SMM638 Network Analytics Mid-term project

Network Flights in Brazil Before, At, and After Covid Pandemic

Group.10
Maggie Yaw: Maggie.Yaw@bayes.city.ac.uk
Tian Sun: Tian.Sun@bayes.city.ac.uk
Yuyu Jiang: Yuyu.Jiang@bayes.city.ac.uk
Kuan-Ting Liu: Kuan-Ting.Liu@bayes.city.ac.uk

Network Flights in Brazil Before, At, and After Covid Pandemic

Overall of the network(social relevance of the relationship), the network plot

Due to the covid pandemic impact, the number of flight passengers in Brazil hugely dropped by 82% from 2020 April with a year-on-year basis at the same time. We are therefore curious whether the Network of flights in Brazil would have any shape of changing about its structure, distribution, or core airports so on and so forth. We use the dataset below to plot the network out in two different ways: the dynamic geographic map network and the networkX Graph to detect core-periphery and community structure. The plot shows before, at, and after the pandemic of network differences. GIF will be provided in the compressed folder.

Two datasets are used in our flight network analysis.

1. Compilation of all flights in Brazil from Jan 2019 to Jun 2021 from Kaggle.
2. The Global Airport Database from partow.net

The first dataset of flights has some missing values on Jul 2020, Sep 2020, and Jan 2021. In order not to output invalid analysis, when separating the time trend to see the pandemic influence, we only chose the month from Apr to Jun and compared from 2019 to 2021. The reason is that Apr 2020 is the most serious flight cutting down in Brazil. We could also avoid the problems of slack and peak season if we choose the same months but with different years, and avoid the missing values from the dataset. In order to have a better understanding of network structure and not be over complex, we excluded line_type = International to see only the flights behaviour in Brazil. One thing to note is airport ICAO code LEPA, UUEE, USSS, UIUU were excluded from our analysis since the four airports merely connect to each other without other regions of the airports.

In the second dataset, we only took out the ICAO Code, Latitude, and Longitude and then merge into the flights in Brazil dataset. Why we need latitude and longitude is to plot it on the geographic map and clearly see the core-periphery structure of the network. However, we did face that some of the small airports are not recorded in the documents. The final output number of recorded airports is around 80, depending on time.

the lexicon of cleaned dataset:

Variable	Description
line_type	"N"=National, "I"=International, we only estimate national flights
origin	ICAO code of The Global Airport Database
destination	ICAO code of The Global Airport Database
real_arr	datetime64, take the realized flights without NA
situation	REALIZADO', 'CANCELADO', 'NÃO INFORMADO'. We take the realized values
origin longitude	longitude from The Global Airport Database to plot geography
origin latitude	latitude from The Global Airport Database to plot geography
origin country	we considered only "BRAZIL"
des longitude	longitude from The Global Airport Database to plot geography
des latitude	latitude from The Global Airport Database to plot geography
des country	we considered only "BRAZIL"

We define the flights as edges regardless of direction. Though In the real world, flights are directed (e.g., A to B is different from B to A); Airports as nodes; National flights as one mode. The weight is equal to the total sum of flights but we only use weighted edges as adjacency matrix plotting to see a trend. Usually, the flight network would look like a scale-free network. A huge percentage of the airports are small and only a few are the transportation hub in a country. However, especially when at covid pandemic we would see the hub or periphery changing apparently. The sections below display an analysis of time-period comparing network analysis.

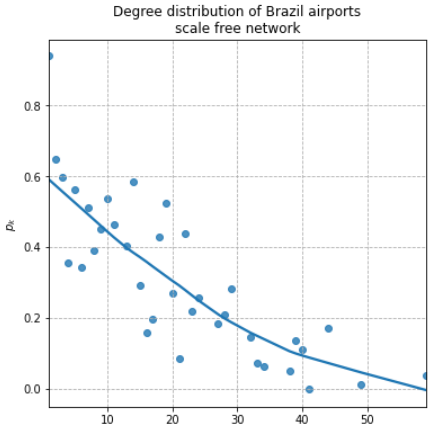
We extract information from the networkx, to see plot shown on the left:
(i) Total sum of national flights in Brazil per month, as known as total weight of edges
(ii) Total number of airports open in Brazil per month, as known as total nodes
(ii) Total degree of airports in Brazil per month, as known as unweight edges

One thing to note is that we could use the plot (iii) numbers dividing plot (ii) to get the average degree per month in Brazil.

