RE-BALANCING FLIGHT ROUTES INEQUALITIES

Group 42,

Vincent Coriou, Pierre Fouché,
Robin Leurent, Alexandre Poussard

A NETWORK TOUR OF DATA SCIENCE

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Outline:

I) Data Analysis

II) Dataset Augmentation

III) Re-Balanced Flight Distribution

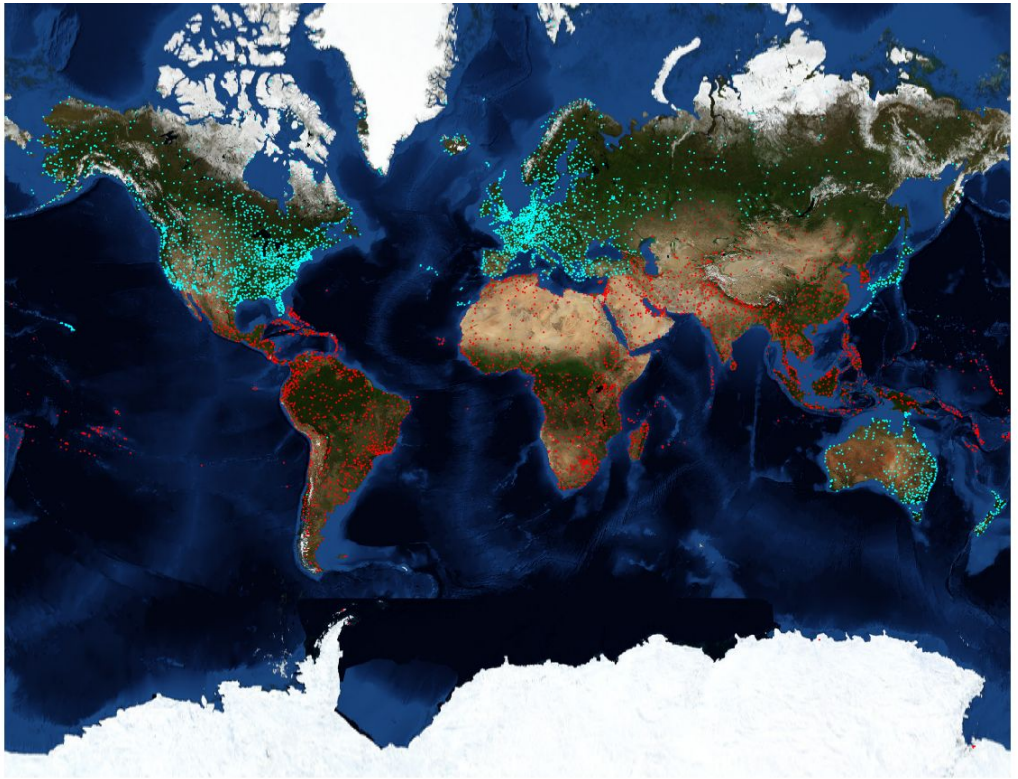
IV) Conclusion

Initial dataset - GIF plot

Data Analysis:

1) Airports:

- Unbalanced distribution
- North = 4103 Airports
- South = 3081 Airports
- Mercator Projection

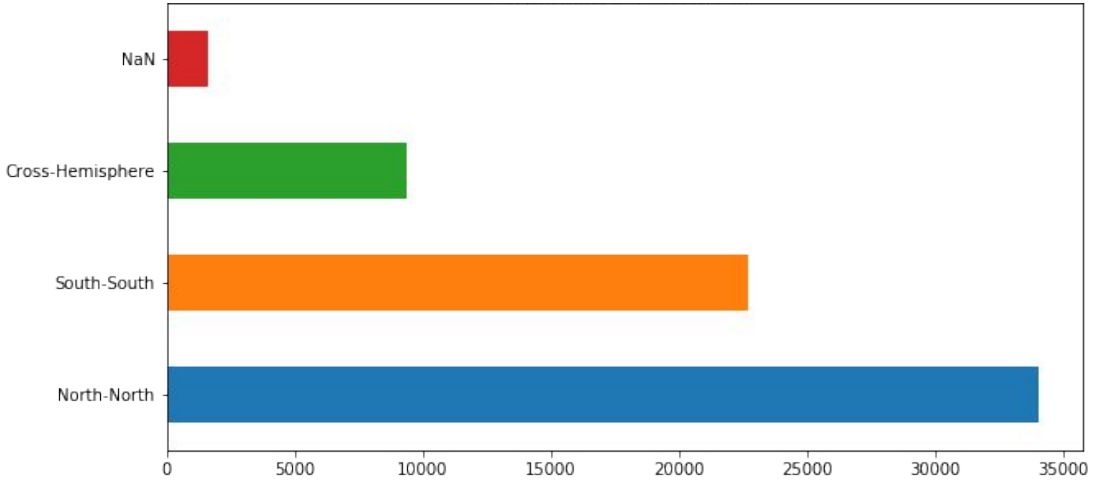


Airports colored by Regions (North=cyan, South=red)

