Fourteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2021)

On-line platform for the short-term prediction of risk of expansion of epidemics

Proof-of-concept based on COVID-19 evolution

Javier García Moreno & Javier Poveda & Óscar Villasante & Pablo Sánchez-Escalonilla
CRIDA
Madrid, Spain
jgamorneo@e-crida.enaire.es

Alfonso Mateos Caballero & Eloy Vicente Cestero
Departamento de Inteligencia Artificial ETSINF
Universidad Politécnica de Madrid
Madrid, Spain
alfonso.mateos@upm.es

Ramon Lorenzo-Redondo
Department of Medicine, Division of Infectious Diseases
Northwestern University
Chicago, IL, USA
ramon.lorenzo@northwestern.edu

***Abstract*— This paper proposes a novel approach for the prediction of the risk of expansion of local epidemics to 3rd regions or countries in the world through the air traffic network. The approach relies on the definition of a new indicator, the Imported Risk, which represents the overall risk of having infected individuals entering an airport from any other airport with connections.**

We performed a proof-of-concept of the proposed approach by using daily data of the air traffic movements on a global scale and of the evolution of the COVID-19 epidemic at the beginning of 2020. For that purpose, we developed a complex network model based on Tagged Graphs to calculate the Imported Risk indicator, together with other complementary indicators showing the centrality of the air traffic network weighted with the Imported Risk.

We implemented our complex network model into an on-line platform which provides the daily risk of expansion of the epidemic to other regions or countries. The platform supports the identification of the components of the network (airports, routes…) that have a major impact on the risk of expansion. The paper also provides findings on how the short-term prediction of diseases' expansion through the Imported Risk indicator allows the identification of effective measures to take control of the virus spread.

Keywords-component; network; epidemics; imported risk; airports; COVID-19; prediction; tagged graphs.

I. INTRODUCTION

Infectious disease epidemics (e.g., influenza virus, Ebola virus, HIV-1) are major threats to human survival. The latest example of these emerging pathogens is the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) the causative agent of Coronavirus disease 2019 (COVID-19) that was declared an international public health emergency on January 30, 2020 by the World Health Organization (WHO).

Previous studies have demonstrated the close relation between the daily passengers' flows through the global air transport network and how COVID-19 epidemic was propagated between countries. Wu et al. [1] compared the risk of spread between international flights and domestic flights. Chinazzi et al. [2] proposed a disease transmission model of a global metapopulation to calculate the impact of travel limitations on the national and international spread of SARS-CoV-2. The results showed that the Wuhan travel quarantine delayed the progression of the epidemic by 3-5 days in mainland China, with a greater international effect. In addition, Bogoch et al. [3] analysed international airline passenger flights from ten Chinese cities. Other studies [4][5] used air travel data from departures in affected provinces of China to estimate the risk of virus importation. Gilbert et al. [4] focused on Africa and estimated the risk of importing SARS-CoV-2 to the African continent. Haider et al. [6] estimated a SARS-CoV-2 transmission risk index from the four largest cities in China based on the number of passengers to destination countries, weighted by the number of confirmed cases in cities of departure provided by WHO. They ordered each country according to the quartile to which it belonged based on the risk index. Nevertheless, although valuable, previous studies did not assess the cumulative risk of COVID-19 import in a country and instead they focused on specific entry points.

Over the last years, the theory of complex networks [7][8] has reached some important achievements and has been used to model and analyse epidemic spreading [1][9][10][11]. While the vast majority of current studies about epidemic propagation on complex networks has focused on static networks [12][13][1], there are other studies that focused on dynamic networks [14][15][16]. In particular, [17] and [18] analysed the complexity of the air traffic network, calculating several centrality metrics associated to the nodes (airports) of the air transport system such as proximity centrality, intermediation and eigenvector. The metrics showed the importance of each airport in the network according to the type of connections that they have with other airports.

Although previous studies succeeded in using the theory of complex network to model epidemic propagation through the air transportation network, they only used their models to get conclusions on how an epidemic was spread in the past. They did not use the theory of complex networks to dynamically calculate the imported risk of all the world's airports as a way of predicting the potential evolution of an infectious disease. This approach makes necessary a detailed characterization of all air transport-related factors impacting the expansion such as the daily number of passengers and duration of each flight, aircraft model and occupancy of each flight and the epidemiological status of the area of influence where the airports are located.

