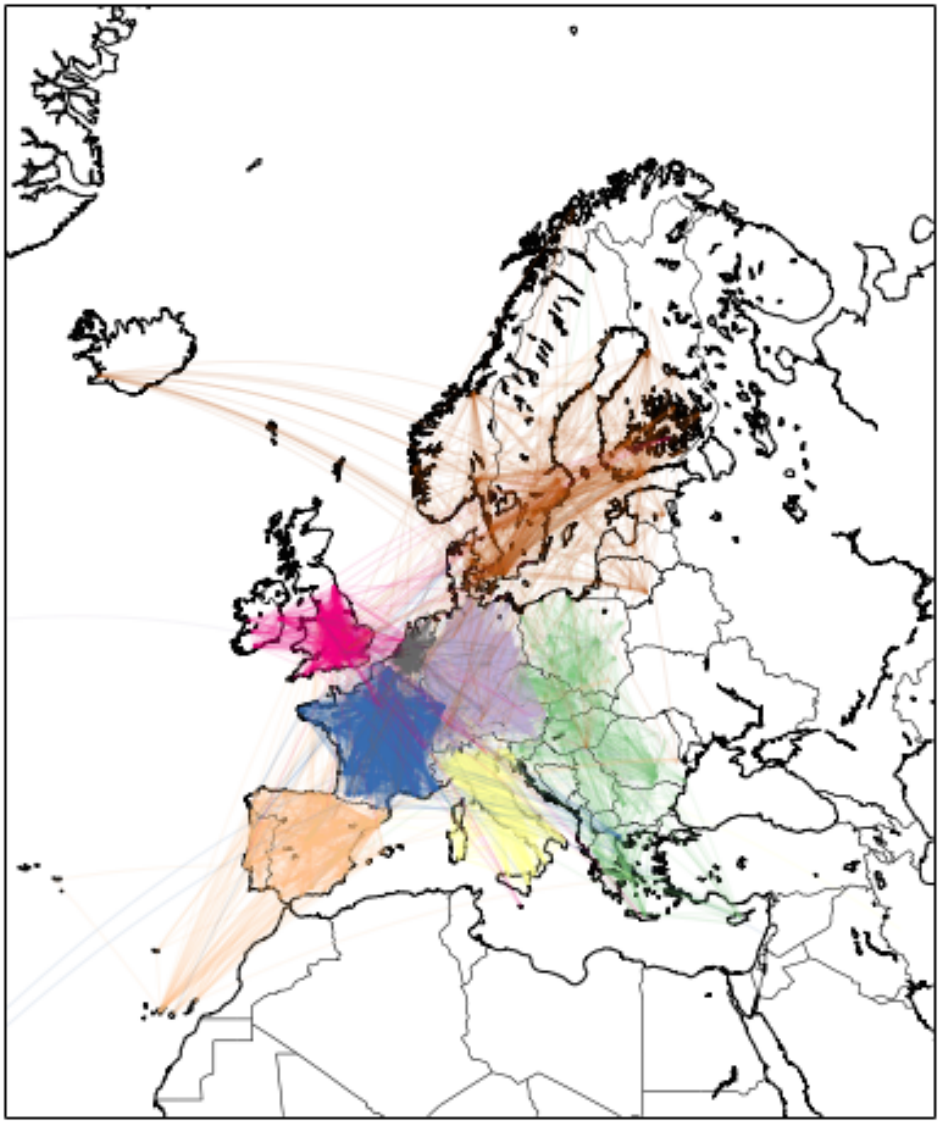


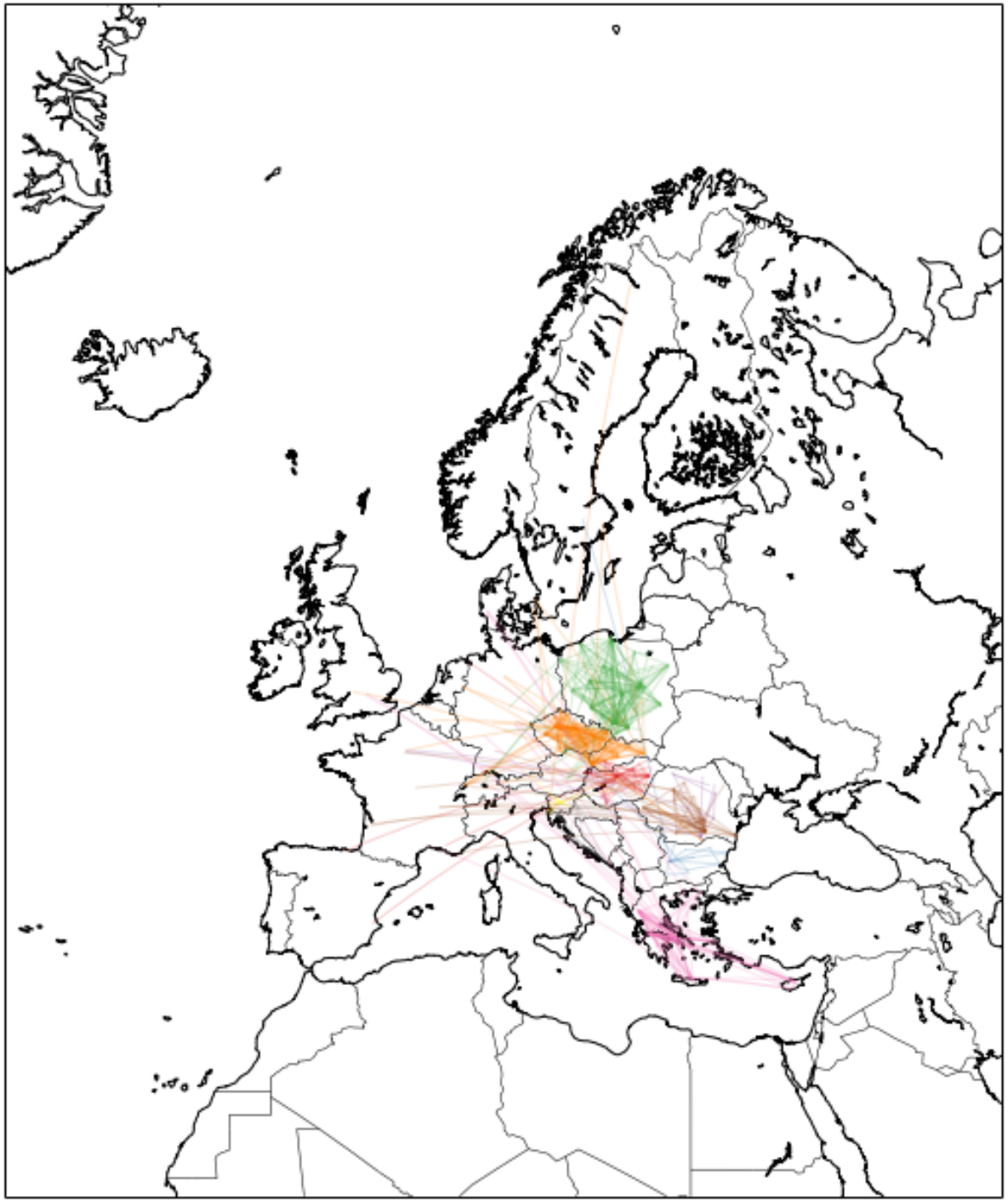
Best partitioning

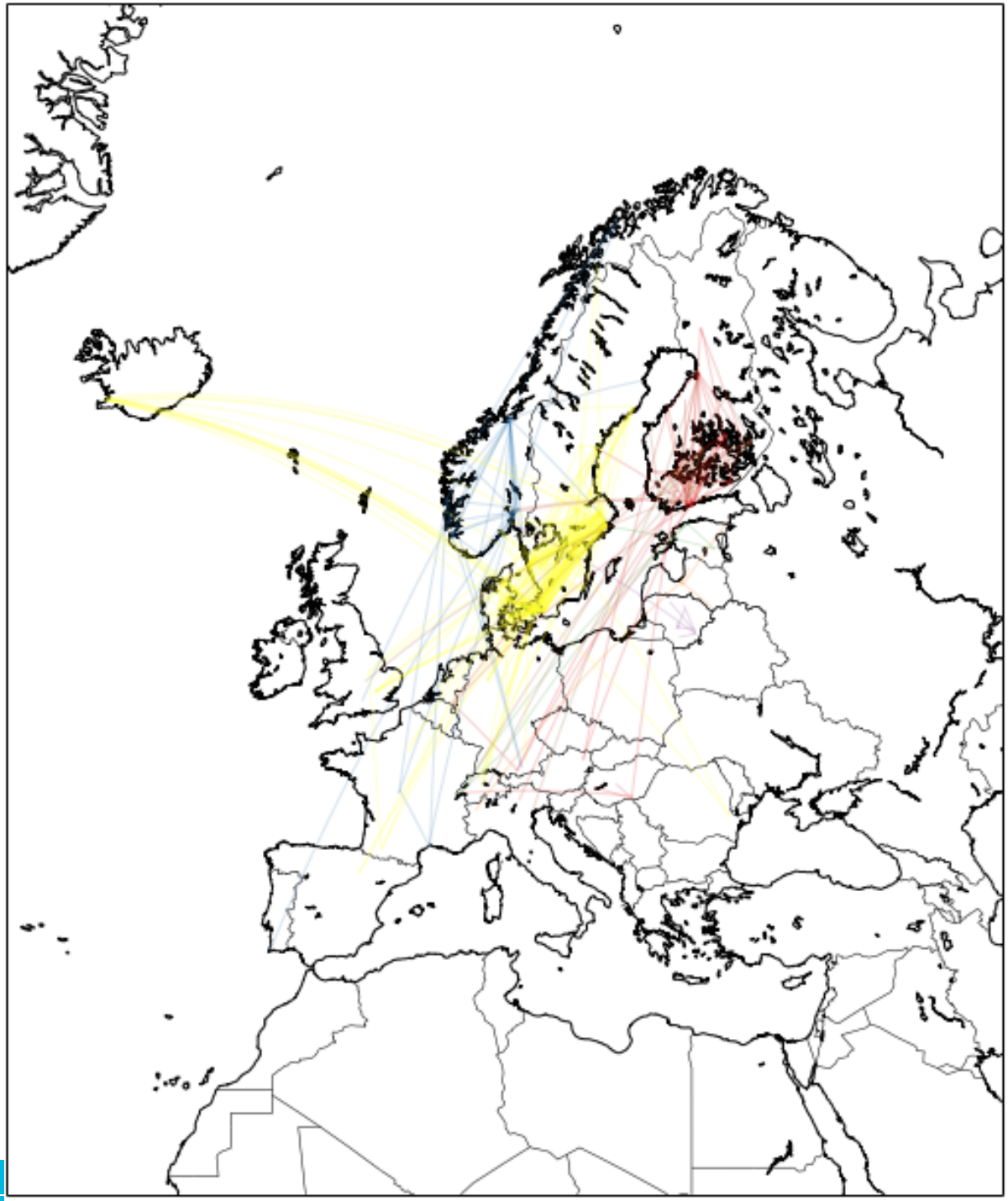