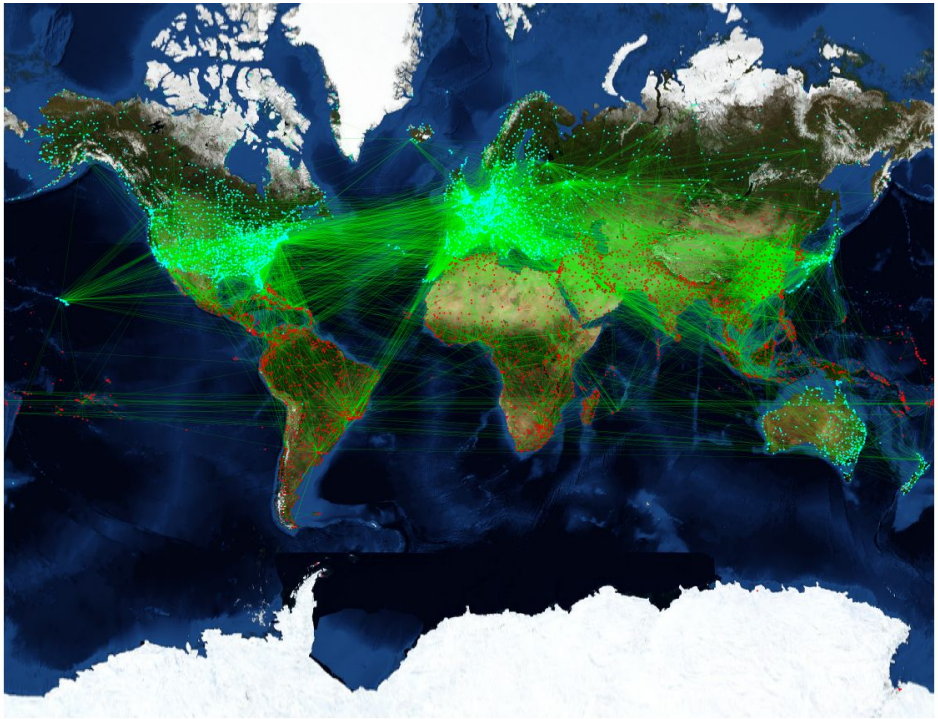
Data Analysis:

2) Flight Routes:

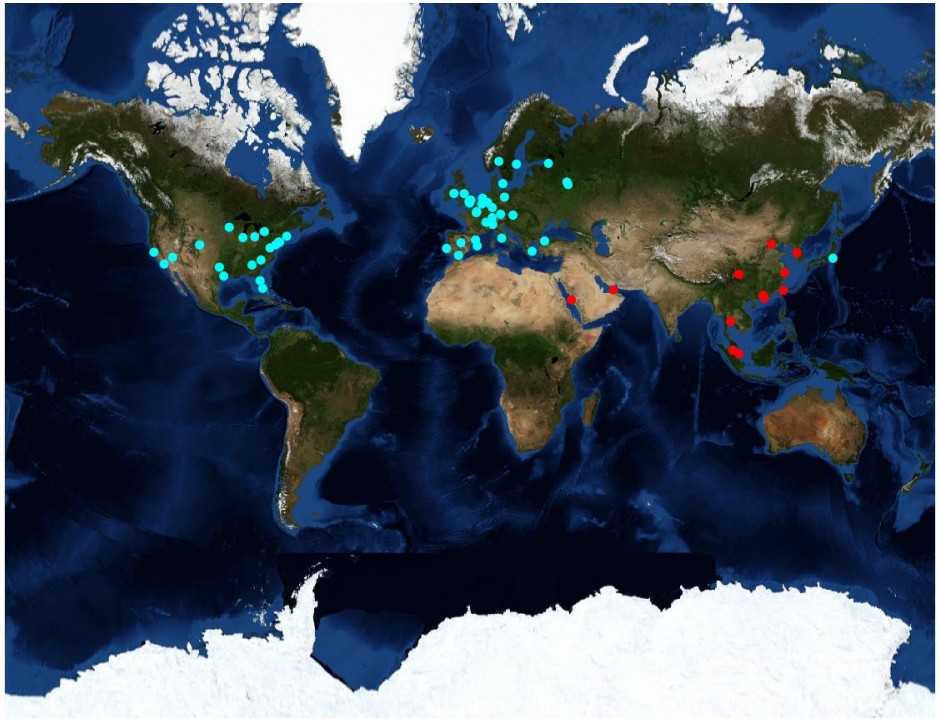
- Connected Airports:
  - 1613 in the North
  - 1566 in the South
- Unbalanced types of routes
- Hubs (degree > 100):
  - 51 the North
  - 12 in the South



Repartition of route types



World Flight Routes

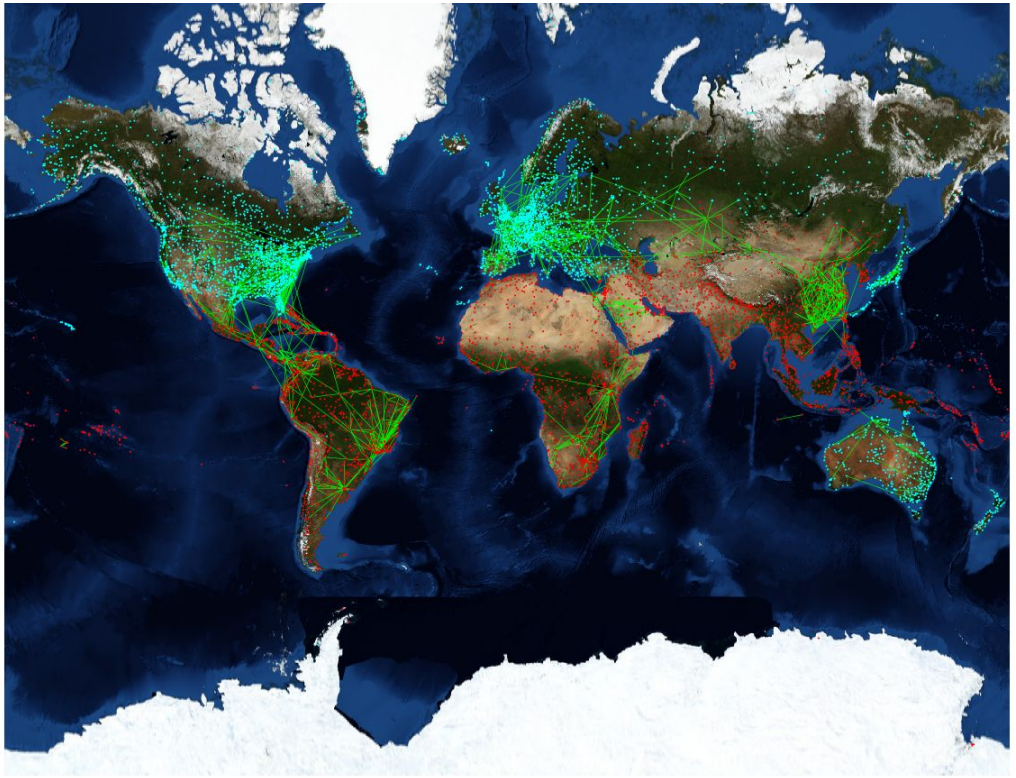


Hubs by Regions

Data Analysis:

