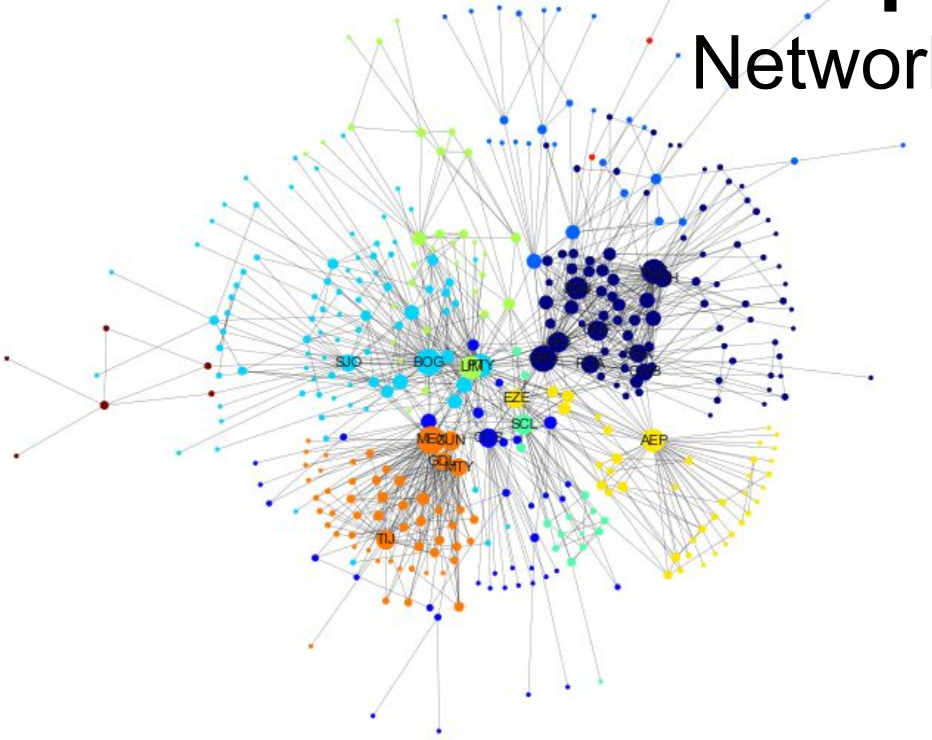
Air Transport Network

Networking Science – NS101
Final Project



Marcelo José Rovai

Universidad del Desarrollo

Master Data Science

March, 2019

Motivation

- ✔ Modeling air transport networks aims airline companies to organize their routes in a cost-efficient way and therefore maximize their profits.
- ✔ Air transport network models are also the tool to investigate system robustness. They help to determine weaknesses of the system in case of various kinds of disruptions.
- ✔ An alternative application is modeling ”human deceases networks”. Air transport network is used by millions of people every day, therefore it plays key role in the spread of some infections, such as Influenza, H1N1 or SARS

Research Questions

- ✔ *How is the Air Transport Network (ATN) structured in terms of nodes (airports) and links (routes)? Is the ATN similar to Social Networks?*
- ✔ *How to calculate the best route between destinations?*
- ✔ *How to hank the most important airports? How to analyze weakness?*
- ✔ *How the ATN can be analyzed in terms of communities?*
- ✔ *How studying ATN can help on human deceases spread prevention?*
- ✔ *How different are the topologies of distinct Air Companies?*

Dataset

1. "OpenFlights Airports Database" (January 2017):
    - Over 10,000 airports (Lat, Long, Alt, IATA code)
    - https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat

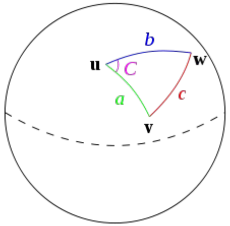
2. "OpenFlights/Airline Route Mapper Route Database" (June 2014):
    - 67,663 routes
    - 3,321 airports
    - 548 airlines
    - https://raw.githubusercontent.com/jpatokal/openflights/master/data/routes.dat

```
routes.sample(2)
```

	airline	source	dest	equipment	dist
57579	UA	SAN	LAX	CR7 EM2	176
4332	A3	DUS	ATH	320	2003

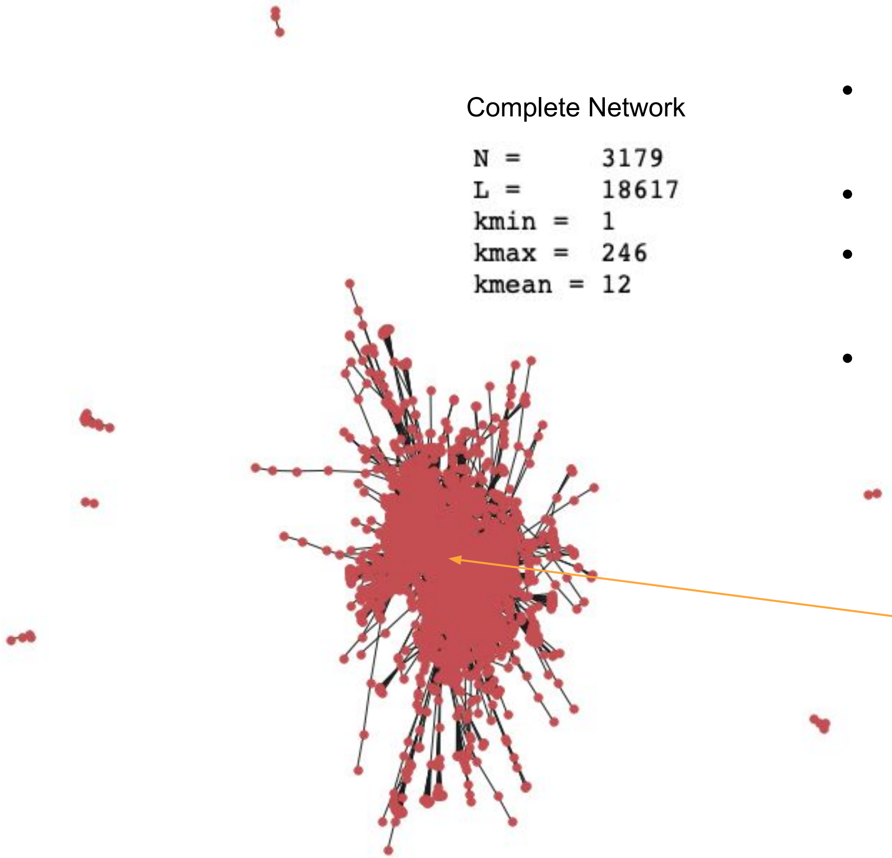
```
routes.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 66056 entries, 0 to 67662
Data columns (total 5 columns):
airline      66056 non-null object
source       66056 non-null object
dest         66056 non-null object
equipment    66038 non-null object
dist         66056 non-null int64
dtypes: int64(1), object(4)
memory usage: 3.0+ MB
```



$$\text{dist} = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$

Creating an Air Transport Network graph



Complete Network

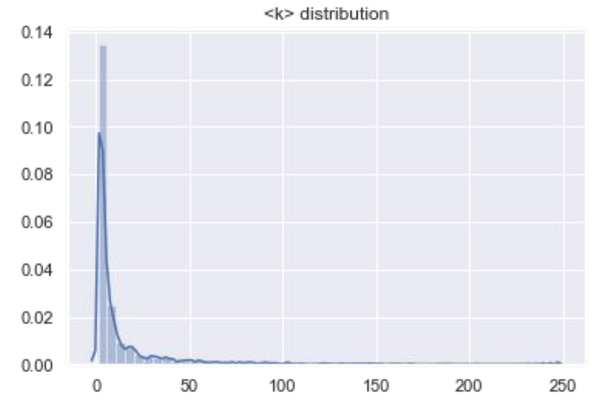
```
N =      3179
L =      18617
kmin =   1
kmax =   246
kmean =  12
```

- Graph is not connected, having a giant component that contains more than 99% of all airports.
- The giant component was used (undirect graph)
- The nodes are airports and its attributes are altitude and geo position.
- The links, are routes between 2 airports and its attributes are airline, distance and aircraft used.