- 3 levels of partitioning
- Modularity ranges from 0.31 to 0.36

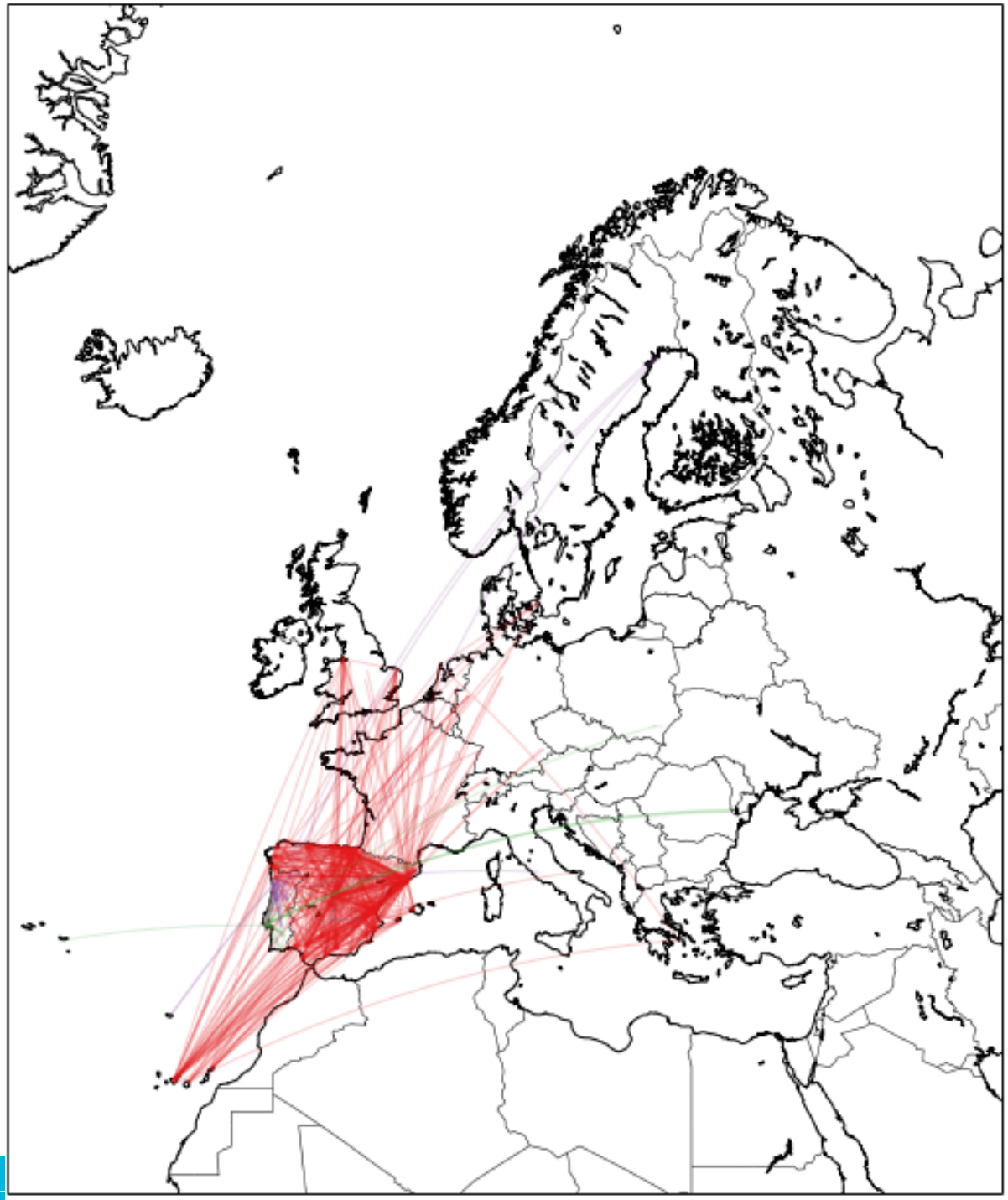
- 8 major partitions in best partitioning

