


Retrieving the continent labels from the air routes structure

Overview

- Context
- Main steps of the study
- Graph Embedding
- K-means clustering
- Labelling results
- Isolated labelled airports
- Conclusion

Context

- Dataset :
 - List existence of routes between airports  **Adjacency Matrix**
 - Merge to get geographical location  Compute distances and create **Weighted Matrix**
 - Obtain and add Continent labels to airports
- Problematic :
 - Retrieve continents from links between airports
 - With and without distances

Main steps of the study

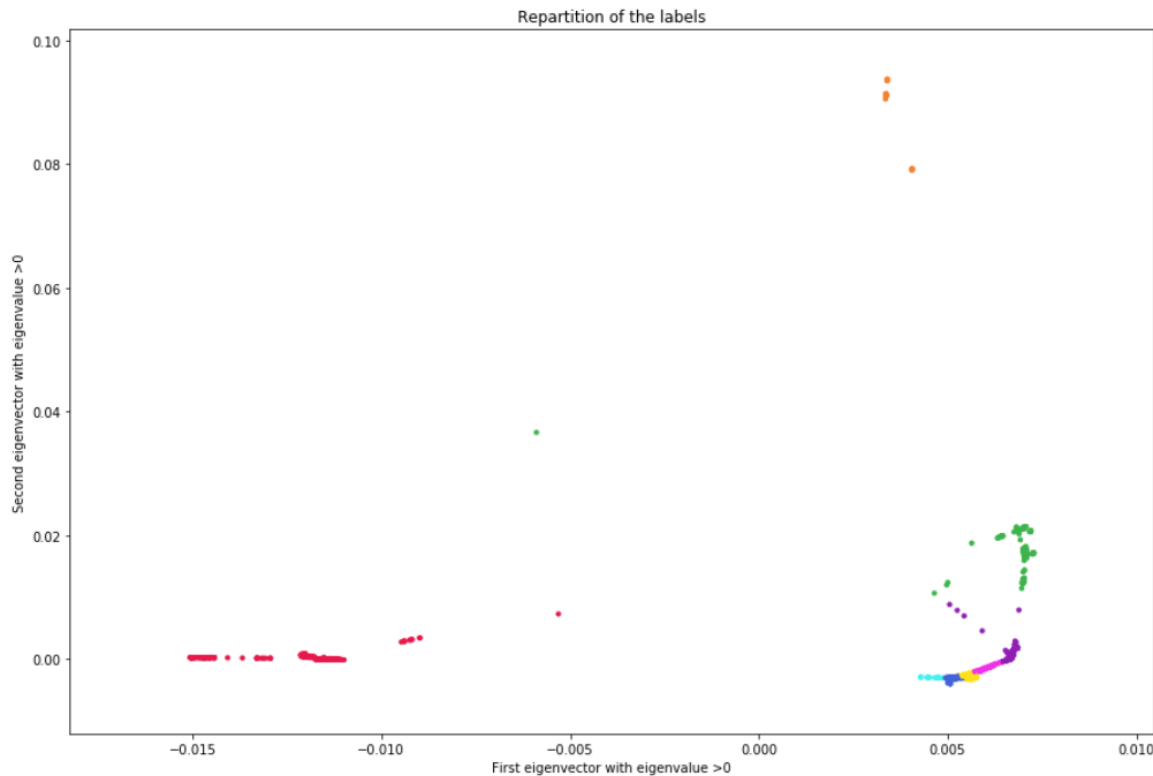
- Embed the graph in some N-dimensional space
- Compute Clustering algorithm
- Assess quality of clustering:
 - See clusters on World Map
 - Use Metrics