We could easily inspect that especially at Apr 2020 to Jun 2020, total flights, open airports, and flights types decreased with a huge number. At Jul 2020 to Jun 2021 is the recovering period to gradually be back to the original structure in Brazil.

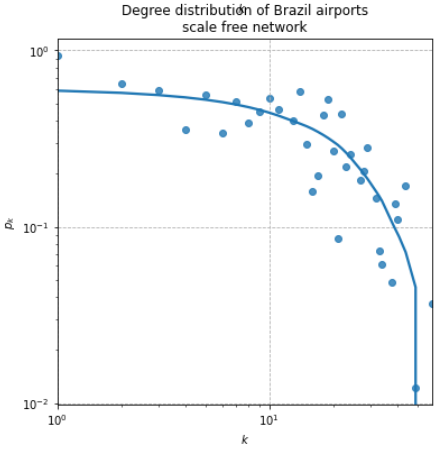
Degree distribution of the network

The degree distribution of all time from 2019 to 2021 is shown right, which is close to a power-law plot following a straight line. We should see some outlier above Pk > 0.6 and that is, makes the line to be a bit non-linear: the grey area between power-law and Poisson distribution. We then estimate another plot by handling the logarithmic axis, the log-log plot, indicating the probability of observing high-degree core nodes, or hub, is a magnitude higher in the network. We could draw the conclusion of the Network of flights in Brazil is scale-free rather than a random one.

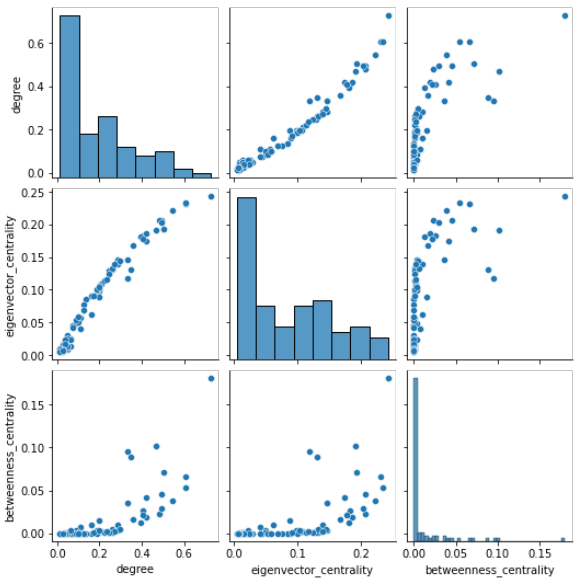


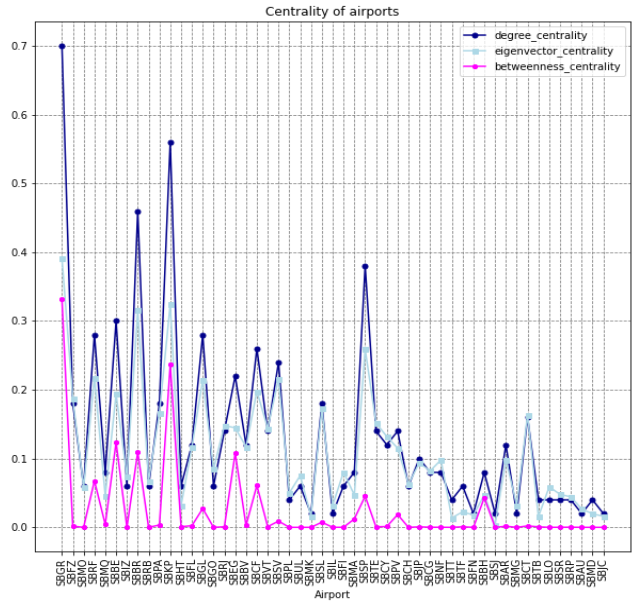
Descriptive statistics about relevant node-level properties

One insight worth mentioning is that the distribution of betweenness centrality in the Network of flights in Brazil of all time from 2019 to 2021 as plot left below. Betweenness centrality is a way to measure the intermediary effect. We could see that most of the betweenness centrality distribution is with a very low value. This indicates that in Brazil majority of flights are direct flights, and there's no need to transfer within a lot of different airports. In other words, the shortest length paths are really short.



