Jonas Hou (Examination number: FPDT7)
COMPGW01: Complex Networks and Web
Coursework 2: Project

Analysis of global air traffic and intercontinental flight connections

0. Abstract
This study analyses the nature of global air traffic and the air traffic between continents in particular by modeling the global air traffic as a directed weighted network. The resulting Global Flight Network is a scale-free small world network with a star-shaped structure consisting of a tightly connected core of intercontinental hubs and a tree-shaped periphery of local nodes. This structure is both reflected by and responsible for the network's overall topology and the anatomy of its intercontinental connections and has implications for the importance of continents and their relationships between each other.

1. Introduction
Global aviation is a major defining factor for mobility in our society with huge political, economic and social implications. According to the figures of the Air Transport Action Group [13], 3,1 billion passengers have been carried by airplanes in 2013. Understanding the nature of global air traffic is therefore beneficial and vital for economies.

This paper investigates the nature of flight connections and in particular, puts an emphasis on flight connections between continents. Studying the nature of intercontinental flights is highly important since it gives a better understanding on how different parts of the world are connected together. For the majority of people, intercontinental flight routes are also the only practical way to move between different continents.

Within this paper, a network representing the global air infrastructure referred to as the "Global Flight Network" or "GFN" is created based on publicly available data. This GFN will then be analysed using methods from network research to gain insights about its structure.

2. Research questions:
- 1. Analysis of the large scale structure of the Global Flight Network
    - What are the key network metrics?
    - How can the the network be classifications?
- 2. Analysis of intercontinental flight connections
    - How are continents connected with each other?
    - What does this say about the relationship between different continents?

3. Data and Methodology

Datasets

The data for this study is taken from openflights.org [8], an online database to search for flight connections and track flight history. Openflights.org provides the following datasets:

(1) airports.dat: A list of 8107 airports containing the following attributes: airport name, city, country, geographic position expressed in latitude and longitude, time zone
(2) routes.dat: A list of 67663 flight connections containing: Airline operating the connection, source airport, destination airport, used machine. All listings are unique combinations of airline, source and destination airport.

Network Model

Through combining the information from both datasets, a network of the global air traffic ("GFN") is created which contains every airport as a node and every flight connection as a directed weighted edge. Nodes which end up having a degree of 0 were not included into the GFN since these nodes didn't contribute to the network topology. The GFN was conceived as a directed network because the flight connections in the datasets are directional in nature.

Since one aim of the study is to investigate the nature of flight connections between different continents, every node was assigned to one of the following continents: North America, South America, Europe, Africa, Asia, Australia and Oceania based on the widely used 7 continents convention [14].

The edges within the GFN have the following properties:
- Every edge in the network is a distinct route which is a unique combination of source and destination airport.
- The route from airport A to airport B and the route from B to A are expressed as two different edges.
- The edge weight represents how many different flight connections exist for the same route and serves as a rough estimation about the popularity of a route.

This leaves different possibilities to interpret the meaning of a node's degree:
- The un-weighted in/out-degree of a node represents the amount of different airports to which a particular airport is connected. The GFN has a reciprocated edge ratio of 97,67% which means that almost all connections are bidirectional. As a result, the in and out-degree is identical for almost all airports.
- The weighted degree of a node represents the amount of flights that arrive or depart at a particular airport. The weighted degree can be seen as a rough indicator about the business of an airport and therefore it's importance. It needs to be pointed out however, that the accuracy of this interpretation is highly limited. Due to the lack of information about flight schedules and passenger counts, the GFN might favor certain airports which might turn out to be less important in reality. It is therefore very important to keep the limitations of the models used in this study in mind when evaluating its findings.

Tools

The following tools were used for this study:
- Gephi [9] was used to model the network and doing the majority of analysis. The network visualisations were also created in Gephi using the Geo Layout plugin [10].
- Microsoft Excel and NodeXL [11]: were used to manually compute certain network metrics.
- Plot.ly [12] was used to create the diagrams for the degree distribution.