II. SOLUTION

We propose to define new indicators which are suitable for the prediction of the risk of expansion of local epidemics to other regions or countries in the world, thanks to the exploitation of massive daily data of passengers' flows and epidemic evolution through an air transport model based on the theory of complex networks.

We implemented our complex network model into an on-line platform which provides the daily risk of expansion of the epidemic to other regions or countries. The platform supports the identification of the components of the network (airports, routes…) that have a major impact on the risk of expansion. In addition, the platform allows assessing the impact of measures which could be implemented to prevent the virus spread, assuming that this could be a key support for decision-makers.

The platform processes automatically diverse sources of live data related to the air transport system worldwide (i.e., all daily flights and their characteristics such as occupancy per aircraft type), and daily data associated to the evolution of the epidemic all over the world. These data are mainly the number of infected, recovered, and susceptible people in all regions, and virologic features such as virulence indicators and reproductive indexes.

We performed a proof-of-concept of our solution - the new indicators, the complex network model based on daily data, and the user-oriented platform – by using real data on how the SARS-CoV-2 was expanded at the beginning of the pandemic. Basically, we assessed if the solution could have been used for the early identification of the risk of SARS-CoV-2 spread when the pandemic was in its initial phases – no local transmission of the coronavirus back then –.

This solution is built under the assumption that the detailed knowledge of the passengers' flows and connections in the network could allow predicting how an epidemic would expand from its origin in one single country to the rest of the world. This paper assumes that the air transport was the main mode of transmission of SARS-CoV-2 at the beginning of the pandemic in the vast majority of the countries in the world. There could be some exceptions to this rule in those countries which are geographically close to the origin of the pandemic, or have multiple and diverse connections with the main focus (e.g. automobile travel, cruise ships, etc.). These transportation models will be considered in future editions to improve the representativeness of our solution in those countries. We also highlight that our solution is exclusively designed to predict the virus spread at its initial stages, and not in the next phases in which local transportation models (public transport, short-distance car travel, etc.) became the most relevant way of local transmission.

Section III describes our methodology to develop the proposed indicators. Section IV details how the complex network model was designed in the case of COVID-19 and the process that was followed for its validation. Section V includes the results, which are divided in three parts: feasibility of the proposed indicators for the short-term prediction of epidemics expansion in third countries, applicability for one specific use case - the expansion of the SARS-CoV-2 in Spain -, and finally, assessment of the main functionalities of the developed platform and how it could be used by decision makers in epidemic crisis. Section VI details the main conclusions and the next steps.

III. METHODOLOGY

In this section, we describe how to calculate the daily imported risk into an airport from a third party. This is used to obtain the overall imported risk into an airport from all airports with connexions. Complementary to this indicator, other metrics associated to the theory of complex networks can also be implemented, showing the centrality of the air traffic network weighted with the imported risk previously calculated.

The imported risk by airport j on a given day, $Risk_j$, is calculated as the sum of the risk of each flight (k) from all airports which have operations with A_j on that day, $Risk_{ijk}$. This value is highly dependent on the epidemiological status of the catchment area of the airports of origin i and the risk of contagion that occurs during flight k. This depends on the duration of the flight and the separation between the passengers (which is a variable of the occupancy ratio of the flight). Mathematically,

$$Risk_j = \sum_{i \in A_j} \sum_{k \in V_{ij}} Risk_{ijk}$$

where A_j represents the set of all airports that fly to airport j and V_{ij} the set of flights that exist during the day from airport i to airport j.

The initial epidemiological status of flight k departing from airport $i \in A_j$ depends on the epidemiological status of the catchment area of airport i and the number of passengers traveling on flight k. We assume that the passengers at airport i is a simple random sample of the inhabitants of the airport's catchment area. Therefore, the epidemiological data for flight $k \in V_{ij}$ is calculated as follows.