Furthermore, let us look at the eigenvector centrality distribution, a measure of the influence of nodes. If one node is connected to other nodes with high degrees, the scoring would be high. This indicates that if core-periphery structure exists in the network, periphery nodes will remain to have low scores and core nodes will have relevant high scores. To be more specific, the distribution of degree and eigenvector centrality would look the same as we could see in the graph or to inspect on the right below graph.





We would like to check three time periods:(i) Before pandemic Apr to Jun 2019, (ii) At pandemic Apr to Jun 2020, (iii) After pandemic Apr to Jun 2021, from right to left of the below graph.

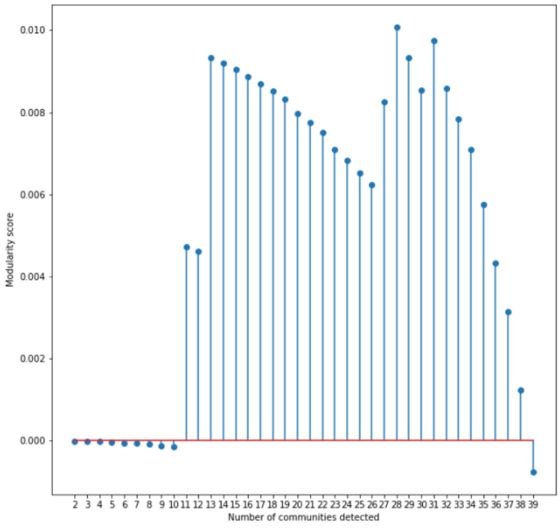
Before and After pandemic distributions look very familiar. However, At the pandemic period shown in the middle plot, we could see the betweenness centrality distribution with a low score is thicker. Theoretically thinking, we believe that some nodes from the periphery were not open during the pandemic period. As for eigenvector centrality, the distribution is not as skewed as before and after the pandemic, indicating that even though core-periphery structure exists, it is not as strong, or a huge degree gap, as before and after the pandemic period.

Community modular structure and performance changing

Consider our airport network at three stages which are before, during, and post the covid pandemic, we use modularity to decide which of the many partitions predicted by the hierarchical method offers the best community structure by choosing the one for which M is maximal. One of the key properties of modularity is that the higher is the modularity for a partition, the better is the corresponding community structure.

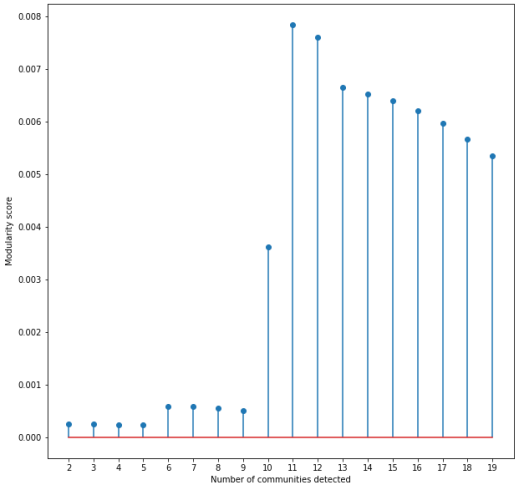
(A)Before-Pandemic, Apr to Jun, 2019

Before the pandemic, the airport network comprises 76 nodes and 398 edges with a low modularity score being detected within the range from -0.001 to 0.01. This implies that the network has dense connections between the nodes with modules as well as dense connections between nodes in different modules. When air travel hasn't been affected, both international and regional airports of Brazil operate as normal under a situation whereby the disruption to the flight frequencies of each individual airport is absent. There are strong connections between within communities and outside communities of the network with 28 communities being detected at partition.

