Investigating air travel network changes in Canada, USA and Europe during COVID-19 using open source data

Adriana-Stefania Ciupeanu
University of Manitoba
Ashani Wickramasinghe
University of Manitoba
Saman Muthukumarana
University of Manitoba
Julien Arino (julien.arino@umanitoba.ca)
University of Manitoba

Research Article

Keywords: Open source data, network analysis, air transportation

Posted Date: January 3rd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2426208/v1

Additional Declarations: No competing interests reported.

Investigating air travel network changes in Canada, USA and Europe during COVID-19 using open source data

Adriana-Stefania Ciupeanu[a,b], Ashani Wickramasinghe[b], Saman Muthukumarana[b,c], Julien Arino[a,c,*]

[a]*Department of Mathematics, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada*
[b]*Department of Statistics, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada*
[c]*Data Science NEXUS, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada*

Abstract

We consider the evolution of the global air transportation network during the COVID-19 pandemic, using publicly available data from the OpenSky Network.

Keywords: Open source data, network analysis, air transportation.

1. Introduction

The COVID-19 pandemic, caused by infection with the SARS-CoV-2 virus, has become one of the most severe and deadly pandemics in recent history. By December 2022, three years after its first known outbreak in December 2019, the World Health Organisation (WHO) reported over 6.6 million COVID-19 deaths and 649 million confirmed COVID-19 cases [39], while the total number of people infected with and having died from COVID-19 is believed to be much greater. It may take many years to fully ascertain the health burden and socioeconomic impact of the pandemic. The COVID-19 pandemic has involved multiple waves (see Figure 1). These waves were

*Corresponding author
Email addresses: ciupeana@myumanitoba.ca (Adriana-Stefania Ciupeanu), Ashani.Wickramasinghe@umanitoba.ca (Ashani Wickramasinghe), Saman.Muthukumarana@umanitoba.ca (Saman Muthukumarana), julien.arino@umanitoba.ca (Julien Arino)

mainly due to the emergence of new variants of concern (VOCs) and the implementation and relaxation of non-pharmaceutical interventions.

An important teaching of the COVID-19 pandemic is twofold. Firstly, warnings by public health authorities worldwide, in particular since the beginning of the 21st Century, about the inevitability of a pandemic were founded. Several events of lesser importance such as the SARS-CoV-1 epidemic of 2003, the H1N1 influenza pandemic of 2009 or resurgent but localised outbreaks of Ebola or MERS somewhat dulled the public and political perception of these risks, but COVID-19 has proved that they are real. Secondly, such events are bound to repeat because of two main factors: the ever increasing interactions between animal reservoirs of pathogens due to encroaching human settlements and range changes due to climate change; and the unprecedented level of human movement at all ranges.

In view of these considerations, quantifying and understanding the impact that the COVID-19 pandemic had on different industries plays an important role in refining containment measures. Across all industries, the aviation sector was probably among the hardest hit, at least during the first year and a half of the pandemic. As a major contributor to the globalised spread, the aviation industry also had a major role to play in efforts to curtail the spread of infection.

Unprecedented country-wide flight bans led to a dramatic change in travel numbers, destinations and flight patterns. While air mobility has allowed more passengers to fly to more and remote destinations within a few hours at affordable prices and cargo shipments to more and more destinations, the aviation sector plays an important role in the spread of diseases, by helping local epidemics to turn into global epidemics [18]. This phenomenon has been observed already for several infectious diseases, such as, Ebola [17], Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS) [40] and influenza [27]. For earlier diseases their impact could mostly be reduced to the regional level and public health authorities were able to cut transmissions of the disease at early stages, which prevented a full pandemic. COVID-19, on the other hand, had a different outcome. Many countries imposed border control measures aimed at reducing the risk of COVID-19 importation. A variety of approaches have been used globally to contain the importation of COVID-19.

As mentioned, COVID-19 has driven home, if that was needed, the point that the issue is not whether a pandemic will happen in the future but to understand when it will occur and what form it will take. As a consequence,

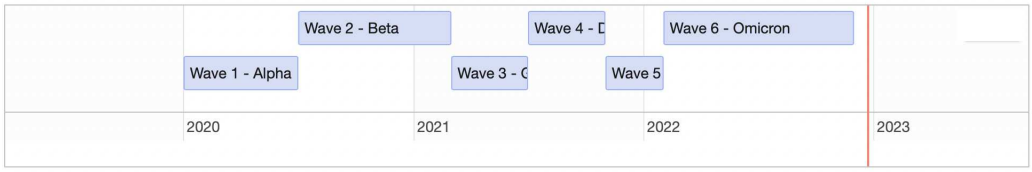


Figure 1: COVID-19 waves in Canada.

planning for the next pandemic is an extremely high priority for any government. Less obvious is the fact that, in the well-connected world of the 21st century, no country is isolated from the potential spread of infection. Thus, there is a pressing need to study the global spread of COVID-19 to understand the impact of a global epidemic. Even though travel restrictions do little to directly reduce the burden of the pandemic, they can buy time to develop a public health response to the pandemic. Travel restrictions can directly decrease the influx of new infected persons into an area. More importantly, restrictions reduce the probability of an infected individual leaving the area in which an outbreak is developing.

Aims and organisation of the paper

Our first aim is to develop a methodology for dealing with an open source dataset of flight information. The type of problems we consider here have typically been addressed using costly datasets such as those obtained from IATA, OAG or other commercial companies such as FlightAware or Flightradar24. While these datasets are more exhaustive, they have the inconvenient that they are expensive and *closed*, in that they can be used in research under restrictive conditions. By opposition, the dataset used here is openly accessible, but requires a certain amount of preprocessing and provides data that is incomplete and is made available monthly (and can even be obtained in quasi-real time). This provides an alternative to the work carried out by [15] using OAG data.

In this context, we provide a methodology for preprocessing this data in the perspective of obtaining network level information, and investigate the shortcomings of the resulting data compared to more extensive paying closed dataset.

The second aim of this paper is to develop a methodology for dealing with temporality in networks, in order to make use of the resulting data to

evaluate the real effect of measures like flight bans. This methodology could, with minimal adaptation, be used on other datasets of the same nature. In the well-connected world of the 21st century, no country is isolated from the potential spread of infection. Travel restrictions can directly decrease the influx of new infected persons into an area. More importantly, they reduce the probability of an infected individual leaving the area in which an outbreak is developing. However, quantifying the precise effect of these measures is complicated. Motivated by this observation, we investigated what impact travel restrictions in US, Canada and Europe had on the spread of COVID-19 in these regions. We collected data about travel restrictions, flights between US, Canada and Europe and COVID-19 cases. Using this data we studied the changes in the flight network and tried to infer the potential effect travel restrictions had on the spread of the disease.

2. Description of the data

The flight data comes from Automatic Dependent Surveillance–Broadcast (ADS-B) data. ADS-B data is transmitted automatically by aircrafts and contains information about their position and identification. The data can be received by simple ground-based receivers and has led to the development of a community, including many members of the general public, who receive and share the data they receive through instances like the OpenSky Network [36].

As well as more subtle limitations to the data, which we address later, there are structural limitations to the data that are worth mentioning at this point. ADS-B, as a system, is being progressively mandated by law in different national and transnational jurisdictions. As a consequence, data coverage is good in jurisdictions that require or will soon require aircrafts to be equiped, e.g., Canada [26] and the USA [11], Europe[21], but is much more patchy in other locations. If an aircraft only flies between countries where ADS-B equipment is not mandatory, for instance, it will most likely not be equiped with the equipment and is therefore absent from the database.

The datasets we use were generated from the OpenSky Network data by the authors of [32]. That data is updated monthly and made available on Zenodo; it covers flight information starting in January 2019 and ongoing at the time of writing (December 2022). Note that there are two versions of the data distributed by the authors: a version under Creative Common license that has some fields anonymised as well as one covered by the OpenSky

Network license, usable freely for research purposes but otherwise limited. It is the latter version we use here.

For this study we concentrated on flight data covering 2019, 2020 and 2021. For each month, the data consists in a csv file; see Table 2 for a sample. The file has the columns in Table 1.

Variable	Meaning
callsign	Flight identifier
number*	Commercial number of the flight
icao24	Transponder unique identification number
registration*	Aircraft tail number
typecode*	Aircraft type
origin*	ICAO code for the origin airport
destination*	ICAO code for the destination airport
firstseen	UTC timestamp of the first message received by OSN
lastseen	UTC timestamp of the last message received by OSN
day	UTC day of the last message received by OSN
latitude_1	First detected position of the aircraft
longitude_1	First detected position of the aircraft
altitude_1	First detected position of the aircraft
latitude_2	Last detected position of the aircraft
longitude_2	Last detected position of the aircraft
altitude_2	Last detected position of the aircraft

Table 1: Variables in the data and their meaning. Starred variables, e.g., origin*, can be empty. OSN: OpenSky Network.

