COVID-19 and the Global Flight Network

Daniel Tang

May 2020

Introduction

Our motivation was to visualise worldwide flight traffic as a network in tandem with the number of COVID-19 cases confirmed globally to see if our network model provided insight regarding virus transmission both internationally and domestically.

Data Engineering

Data Sources

1. Historical Flight Data
   - aviationstack.com
2. Information on Airports and Airlines
   - aviationstack.com
   - aviation-edge.com
   - icao.int
3. COVID-19 Confirmed Cases Data
   - John Hopkins University (github.com/CSSEGISandData/COVID-19)
4. Geographical Data to Link Airports to State or Country Locations
   - geonames.org (Open Geographical Database)
5. Geographical Map Object (shapefile) Sources
   - naturalearthdata.com; gadm.org

Data Processing

Python Packages
requests
datetime
networkx
numpy
pandas
matplotlib
cartopy

- Data engineering primarily programmed in Python
- Download all flight data (up to 200k flights per day) and auxiliary airport and location information
- Download COVID-19 case data
- Data cleaning performed for both data sets (raw flight data had a non-trivial number of errors)
- Data cleaning and integrity checking was done to ensure both data sets had valid geographical information hat could map them to states, countries, etc.

Data Processing

- Extensive data preparation and processing was required to merge all data and handle a number of complexities
- Large number of flights per day - had to handle memory constraints
- Computational efficiency concerns computing large number of rows and dimensions - vectorised operations
- Likewise had to optimise code for animation
- Network state over time stored as a list of `NetworkX` graphs
- Timezones had to be handled properly due to source error
- Codeshares for flights weren't entirely clean - had to reverse engineer unique transits
- It was interesting to see different political naming conventions in the geographical data, but this meant manual editing to match locations.

Network Analysis

Network Analysis

Measuring the degree of connectivity in a directed weighted graph:

- Typical calculation for degree is for undirected, unweighted graphs
- We calculate both ”in degree” - the number of inbound edges per node and ”out degree” - the number of outbound edges per node
- Each edge is weighted by the number of transits on the edge daily

Global Mean Degree

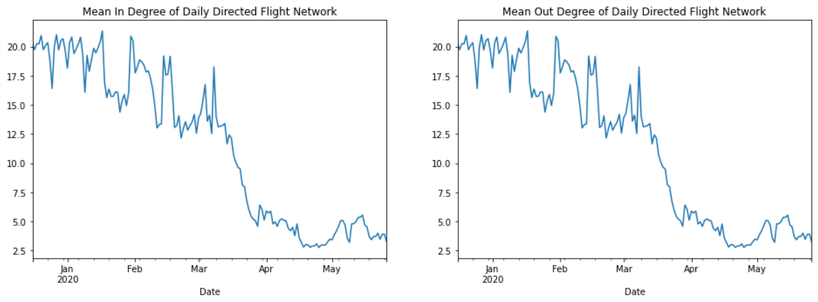

Figure 1: In degree and out degree over time averaged across all airports globally.

We also looked at averaging the in degree each week across all nodes to smooth out the curve.

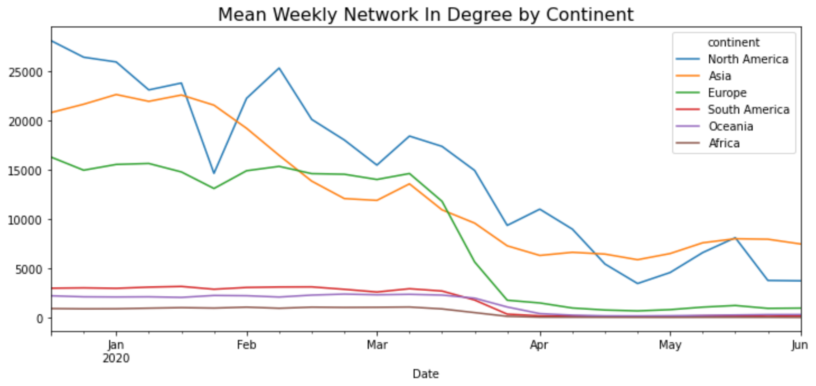

Figure 2: Weekly in degree by continent

Percentage change for in degree since 22nd January 2020

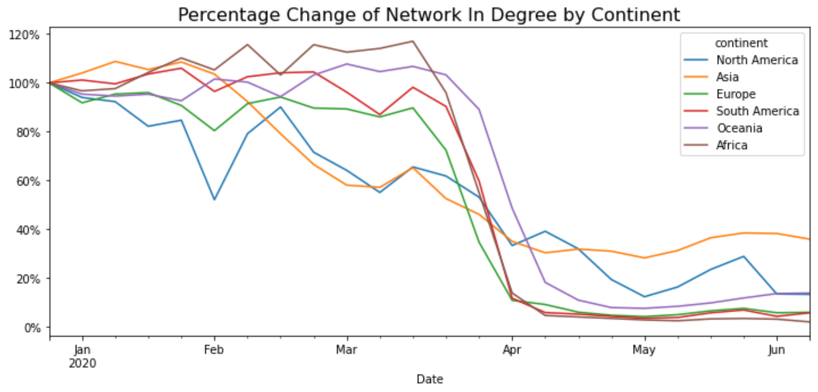

Figure 3: Percentage change of in degree by continent

Degree Distributions

Degree Distribution

Examining the degree distributions of the network over time reveals that the degree of nodes has a heavy tailed distribution.