The number of *susceptible passengers*, $S_{ijk}(0)$, *infectious passengers*, $I_{ijk}(0)$, and *recovered passengers*, $R_{ijk}(0)$, on flight k at the moment of departing from airport i to j are computed as:

$S_{ijk}(0) = n_{ijk} \frac{S_i(0)}{h_i}$, $I_{ijk}(0) = n_{ijk} \frac{I_i(0)}{h_i}$ and

$R_{ijk}(0) = n_{ijk} \frac{R_i(0)}{h_i}$, respectively.

That is, the number of susceptible, infectious and recovered passengers on flight k that departs from airport i to j is computed

as the average number of susceptible, infectious and recovered passengers, obtained as the product of the number of passengers on flight *k* from airport *i* to *j* and the probability that a passenger is susceptible, infectious and recovered, respectively, where the probability is computed as the total amount of people susceptible, infectious and recovered, respectively divided by the number of people living in the influence area (h_i).

For the case of COVID-19, the risk of a passenger being infected is estimated in several steps:

i) Calculate the number of new confirmed COVID-19 infections in the airport's catchment area in the last 7 days, $I_i(0)$. Seven days is chosen because it is the approximate duration of the contagion period for someone who has SARS-CoV-2 [19];

ii) Multiply $I_i(0)$ by ten to approximate the current number of new infections [20], as in first stages of a new disease, the reliability of the source is low;

iii) We assume that people with SARS-CoV-2 who take a plane are asymptomatic, pre-symptomatic or partially symptomatic (those with severe symptoms are unlikely to fly). Given that asymptomatic COVID-19 carriers make up approximately 40% of all carriers and they are only about 40% as contagious as others with more severe symptoms [21], the above product is multiplied by a factor of ¾.

iv) Multiply by a factor of about ½ to reflect the premise that flying passengers are generally affluent (less likely to encounter COVID-19 risks) than the general public [26].

Therefore, the number of infectious passengers is:

$$I_{ijk}(0) = n_{ijk}\frac{10\times\frac{3}{4}\times\frac{1}{2}I_i}{h_i} = 3.75n_{ijk}\frac{I_i(0)}{h_i}$$

Once the initial epidemiological status of flight *k* departing from airport $i \in A_j$ is calculated, we proceed to calculate the final epidemiological status of flight *k* departing from airport $i \in A_j$.

The epidemiological status at the end of flight *k* departing from airport $i \in A_j$ depends on the initial epidemiological status of flight *k*, the duration of the flight and the security measures adopted in the aircraft, which will be calculated by applying the SIR method (Susceptible-Infectious-Recovered) as it is suitable for diseases whose infectious agents are virus of the SARS-CoV-2 type.

Taking time *t* in minutes, we will consider the following classes of passengers: $S_{ijk}(t)$, $I_{ijk}(t)$ and $R_{ijk}(t)$ denote the number of susceptible, infectious and recovered passengers of flight *k* departing from airport *i* to *j* at minute *t*, respectively. In this way we have that $n_{ijk} = S_{ijk}(t) + I_{ijk}(t) + R_{ijk}(t)$ for every time instant *t.*

We assume the hypothesis of a homogeneous mixture of passengers [22]. That is, any infectious passenger can contact and infect any susceptible individual. Under this hypothesis, the susceptible individuals which could be infected are modelled by the term: $\beta S_{ijk}(t)\frac{I_{ijk}(t)}{n_{ijk}}$, where $\beta > 0$ is the transmission rate.

An infectious passenger may have mild or no symptoms and, after a while, recover. This step can be modelled using the linear term: $\alpha I_{ijk}(t)$, where $\alpha > 0$ is the *recovery rate*. That is, infectious individuals spend an average of $1/\alpha$ units of time in state *I* before recovering from the disease.