Embedding

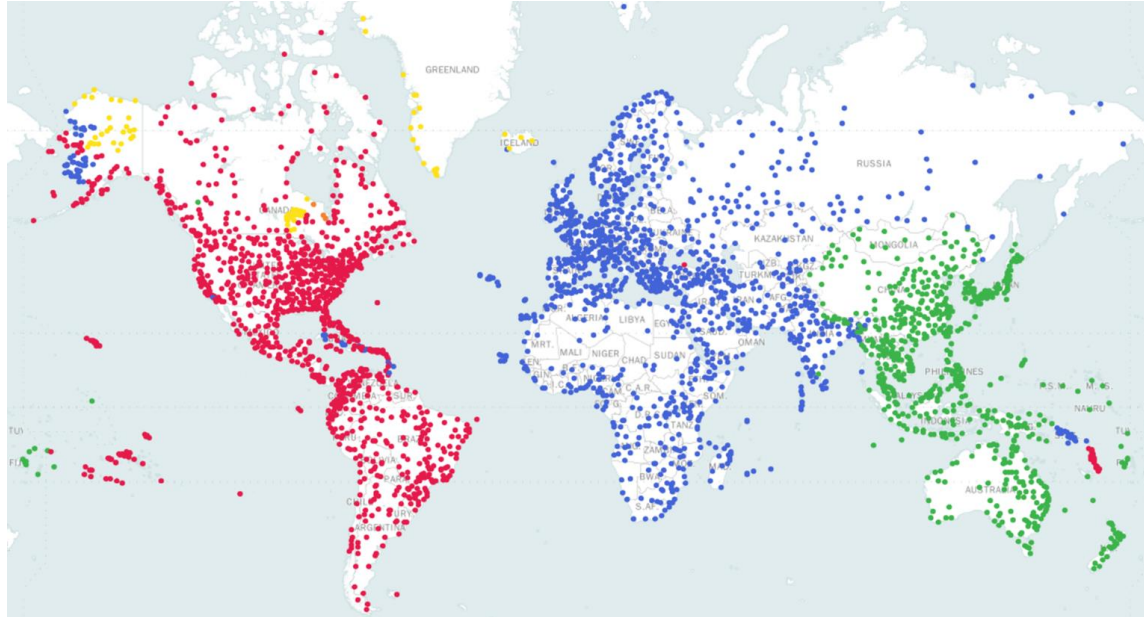
- Weighted matrix W : Heat Kernel on distances
- Laplacian Eigenmap
 - Compute Laplacian and get eigenvectors of the generalized problem
 - $L = D - W$
 - $Lu = \lambda Du$
 - Sort eigenvalues and get n first eigenvectors

K-means clustering

- Clustering made in 2 dimensions
- Visualisation in 2 dimensions
- Worked on initialization to improve clustering