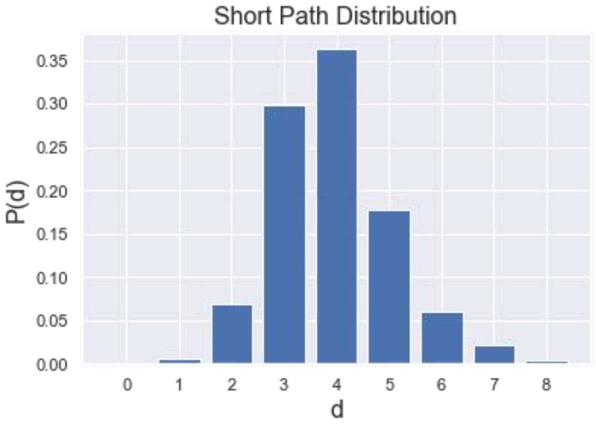
Giant Component

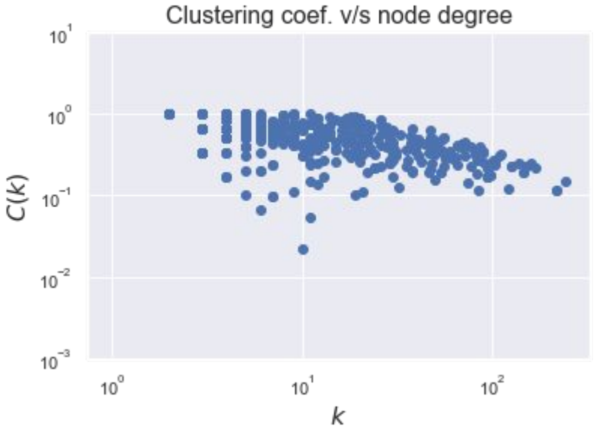
```
Number of Nodes (N):   3154
Number of Links (L):   18593
Min Degree     kmin:   1
Max Degree     kmax:   246
Mean Degree     <k>:   12
Ave Short Path  <d>:   4.0
Diameter       dmax:   9
Ave Clustering  <C>:   0.48
```

Networking Analysis



- There are a few airports with hundreds of routes
- More than 1/3 of all routes has 3 or less legs, being 4 the average.

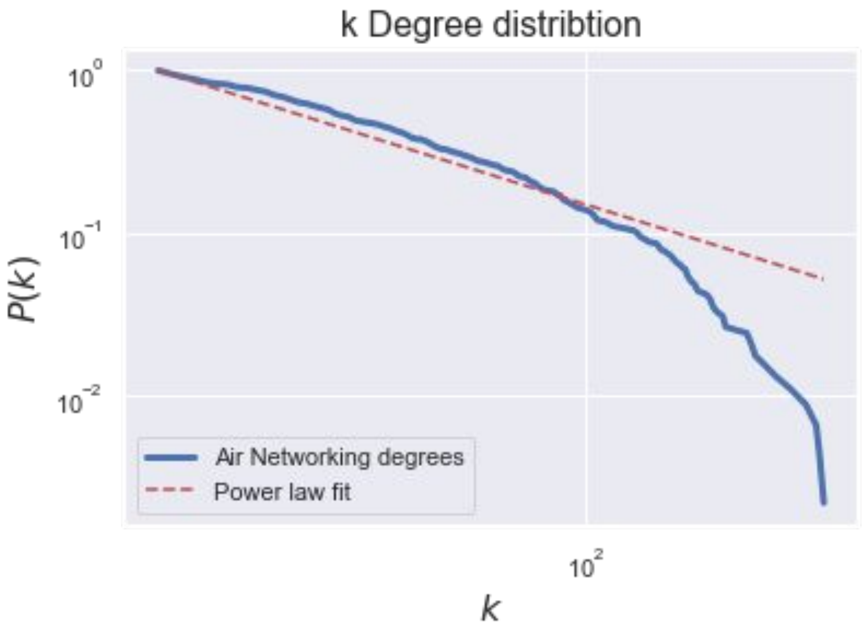




- The clustering coefficient C, is defined as the probability that two cities that are directly connected to a third city also are directly connected to each other.
- C is typically larger for the air transportation network than for a random graph and that it decreases with K size, but slower than d, for example.
- These results are consistent with the expectations for a small-world network but not with those for a random graph.

Networking Analysis

Ultra-Small-World Network



- Few HUBs (airport with high number of connections) appears and the pattern follows a "Power Law", for great part of distribution.
- The network shows a scale-free feature
- Hubs in scale-free networks affect the small world property, Barabasi calls networks in this regime ultra-small, as the hubs radically reduce the path length.
- They do so by linking to a large number of small-degree nodes, creating short distances between them."

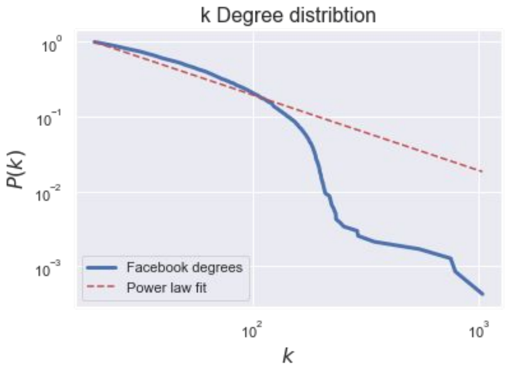
Networking Analysis

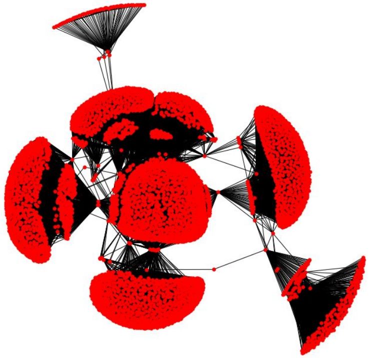
Comparing models

Network				
	ATN	WS	BA	FB
N	3,154	3,154	3,154	4,039
L	18,593	19,924	18,888	88,234
Kmin	1	7	6	1
kmax	246	17	229	1,045
<k>	12	12	12	44
<d>	3.92	4.60	3.13	3.73
Diameter	8	6	4	8.00
<C>	0.48	0.46	0.02	0.60
Hubs	Yes	No	Yes	Yes
Small-Word	Yes (*)	Yes	Yes	Yes
Power-Law	Yes	No	Yes	Maybe

Conclusion: ATN is a "Scale-Free" and "Small World" type network, where its behavior is very similar to Social networks as Facebook.

- Watts-Strogatz (WS) model is intended to model networks in the natural and social sciences ("Small-World"), but fails on degree (k) distribution (not prevent the appearance of Hubs).
- The BA model is better than the WS model at reproducing the degree distribution (Power-Law), but fails on C.
- FaceBook has huge hubs (k > 1,000), but for k < 200, similar to ATN.