3) Airplanes Manufacturers & Airline Companies:

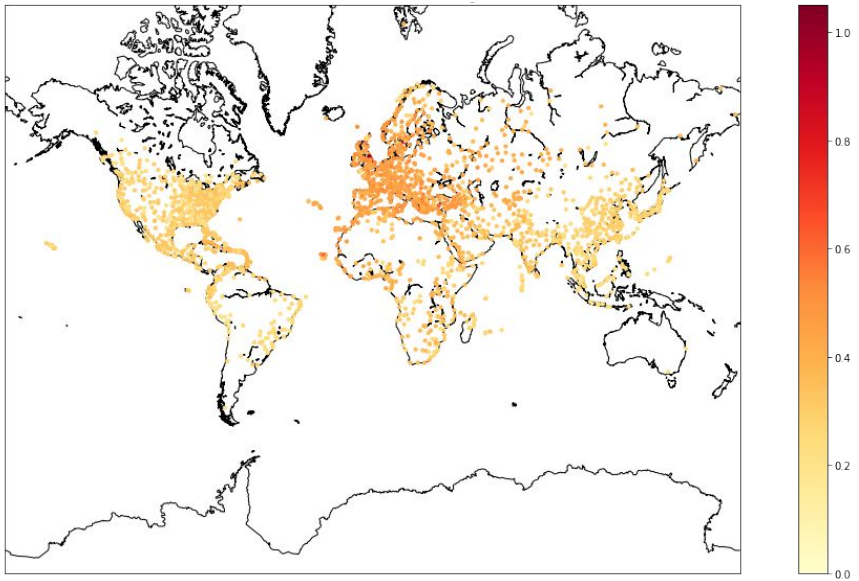
- Airbus > Boeing > Embraer
- Airbus + Boeing = all market
- Embraer = regional flights
- Top companies:
  - North = 24
  - South = 6



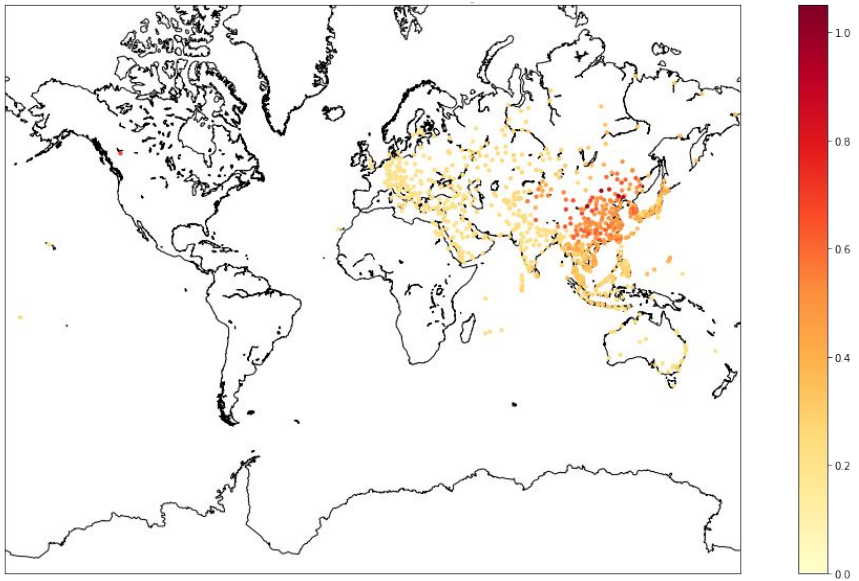
Route flights with Embraer Airplanes

Data Analysis:

4) Diffusion:



Heat Diffusion from Amsterdam

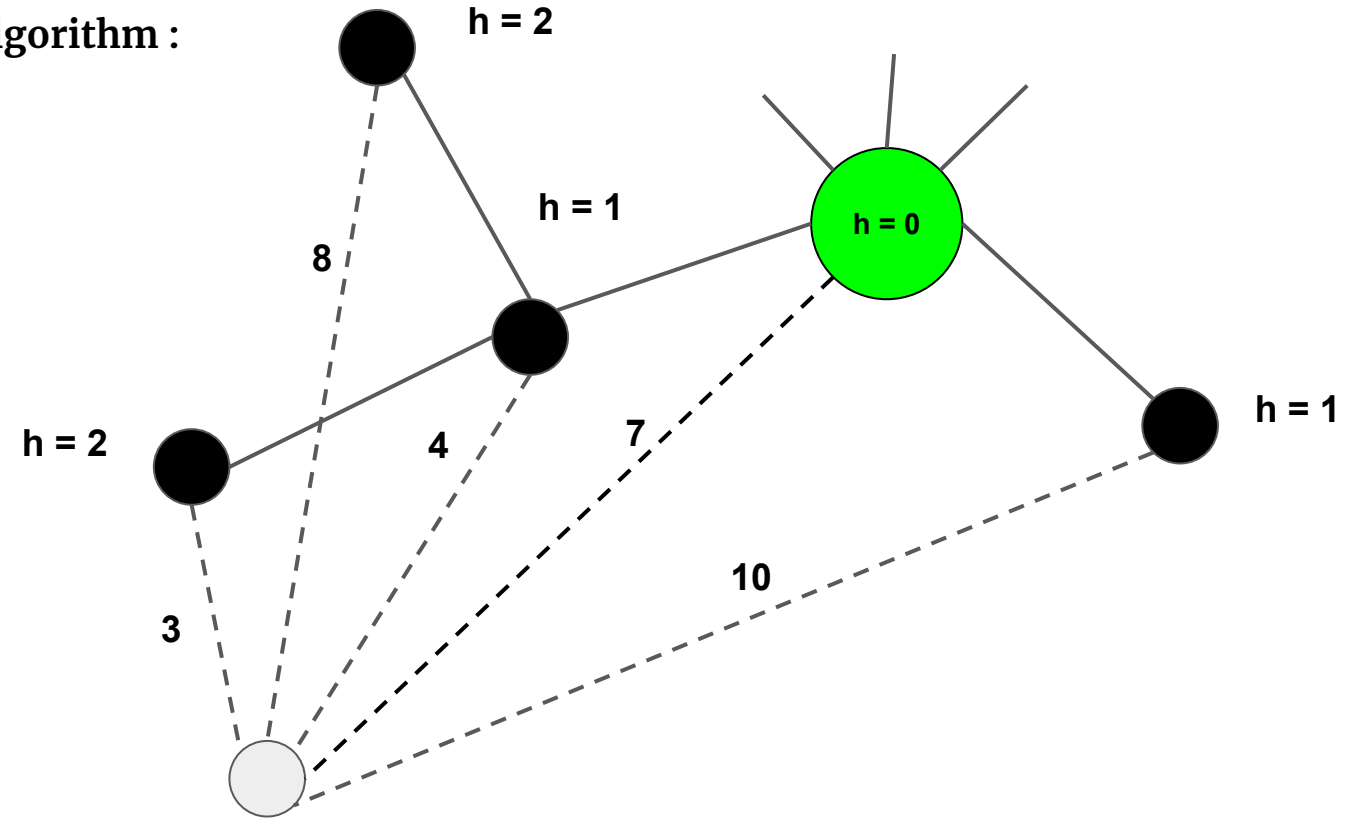


Heat Diffusion from Beijing

Dataset Augmentation:

- 7184 airports / 4005 disconnected
- Predict missing routes → re-balanced flight routes inequalities ?
- Minimize number of hops to the closest hubs.
- Minimize distance of the route.
- Customized preferential attachment algorithm :
  - Cost function: (delta hyperparameter) $L(j) = \delta d_{ij} + h_j$
  - Find the m nodes minimizing: $k_l = \underset{j \ in \ V \setminus \{k_p : 0<p<l\}}{\arg\min} L(j) \qquad l = \{1, ..., m\}$
  - Connect to these nodes

Distance attachment algorithm :

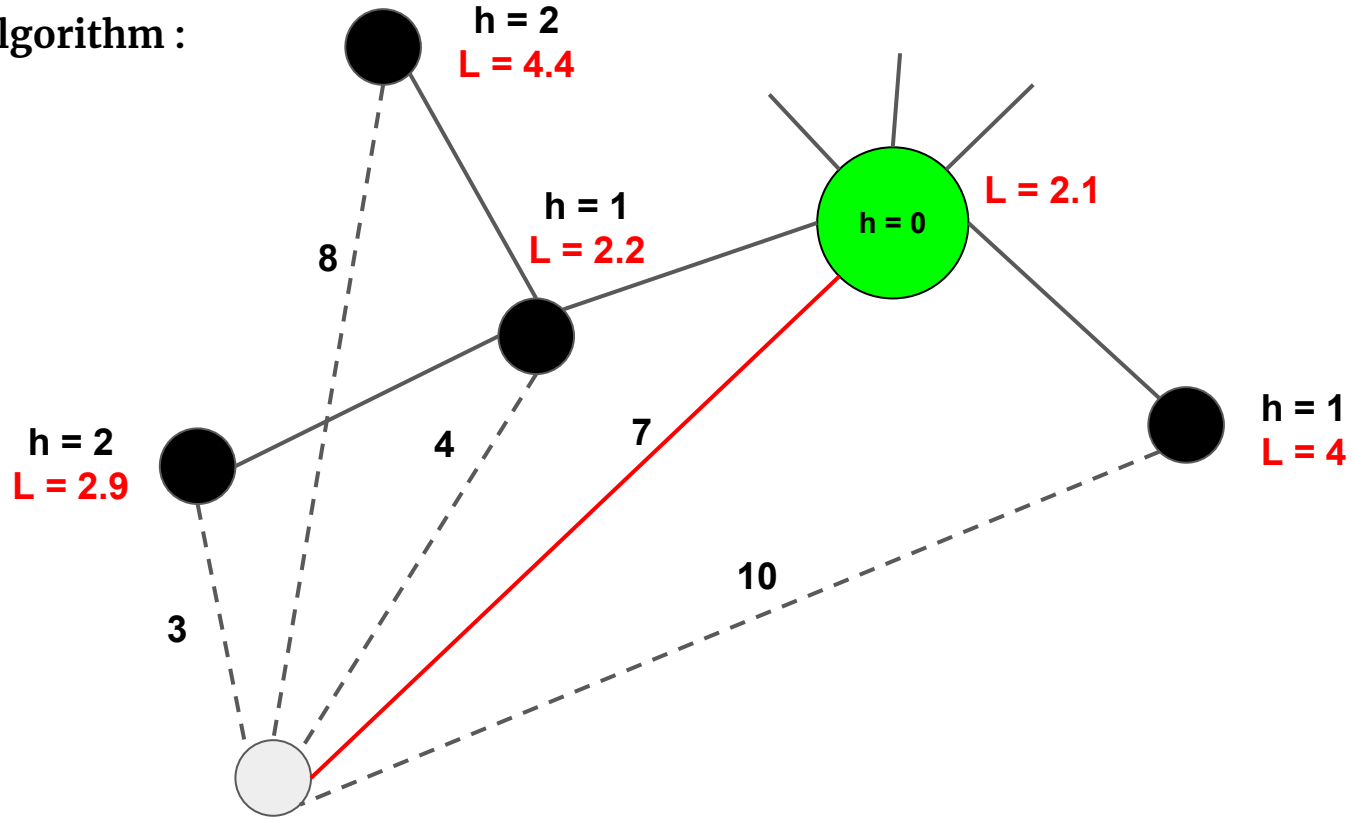


$$L(j) = \delta d_{ij} + h_j$$

h: # hops to the closest hub
δ: hyperparameter
L: value of the cost function

Distance attachment algorithm :