Thus, the following equations describe the transmission dynamic of virus according to the SIR method, with time *t* minutes,

$$S_{ijk}(t+1) = S_{ijk}(t) - \beta S_{ijk}(t)\frac{I_{ijk}(t)}{n_{ijk}} \quad (1)$$

$$I_{ijk}(t+1) = I_{ijk}(t) + \beta S_{ijk}(t)\frac{I_{ijk}(t)}{n_{ijk}} - \alpha I_{ijk}(t) \quad (2)$$

$$R_{ijk}(t+1) = R_{ijk}(t) + \alpha I_{ijk}(t) \quad (3)$$

Taking into account that flight *k* from airport *i* to *j* has a duration of m_{ijk} minutes, $I_{ijk}(m_{ijk})$ represents the average number of infected passengers arriving at airport *j* from airport *i* on flight *k*:

$$Risk_{ijk} = I_{ijk}(m_{ijk})$$

Therefore, the imported risk by airport *j* on a given day is the following, which only depends on two parameters, α and β.

$$Risk_j = \sum_{i\in A_j}\sum_{k\in V_{ij}} I_{ijk}(m_{ijk})$$

where A_j represents all airports that have flights to airport *j* and V_{ij} the flights that exist during the day from airport *i* to airport *j*.

Two additional parameters must be determined to calculate the imported risks at destination, the recovery rate (α) and the transmission rate (β). We describe in the following paragraphs how these parameters are calculated.

The recovery rate indicates the proportion of infected persons which are recovered per day, that is, $\alpha = 1/D$, where *D* is average number of days it takes a person to go from infected to recovered.

For COVID-19, $\alpha = 1/9$.

The transmission rate (β) is estimated by assuming that all passengers wear masks during the flight. Chu et al. [23] estimated that the use of masks reduces the risk of transmission in 82%. Therefore, the probability of a mask failure resulting in a transmission is:

$$P_M = 1 - 0.82 = 0.18.$$

In addition, we also need to calculate the conditional probability that a contagious passenger transmits SARS-CoV-2 to an uninfected one if the mask fails. Two situations are considered: full plane (FP) and with empty middle seat (EMS) on a Boeing 737 or Airbus 320 jet. Both planes have six seats per row: *A* (window), *B* (middle) and *C* (aisle); and *D* (aisle), *E* (middle) and *F* (window). We assume that the primary infection risk for a passenger flying alone comes from passengers sitting in the same row. In addition, we assume additional risks from passengers seated in the front and back rows. The simulations

done in [25] show that contagions are negligible in further rows for droplet transmission diseases like COVID-19.

A passenger can be infected by droplets from an infected passenger in the same row, depending on the distance between the two passengers. Chu et al. [23] estimates the risk of infection from physical contact with the infected person at 13%.and that this is reduced by half if the distance between people increases in one meter. The equation that reflects this exponential decrease pattern is $R_T = 0.13 \times e^{-0.69d}$.

In each row, the individual seats are 0.457 m. apart, while the aisle width is 0.762 m. R_T is the transmission risk between passengers without masks, and P_M is the probability of failure of the masks. Then, we estimate the risk of transmission of a passenger sitting in seat A *as* $P_L(A) = R_T(A,C) + R_T(A,D) + R_T(A,F)$ (when EMS) and $P_L(A) = R_T(A,B) + R_T(A,C) + R_T(A,D) + R_T(A,E) + R_T(A,F)$ (when FP), where $R_T(A,X) \approx 0.13 \times e^{-0.69d(A,X)}$ is the probability of transmission without masks given by a contagious passenger in chair X of the same row of an uninfected passenger in the chair A and $d(A,X)$ is the distance between the heads of both passengers.

On the other hand, several studies [24] [25] considered that, in addition to the potential transmission between passengers in the same row, passengers within the 2-rows of infected individuals can also be infected by droplets diseases. We should highlight that, at the time of doing this study, no final conclusions were published about the form of transmission of COVID-19, and droplets were thought as the main way of transmission.

Barnett [26] indicates that, although backups can block some drops of a contagious passenger, they are less effective than Plexiglass which almost eliminates transmission. Barnet estimates that seatbacks are about ¾ as effective as Plexiglass. This is another argument justifying that the possibility of infection is negligible from row 2 onwards.