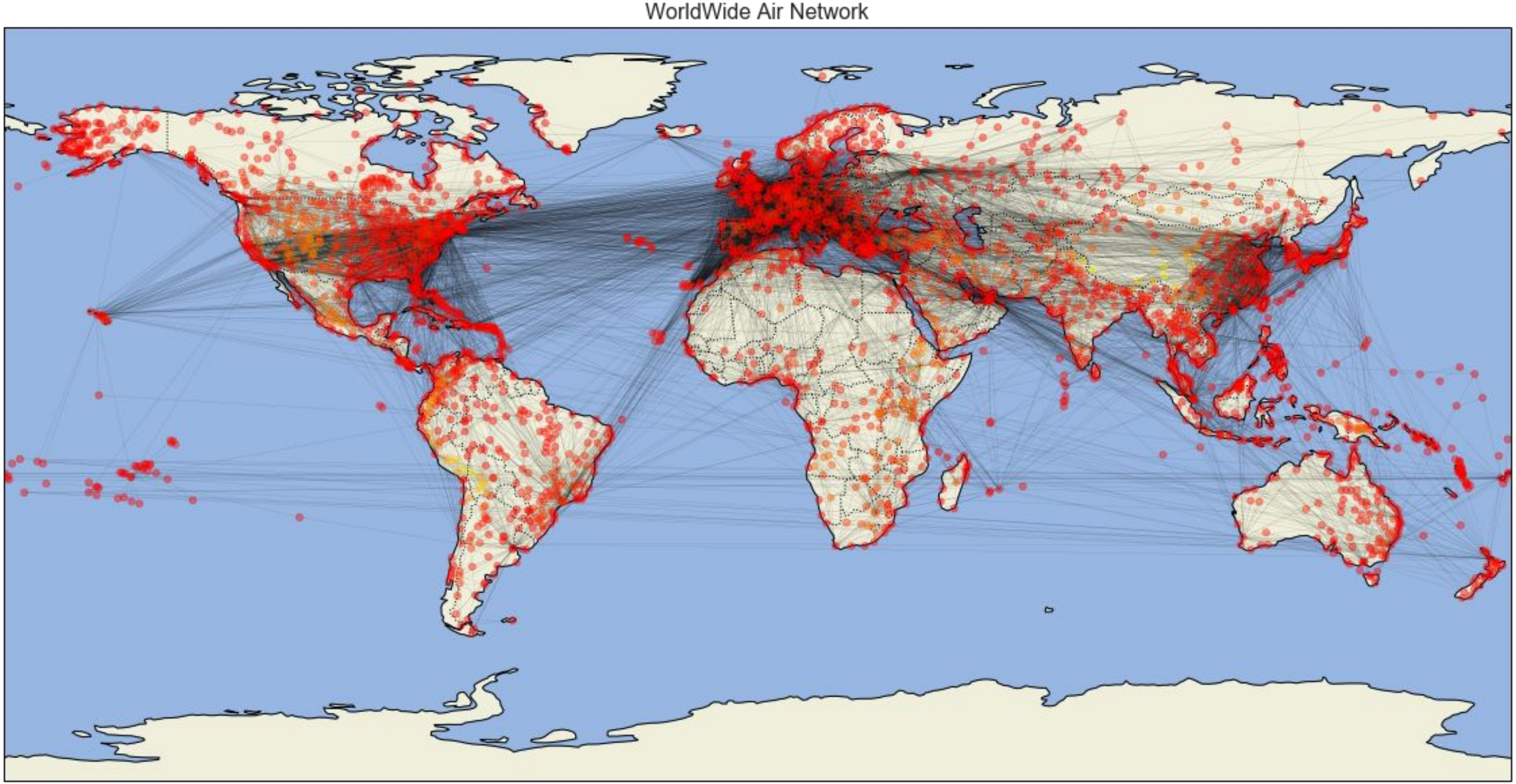




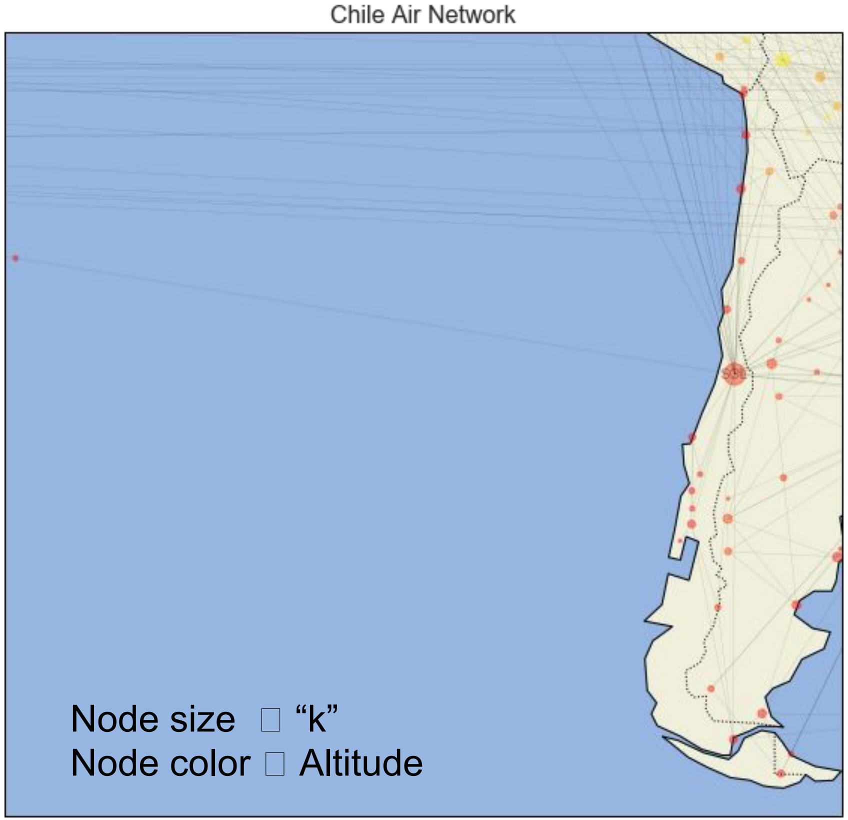
Research Questions

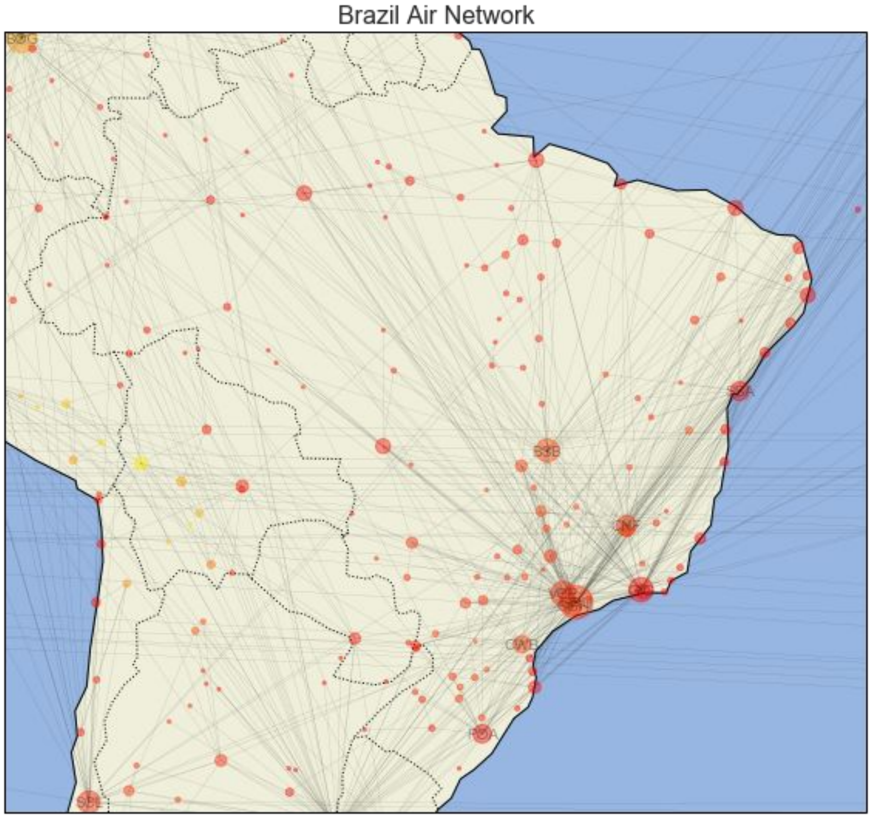
- ✔ *How is the Air Transport Network (ATN) structured in terms of nodes (airports) and links (routes)? Is the ATN similar to Social Networks?*
- ✔ *How to calculate the best route between destinations?*
- ✔ *How to hank the most important airports? How to analyze weakness?*
- ✔ *How the ATN can be analyzed in terms of communities?*
- ✔ *How studying ATN can help on human deceases spread prevention?*
- ✔ *How different are the topologies of distinct Air Companies?*