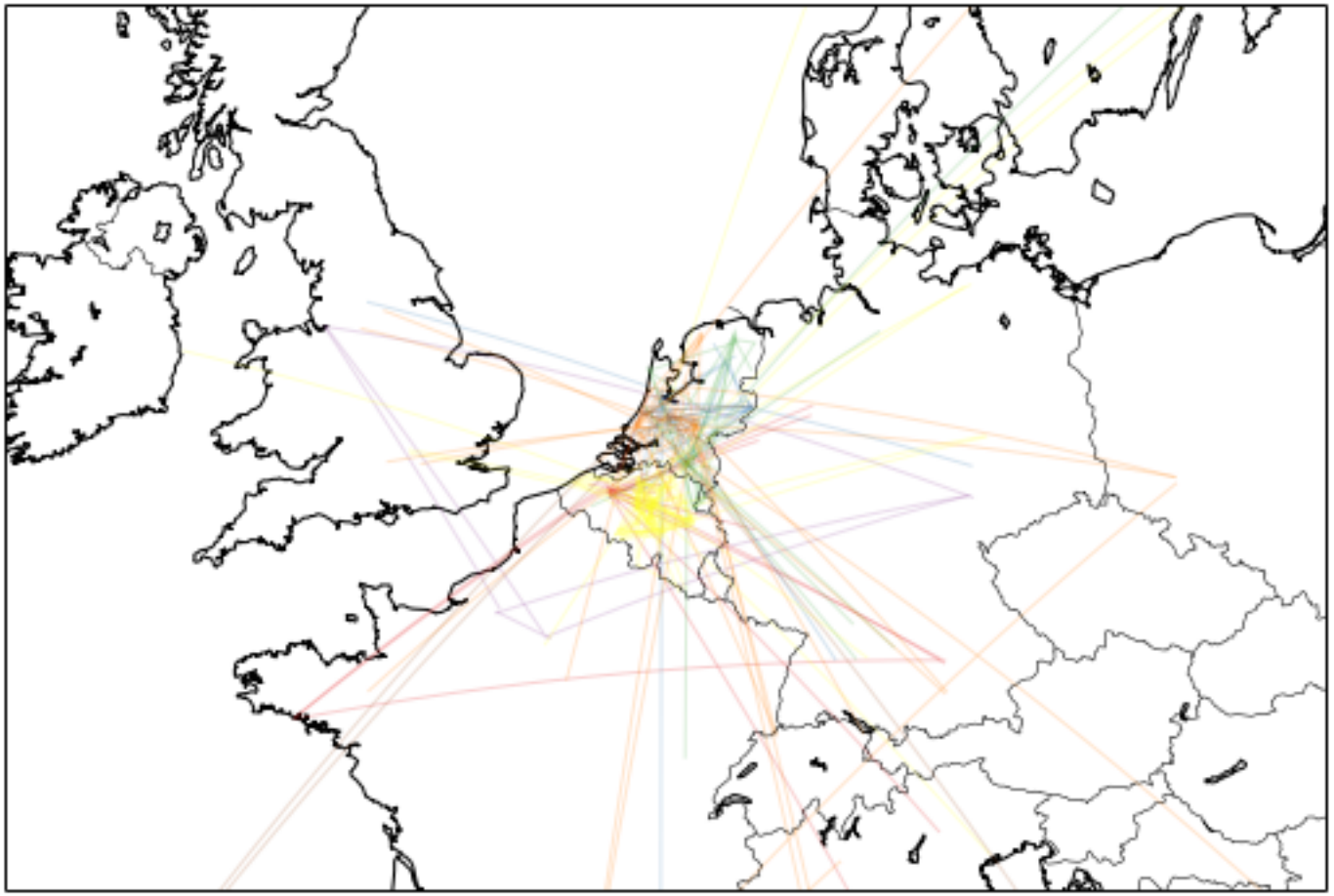
- Clear geographic structure to collaboration network











Conclusion

- Collaboration communities appear to be structures by nations
- Some evidence for communities at regional level

- Why?
  - Language,
  - national funding, systemic proximity
  - What about e.g. EU funding?

- Is the fact that nations dominate the community structure due to the aggregation by city?
  - i.e. are we not aggregating epistemic structure away