4. Results

4.1. Topology of the Global Flight Network ("GFN")

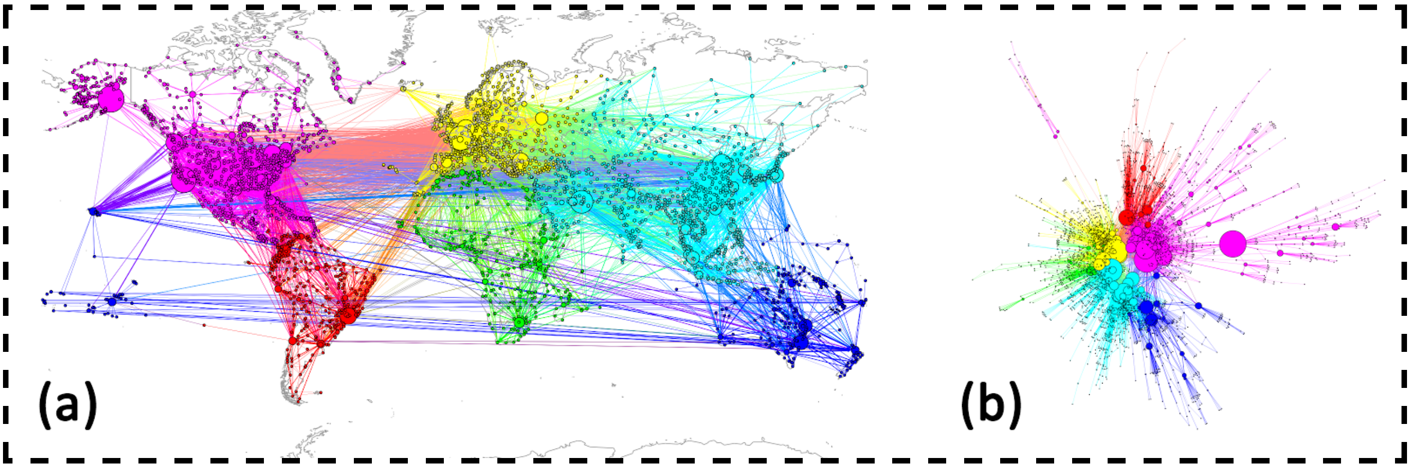


Figure 1: Visualisation of GFN (a) with accurate geographic positions (b) with Yifan Hu [3] layout algorithm. Size of nodes correspond to betweenness centrality. Color corresponds to continents.

The GFN created for this study consists of 3281 nodes and 67202 weighted edges (flight connections) or 37181 unweighted edges which translate to an average weighted average node degree of 40,96 or an average unweighted in/out-degree of 11,33. Compared to the amount of possible edges to achieve a fully connected graph (3281*3280 = 10761680), the GFN is sparsely connected with a density of 0,00345. This fact is not surprising since not every connection between any two nodes is desirable or commercially viable. At the same time, the connections of the GFN are highly redundant. The average edge weight is 1,8 which means that every route is operated by 3,6 airplanes on average.

The GFN is visualised in Figure 1. Fig.1a shows how the flight connections are laid out on the world map while Fig.1b depicts the topology of the network using a layout algorithm by Yifan Hu [3]. It can be inferred from Fig.1b that the GFN is shaped like a star consisting of a highly connected core in the center and a loosely connected tree-shaped periphery similarly to the structure of the internet [7]. It also appears that you are very likely to visit a node in the center if you want to move from one continent to another since all paths go through the highly connected core.

The importance of the betweenness centrality of a node also becomes clear when analysing Fig.1 where the betweenness centrality is represented by a node's size. The betweenness centrality is a measure of how many shortest path go through a certain node. Nodes that act as bridges between two isolated parts A and B of a network (often referred to as "hub") tend to have high betweenness centrality since all paths between A and B have to go through those particular hub nodes. In Fig.1b it can be seen that such hubs are either located in the center of the network or are connecting the center with the tree-shaped periphery. In Fig.1a it can be seen that the hubs are evenly distributed across the globe which raises the assumption that every continent has a small amount of hubs which connect the continent with the rest of the world.

The degree distribution is an important key figure to understand the structure of a network. An analysis of the in-degrees reveals that the degree distribution of the GFN follows a power law

$$P(k) \propto k^{-\gamma} \tag{1}$$

with $\gamma \approx 1,5$. This confirms that the GFN is in fact a scale-free network which are said to be especially resilient against random node failures [6] and also correspond to the findings from an independent study conducted by Guimerà et al. [1]. The degree distribution of the GFN can be seen in Figure 2.