The dataset has some limitation, because the origin and destination are computed using ADS-B (Automatic Dependent Surveillance–Broadcast) trajectories (see for a [14]) on approach and takeoff and they are empty when no airport can be found. Furthermore, no crosschecking with external sources of data has been conducted. The aircraft information comes from OpenSky dataset and the fields typecode and registration are empty when the aircraft is not present in the OpenSky dataset. Because not every flight has an aircraft type, we can only get lower and upper bounds for volume.

callsign	icao24	registration	typecode	origin	destination
HVN19	888152			YMML	LFPG
CES219	780b7e	B-5936	A332	YSSY	EDDF
TGW700	76bcca	9V-OFJ	B788		RJBB
CSN609	781364				KLAX
SVA840	710411			WMKK	WMKK
LAN600	e8027b	CC-BBG	B788	SKBO	KLAX
HVN55	8880f8	VN-A868	B789	YSSY	EGLL
AAR551	71bf94	HL7794	A333		LTBA
CPA343	789202	B-LRU	A359	YMML	EGKK
AAL126P	a999d2	N718AN	B77W	KLAX	KDFW
LAN706	e80450	CC-BGJ	B789	KJFK	LEMD
CCA985	780cb8	B-2487	B748		KSFO

Table 2: Sample rows in the dataset. Flight number (usually a very small variation on the callsign), location information (latitude, longitude and altitude) as well as date and time are omitted.

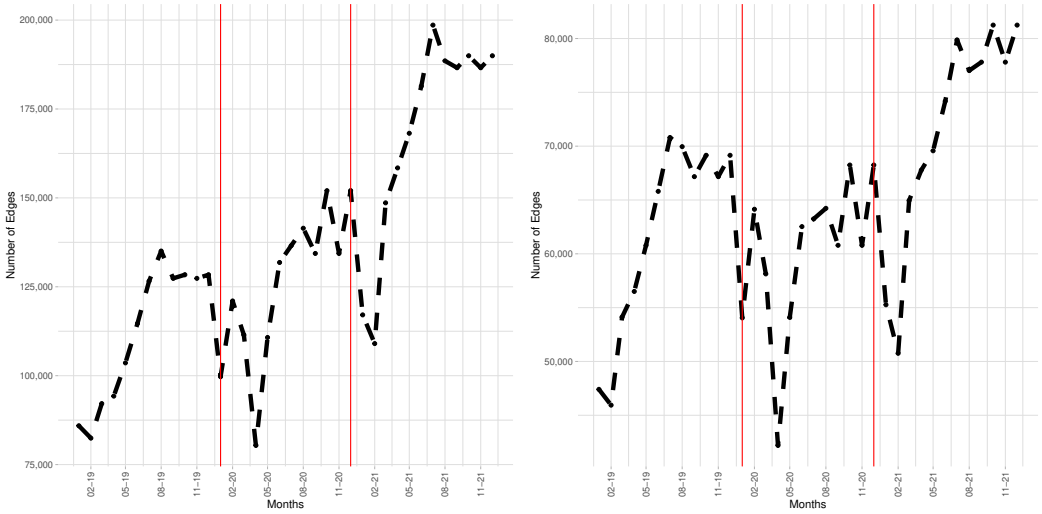
3. Data cleaning and preprocessing

3.1. Data cleaning

The data as posted by the authors of [32] is already cleaned to a large extent, so the cleaning steps are quite limited.

1. Select rows in which both origin and destination are non-empty. Cross-linking with external sources should be possible in some instances, given for instance the tail number or the callsign and flight number, but this is an entirely different project and was not undertaken here.
2. Exclude rows in which the origin and destination airports are identical. These correspond often to leisure personal flights, mostly in the USA. These flights have no consequence for the global spread of infectious diseases and also have no impact on the overall dynamics of the network, since they are not transport flights.

To illustrate the effect of these initial cleaning steps, let us illustrate with the data from January 2019. The data initially had 2,660,901 rows. Of these, 1,341,646 rows were excluded in the cleaning steps because they had either



(a) Number of edges prior to the second preprocessing step.

(b) Number of edges after the second preprocessing step.

unknown origin or destination or had same origin and destination. The rest of the months are similar; see Table D.5.

3.2. Data preprocessing

Once the data has been cleaned, we submit it to preprocessing steps.

1. We add country, continent and country region information for each flight for both the origin and destination, using the data in [33].
2. Using an aircraft capacity dataset [1], we add information about the flight capacities. This provides upper bounds for the number of passengers on each flight. For flights that do not have an entry for the aircraft type, we assign a volume of 2 passengers, the reasoning being that many aircrafts in the database are small personal planes in the USA.
3. To be able to perform pairwise comparisons between the networks, we further process the data so that the list of airports is the same each month. This means that we remove airports that do not appear in all monthly datasets.

3.3. Specialising to the countries under consideration

Using the preprocessed data, we consider a subset of airports including airports in Canada, the United States of America and Europe. For the

exact list of countries used in Europe, see Appendix A. Note that for USA territories outside of territorial USA, only Puerto Rico is used; for European countries, only territories close to the continent are included.

4. Evolution of the network

The techniques used are detailed in Appendix C. We consider the evolution of the network from two different perspectives: changes to the characteristics of individual nodes and changes to the global topology of the network.

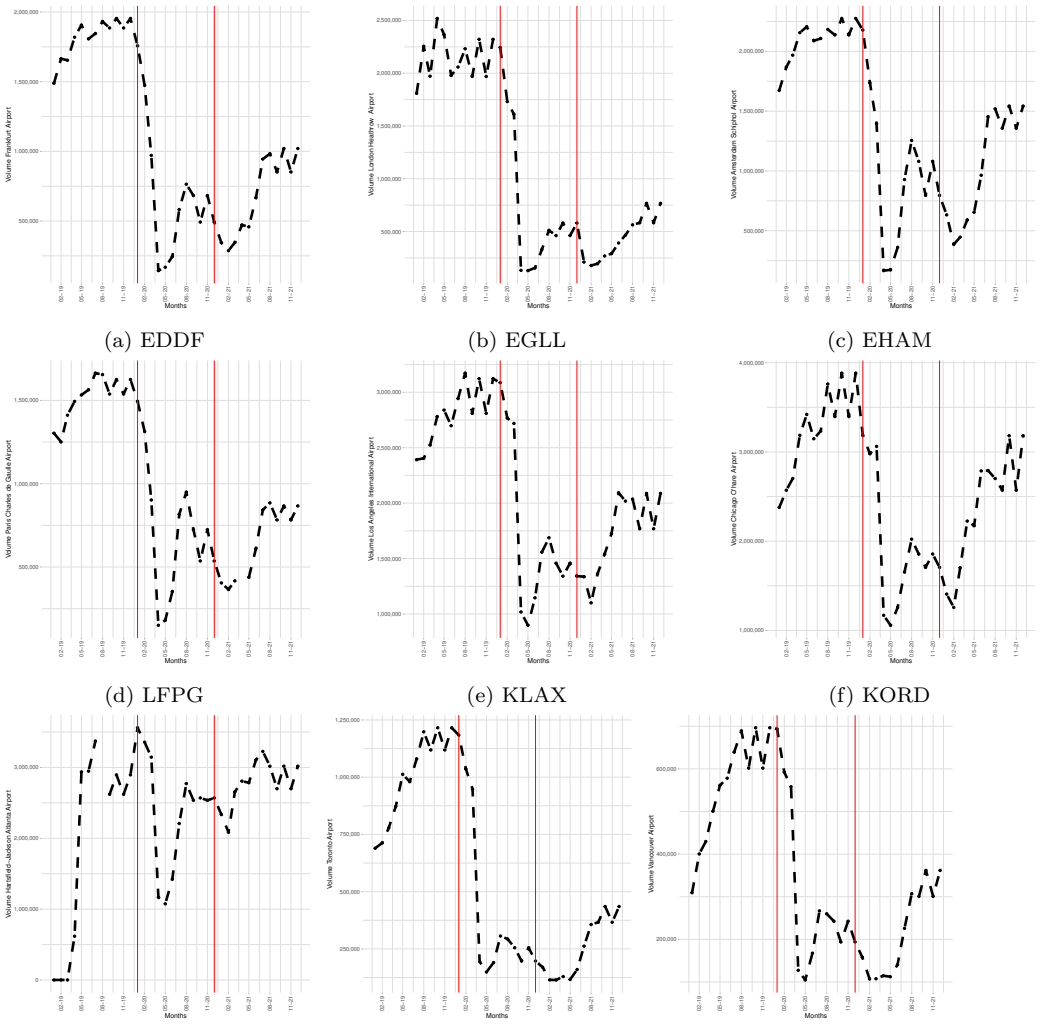
4.1. Airport-based evolution of the network

First, we consider nine major airports in Canada, the United States of America and Europe: Amsterdam Airport Schiphol (ICAO: EHAM), Hartsfield-Jackson Atlanta International (KATL), Chicago O'Hare International (KORD), Frankfurt (EDDF), Los Angeles International (KLAX), London Heathrow (EGLL), Paris Charles de Gaulle (LFPG), Toronto Pearson International (CYYZ) and Vancouver International (CYVR).