Labelling result - without weights

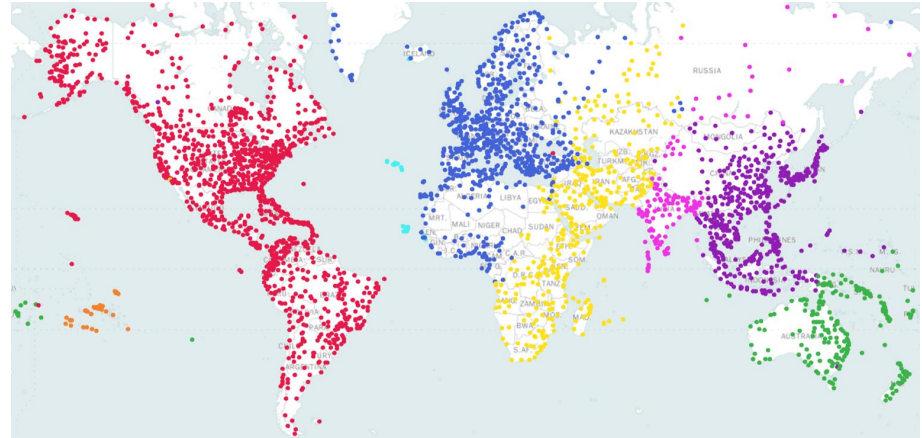


Labelling with initial nodes, no weights

Labelling result - airport continents

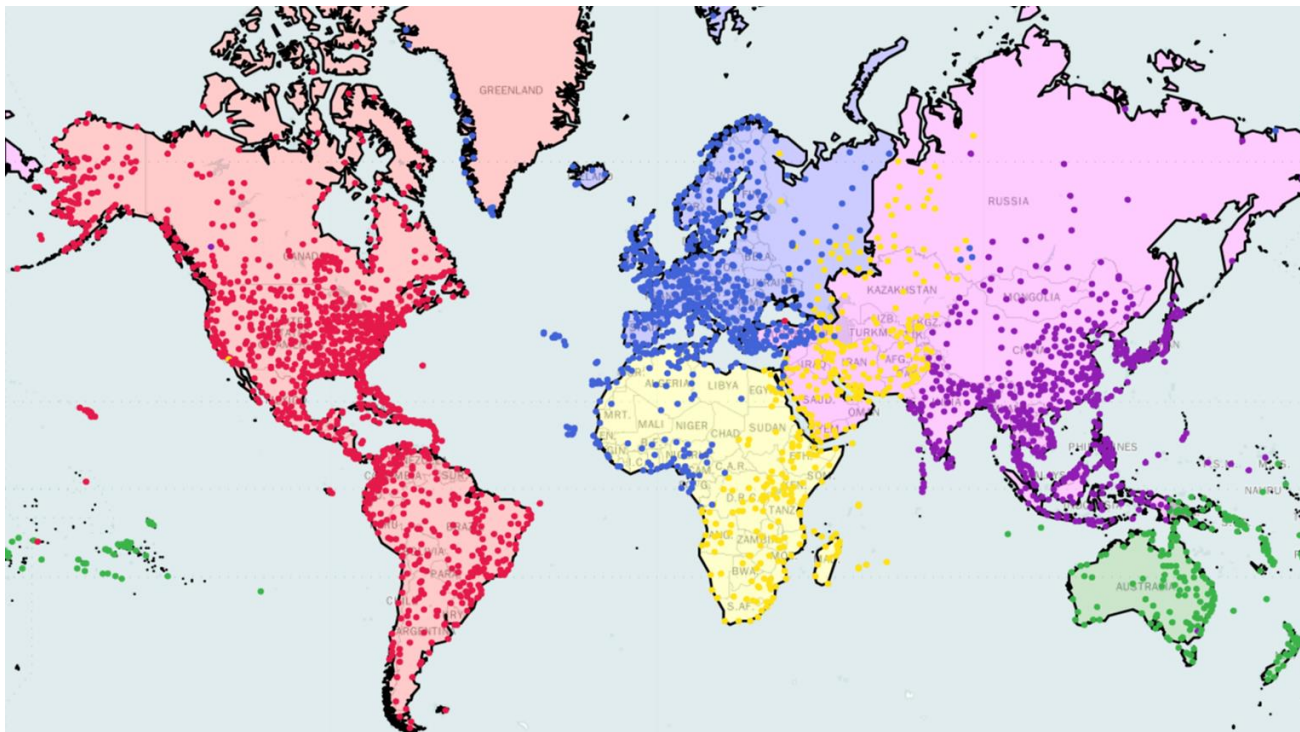


Ground truth labels



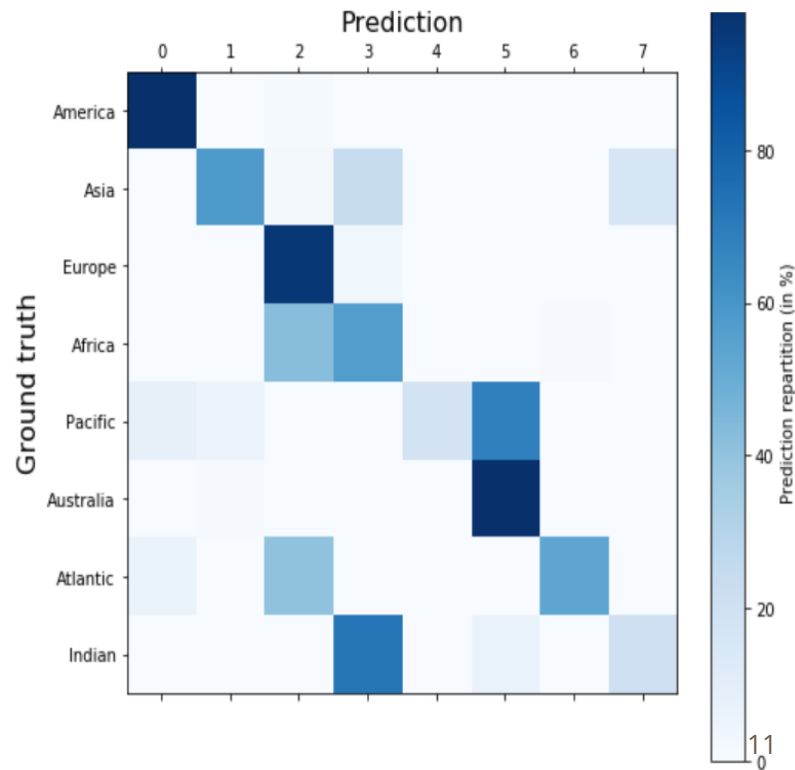
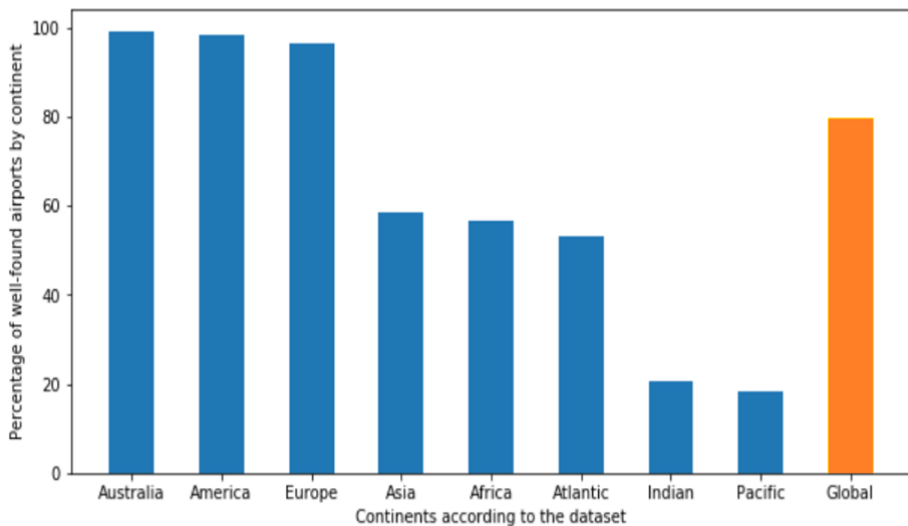
Labelling with two eigenvectors and initial nodes