“Geo mapping” the Air Network



“Geo mapping” the Air Network





Node size □ “k”
Node color □ Altitude

Calculating the best route

Using Dijkstra's algorithm

Examples of best route calculation using Dijkstra's algorithm, having (a) distance as a “weight” and (b) minimum number of routes (no weight)”:

a.

```
# Mount Pleasant, Falkland Islands: (MPN) to Kioto, Japan (ITM)
source = 'MPN'
target = 'ITM'
plot_short_route(g, source, target)
```

executed in 245ms, finished 15:10:10 2019-03-18

```
Shortest path: ['MPN', 'PUQ', 'SCL', 'IPC', 'PPT', 'NRT', 'ITM']

Track 1: [MPN-PUQ]   Distance: 851 Km     Arline: LA / Plane: 320
Track 2: [PUQ-SCL]   Distance: 2180 Km    Arline: LA / Plane: 320
Track 3: [SCL-IPC]   Distance: 3752 Km    Arline: LA / Plane: 763
Track 4: [IPC-PPT]   Distance: 4249 Km    Arline: LA / Plane: 763
Track 5: [PPT-NRT]   Distance: 9445 Km    Arline: TN / Plane: 343
Track 6: [NRT-ITM]   Distance: 462 Km     Arline: NH / Plane: 767 738

Total distance: 20939 Km
```

Shortest path between MPN and ITM

b.

```
# Mount Pleasant, Falkland Islands: (MPN) to Kioto, Japan (ITM)
source = 'MPN'
target = 'ITM'
plot_short_route(g, source, target, None)
```

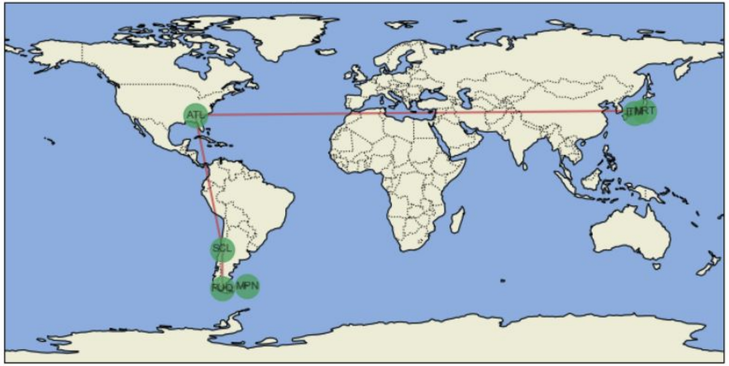
executed in 235ms, finished 15:12:42 2019-03-18

```
Shortest path: ['MPN', 'PUQ', 'SCL', 'ATL', 'NRT', 'ITM']

Track 1: [MPN-PUQ]   Distance: 851 Km     Arline: LA / Plane: 320
Track 2: [PUQ-SCL]   Distance: 2180 Km    Arline: LA / Plane: 320
Track 3: [SCL-ATL]   Distance: 7588 Km    Arline: KE / Plane: 76W
Track 4: [ATL-NRT]   Distance: 11002 Km   Arline: DL / Plane: 744
Track 5: [NRT-ITM]   Distance: 462 Km     Arline: NH / Plane: 767 738

Total distance: 22083 Km
```

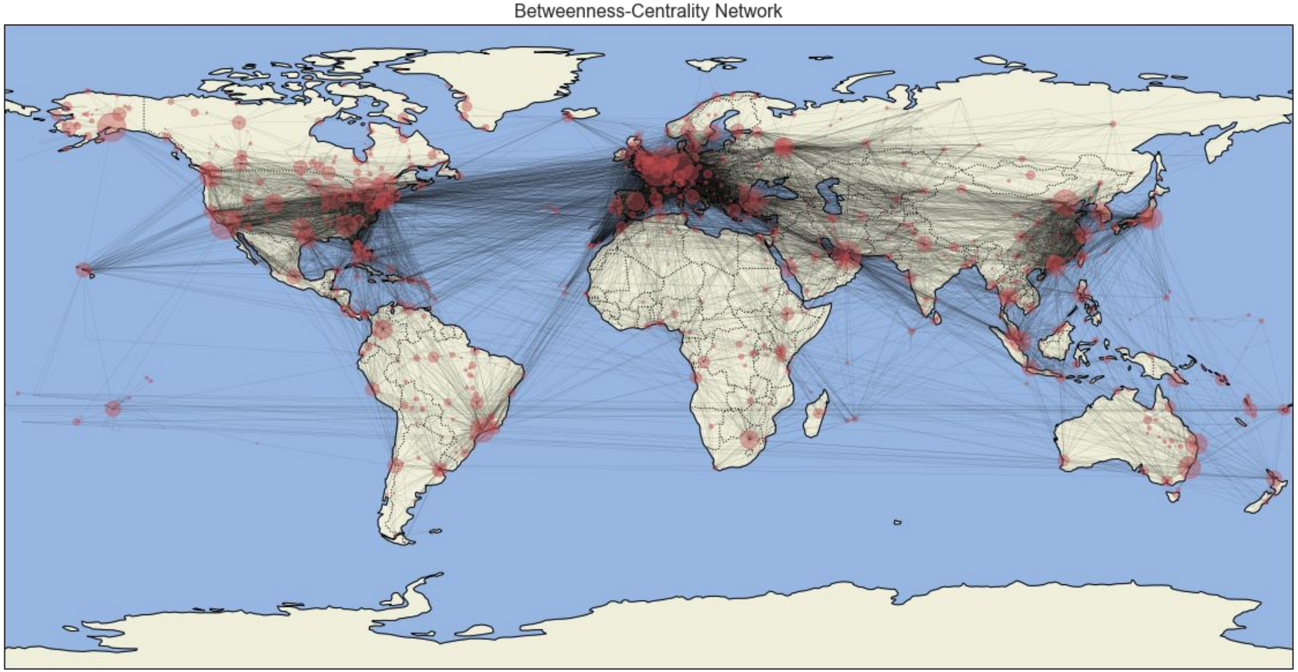
Shortest path between MPN and ITM



Research Questions

- ✔ *How is the Air Transport Network (ATN) structured in terms of nodes (airports) and links (routes)? Is the ATN similar to Social Networks?*
- ✔ *How to calculate the best route between destinations?*
- ✔ *How to hank the most important airports? How to analyze weakness?*
- ✔ *How the ATN can be analyzed in terms of communities?*
- ✔ *How studying ATN can help on human deceases spread prevention?*
- ✔ *How different are the topologies of distinct Air Companies?*