In Figure 3, we observe that the minimum travel volume was attained at different times, where the travel volume is the sum of the inbound and outbound volume. For European airports, April 2020 had the lowest number of passengers while for the airports in the United States of America and Canada, May 2020 had the smallest of volume. No airport had bounced back to their pre-pandemic volumes in the period of study (2020-2022). Moreover, Canadian airports had the slowest recovery out of the nine studied airports.

Hartsfield-Jackson Atlanta International airport is an airport which illustrates some of the limitations of the dataset. Because of the way the data is collected, KATL appears to have a volume of virtually zero for January, February and March 2019. Indeed, inspecting the data, KATL appears as origin once in Jan 2019, three times in February 2019 and seven times in March 2019. This is impossible since this airport is one of the biggest hubs in the USA.

Figure 4 shows the evolution of the in-degree centrality, betwenness centality and in-closness centralities for Chicago O'Hare and Toronto Pearson airports. The in-degree of centrality quantifies the number of non-stop pathways into a given airport from points of origin and is often interpreted as an indication of the popularity of a location. Cities with higher values have a greater number of pathways through which passengers and consequently infectious diseases may arrive. Recall that because of the preprocessing and



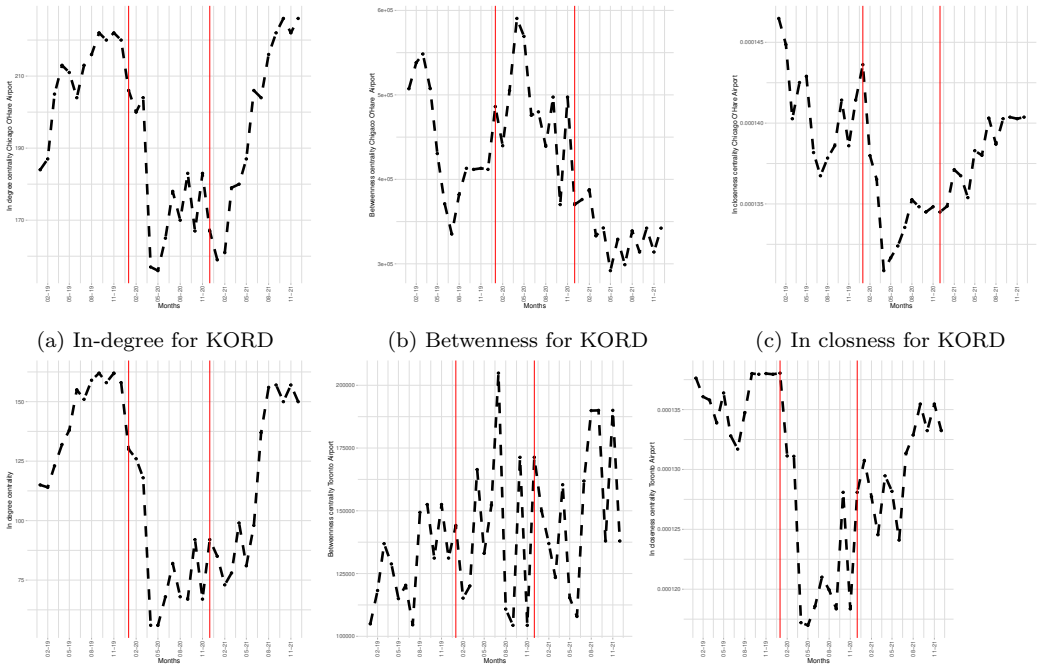
(a) EDDF (b) EGLL (c) EHAM

(d) LFPG (e) KLAX (f) KORD

(g) KATL (h) CYYZ (i) CYVR

Figure 3: Evolution of travel volumes at the nine airports selected for the airport-based analysis. Red lines separate the years.

specialisation to Canada, the US and Europe, the in-degree (as well as most other measures) is smaller than it would be were data about the whole network be available. Another important metric is betweenness centrality. This metric is a measure of how frequent an airport falls along the shortest path between other airports throughout the network. Betweenness centrality is important because it highlights the potential for airports to act as a channel



(a) In-degree for KORD (b) Betwenness for KORD (c) In closness for KORD

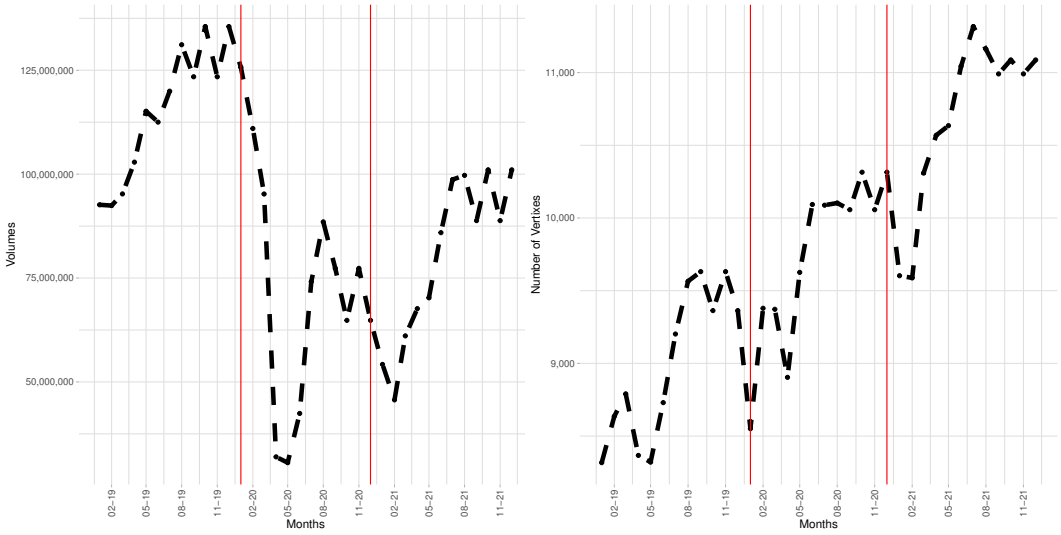
(d) In degree for CYYZ (e) Betwenness for CYYZ (f) In closness for CYYZ

Figure 4: Evolution of different centrality measures Chicago O'Hare and Toronto Pearson for the airport-based analysis. Red lines separate the years

for the spread of infectious diseases, since airports with high betweenness tend to be important transit points for passengers (and consequently infectious diseases) en route to their final destination. Finding airports with high betweenness centrality is important; this information can be used to pinpoint airports where non-pharmaceutical interventions (for example, surveillance, enhanced diagnostics for dangerous pathogens) are likely to disrupt the international and national spread of an infectious disease. Closeness is another centrality metric used, it represents the average length of shortest paths between a given airport and all others throughout the airline travel networks. These means that closeness centrality can be used to measure how accessible an airport is from other airports. Similarly to what is observed for travel volumes, there is a precipitous drop in connectivity. Here, however, the drop happens in all nine airports in April 2020. For betweenness centrality, we see that both KORD and CYYZ airports have an increase in centrality in April 2020.

4.2. Global network evolution

The aviation industry has suffered greatly, as seen from Figure 5a, which shows a dramatic decrease of the number of passengers in March 2020, not just because of a decline in transportation needs due to the pandemic but also because of the non-pharmaceutical interventions curtailing travel implemented by countries. Furthermore, as seen from 5b, there was also a sharp decrease in the number of active airports in March 2020. Note that that number is obtained by considering the number of airports prior to the second preprocessing step.



(a) Total number of travellers.