δ = 0.3

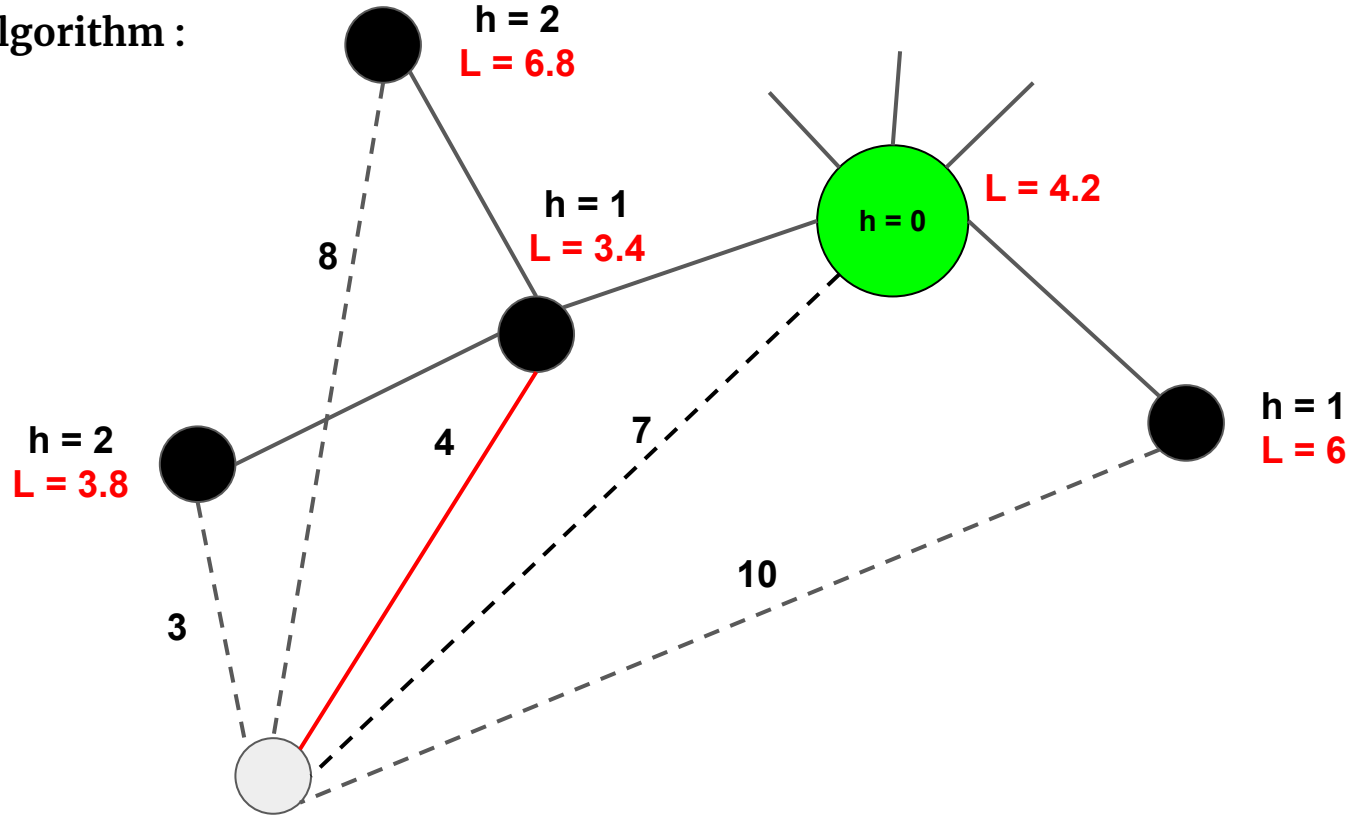


$$L(j) = \delta d_{ij} + h_j$$

h: # hops to the closest hub
δ: hyperparameter
L: value of the cost function

Distance attachment algorithm :