Barnett [26]assumes that: *i)* When a flight is full, the six passengers in the row in front and behind of the uninfected passenger have a transmission risk of ¼ of the five passengers in the same row; *ii*) When the flight has the empty middle seat, but the rest are full, the four passengers on the front and back row of the uninfected passenger have a 2/3 risk of transmission of the six passengers in that row if the flight is full. In conclusion, the probability that an infected passenger transmits the disease to an uninfected one is

$$P_L(\text{FP}) = \left(1 + \tfrac{1}{4} + \tfrac{1}{4}\right) P_L(\text{FP}, same\ row),$$

$$P_L(\text{EMS}) = P_L(\text{EMS}, same\ row) + \frac{2}{3}\frac{1}{2} P_L(\text{FP}, same\ row).$$

Then, $P_L = \begin{cases} 0.402 & \text{FP} \\ 0.224 & \text{EMS} \end{cases}$

The infection rate or transmission rate (β) of SARS-CoV-2 indicates the number of contacts sufficient for the transmission of a person per unit of time. That is, the parameter β is the product of average number of contacts per individual and unit of time and the probability of contagion in a single contact between an infected individual and a susceptible individual. Therefore, $\beta = PS$, where P is the probability that a relationship between two people is contagious, and S is the average number of significant relationships that one person has. That is to say, $P = P_L P_M$, where P_L is the conditional probability that a contagious passenger transmits SARS-CoV-2 to the uninfected passenger if the mask fails and P_M the probability that the mask fails to prevent transmission of SARS-CoV-2. Substituting each variable for its values we obtain:

$$\beta = PS = P_L P_M S = \begin{cases} 0.402 \times 0.18 \times 3.5 = 0.253 & \text{FP} \\ 0{,}224 \times 0{,}18 \times 2 = 0{,}081 & \text{EMS} \end{cases}$$

where $3.5 = 2 + \frac{1}{4} \times 6$ (two passengers in the same row and 3 in front and 3 behind), $2 = 1 + \frac{1}{4} \times 4$ (1 passenger in the same row and 2 in front and 2 behind).

In conclusion, the daily imported risk for each airport is calculated as the sum of the risks of the flights arriving at that airport. The risk of each flight is calculated based on the epidemiological status of the catchment area of the departure airport, the duration of the flight, the number of passengers and whether they wear masks or not. These calculations rely on a detailed characterization of the recovery rate (α) and the transmission rate (β) of the infection disease.

IV. Description of the Proof of Concept (COVID-19)

We performed a proof-of-concept of the solution described in Section III. The air transport network and the COVID-19 expansion along the world were modelled through graph database platform Neo4j. Complex network models based on Tagged Graphs [27] were used to calculate the daily risk of importing the COVID-19 in each country, named the Imported Risk in this document.

Additionally, we calculated two complementary indicators for all airports and countries in the world, the Eigenvector and the Page Rank, showing the centrality of the air traffic network weighted with the Imported Risk previously calculated. These indicators determine how relevant a component (airport, route…) was in the expansion of the virus through the air transportation network.

The solution was tested against real data of how the COVID-19 pandemic was spread all over the world through the air transport network from the 22nd of January 2020 to the 31st of March 2020 - no official COVID-19 data at world scale was provided by the Johns Hopkins University before those dates -.

We tested if the proposed indicators fulfil the objective of being able to predict the propagation of the disease through the air transport network. Our hypothesis was that the indicators should present high correlation with the number of infected people in the short, medium, and long term. Thus, countries with higher values for a specific day should have higher impact of the pandemic 15 or 30 days later. On the other hand, countries with lower imported risk should have a slight evolution of the pandemic.

To validate the previous hypothesis, we prepared and cleaned the data and then, checked the correlation between the COVID-19 evolution in all countries and the proposed indicators.

A. *Data sources and cleaning*

To compute the Imported Risk indicator as described in section III, we need information at worldwide level. The data sources selected were:

1. Flight Information from *Flight Radar 24*: Contains the flight information at a worldwide level, captured from Automatic Dependent Surveillance - Broadcast (ADS-B) system. The fields in this data source include flight identification number, callsign, aircraft type, Actual Take-Off Time (ATOT), Actual Landing Time (ALDT), origin airport and destination airport in ICAO (International Civil Aviation Organization) format.
2. Disease information from *John Hopkins* database: Contains the information of the confirmed infected people, deaths and recovered from COVID-19 pandemic per day, country and/or region in the world.
3. Flight Occupancy from *International Air Transport Association (IATA)*: Contains the information of the percentage of occupancy of the flights grouped by continent in a monthly base.