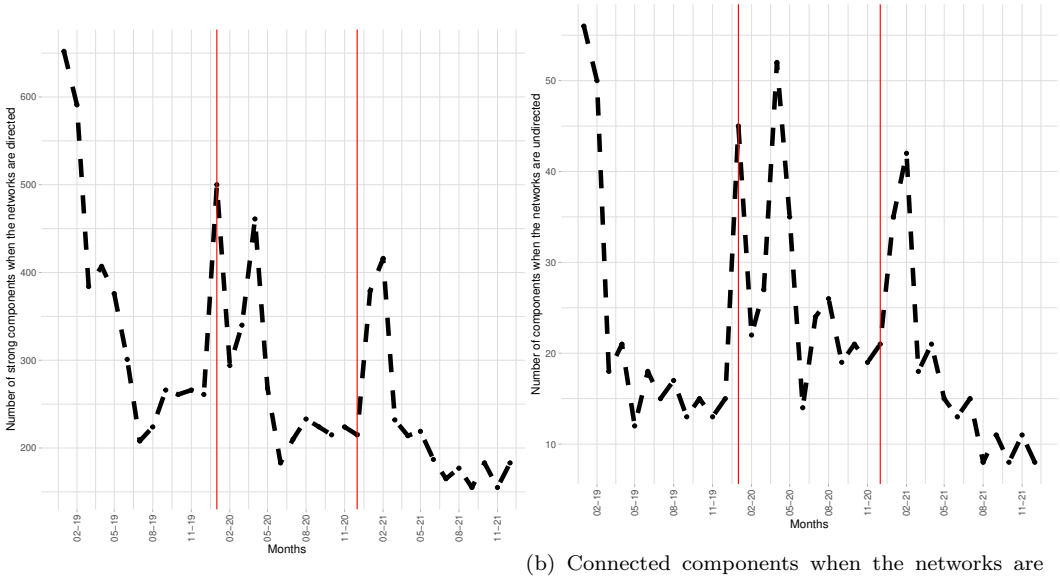
(b) Number of active airports.

Figure 5: Different changes in the network during 2019-2021

Note that the networks are not strongly connected, Figure 6a shows the number of strongly connected components for each month.

In social network analysis, a *community* is a subset of nodes within the graph that have a higher probability of being connected to each other than to the rest of the network. In the following, we use three algorithms, the so-called Louvain, Leiden and Infomap algorithms.

In Figure 7, we observe that based on the Louvain algorithm, the Leiden algorithm and the Infomap algorithm, the number of communities changes over time; the size of the maximum community also changes. Interestingly, the three algorithms detect smallest sizes of maximal communities at different



(a) Strong connected components

(b) Connected components when the networks are undirecetd

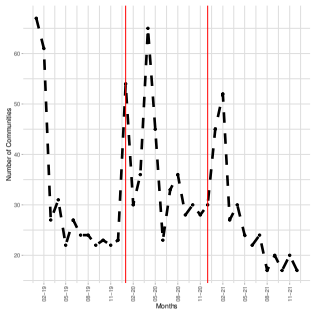
Figure 6: Number of connected components

times: April 2020 for Louvain, May 2020 for Leiden and August 2020 for Infomap.

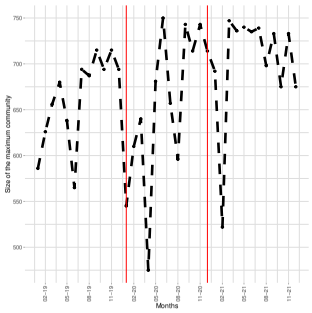
We then proceeded to a pairwise comparison of the results of some of the community detection algorithms; see Figure 8. In Figure 8a, we see the the pairwise comparison of the Louvain and Leiden algorithms using Variant of Information method and in Figure 8f we used Adjusted Rand Index (ARI). It is apparent that these two community detection methods provide different results. See Appendix Appendix C for a overview of the Variant of Information and ARI. Similar results where observed when comparing other community detection algorithms.

In Figure 9a, we can see United States airports (KORD is Chicago O'Hare airport and KADS is Addison Airport in Dallas, Texas) have the highest between centrality in 2020, furthermore, from 2019-2021 United States airports have the highest centrality. Table D.3 illustrates the 10 airports which have the highest betweeness centrality values for March 2019, April 2019-2021 and May 2019-2021. Values have been normalised between zero (lowest betweeness) and one (highest betweeness) to aid interpretation.

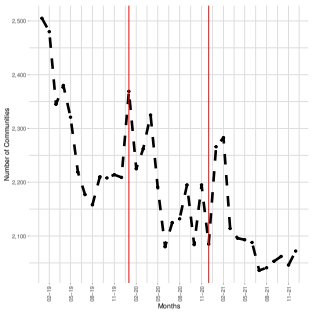
Note that centrality metric in this paper are based on the global airline flight networks, and they represent the potential for flow through the net-



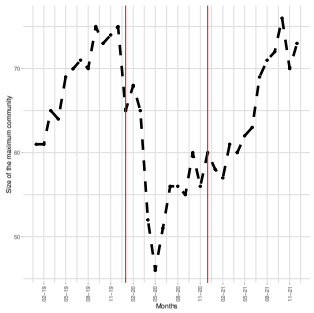
(a) Number of communities for Louvain detection algorithm



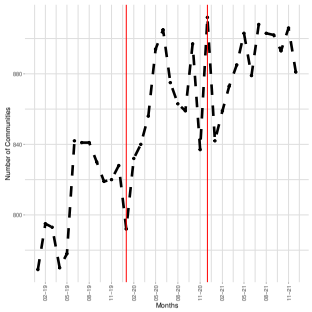
(b) Size of the maximum community for Louvain detection algorithm



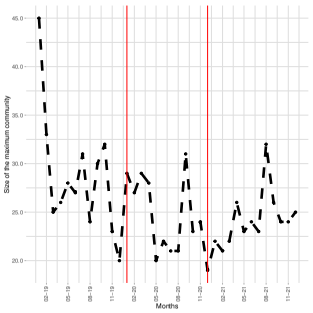
(c) Number of communities for Leiden detection algorithm



(d) Size of the maximum community for Leiden detection algorithm

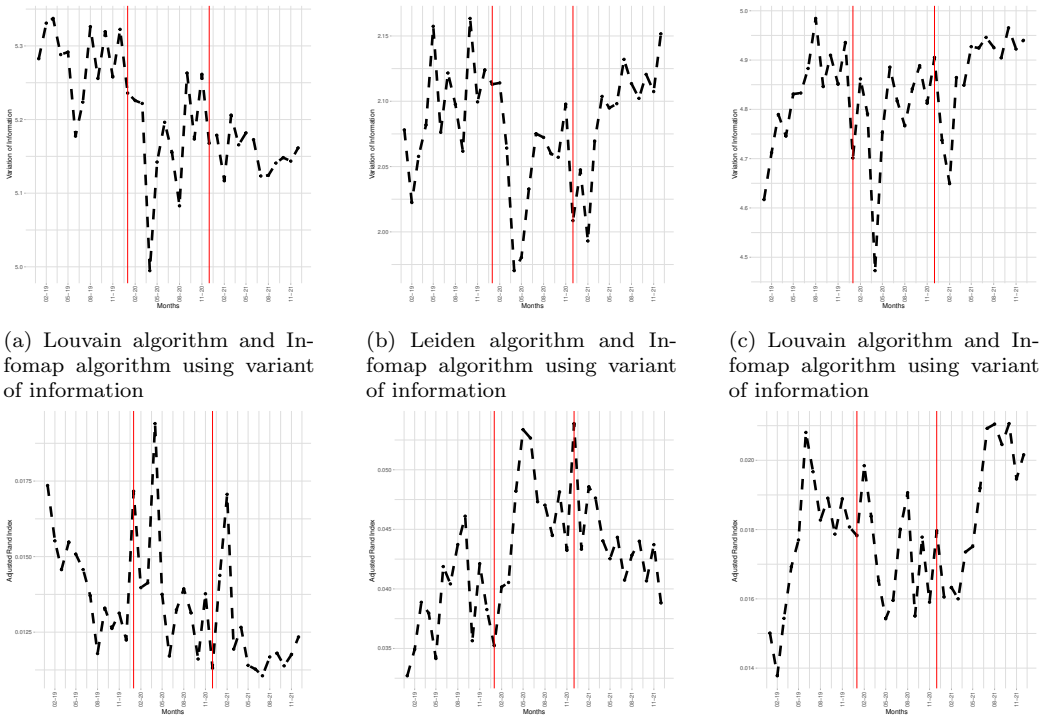


(e) Number of communities for Infomap detection algorithm



(f) Size of the maximum community for Infomap detection algorithm

Figure 7: Different metrics from community detection algorithms.