δ = 0.6

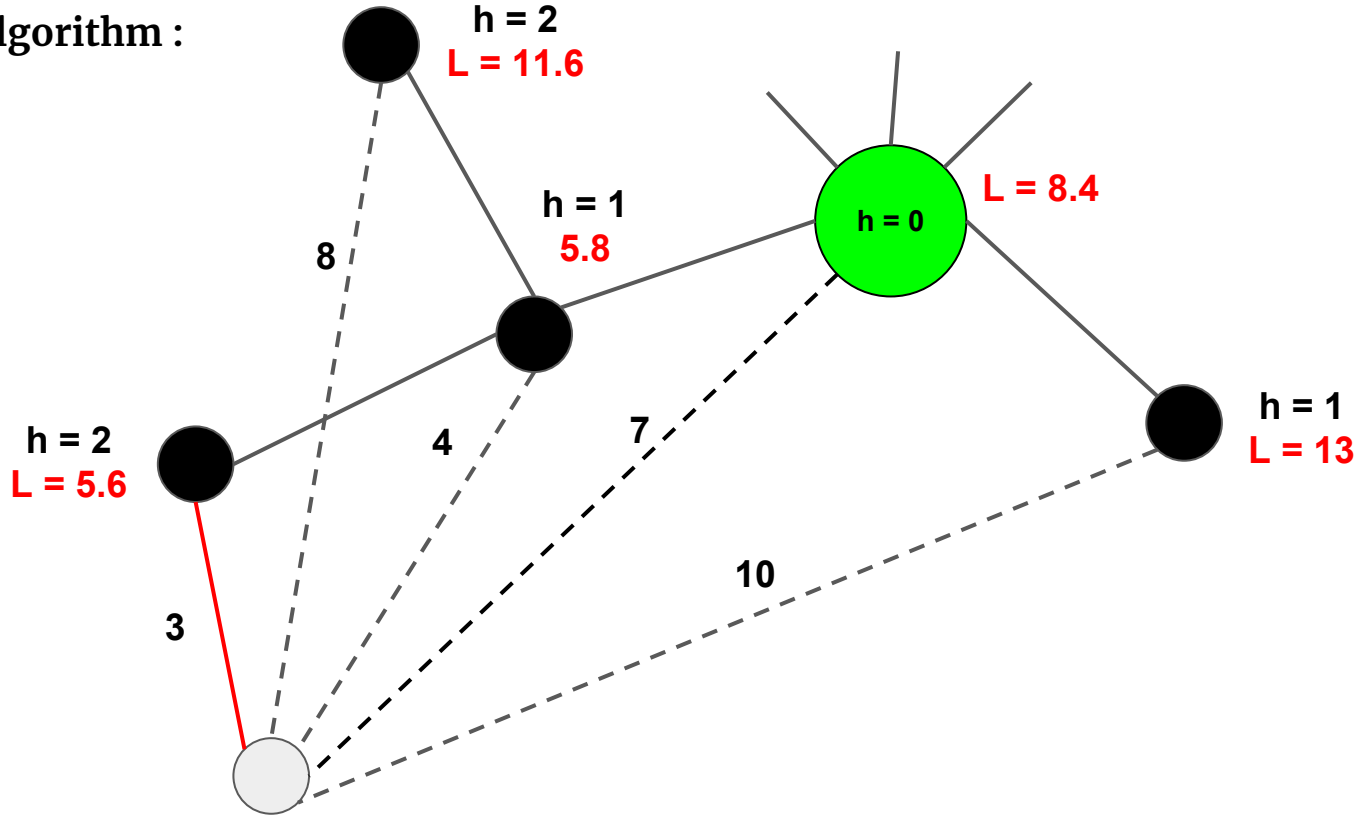


$$L(j) = \delta d_{ij} + h_j$$

h: # hops to the closest hub
δ: hyperparameter
L: value of the cost function

Distance attachment algorithm :