Some data sources had limitations, for example ADS-B information does not provide full coverage of all flights at worldwide level and we did not have information on the occupancy of each single flight. However, data was considered representative enough to elaborate a proof of concept.

All this data was stored across a graph database using Neo4j, fitting the scheme presented in Figure 1 with all flight operations (departures and arrivals) by airport linked with its geopolitical hierarchy (region, country, province state) and these with their pandemic reports for every day (infected, recovered, deaths).

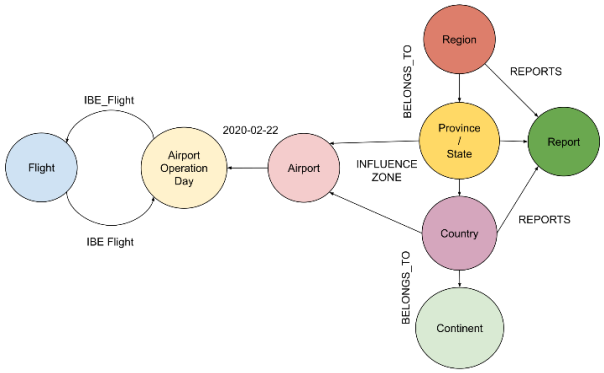

Figure 1. Graph database structure for the proof of concept using Neo4j.

After the information was merged using the graph database, the proposed indicators were computed. From the graph database, the variables included in TABLE I. were selected.

TABLE I. TABLE OF THE DATASET FOR COVID-19 PROOF OF CONCEPT

Variable Name	Definition	Example
Name	Country name.	Spain
importedRisk	Imported risk by the airport/country j ($Risk_j$).	21.17
InfectedActual	Infected people for the selected date.	12,567
Infected_15Days	Accumulated of infected people from the selected date to 15 days later.	1,260
Infected_30Days	Accumulated of infected people from the selected date to 30 days later.	8,909
scorePR_ALLF	Page Rank indicator for all flights.	1.34
scoreE_ALLF	Eigenvector indicator for all flights.	0.15

A statistical analysis of the previous variables was performed. The first step was the detection and filtering of outliers. To reduce the number of dimensions, we selected the Mahalanobis distance (MD) technique. Given that MD considers the correlation in the data, its values for the outliers will be always larger than those of other distance-based techniques such as the Euclidian distance (ED).

The MD distances were plotted versus the estimated quantiles or percentiles for a sample of size n from a chi-squared distribution, with k degrees of freedom. After applying the chi-squared at the standard quantiles (99.5%, 97.5%, and 95%), we selected 97.5%. Consequently, the values above quantile 97.5% of the chi-squared were filtered from the dataset, as they were tagged as outliers. This process was performed for every day in the dataset. Between 2 to 4 countries were usually filtered per day. Surprisingly, China is one of the detected outliers in most of the days analysed. This is due to the fact that China applied rapidly restrictive measures before other countries started having cases. Then, it can be observed a different behaviour in the evolution of the COVID-19 epidemic in China in comparison with the rest of the countries.

B. *Correlation indicators*

In this study, three correlation indicators were selected: Pearson product-moment correlation coefficient, also known as Pearson's or lineal correlation, and two other rank correlation coefficients, the Spearman's rank correlation coefficient and the Kendall's Tau rank correlation coefficient. The Pearson coefficient provides the measure of the linear relationship between two variables, while rank correlations provide a measure of ordinal association. Kendall's Tau calculations are based on concordant and discordant pairs, whereas Spearman's Rho calculations are based on deviations. Kendall's Tau usually provides smaller values than Spearman's Rho which is the most widely used rank correlation coefficient.

Once the outliers were removed from the dataset, the correlations were computed for every day. Then, after transforming the data to logarithmic scale, the correlation coefficients were represented in a box plot for all the period analysed. This representation was used to assess our hypothesis of high correlation between the indicators and the number of infected people in the short, medium, and long term.