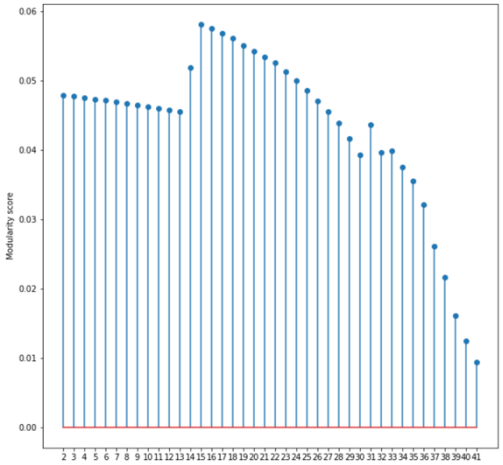


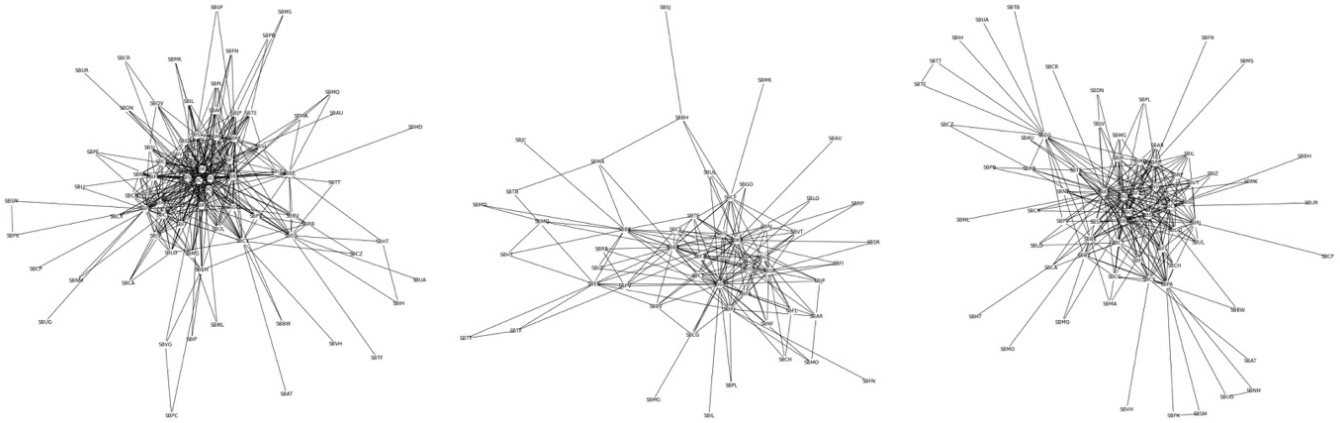
(B)At Pandemic, Apr to Jun, 2020

During the pandemic, the airport network now comprises 51 nodes and 172 edges which have substantially decreased as compared to the situation before the pandemic resulting from the disruption to the airline industry. As part of the covid containment measures of the Brazilian government, Brazil had announced the short-term closures of airports to certain high-risk countries. It is supported by an official announcement to prevent foreigners from temporarily entering the country from its airports on 27th March 2020. Modularity score has decreased at nearly 0.008 when compared to the score at pre-pandemic. The network does not have dense connections between the nodes within modules but sparse connections between nodes in different modules resulting from the temporary shutdown of airports.