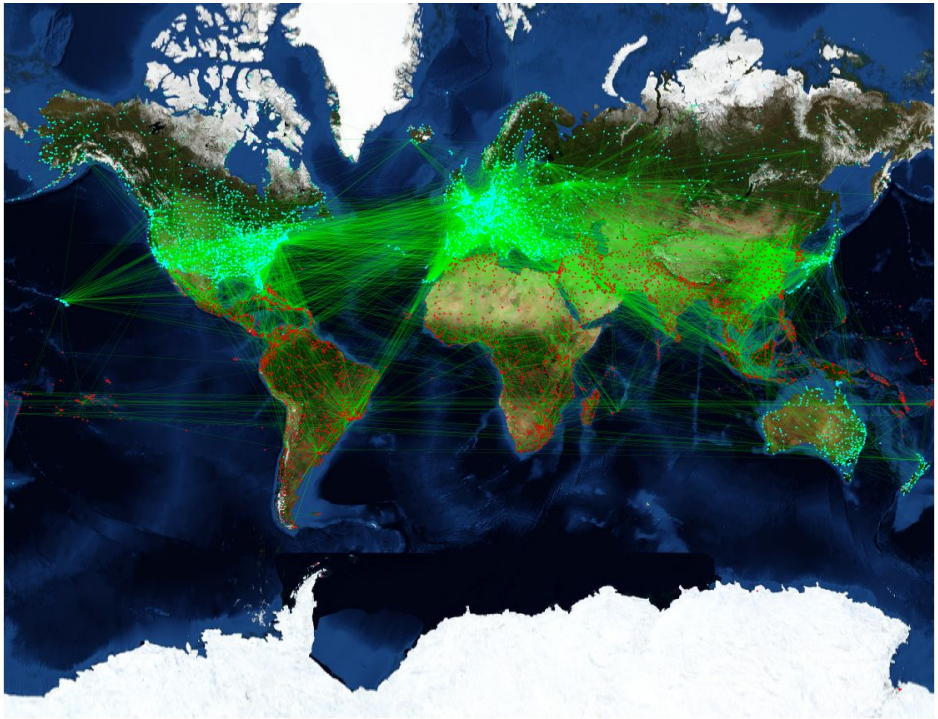
δ = 1.2

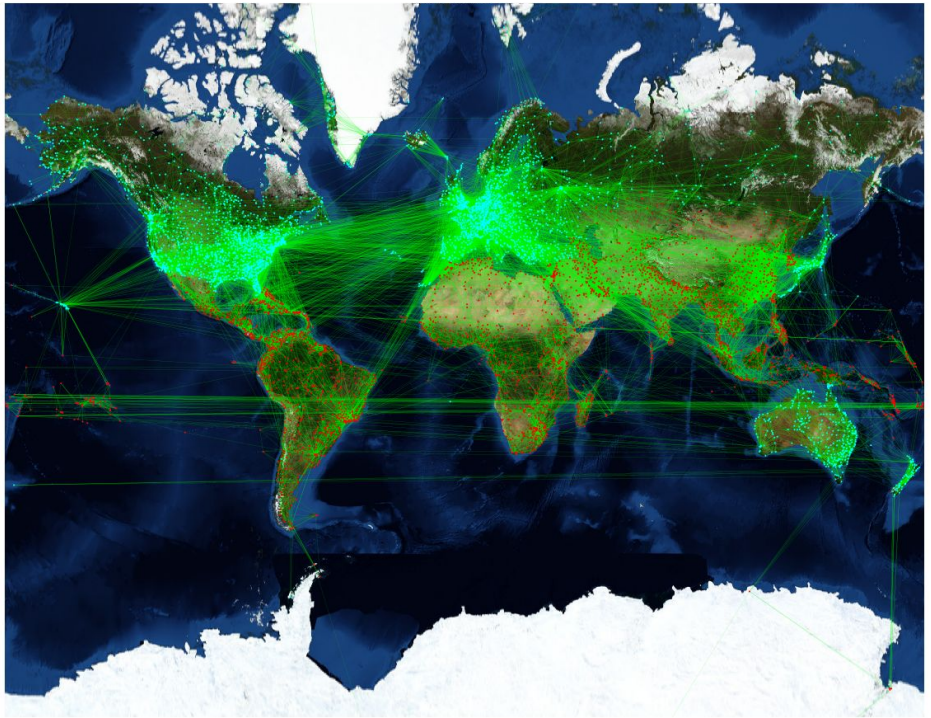


$$L(j) = \delta d_{ij} + h_j$$

h: # hops to the closest hub
δ: hyperparameter
L: value of the cost function

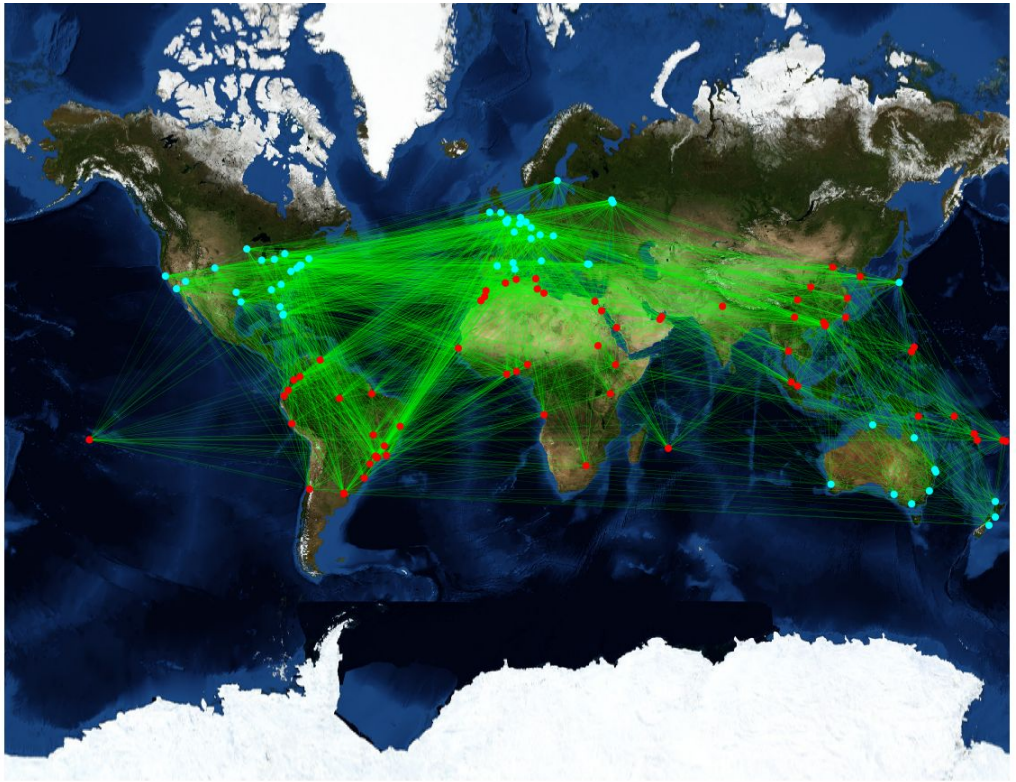
Distance attachment algorithm :



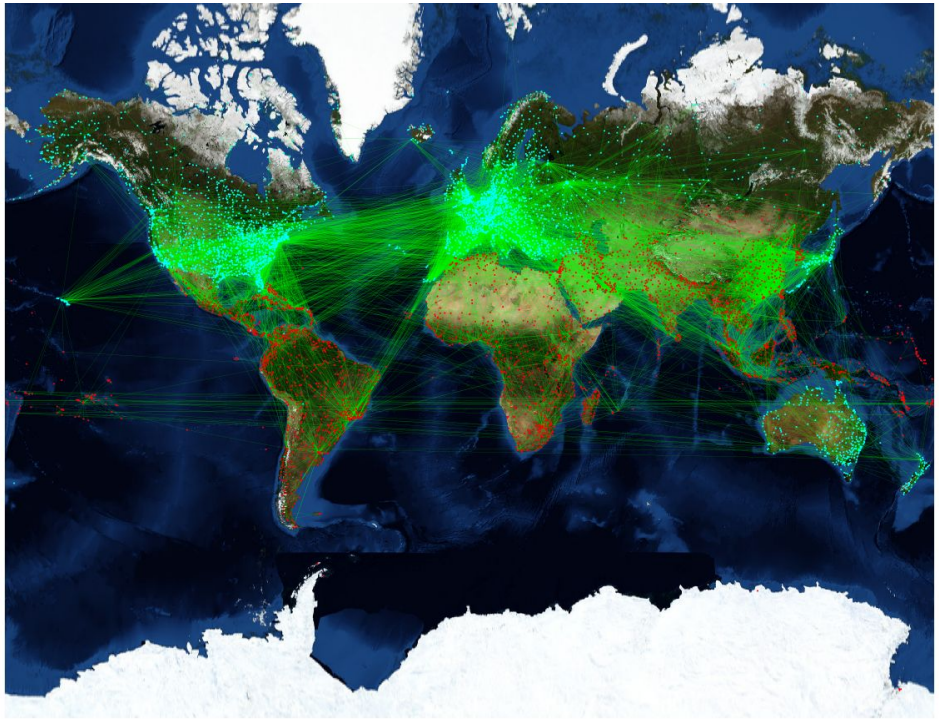


Re-Balanced Flight Distribution:

- Select the 20 biggest airports in each continent (= new hubs).
- Keep subgraph of hubs disconnect everything else
- Add hubs' routes if distance < 10'000 km
- Run previous algorithm.