(a) Louvain algorithm and Infomap algorithm using variant of information

(b) Leiden algorithm and Infomap algorithm using variant of information

(c) Louvain algorithm and Infomap algorithm using variant of information

(d) Louvain algorithm and Infomap algorithm using adjusted Rand index

(e) Leiden algorithm and Infomap algorithm using adjusted Rand index

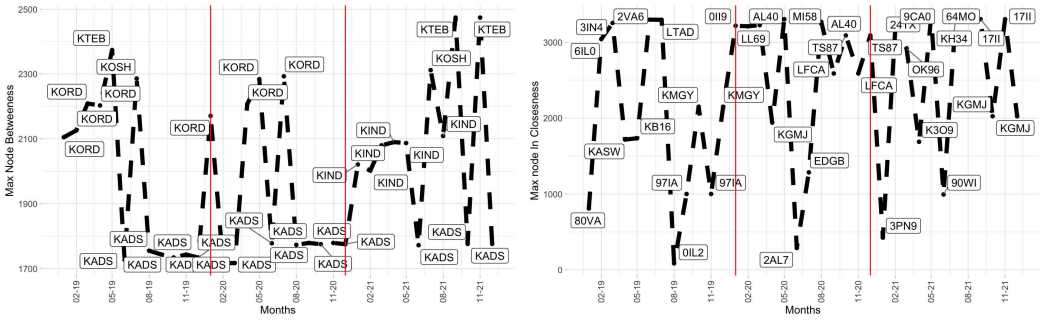
(f) Louvain algorithm and Leiden algorithm using adjusted Rand index

Figure 8: Pairwise comparison of different community detection algorithms using Variant of Information and Adjusted Rand Index.

works and although airports with high betweeness centrality are important transit points for passengers, in reality this may not be the case.

Note that, Figure 9 illustrate another limitation of the data, since private and small airports appear 23 times as having the highest in-closness centrality.

Figure 10a shows the the edge formation for 2019, 2020 and 2021, from this we can see that April 2020 has a significant drop of edge formations compared to April 2019 and April 2021. Note that the function used evaluates a network object at multiple time points and return counts of the number of edges forming (edge onset at time point). This function provides a descriptive stats about momentary rate of change in the network. Figure 10b shows the edge disolution for our networks, and again we can see that April 2020 has a significant drop of edge dissolution compared to April 2019 and April 2021.



(a) Betweenness. (b) Closeness.

Figure 9: Highest monthly centralities for the years 2019-2021.

Note that the function used evaluates a network object at multiple time points and return counts of the number of edges dissolving (edge onset at time point).

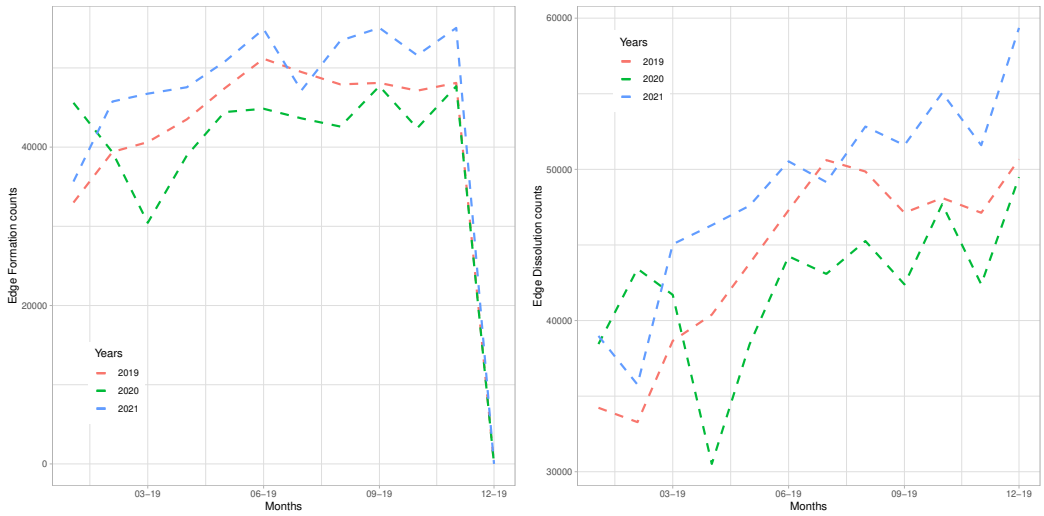
5. Discussion

In this paper, we had several aims. The first was to use of a publicly available dataset to answer questions about air travel that have typically been addressed using costly datasets such as those obtained from IATA. We provided a methodology for preprocessing this data in the perspective of obtaining network level information, and investigated the shortcomings of the resulting data compared to more extensive paying closed datasets. The second aim was to make use of the resulting data to evaluate the effect of measures like flight bans. We collected data about travel restrictions, flights between US, Canada and Europe. Using this data, we studied changes in the flight network, characterising a precipitous drop in April 2020, followed by a progressive recovery.

In future work, we will compare the information obtained here with information obtained from other datasets. Such comparison could allow to derive rules for inferring travel data from this free resource.

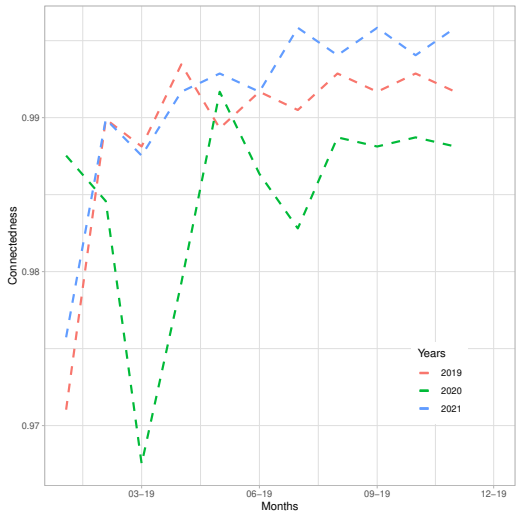
Appendix A. List of countries in "Europe"

In our analysis, the countries used under the term Europe are the following: Albania, Andorra, Armenia, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia,

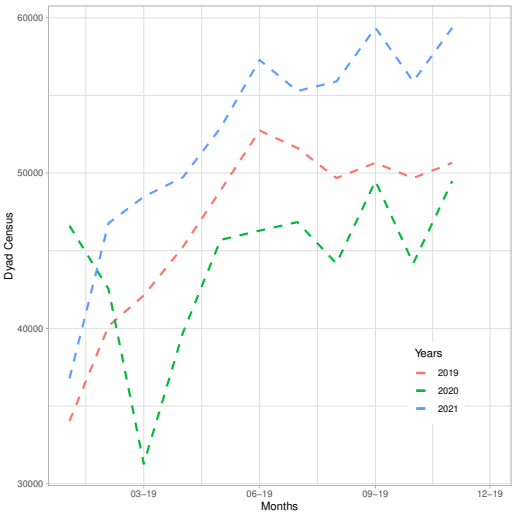


(a) Edge Formation counts

(b) Edge Dissolution count



(c) Connectedness



(d) Dyad Census

Figure 10: Evolution of different measures of the network.

Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Moldova, Monaco, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Russia, San Marino, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Türkiye, Ukraine, United Kingdom.

Appendix B. Timeline of travel restrictions

On 17 March 2020, the European Union imposed a ban on incoming travel for citizens from countries not in the European Union, European Economic Area, Switzerland and the United Kingdom; the only country not participating in the travel suspension was Ireland [23]. Starting 30 June 2020, the restrictions were gradually lifted from global restrictions to non-global restrictions [22]. On 27 March 2020, Republic of Türkiye suspended all international flights [12] and on 4 May 2020, they started to ease up the travel restrictions [8]. Serbia imposed a travel ban on 15 March 2020 [5] and started to ease up restrictions on 10 June 2020 [4]. Lastly, on 14 March 2020, Ukraine imposed a travel ban [10] which was lifted on 28 September 2020 [9].

In mid-March 2020 Canada closed its border for all travelers who were not citizens, permanent residents or U.S. citizens. Starting 6 January 2021, Canada started to gradually lift its travel restrictions, then on 31 January 2021, it imposed further restriction to funnel scheduled international commercial passenger flights into four Canadian airports: Montréal-Trudeau International Airport, Toronto Pearson International Airport, Calgary International Airport, and Vancouver International Airport [3]. These restriction were lifted and currently international travelers can enter Canada only if are vaccinated [2].

The USA did not start with a global travel suspension, however they suspended travels from countries that had a high amount of infections. Starting 31 January 2020, people other than citizens, permanent residents and their immediate family were prohibited from entering the U.S. within 14 days of being in China [7], followed by Iran on 2 March 2020, the Schengen area on 13 March 2020 and Brazil on 29 May 2020 [6].

Appendix C. The air transportation network as a social network

Social network analysis is an analytical tool used to map and measure social relationships. This multidisciplinary area involves social, mathematical, statistical and computer sciences. A social network is defined as a set of social "actors" and a social relationship between each pair of actors. These actors, called *nodes*, can be individuals, families, households, villages, communities, regions, etc. In our case, the nodes are the airports and the social relationship between nodes indicates the existence of a flight from one airport to another.

The network is *oriented* because the direction of flights is important and, although rare, there are cases when it is not symmetric. We call *arcs* the links between nodes in this case. An oriented network can be transformed into a *non-oriented* one by assuming two nodes are in relation if there is an arc between them in any direction. In this case, the link is called an *edge*. Multiple connections between the same two nodes can be represented by putting a weight on the arc, obtaining a *weighted network*. A weighted network is also easily transformed into an unweighted network by assuming that there is an arc if the unweighted network if there is a nonzero weight in the weighted one and no arc otherwise.