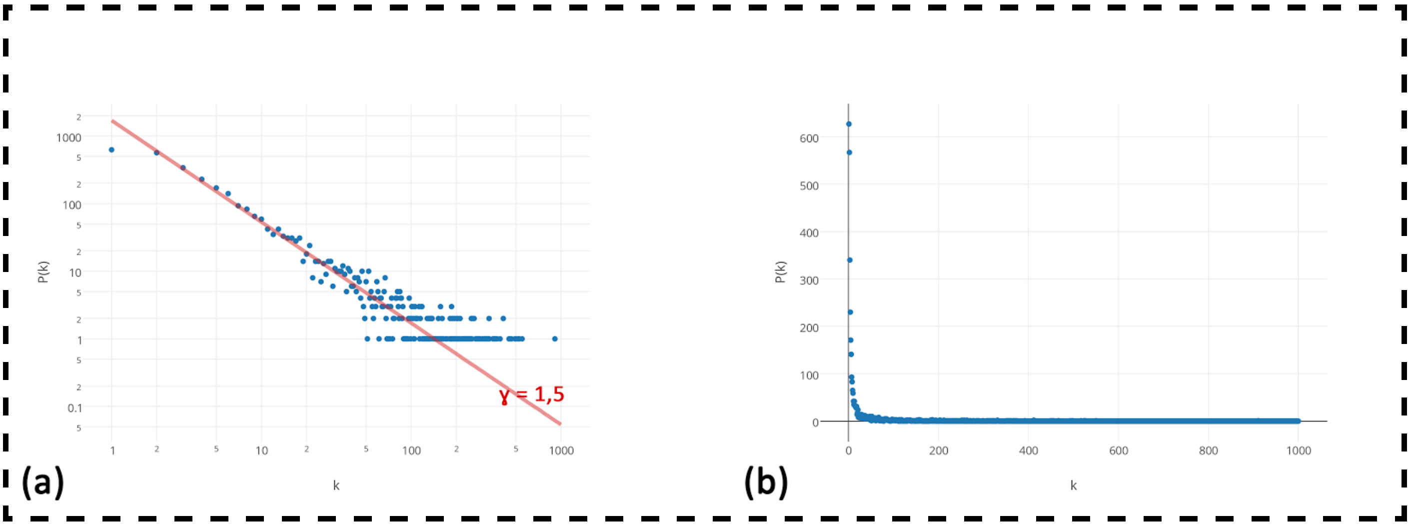


Figure 2: Degree distribution of GFN (a) in log scale (b) in linear scale

Another important property to classify networks is the small world property which is characterised through short average paths and high avg. clustering coefficients [5]. The average undirected path length inside the GFN is 4,01 while the average clustering coefficient is 0,47. In contrast, a randomly generated network following the Erdős–Rényi (ER) model [4] with p = <k>/(n-1) = 11,33/3280 = 0,00345 has an average path of ln(n)/ln(<k>) = 3,3 and an avg. clustering coefficient of p = 0,00345. Comparing these metrics shows that the GFN has comparatively short avg. paths and a much higher avg. clustering coefficient than the random network which confirms that the GFN is a small world network according to the principles taught in the lecture. This again confirms Guimerà et al.'s [1] findings. .