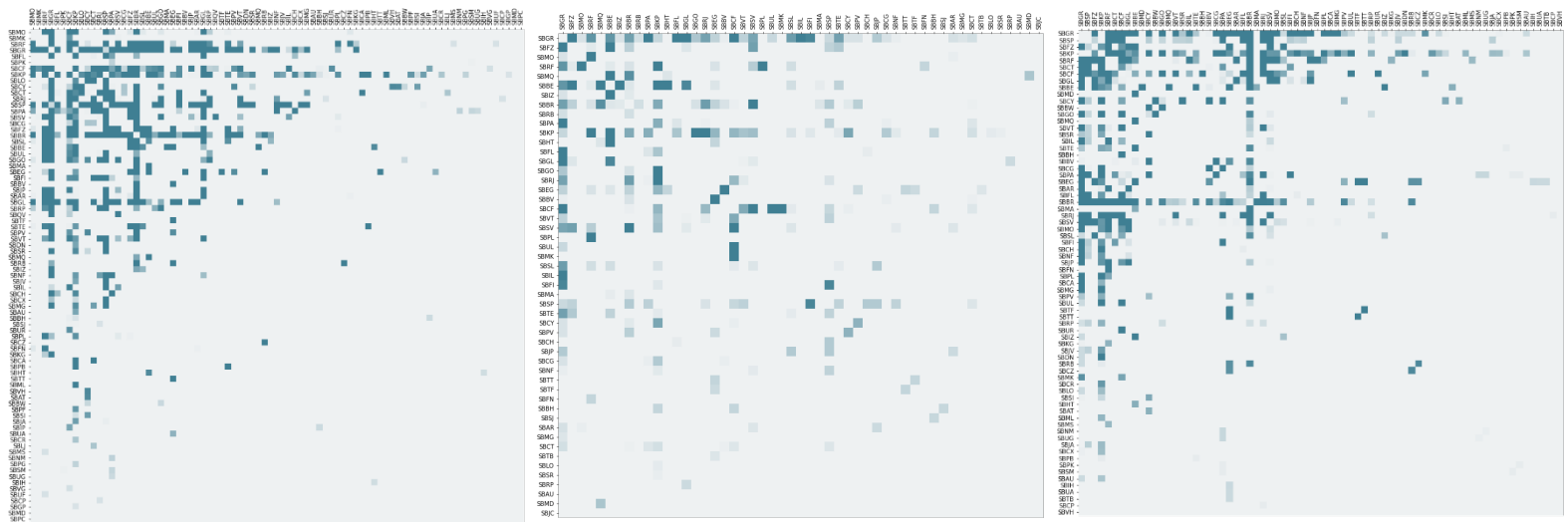
(C)After Pandemic, Apr to Jun, 2021
While nodes and edges remain unchanged post-pandemic, with 69 nodes and 268 edges.
As part of cost-saving measures, operation or airports has since adjusted to the mode as during the pandemic. With the implementation of travel restrictions, the operation mode of airports is likely to continue even after covid.
The modularity score is close to that during pandemic with a peak at nearly 0.06. The network has dense connections between the nodes within modules but sparse connections between nodes in different modules.





The above network plots show (A), (B), (C) Network structures from left to right, which (B) periphery airports shutdown indicates that only core structure is left in the network.

Core-periphery structure and performance changing



After Modularity score and community detection, we now see a clear picture of Network Flights in Brazil is a core-periphery structure network regardless of time period. Three plots shown above are the adjacency matrix before, at, and after the covid pandemic. The heavy of colour represents the weight of edges is heavy. The plots are a little hard to recognize since

nodes are too many and we'll get different from the networkx. Thus, x and y tick would not always be the same. However, we could see trend that:

(i) Before pandemic the core nodes are quite strong.
(ii) At pandemic the core-periphery is not so different, and do explain why the modularity score is lower and community detection with merely 10.
(iii) After pandemic the network seems to reform another structure with two cores.