Appendix C.1. Implementing and analysing the network

We create a network for each month of 2019, 2020 and 2021, leading to 36 different networks, which we call the *monthly networks*. Recall that in the preprocessing stage, we select nodes that are present in all networks; this implies that the 36 networks considered share the same nodes, although arcs and weights vary month to month. The network is typically weighted, with weight the number of passengers on a given route for a given time period (one month here, although further work might consider smaller time steps).

The analyses described in the following sections are then run on each monthly network and the results are collated.

Appendix C.2. Centralities

Centralities are measures computed at individual nodes in the network, representing how they are connected to other nodes in the network, describing influential nodes in the connected structure of a graph [30]. In our computational work, we use [13, 19] to compute centralities.

In a directed network, each node has an in-degree (the number of arcs terminating in the node), an out-degree (the number of arcs originating in the node) and a degree (sum of in- and out-degree). Because the cleaning step removes self-connections, the in- and out-degrees are exactly the number of airports directly connected (inbound or outbound) to an airport in the month under consideration.

The *shortest path* between two airports is the minimal number of flights one has to take to go from one airport to another; we assume the distance is infinite if there does not exist a shortest path. The *out-eccentricity* of a node is the length of the longest shortest path out of a node; the *in-eccentricity* is the length of the longest shortest path into a node.

Other types of centralities considered on monthly networks are *betweenness* and *closeness* centralities.

Betweenness centrality measures how often a given node is on shortest paths not initiated or terminating there. A node (airport) with high betweenness centrality has more control over a network and plays an important role in ensuring overall connectivity of the network; if it is removed from the network, the risks of overall disconnection are higher.

Closeness centrality is the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph. Closeness captures how efficiently the entire network can be traversed from a given node; a node is central in that sense if it is close to many other nodes.

Appendix C.3. Network level properties

The network *diameter* is the maximum value of the eccentricity, while the *radius* is the minimum value. The *average path length* refers to the average number of steps along the shortest paths between all pairs of nodes in the graph. Another network-level property used is *density*, which is the ratio of the number of arcs present in the network to the total number of arcs that a *complete* network would have, where a complete network has all nodes connected to all other nodes.

Appendix C.4. Communities

A network has a *community structure* if the nodes can be grouped into sets such that each set is densely connected and loosely connected to vertices in the other communities. Note that nodes in networks can have overlapping and non-overlapping community structures. However communities partition the set of nodes; each node belongs to a single community. Nodes in a given community behave more like other nodes in the same community, so each community can be considered as a meta-node in a smaller graph, simplifying the analysis.

Finding communities can be a difficult task; usually the number of communities is unknown and the size of the communities is unequal. There are several methods to find the community structure of a network. There are two primary types, agglomerative and divisive methods [34].

Agglomerative methods start with an empty network having only vertices and no edges, we think of each vertex as its own community. Then at each step of the algorithm, edges are added to merge the two closest communities

to create new communities. The division method starts with a network having both vertices and edges. Initially we have one big community containing all vertices. Then we start removing edges to partition the vertices into similar communities using sum of square errors of each community and keeping the ones with the largest value; after a certain number of steps, communities of densely connected vertices are obtained.

Five community detection algorithms are used in this paper: Louvain, Leiden, Infomap, Cluster leading eigenvector, Clique-based networks.

Appendix C.4.1. Louvain community detection algorithm

The Louvain algorithm was proposed as a fast community unfolding method for large networks [16]. This algorithm tries to maximise the difference between the actual number of edges in a community and the expected number of edges. Each network vertex starts in its own community. The algorithm moves individual vertices from one community to another to find a partition. After a partition is found, an aggregate network is created with each individual vertex. These steps are repeated until the quality function cannot be increased further [37].

Appendix C.4.2. Leiden community detection algorithm

The Louvain algorithm has a tendency to discover communities which are internally connected and weakly connected communities [37] and another community detection algorithm, called *Leiden* was proposed [37], which guarantees that communities are well connected.

In addition to the phases of the Louvain algorithm, the Leiden algorithm has one extra phase of aggregation of the network based on the refined partition, using the non-refined partition to create an initial partition for the aggregate network. The algorithm starts from one partition, then individual vertices are moved from one community to another to find a partition. In the second phase, refined partitions from the partitions proposed in the first phase are sought. Communities from the first phase may be split into multiple other partitions in the second phase. An aggregate network is created based on the refined partition from the second phase, using the non-refined partition to create an initial partition for the aggregate network. The algorithm then moves individual nodes in the aggregate network. This repeats until no changes can be made [37].

Appendix C.4.3. Infomap community detection algorithm

The Infomap community detection algorithm approaches the problem like a data compression problem: its goal is to use the least amount of data to represent communities, using the partitions of the network as a Huffman code. To do this, Infomap finds the smallest number of communities needed to describe a random walk across every vertex in a network. A random walk across every vertex in a network can be described by storing the vertices that the walk traverses at each step. Infomap first identifies communities using the random walk, then assigns each community a name so that when the walker enters a community the algorithms can store the name and the vertices of the community. In most real-life networks there are regions of the network that once entered by the random walker, will keep the walker for a longer time; furthermore, movement between different regions is rare. The algorithm seeks an optimal partition that assigns nodes to communities such that the information needed to compress the movement of the random walkers is minimised. As the number of communities of a network becomes larger and larger with the number of nodes, InfoMap uses a variation of the Louvain algorithm to help find the communities, in order to help effectively explore partitions.

Appendix C.4.4. Cluster leading eigenvector community detection algorithm

This algorithm finds densely connected subnetworks by calculating the leading nonnegative eigenvector of the modularity matrix of the network. The algorithm optimises the modularity function; at each step, the network is split into two parts such that separation increases modularity. The leading eigenvector of the modularity matrix is used to determine the parts of the graph. Note that tightly connected groups cannot be split [31].

Appendix C.4.5. Clique-based networks community detection algorithm

A clique is a subset of nodes of an undirected network such that every two distinct nodes in the clique are adjacent, i.e., directly connected to one another. Nodes can be members of more than one clique and thus a node can be a member of more than one community. A maximal clique of a network, G, is a clique which can not be increased by including one more adjacent node. A maximum clique of a network, G, is a clique such that there is no clique with no more vertices. Furthermore, the clique number of a network is the number of nodes in a maximum clique [24].

Appendix C.5. Comparing community structures

Several algorithms can be used to compare communities as detected by the methods of Appendix C.4. Note, however, that the comparisons are only of two networks at a time. These algorithms are the variant of information [29], the normalized mutual information measure [20], Rand index [35] and the adjusted Rand index [25]. Note that the vectors being compared need to be of the same length for these methods to used.

Appendix C.5.1. Variant of information

Let D be a dataset with m points (the nodes) and $C_1, \ldots, C_N$ the communities. D can be partitioned so that $C_k \cap C_\ell = \emptyset$ and $\cup_{n=1}^{N} C_n = D$.

Choose a point (node) in D. Assume that each point has an equal probability of being chosen. The probability of the point being in the community C_n is

$$P(n) = \frac{m_n}{m},$$

where m_n is the number of nodes in C_n. This gives a discrete random variable taking N values, associated with the communities $\mathcal{C} = \{C_i\}_{i=1}^{N}$. There is some uncertainty regarding the cluster that the chosen point belongs to. This uncertainty is

$$H(\mathcal{C}) = -\sum_{n=1}^{N} P(n) \log[P(k)].$$

Note that $H(\mathcal{C})$ is always non-negative, takes the value 0 only when there is a single community and does not depend on the number of points in D.

Mutual information between two communities is then defined as follows. Let $P(n)$, $n = 1, \ldots, N$ and $P'(n')$, $n' = 1, \ldots, N'$ be the random variables associated with the communities $\mathcal{C}$ and $\mathcal{C}'$, respectively. Let $P(n, n')$ be the probability that a point belongs to both C_n in community $\mathcal{C}$ and $C'_{n'}$ in community $\mathcal{C}'$,

$$P(n, n') = \frac{|C_n \cap C'_{n'}|}{m}.$$

Define the mutual information $I(\mathcal{C}, \mathcal{C}')$ between $\mathcal{C}, \mathcal{C}'$ to be

$$I(\mathcal{C}, \mathcal{C}') = \sum_{n=1}^{N} \sum_{n'=1}^{N} P(n, n') \log \left[\frac{P(n, n'')}{P(n)P(n'')} \right].$$

The mutual information function is nonnegative, symmetric and cannot exceed $\min(H((C), H(\mathcal{C}')$. In [29], Variant of information (VI) between two communities is defined as

$$VI(\mathcal{C}, \mathcal{C}') = H(\mathcal{C}) - I(\mathcal{C}, \mathcal{C}') + H(\mathcal{C}') + I(\mathcal{C}, \mathcal{C}').$$

The first and second terms in the equation measure how much information about $\mathcal{C}$ is lost and how much information is gained about $\mathcal{C}'$ when going from $\mathcal{C}$ to $\mathcal{C}'$, respectively. Two community structures of a network are then called *similar* if $VI(\mathcal{C}, \mathcal{C}')$ is close to 0.