Labelling result - true continents



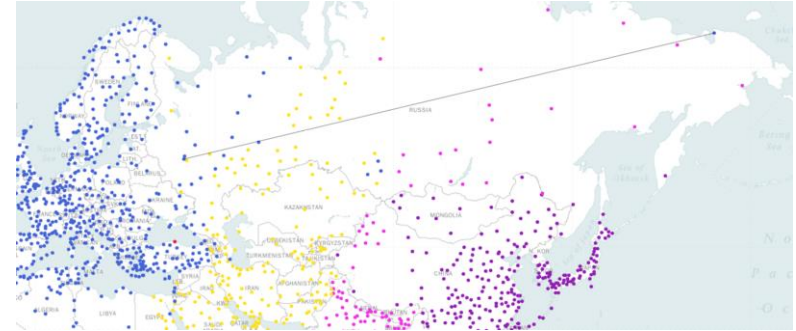
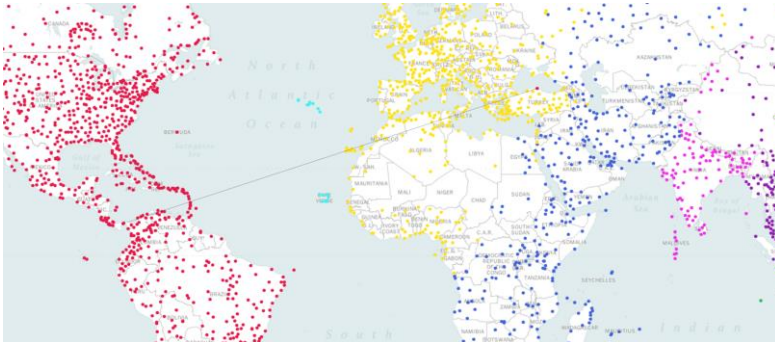
Labelled map over the continent map, 87% of the airports well labelled

Evaluation of the clustering



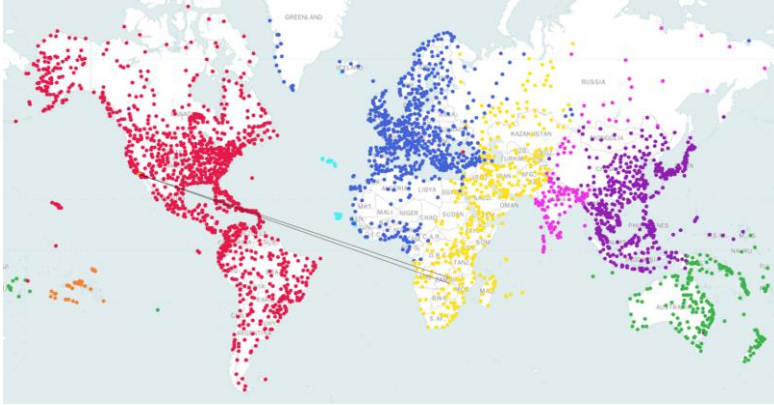
Isolated labelled airports

Sinop airport, Turkey



Pevek airport, Russia

Isolated labelled airports



Los Alamitos Army Air Field, United States

Mackenzie airport, United States



Conclusion

- 80% of the airports well labelled
- Up to 87% if we consider the 6 true continents
- Improvement of our geographic skills

Any question ?