Top 20 Degrees of Core National Airports in Brazil

Before Pandemic 04/2019~06/2019			At Pandemic 04/2020~06/2020			After Pandemic 04/2021~06/2021		
index	airport	degree	index	airport	degree	index	airport	degree
101	SBMA	47	201	SBSV	35	301	SBVT	41
102	SBUL	43	103	SBSP	28	302	SBSP	40
103	SBSP	42	203	SBBE	23	303	SBEG	31
104	SBCY	35	204	SBCT	19	304	SBPL	28
105	SBLJ	33	205	SBMO	15	305	SBJV	25
106	SBJP	30	206	SBCF	14	306	SBTT	21
107	SBFN	29	207	SBTF	14	307	SBNM	21
108	SBUR	27	208	SBEG	13	308	SBCX	19
109	SBGL	26	101	SBMA	12	309	SBMD	18
110	SBPF	23	210	SBHT	11	310	SBAT	17
307	SBNM	23	309	SBMD	9	311	SBPV	16
112	SBBV	21	212	SBBH	9	312	SBNF	16
113	SBBW	20	118	SBKP	9	106	SBJP	15
114	SBFL	18	214	SBSL	8	314	SBPA	14
308	SBCX	18	105	SBJP	7	113	SBBW	13
116	SBFZ	18	107	SBFN	7	204	SBCT	13
117	SBRP	17	217	SBTB	7	317	SBCA	9
118	SBKP	16	218	SBJC	7	120	SBTE	9
205	SBMO	15	116	SBFZ	6	319	SBMS	8
120	SBTE	14	114	SBFL	6	117	SBRP	8

In order to have better understanding of core-periphery airports changing in Brazil, we preform the columns above. We could easily see:

(i) Airport SBMD is originally from non-core position to climb up to the top 10 of core airports in Brazil.
(ii) Airport SBSP remains the same as core airport.
(iii) Airport SBEG is firstly outside the top 20 airport with largest degree but eventually rank at third highest degree airport in Brazil.

The trend indicates above shows that Flights Network in Brazil after covid pandemic is actually producing and transformed into different hubs based on core-periphery structure.

Summary

1. We could easily inspect that especially at Apr 2020 to Jun 2020, total flights, open airports, and flights types decreased with a huge number. At Jul 2020 to Jun 2021 is the recovering period to gradually be back to the original structure in Brazil.
2. Network of flights in Brazil is scale-free rather than a random one.
3. most of the betweenness centrality distribution is with a very low value. This indicates that in Brazil majority of flights are direct flights, and there's no need to transfer within a lot of different airports. In other words, the shortest length paths are really short.
4. At the pandemic period shown in the middle plot, we could see the betweenness centrality distribution with a low score is thicker. Theoretically thinking, we believe that some nodes from the periphery were not open during the pandemic period.
5. At pandemic, modularity score has decreased at nearly 0.008 when compared to the score at pre-pandemic. The network does not have dense connections between the nodes within modules.

6. Adjacency matrix after pandemic the network seems to reform another structure with two cores.
7. Flights Network in Brazil after covid pandemic is actually producing and transformed into different hubs based on core-periphery structure.

Reference and original dataset

1. Compilation of all flights in Brazil: https://www.kaggle.com/ramiromelo/flights-in-brazil-compilation-jun2019-jun2021
2. The Global Airport Database: http://www.partow.net/miscellaneous/airportdatabase/
3. Brazil flights impact news: https://www.flyflapper.com/stories/the-impact-of-coronavirus-outbreak-on-aviation-sector-in-brazil/
4. Brazil closes airports news:https://labsnews.com/en/news/travel/brazil-airports-covid-19/

Group Coursework Submission Form

Specialist Masters Programme

Please list all names of group members: (Surname, first name) 1. Maggie Yaw 2. Tian Sun 3. Yuyu Jiang 4. Kuan-Ting Liu	5. 6. 7. **GROUP NUMBE** **10**
MSc in: MSc Business Analytics (2021/22)	
Module Code: SMM638	
Module Title: Network Analytics (PRD1 A 2021/22)	
Lecturer: Dr. Simone Santoni — simone.santoni.1@city.ac.uk	**Submission Date:** November 15, 2021

Declaration:

By submitting this work, we declare that this work is entirely our own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the coursework instructions and any other relevant programme and module documentation. In submitting this work we acknowledge that we have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. We also acknowledge that this work will be subject to a variety of checks for academic misconduct.

We acknowledge that work submitted late without a granted extension will be subject to penalties, as outlined in the Programme Handbook. Penalties will be applied for a maximum of five days lateness, after which a mark of zero will be awarded.

Marker's Comments (if not being marked on-line):

Deduction for Late Submission: **Final Mark:** %