Appendix C.5.2. Rand index

The Rand index is a measure of similarity between two community structures of a network [25]. Let D be the set of (n) nodes in the network and $\mathcal{X} = X_{i i=1}^{k}$ and $\mathcal{Y} = Y_{i i=1}^{l}$ be two community structures. The Rand index is defined as

$$R = \frac{a+b}{\binom{n}{2}},$$

where

- a is the number of times a pair of elements belongs to the same community across two community methods,
- b is the number of times a pair of elements belongs to different community across two community methods,
- $\binom{n}{2}$ is the number of unordered pairs in a set of n elements.

The Rand index takes values between 0 and 1, where 0 indicates that the two community structures do not agree on the community partition of any pair of elements and 1 indicates that the two community structures agree on the community partition of every pair of elements.

Appendix C.5.3. Adjusted Rand index

The adjusted Rand index [38] is the corrected for change version of the Rand index, which measures the similarity between two community structures of a network. The adjusted Rand index can have negative values if the index is less than the expected index.

Let D be the set of n vertices of the network and $\mathcal{X} = X_{i i=1}^{k}$ and $\mathcal{Y} = Y_{i i=1}^{\ell}$ be two community partitions. The information overlap between X and Y

$X \setminus Y$	Y_1	Y_2	$\cdots$	Y_ℓ	Sums
X_1	n_{11}	n_{12}	$\cdots$	$n_{1\ell}$	a_1
X_2	n_{21}	n_{22}	$\cdots$	$n_{2\ell}$	a_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
X_k	n_k1	n_k2	$\cdots$	$n_{k\ell}$	a_k
Sums	b_1	b_2	$\cdots$	b_ℓ	$\sum_{ij} n_{ij}$

can be summarised using the following contingency table, where n_{ij} denotes the number of objects that are common to X and Y

The adjusted Rand index (ARI) is defined as follows:

$$ARI(X,Y) = \frac{\sum_{ij}\binom{n_{ij}}{2} - \frac{\left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]}{\binom{n}{2}}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right] - \frac{\left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]}{\binom{n}{2}}}.$$

The ARI has expected value zero for independent communities and is bounded above by 1 (for identical communities)

Appendix D. Other measures

Global measures exist describing the level of connectedness and the hierarchy of the network's structure. They can be evaluated using Krackhardt's axiomatic approach [28].

Appendix D.1. Connectedness

A *strongly connected component* is a subnetwork in which each node can reach every other node. A network is *disconnected* if there are at least two strongly connected components, otherwise the network is *strongly connected*. The degree in which a graph is disconnected is a function of the number of vertices being unable to reach another vertex in the graph; to describe this, one uses *connectedness*, defined as

$$C_K = \frac{D}{n(n-1)/2},$$

where D is the number of pairs of nodes that are not mutually reachable and $N(N-1)/2$ is the maximum number of nodes unable to reach another point in the graph. Connectedness scores range from 0 for a null network to 1 for a weakly or strongly connected network [28].

Appendix D.2. Efficiency

The *efficiency* of an undirected network G is defined using $G_1, \ldots, G_m$, the weak components of a network G. Let $|V(G)| = N$ be the cardinality of G and $|V(G_i)| = N_i$, $i = 1, \ldots, m$, be the cardinality of the weak components. Then the efficiency of G is

$$E_K = 1 - \frac{|E| - \sum_{i=1}^{m}(N_i - 1)}{[N(N-1)/2] - \sum_{i=1}^{m}(N_i - 1)},$$

where $|E|$ is the number of edges in G. A network with an efficiency of 1 has precisely as many edges as are needed to connect its components; as additional edges are added after that, the efficiency gradually falls towards 0 [28].

Appendix D.3. Hierarchy

Hierarchy measures quantify quantify the asymmetry of the structure of the network. The Krackhardt hierarchy score measures the degree of asymmetry of a directed network and is the fraction of node pairs in the reachability graph that are asymmetric [28].

Appendix D.4. Least upper bound (LUB)

Let G be a network. A node V_k is the *upper bound* for two nodes V_i and V_j if the directed paths V_k to V_i and V_k to V_j belong to the graph G. If V_k does not exists then V_i and V_j does not have an upper bound. A node V_ℓ is a *least upper bound* for V_i and V_j, if for all upper bounds V_k of V_i and V_j, V_ℓ belongs to at least one of the V_k to V_i paths and at least one of the V_k to V_j paths. If all vertices have a least upper bound, then Krackhardt's LUBness is 1; in general, it approaches 0 [28].

2019-04		2019-05		2020-04		2020-05		2021-04		2021-05	
Airport	Value	Airport	Value	Airport	Value	Airport	Value	Airport	Value	Airport	Value
KORD	1.00	KTEB	1.00	KORD	1.00	KORD	1.00	KIND	1.00	KIND	1.00
KADS	0.91	KORD	0.97	KADS	0.85	KADS	0.88	KADS	0.90	KADS	0.95
KTEB	0.90	KADS	0.93	KIAD	0.64	KIAD	0.66	KTEB	0.80	KTEB	0.92
KPHL	0.73	KIAD	0.79	KPHL	0.63	KDAL	0.63	KDAL	0.76	KHPN	0.72
KIAD	0.69	KIND	0.78	KAPA	0.56	KAPA	0.55	KPWK	0.76	KIAD	0.67
KIND	0.68	KBFI	0.77	KDFW	0.53	LSZH	0.51	KORD	0.74	KPWK	0.66
KBFI	0.65	KATL	0.71	KDAL	0.49	KSDL	0.49	KBFI	0.63	KORD	0.64
KDAL	0.63	KPHL	0.68	KATL	0.46	KBFI	0.46	KAPA	0.62	KBFI	0.63
KLAS	0.61	KDAL	0.66	KBFI	0.44	KDFW	0.45	KDFW	0.61	KDAL	0.63
KRDU	0.58	KFTW	0.65	PANC	0.44	KPWK	0.45	KSDL	0.59	KAUS	0.62

Table D.3: Top 10 Airports Global Betweeness Centrality

2019-04		2019-05		2020-04		2020-05		2021-04		2021-05	
Airport	Value	Airport	Value	Airport	Value	Airport	Value	Airport	Value	Airport	Value
KADS	1.00	KTEB	1.00	KADS	1.00	KADS	1.00	KDAL	1.00	KTEB	1.00
KORD	1.00	KPHL	0.99	KORD	0.79	KDAL	0.79	KIND	0.98	KIND	0.99
KIAD	0.95	KORD	0.98	KPHL	0.68	KSNA	0.71	KADS	0.92	KHPN	0.98
KPHL	0.92	KIAD	0.96	KATL	0.67	KAPA	0.70	KPWK	0.90	KPWK	0.94
KTEB	0.91	KADS	0.92	KDFW	0.66	KSDL	0.70	KHPN	0.88	KADS	0.92
KHHR	0.88	KMDW	0.91	KPHX	0.66	KORD	0.68	KTEB	0.88	KDAL	0.89
KLAS	0.85	KHPN	0.90	KFTW	0.63	KFTW	0.64	KPDK	0.86	KIAD	0.86
KRDU	0.85	KIND	0.90	KSNA	0.62	KIAD	0.63	KSDL	0.86	KSDL	0.85
KIND	0.84	KLAS	0.88	KGKY	0.61	KPWK	0.62	KAPA	0.82	KRDU	0.84
KDAL	0.81	KBFI	0.86	KIAD	0.61	KDTO	0.61	KRDU	0.82	KMDW	0.83

Table D.4: Top 10 Airports Global Degree Centrality

References

[1] http://www.lsv.fr/~sirangel/teaching/dataset/, 2022. Accessed: 2022-06-10.

[2] COVID-19 Boarding flights and trains in Canada. https://travel.gc.ca/travel-covid/travel-restrictions/domestic-travel, 2022. Accessed: 2022-06-10.

[3] Government of Canada introduces further restrictions on international travel. https://www.canada.ca/en/transport-canada/news/2021/01/government-of-canada-introduces-further-restrictions-on-international-travel.html, 2022. Accessed: 2022-06-10.

[4] Serbia Has Reopened To Tourists From All Countries. https://www.traveloffpath.com/serbia-has-reopened-to-tourists-from-all-countries/, 2022. Accessed: 2022-06-10.