4.2. Nature of Intercontinental Connections

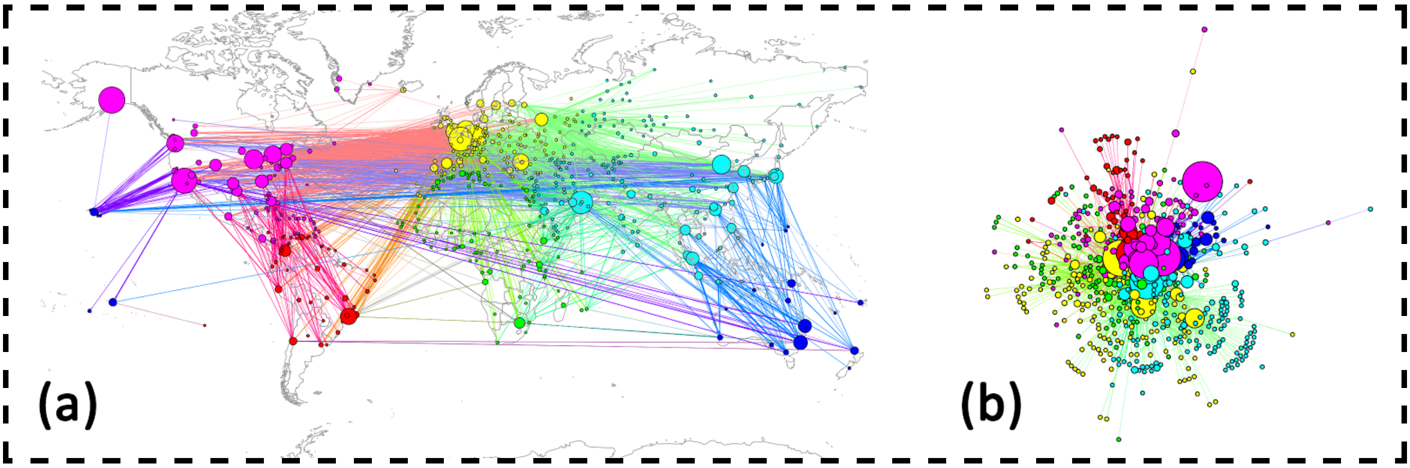
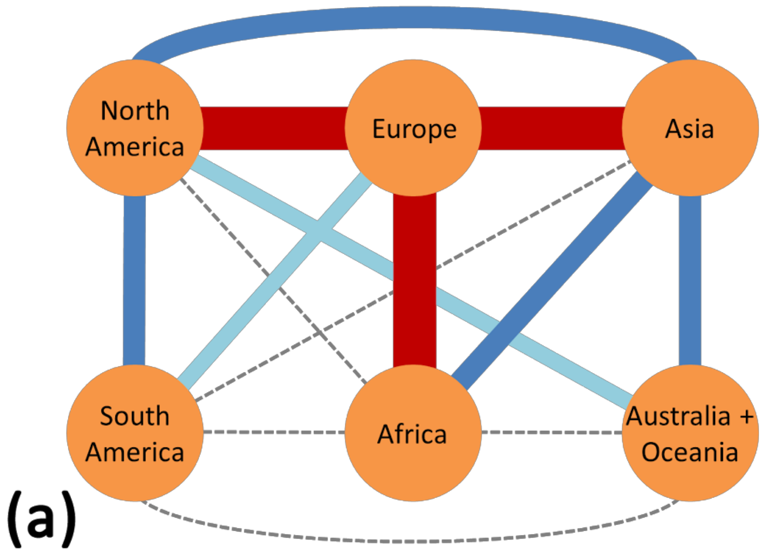


Figure 3: Visualisation of GFN showing only intercontinental flight connections
(a) with accurate geographic positions (b) with Yifan Hu [3] layout algorithm.

This section will explore how continents are connected between each other and thus will only focus on nodes which are connected to at least one airport from a different continent and the edges that connect these airports. A visualisation of this subset of nodes and edges can be seen in Figure 3. It turns out, that only a small fraction of nodes are connected to a different continent. Precisely, 605 nodes can be classified as intercontinental airports which is 18,44% of the total amount of nodes. The average betweenness of these intercontinental nodes (without recalculating the values) is 42768 which is considerable higher than the average betweenness centrality of the full GFN (9707). This means that nodes with high betweenness centrality also have the tendency to have intercontinental connections which confirms the assumption made in section 4.1 that the GFN consists of a core connecting all continents together.

The intercontinental airports are connected through 5276 unweighted intercontinental edges or 9769 weighted intercontinental edges. this is again a considerably small number compared to 31905 unweighted or 57433 weighted intracontinental connections and reassures the conclusion, that the most flights in this world happen within a continent. This makes sense since firstly, flights within a continent are usually shorter and therefore more profitable for airlines and secondly, places within the same continent tend to have stronger cultural or economic bonds due to spatial proximity.



(b)

Weighted intercontinental Connections			
Europe	-	Asia	3072
North America	-	Europe	2009
Europe	-	Africa	1704
Asia	-	Africa	649
North America	-	South America	603
North America	-	Asia	580
Asia	-	Australia	579
North America	-	Australia	290
Europe	-	South America	196
North America	-	Africa	45
South America	-	Africa	14
South America	-	Australia	10
Africa	-	Australia	10
Asia	-	South America	8
Europe	-	Australia	0

Figure 4: Distribution of intercontinental flight connections
(a) expressed by a picture, (b) expressed by figures

A closer look over how the continents are connected with each other in detail (Figure 4) reveals a number of interesting findings. First of all, the intercontinental connections in this world are very unevenly distributed. While the continents of the northern hemisphere are tightly connected with each other, the continents of the southern hemisphere are only very loosely connected. In fact, all connections between the three continents on the southern hemisphere are maintained by a mere total number of 15 airports. A number of 10 weighted intercontinental connections means that often times two continents on the southern hemisphere are only connected by 2-3 airports on each continent. This type of connection is very fragile shutting down a handful of key intercontinental airports would be sufficient to separate two entire continents from each other, forcing passengers who wish to travel between these continents to pass through a third continent in between.