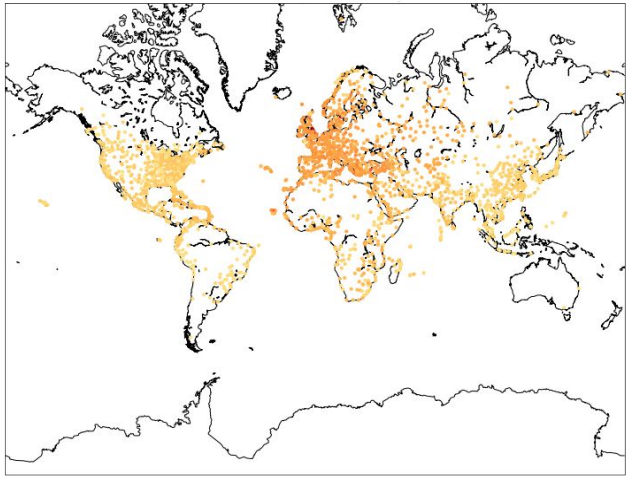
Best hubs for re-balance



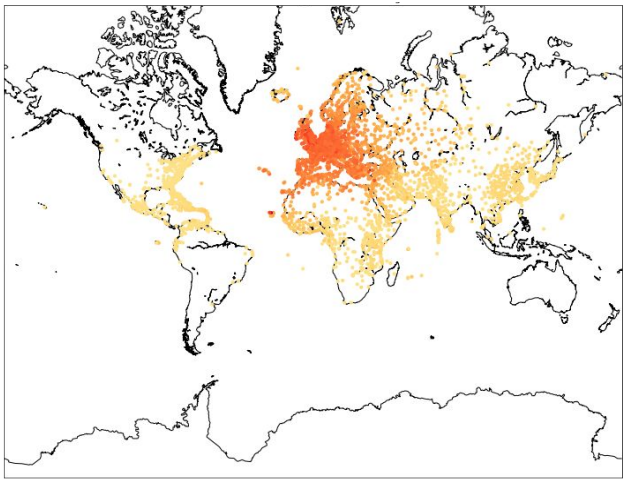
Original

Predicted network

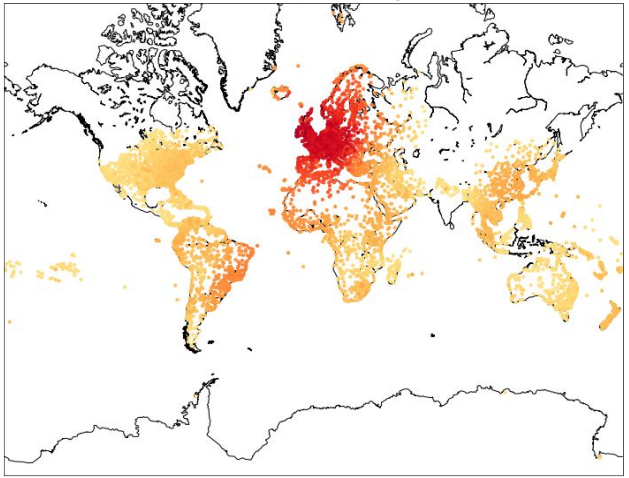
Re-Balanced Flight Distribution:



Original



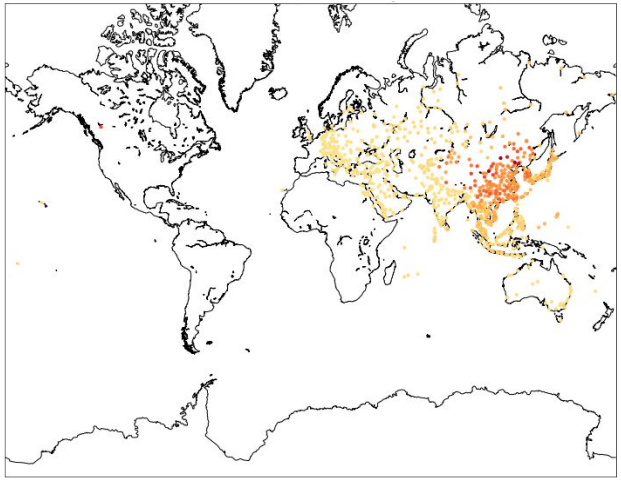
Dataset augmentation



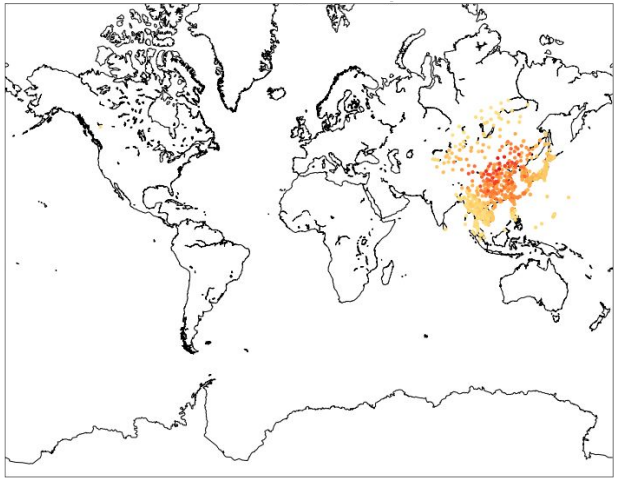
Re-Balances Flight Distribution

Heat Diffusion from Amsterdam

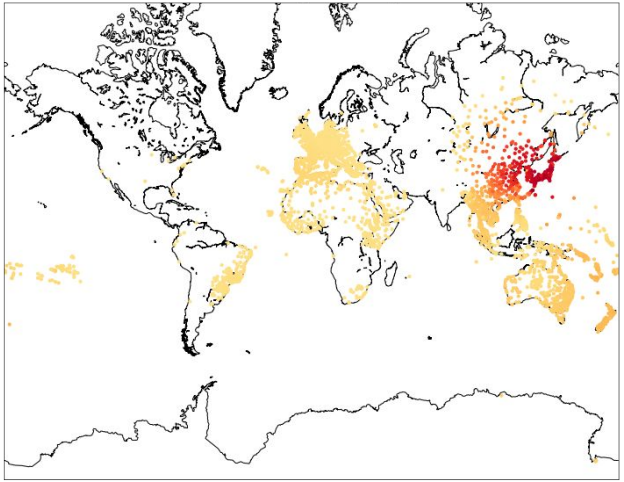
Re-Balanced Flight Distribution:



Original



Dataset augmentation



Re-Balances Flight Distribution

Heat Diffusion from Beijing

Conclusion:

- Close network properties:
  - Average degree
  - Number of edges
  - Diameter
- But more South and cross-hemispheres routes than before.
- Evenly spread flight routes around the world.
- May reveal emerging airports and world regions.

Re-Balanced dataset - GIF plot

Thank you

Questions ?