Centrality



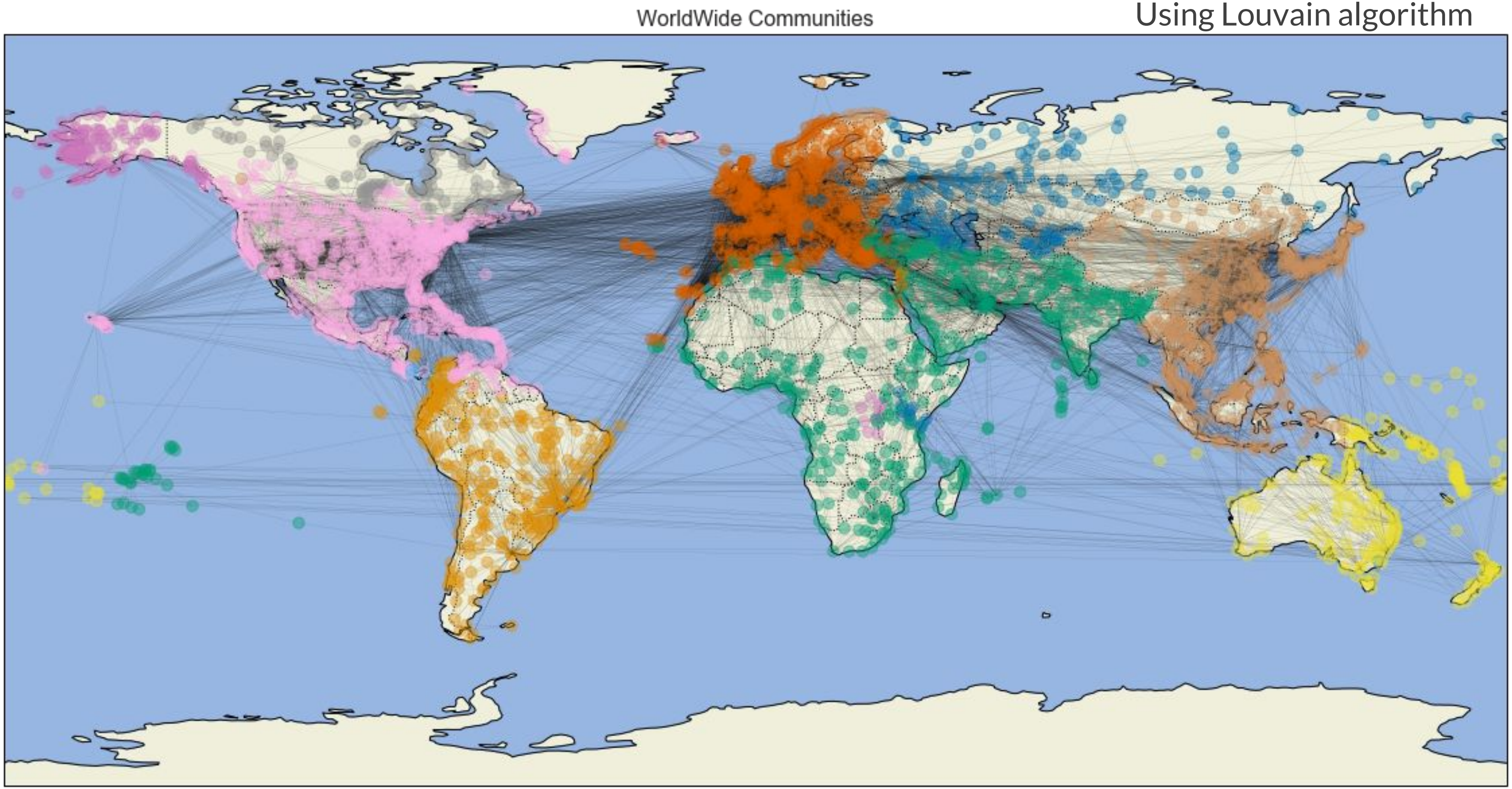
Degree	Betweenness	Closeness	Eigenvector
AMS	CDG	FRA	AMS
FRA	LAX	CDG	FRA
CDG	DXB	LHR	CDG
IST	ANC	AMS	MUC
ATL	FRA	DXB	FCO
ORD	AMS	LAX	LHR
PEK	PEK	JFK	BCN
MUC	ORD	YYZ	IST
DFW	YYZ	IST	ZRH
DME	IST	MUC	MAD
DXB	GRU	ORD	BRU
LHR	LHR	PEK	DUB
DEN	NRT	FCO	DUS
IAH	SYD	NRT	LGW
LGW	SEA	EWR	MAN

Research Questions

- ✔ *How is the Air Transport Network (ATN) structured in terms of nodes (airports) and links (routes)? Is the ATN similar to Social Networks?*
- ✔ *How to calculate the best route between destinations?*
- ✔ *How to hank the most important airports? How to analyze weakness?*
- ✔ *How the ATN can be analyzed in terms of communities?*
- ✔ *How studying ATN can help on human deceases spread prevention?*
- ✔ *How different are the topologies of distinct Air Companies?*

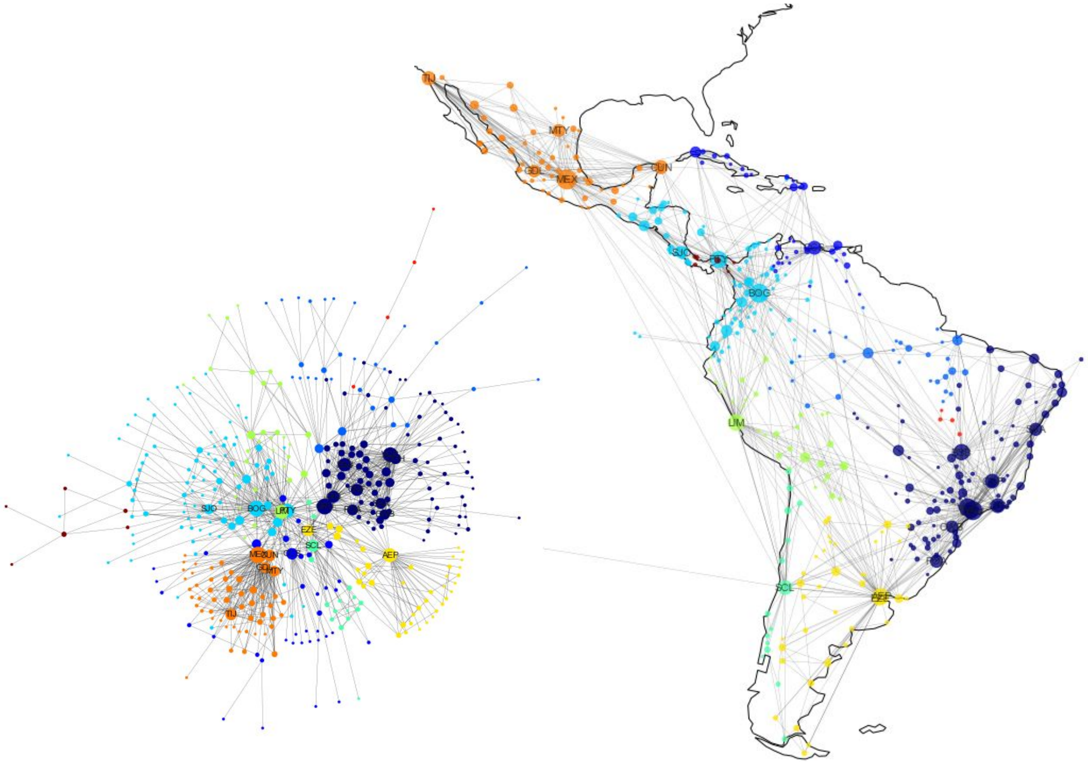
Community detection

Using Louvain algorithm



Community detection

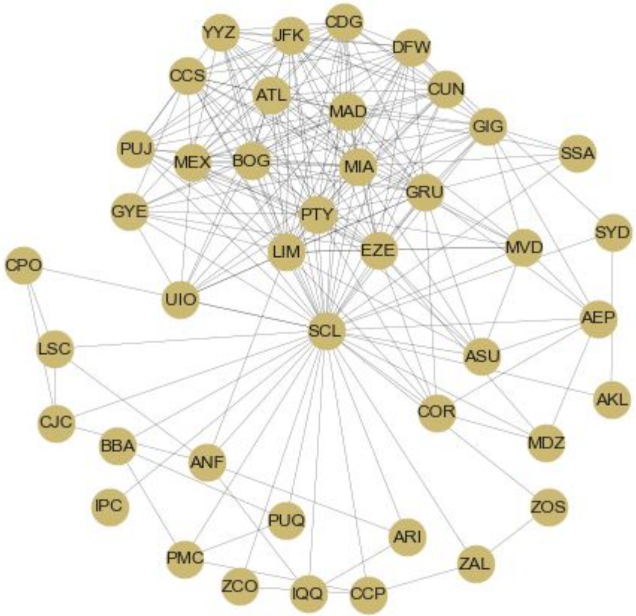
On previous map was possible to confirm that the airports are connected in "communities" with some geographic correlation. Even though geographical distance plays a clear role in the definition of the communities, the composition of some of the communities cannot be explained by purely geographical considerations. For example, India is mostly grouped with the Arabic Peninsula countries and with countries in Africa. These facts are consistent with the important role of political factors in determining community structure [4].These facts are consistent with the important role of political factors in determining community structure, as we can see on the Latin America case (Cuba & Venezuela):