On the contrary, the continents of the southern hemisphere are best connected with the closest continent to them on the northern hemisphere. This again gives the network a star-like shape with a tightly connected core consisting of North America, Europe, Africa and Asia and South America and Australia & Oceania in the periphery. The nature of intercontinental connections have a significant impact on the role of each continent in the GFN. Europe for example assumes a central position in the network which means that it's very likely to you need to pass Europe in order to move from one point in the world to another. This gives Europe a high degree of importance and influence for air traffic as a whole. On the other hand, a continent like Australia & Oceania can more or less only be accessed through North America or Asia, making Australia & Oceania highly dependant on these two continents.

The analysis of intercontinental connections also yields a couple of surprises. For instance, the number of connections between two continents is only weakly correlating to the distance between these continents. A good example for that is the relationship between North America and Asia which have a comparatively small number of connections between each other despite their spatial proximity and well established political and economic relationships.

	Weighted Intraedges	Weighted Interedges
North America	15812	3527
Europe	17159	6981
Asia	18214	4888
South America	2380	831
Africa	2108	2422
Australia	1760	889
total	57433	9769

Figure 5: Number of intercontinental edges connected to every continent

Another surprise lies in the fact how Africa is connected with the rest of the world. A look at Figure 5 reveals that Africa has more intercontinental connections than intracontinental connections. This means that Africa is in fact the only continent in the world that is better connected to the outside than it is connected on the inside. This also means that Africa is the only continent that could not be classified as a community inside the GFN. Fig.1a provides an explanation for this special case: Most continents have a couple of central nodes which are connected to a lot of airports inside the continents in a star-shaped structure and which also act as a hub to other continents. These hubs however are mostly absent in Africa probably because of economic factors and because airports inside Africa are only loosely connected with each other in the first place. As a result, most airports in Africa are directly connected to another continent instead of being connected to a central hub in between.

5. Discussion

In this paper both the large scale structure of the GFN and the connections between continents are analysed. The GFN is a scale-free small world network with sparse but redundant connections. The GFN also has a star-shaped structure consisting of a tightly connected core and a tree-shaped periphery similar to the internet. Nodes with high betweenness centrality pay a huge role in the network since they act as a bridge between different parts of the network. The connections between continents are fewer than the connections inside continents except for Africa and again display a star-shaped structure. Some continents can almost only be reached through other continents leaving these continents dependant on their connecting continents. A lot of the characteristics of the GFN can be explained through real world constraints such as political, social or economic circumstances and reflect the role of different continents in the world and their relationships between each other.

Future research could divide the GFN into finer grained communities or focus on the role of air traffic for particular continents or economies. The global air traffic could also be further analysed by refining the model with flight schedule or passenger count data to produce more accurate results. Another approach could be to investigate human mobility as a whole by also considering the traffic network for land or sea vehicles.

6. References

All webpages have been accessed on the 06.01.2015

Papers

[1] Guimerà, R., Mossa, S., Turtschi, A., & Amaral, L. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. Proceedings of the National Academy of Sciences, 102(22), 7794-7799.

[2] Verma, T., Araújo, N. A., & Herrmann, H. J. (2014). Revealing the structure of the world airline network. arXiv preprint arXiv:1404.1368.

[3] Hu, Y. (2011). Algorithms for visualizing large networks. Combinatorial Scientific Computing, 5(3), 180-186.

Lecture notes from "Complex Networks and Web"

[4] Slide set 4: Random Graph

[5] Slide set 5: Small World

[6] Slide set 6: Power law network

[7] Slide set 11: Internet New

Data:

[8] Openflights.org: http://openflights.org/data.html

Tools:

[9] Grephi: https://gephi.github.io/
[10] Geo Layout Plugin for Gephi: https://marketplace.gephi.org/plugin/geolayout/
[11] NodeXL: http://nodexl.codeplex.com/
[12] Plot.ly: https://plot.ly/plot

Others

[13] Air Transport Action Group: http://www.atag.org/facts-and-figures.html

[14] Continents: "Continents: What is a Continent?". National Geographic. (2009).