[5] Serbia launches state of emergency to counter coronavirus. https://www.reuters.com/article/us-health-coronavirus-serbia-idUSKBN21215E, 2022. Accessed: 2022-06-10.

[6] Suspension of Entry as Immigrants and Nonimmigrants of Certain Additional Persons Who Pose a Risk of Transmitting 2019 Novel Coronavirus. https://www.federalregister.gov/documents/2020/03/16/2020-05578/suspension-of-entry-as-immigrants-and-nonimmigrants-of-certain-additional-persons-who-pose-a-risk-of, 2022. Accessed: 2022-12-19.

[7] Trump Declares Coronavirus A Public Health Emergency And Restricts Travel From China. https://www.npr.org/sections/health-shots/2020/01/31/801686524/trump-declares-coronavirus-a-public-health-emergency-and-restricts-travel-from-c, 2022. Accessed: 2022-06-10.

[8] Turkey: Authorities ease covid-19 travel restrictions may 4 /update 20. https://crisis24.garda.com/alerts/2020/05/turkey-authorities-ease-covid-19-travel-restrictions-may-4-update-20, 2022. Accessed: 2022-06-10.

Table D.5: Data Cleaning - number of rows

Month	Number of Initial Rows	Number of Rows with NA and loops	Number of NA rows with aircraft
2019-06	2,660,901	1,341,646	251,250
2019-07	2,898,415	1,472,035	287,734
2019-08	2,990,061	1,441,288	287,881
2019-09	2,721,743	1,263,726	178,654
2019-10	2,946,779	1,348,042	271,106
2019-11	2,721,743	1,263,726	254,324
2019-12	2,946,779	1,348,042	271,106
2020-01	2,734,791	1,253,919	70,467
2020-02	2,648,835	1,133,469	249,670
2020-03	2,152,157	921,075	203,858
2020-04	842,905	349,828	77,161
2020-05	1,088,267	458,980	109,234
2020-06	1,444,224	622,932	156,686
2020-07	1,905,528	820,004	82,111
2020-08	2,042,040	872,631	81,664
2020-09	1,930,868	819,864	73,794
2020-10	1,985,145	851,915	223,703
2020-11	1,930,868	819,864	73,794
2020-12	1,985,145	851,915	223,703
2021-01	1,783,384	805,621	221,904
2021-02	1,617,845	715,874	203,027
2021-03	2,079,436	901,919	252,124
2021-04	2,227,362	958,959	269,123
2021-05	2,278,298	939,572	258,830
2021-06	2,540,487	1,044,863	295,776
2021-07	2,840,201	1,173,235	349,230
2021-08	2,794,400	1,169,944	363,986
2021-09	2,523,676	1,043,522	328,727
2021-10	2,726,252	1,135,317	364,723
2021-11	2,523,676	1,043,522	328,727
2021-12	2,726,252	1,135,317	364,723

[9] Ukraine Reopens Border For Tourism Again on September 28. https://www.traveloffpath.com/ukraine-reopens-border-for-tourism-again-on-september-28/, 2022. Accessed: 2022-06-10.

[10] Ukraine – Immigration and Travel Policies Instituted to Combat COVID-19. https://home.kpmg/xx/en/home/insights/2020/04/flash-alert-2020-147.html, 2022. Accessed: 2022-06-10.

[11] ADMINCOPA. U.S. ADS-B Mandate. https://copanational.org/u-s-ads-b-mandate/, 2019. Accessed: 2022-07-18.

[12] Ali Murat Alhas. Turkey cancels all int'l flights amid COVID-19 outbreak. https://www.aa.com.tr/en/latest-on-coronavirus-outbreak/turkey-cancels-all-int-l-flights-amid-covid-19-outbreak/1782602, 2022. Accessed: 2022-06-10.

[13] Minoo Ashtiani, Mehdi Mirzaie, and Mohieddin Jafari. CINNA: an R/CRAN package to decipher Central Informative Nodes in Network Analysis. *Bioinformatics*, 35(8):1436–1437, 09 2018.

[14] K.-Y. Baek and H.-C. Bang. ADS-B based Trajectory Prediction and Conflict Detection for Air Traffic Management. *International Journal of Aeronautical and Space Sciences*, 13(3):377–385, 09 2012.

[15] Xiaoge Bao, Peng Ji, Wei Lin, M. Perc, and Jürgen Kurths. The impact of COVID-19 on the worldwide air transportation network.

[16] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008.

[17] Isaac I Bogoch, Maria I Creatore, Martin S Cetron, John S Brownstein, Nicki Pesik, Jennifer Miniota, Theresa Tam, Wei Hu, Adriano Nicolucci, Saad Ahmed, James W Yoon, Isha Berry, Simon I Hay, Aranka Anema, Andrew J Tatem, Derek MacFadden, Matthew German, and Kamran Khan. Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak. *The Lancet*, 385(9962):29–35, 2022/07/18 2015.

[18] Dirk Brockmann and Dirk Helbing. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science*, 342(6164):1337–1342, 2013.

[19] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

[20] Leon Danon, Albert Díaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008, sep 2005.

[21] EASA. Amendment to the Airspace Requirements on ADS-B and Mode S. `https://www.easa.europa.eu/newsroom-and-events/news/amendment-airspace-requirements-ads-b-and-mode-s`, 2022. Accessed: 2022-07-18.

[22] European Council. Council agrees to start lifting travel restrictions for residents of some third countries. `https://www.consilium.europa.eu/en/press/press-releases/2020/06/30/council-agrees-to-start-lifting-travel-restrictions-for-residents-of-some-third-countries/`, 2022. Accessed: 2022-06-10.

[23] European Council. Video conference of the members of the european council, 17 march 2020. `https://www.consilium.europa.eu/en/meetings/european-council/2020/03/17/`, 2022. Accessed: 2022-06-10.

[24] T S Evans. Clique graphs and overlapping communities. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(12):P12037, dec 2010.

[25] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985.

[26] Ken Pole. Canada's ADS-B mandate raises collective concerns among industry. `https://skiesmag.com/news/canadas-ads-b-mandate-raises-collective-concerns-among-industry/`, 2022. Accessed: 2022-07-18.

[27] K. Khan, J. Arino, W. Hu, P. Raposo, J. Sears, F. Calderon, C. Heidebrecht, M. Macdonald, J. Liauw, A. Chan, and M. Gardam. Spread of

a novel influenza A (H1N1) virus via global airline transportation. *New England Journal of Medicine*, 361(2):212–214, 2009.

[28] David Krackhardt. Graph theoretical dimensions of informal organizations, Computational organization theory. *Computational Organizational Theory, K. Carley, and M. Prietula, Eds. Hillsdale, NJ: Lawrence Erlbaum Associaties, Inc*, pages 89–111, 1994.

[29] Marina Meilă. Comparing clusterings—an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895, 2007.

[30] Christian F. A. Negre, Uriel N. Morzan, Heidi P. Hendrickson, Rhitankar Pal, George P. Lisi, J. Patrick Loria, Ivan Rivalta, Junming Ho, and Victor S. Batista. Eigenvector centrality for characterization of protein allosteric pathways. *Proceedings of the National Academy of Sciences*, 115(52):E12201–E12208, 2018.

[31] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74:036104, Sep 2006.

[32] Xavier Olive, Martin Strohmeier, and Jannis Lübbe. Crowdsourced air traffic data from The OpenSky Network 2020, June 2022. `https://doi.org/10.5281/zenodo.3737101`, Accessed: 2022-11-30.

[33] OurAirport. Open data downloads. `https://ourairports.com/data/`, 2022. Accessed: 2022-06-10.

[34] Mason A. Porter, Jukka-Pekka Onnela, and Peter J. Mucha. Communities in networks. *CoRR*, abs/0902.3788, 2009.

[35] William M. Rand. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

[36] Matthias Schäfer, Martin Strohmeier, Vincent Lenders, Ivan Martinovic, and Matthias Wilhelm. Demonstration abstract: Opensky: A large-scale ADS-B sensor network for research. In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, IPSN '14, page 313–314. IEEE Press, 2014.

[37] Vincent A. Traag, Ludo Waltman, and Nees Jan van Eck. From Louvain to Leiden: guaranteeing well-connected communities. *CoRR*, abs/1810.08473, 2018.

[38] Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11(95):2837–2854, 2010.

[39] World Health Organization. Weekly epidemiological update on COVID-19 - 21 December 2022. `https://www.who.int/publications/m/item/covid-19-weekly-epidemiological-update---21-december-2022`, 2021. Accessed: 2022-12-29.

[40] Ali M. Zaki, Sander van Boheemen, Theo M. Bestebroer, Albert D.M.E. Osterhaus, and Ron A.M. Fouchier. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *New England Journal of Medicine*, 367(19):1814–1820, 2012. PMID: 23075143.