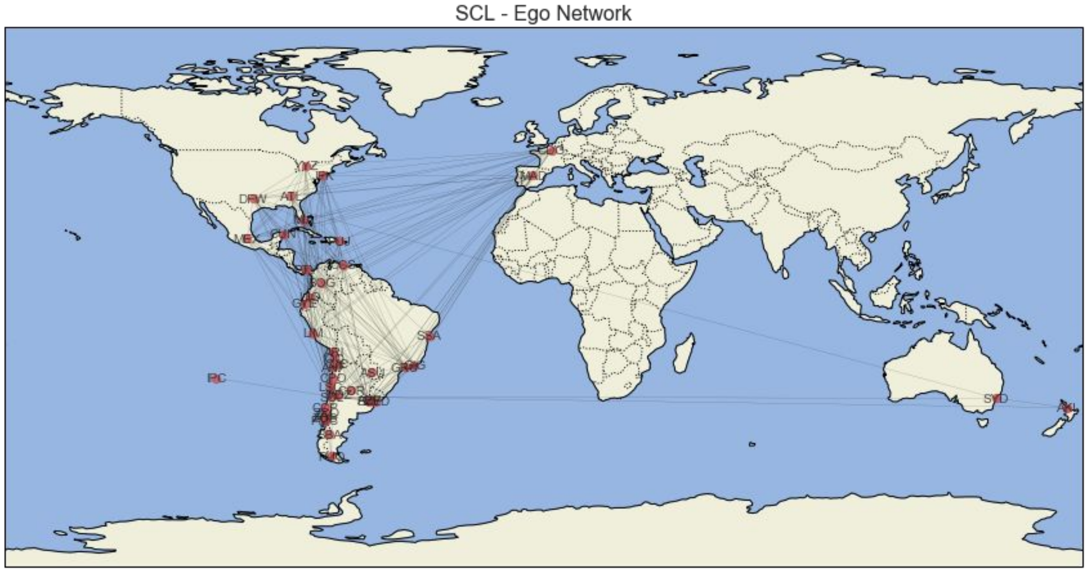


Research Questions

- ✔ *How is the Air Transport Network (ATN) structured in terms of nodes (airports) and links (routes)? Is the ATN similar to Social Networks?*
- ✔ *How to calculate the best route between destinations?*
- ✔ *How to hank the most important airports? How to analyze weakness?*
- ✔ *How the ATN can be analyzed in terms of communities?*
- ✔ *How studying ATN can help on human deceases spread prevention?*
- ✔ *How different are the topologies of distinct Air Companies?*

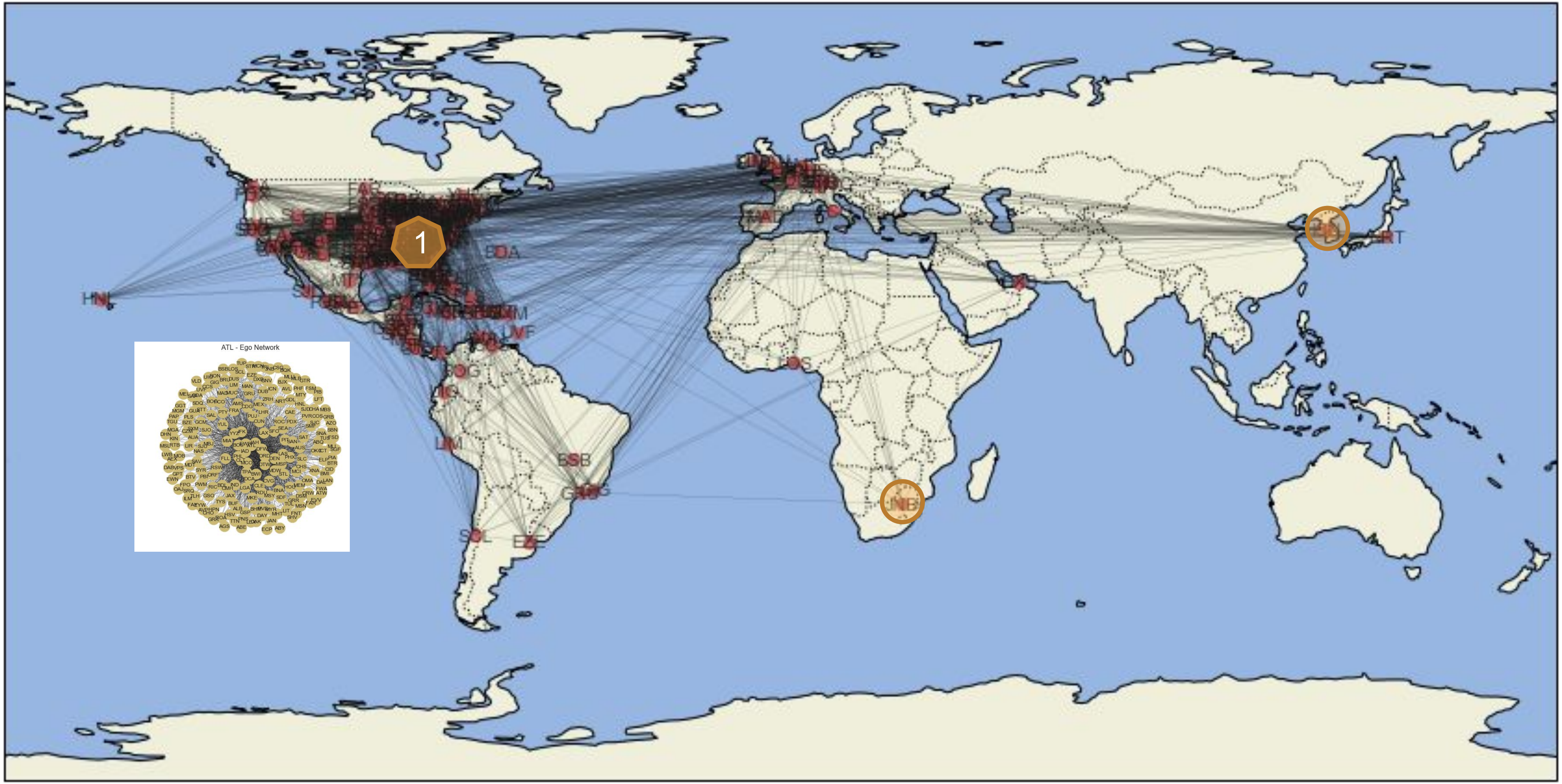
“Geo mapping” the Air Network – “Ego”