We posit that the growth in the airline industry and the preferential attachment to pre-existing airlines and airports creates hubs of transit activity.

Preferential attachment can come about from either consumer preference or business and financial constraints.

Degree Distribution

Structure of heavy tailed distribution in the airline flight network holds even as the number of flights (degree) decreases by approximately 80% worldwide over time.

COVID-19 Case Data

COVID-19 Case Data

- COVID-19 case data was sourced from the John Hopkins University GitHub
- Data tracking starts from 22nd January 2020
- Information is at a country level scale, with state granularity for US, Australia, and province granularity for China

COVID-19 Case Data

We show a static image of COVID-19 cases over time by continent (instead of by country) to reduce visual clutter.

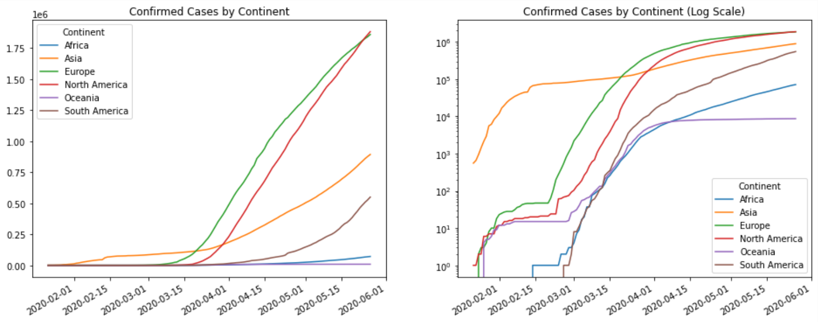

Figure 4: COVID-19 cases worldwide by continent

Flight Network Model and COVID-19

Flight Network Model and COVID-19

We have highlighted only cases from the countries and states that have the highest in degree at the start of the time series.

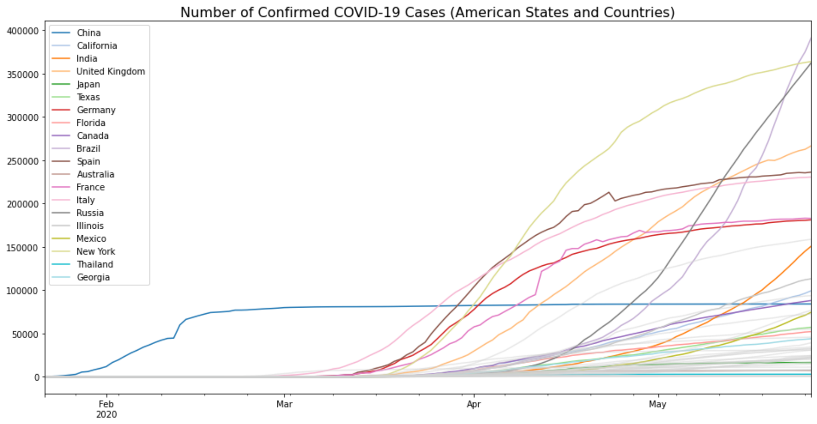

Figure 5: COVID-19 cases worldwide by continent

COVID-19 Animations

Flight Network Model and COVID-19

We can clearly see that the highlighted focus regions with high inbound connectivity are the first (and often substantially severe) cases.

Time evolution of cases also shows it would be unreasonable to expect the virus to spread via any other medium (i.e. ocean travel, etc).

Flight Network Model and COVID-19

Other topic areas for future study of the data:

- More detailed analysis on outbound flights from China around Chinese New Year
- Can we quantifiably measure a negative (cross)-correlation between number of cases and outbound flights?
- Include more details on demographics for each country - i.e. population, flight traffic per capita (we might imagine some countries have bigger expat communities/tourist communities etc)
- Look at network topology in a context separate to COVID19 - Could look at characteristic path length (flight time) and so on. Lots of information in the data
- Oher papers have also studied the worldwide airline network and have found small world properties. They have also done clustering analysis that finds hubs and peripheries in the network [1][2].

Thank you ☺

FirstName LastName.
Revealing the structure of the world airline network.
https://www.nature.com/articles/srep05638, 0000.

The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles.
https://www.pnas.org/content/102/22/7794.