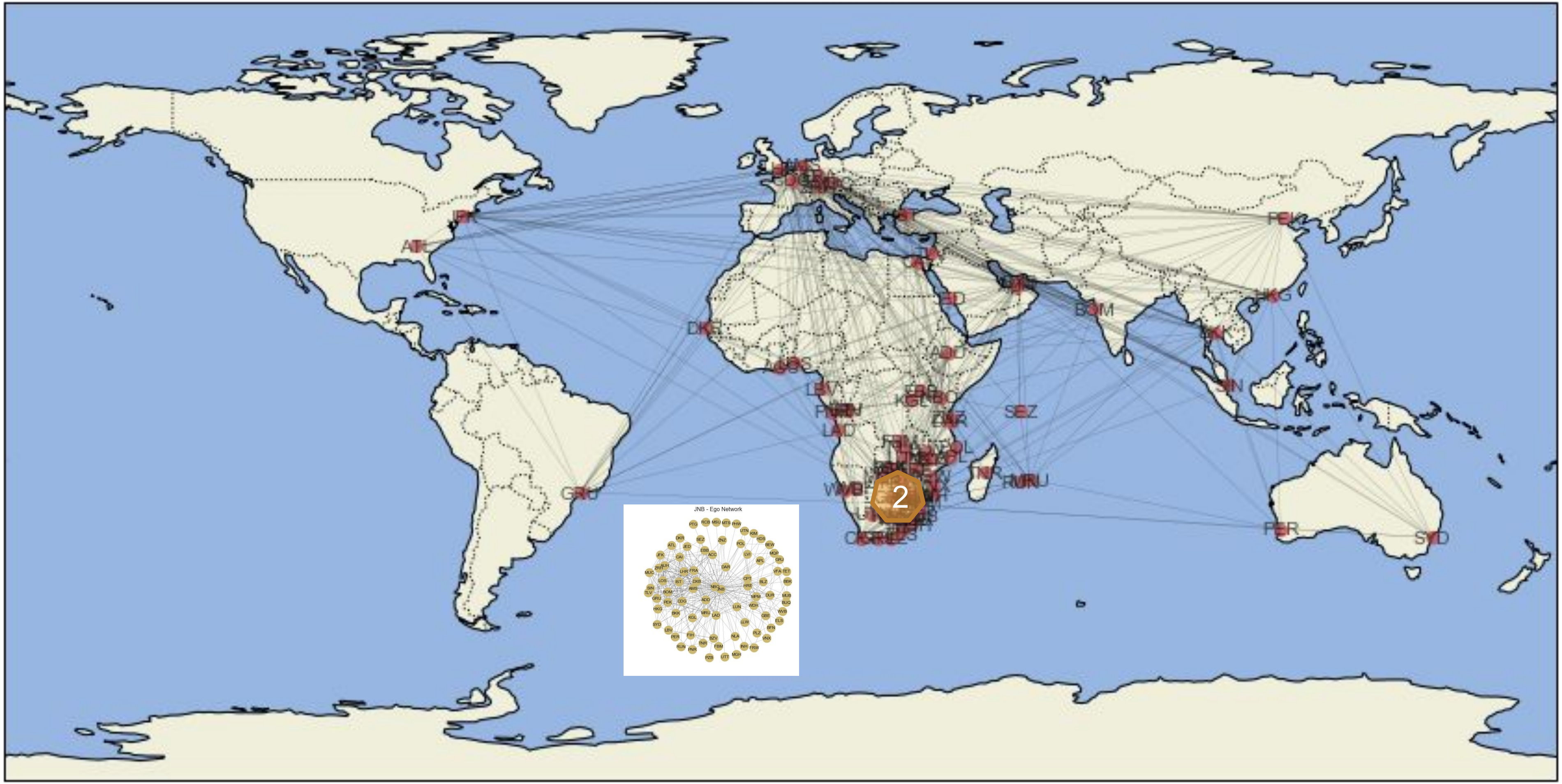


Ego networks consist of a focal node ("ego") and the nodes to whom ego is directly connected to (these are called "alters") plus the ties, if any, among the alters. This is a very important tool on Air Transport Network analysis, for example to study how human decease can spread along the network

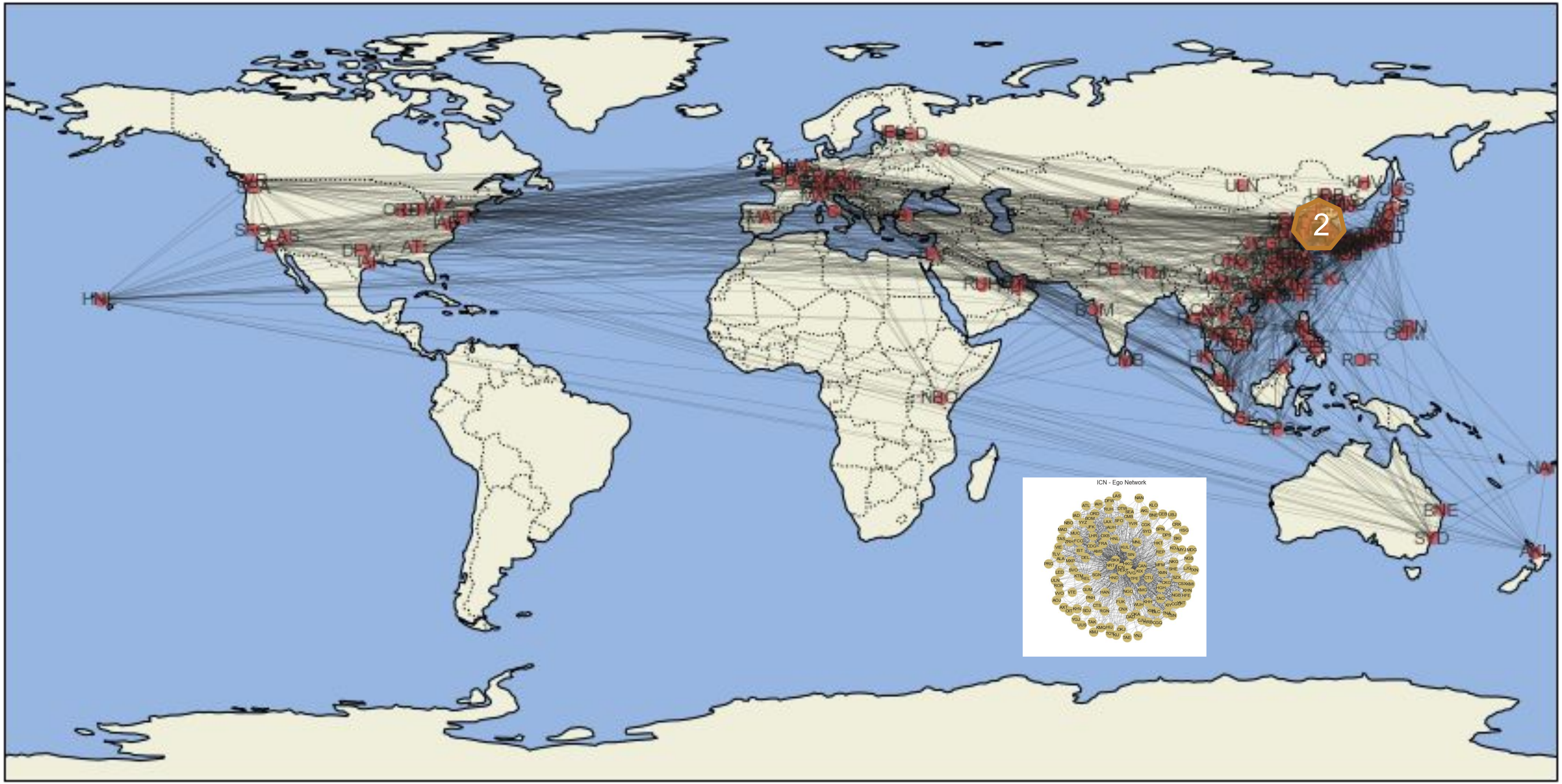
Decease outbreak, having Atlanta as starting point



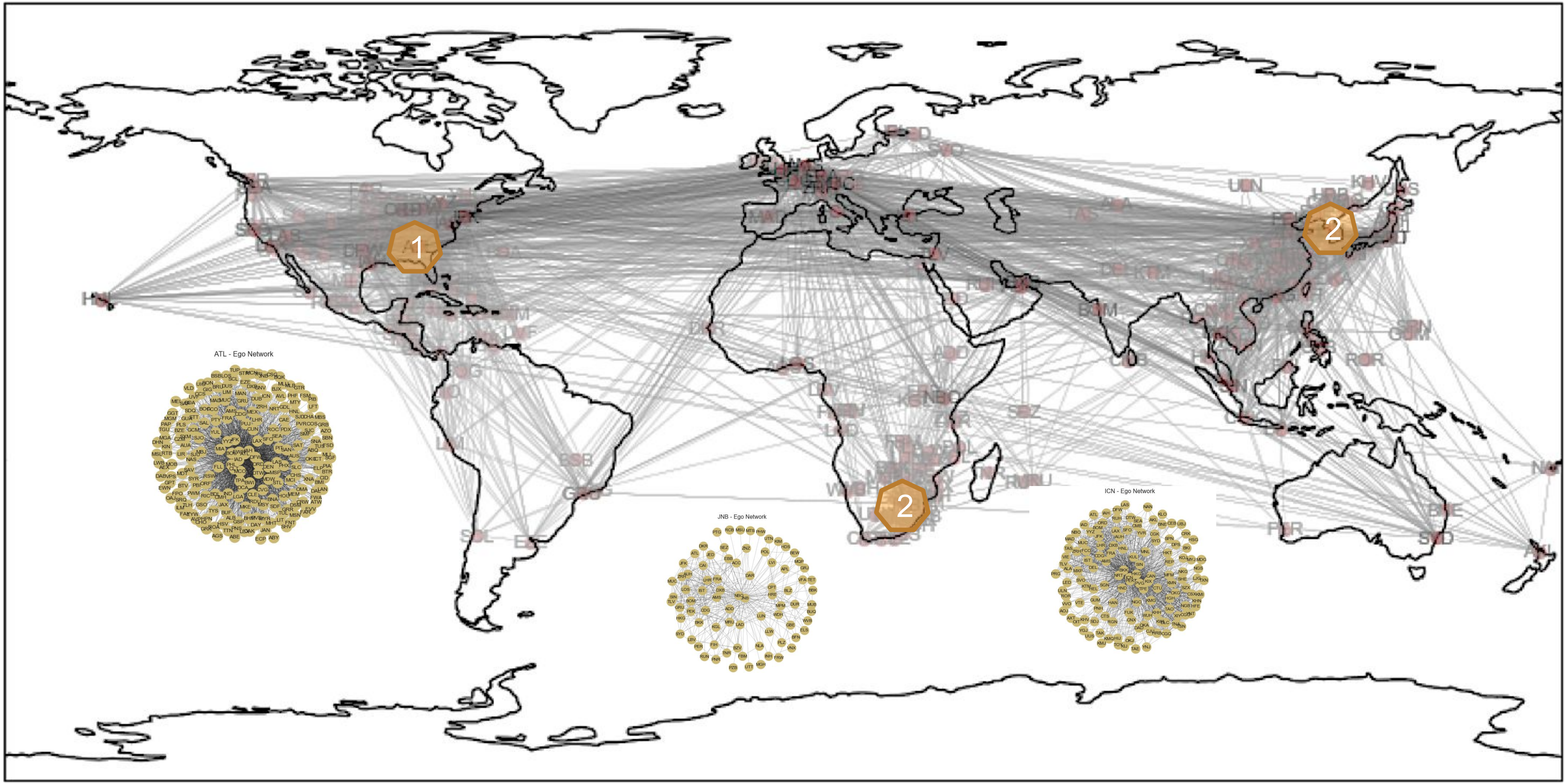
Decease outbreak, having Atlanta as starting point and Johannesburg as next stop



Decease outbreak, having Atlanta as starting point and Seoul as next stop



JNB □ ATL □ ICN



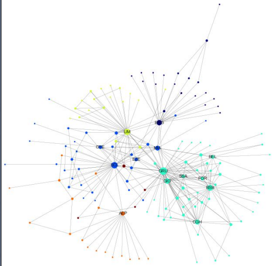
Research Questions

- ✔ *How is the Air Transport Network (ATN) structured in terms of nodes (airports) and links (routes)? Is the ATN similar to Social Networks?*
- ✔ *How to calculate the best route between destinations?*
- ✔ *How to hank the most important airports? How to analyze weakness?*
- ✔ *How the ATN can be analyzed in terms of communities?*
- ✔ *How studying ATN can help on human deceases spread prevention?*
- ✔ *How different are the topologies of distinct Air Companies?*

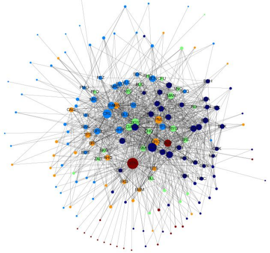
Comparing Air Companies Topologies

LATAM (LAN+TAM) Air Company Network

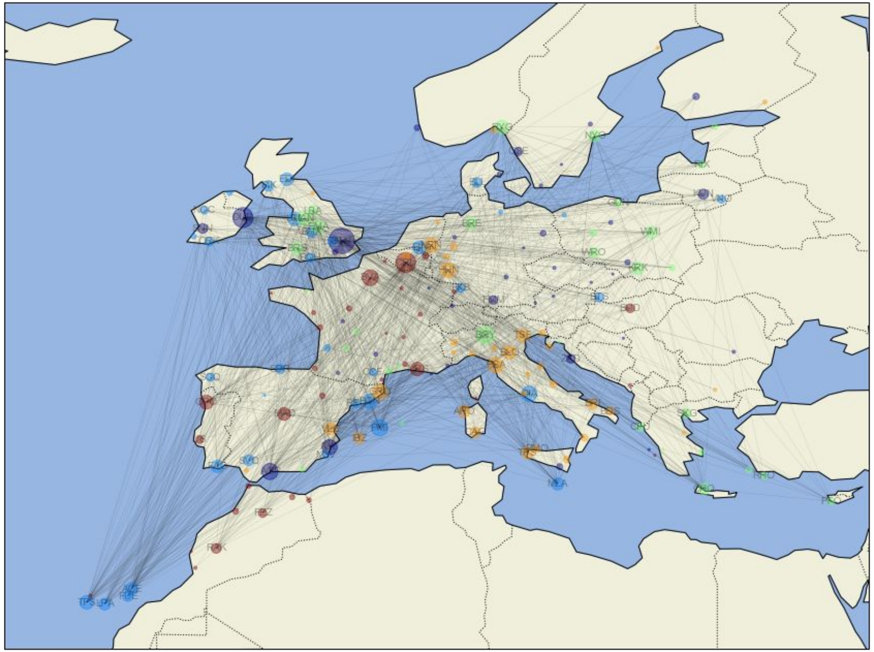
LATAM



RYANAIR



Ryanair Air Company Network



Looking on both air companies network distribution, we can verify what is described on [6], that the structure of of low-cost airlines, as Ryanair, the biggest European low-cost company, has a densely connected core, in opposition to companies as LATAM or other big companies.

Conclusion

On this project, the main idea, was to analyze the structure of the world-wide air transport network, understanding how we could define the best possible routes between two destinations, not only in economic terms, but also to prevent the spread of human deceases.

The study enables us to unveil a number of significant results. The worldwide air transport network is a "Small-World" network (*), similar to social networks as Facebook, in which (i) the number of nonstop connections from a given city and (ii) the number of shortest paths going through a given city have distributions that are scale-free. Surprisingly, the nodes with more connections are not always the most central in the network. We hypothesize that the origin of such a behavior is the metacommunity structure of the network. We find the communities in the network and demonstrate that their structure can only be understood in terms of both geographical and political considerations.

(*) According Barabasi, the presence of hubs in scale-free networks affect the small world property. "Airlines build hubs precisely to decrease the number of hops (stops) between two airports". Barabasi calls networks in this regime "Ultra-Small, as the hubs radically reduce the path length.

Resources

1. Networking Science
2. Think Complexity
3. IPython Cookbook, Second Edition
4. The worldwide air transportation network
5. Revealing the structure of the world airline network
6. Modelling the Air Transport with Complex Networks: a short review
7. Reliability Analysis for Aviation Airline Network Based on Complex Network
8. Finding The Shortest Path, With A Little Help From Dijkstra
9. An Introduction to Graph Theory and Network Analysis
10. Catching that flight: Visualizing social network with Networkx and Basemap
11. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena

Notebook and Datasets

- Datasets:
  - https://github.com/Mjrovai/UDD_Master_Data_Science/tree/master/NS101-WW_Air_Transport_Network/data
- Notebook:
  - https://github.com/Mjrovai/UDD_Master_Data_Science/tree/master/NS101-WW_Air_Transport_Network

Thank you

Marcelo José Rovai

Santiago, Chile **March 19th 2019**