V. RESULTS

A. *Statistical analysis*

Figure 2 shows the correlations between our main indicator, the Imported Risk, and the infected people - actual infected people in the selected day, accumulated infected 15 days later, and accumulated infected 30 days later - for all dates in the dataset. The mean values of the correlations can also be seen in TABLE II. We can see that there is a high degree of correlation

because the coefficient value lies between ± 0.50 and ± 1, then is said to be a strong correlation. In addition, two of the three correlation measures are higher for 15 and 30 days later than for the actual day.

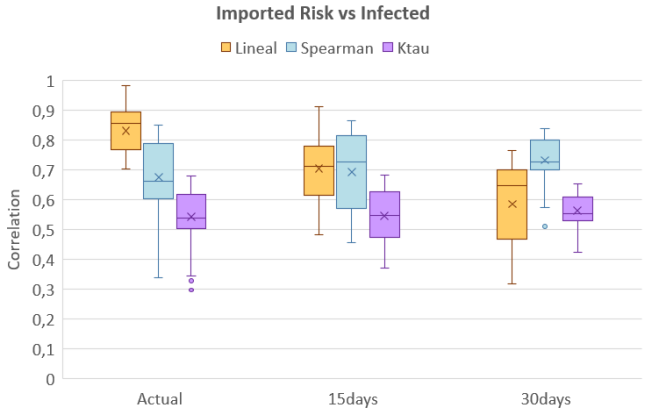

Figure 2. Imported Risk indicator vs Infected people of COVID-19 (actual, 15 days later and 30 days later)

TABLE II. MEAN CORRELATION VALUES FOR IMPORTED RISK **(IR)** INDICATOR VS INFECTED PEOPLE **(I)**

Mean	IR vs I Actual	IR vs I 15 days	IR vs I 30 days
Lineal	0.856	0.710	0.645
Spearman	0.660	0.726	0.726
Kendall Tau	0.537	0.546	0.551

For the 1st centrality indicator, the Page Rank, the correlation values are slightly lower than those of the Imported risk as it is shown in Figure 3. In particular, the results obtained for the case of infected people accumulated 30 days later showed poorer correlation when comparing with Imported Risk correlations. In this specific case, the median values are below 0.5 as it can be seen in TABLE III. For the rest of the values, we can see again that there is a high degree of correlation, highlighting the correlation of the current day that is very close to the perfect correlation for two of the three correlation indicators.

Overall, the correlations for 15 and 30 days later are worse with Page Rank than with Imported Risk Therefore, the Page Rank indicator is recommended to determine the risk on the actual day and Import Risk for the risk forecasts 15 and 30 days later. This is consistent with that fact that the Page Rank indicator is based on the topology of the air traffic network and the Imported Risk is based on the expected number of infected who arrives at a location that will influence the number of infected in the following days.

Finally, for the Centrality Eigenvector Indicator, we can see that the behaviour is worse than Page Rank and Imported Risk indicators, especially for the correlations with the accumulated number of infected people 30 days later, see TABLE IV.

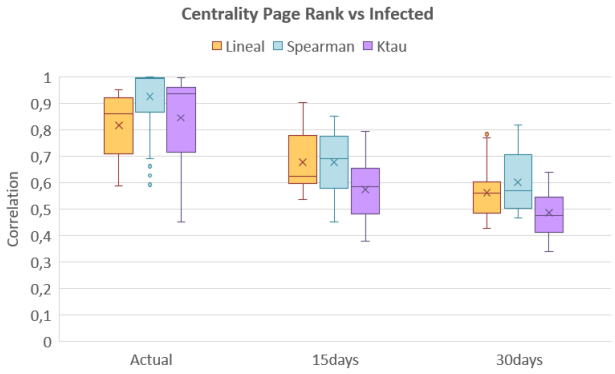

Figure 3. Page Rank indicator vs Infected people of COVID-19 (actual, 15 days later and 30 days later)

TABLE III. MEAN CORRELATION VALUES FOR PAGE RANK INDICATOR (PR) VS INFECTED PEOPLE (I)

Mean	PR vs I Actual	PR vs I 15 days	PR vs I 30 days
Lineal	0.786	0.556	0.495
Spearman	0.992	0.654	0.489
Kendall Tau	0.939	0.546	0.413

TABLE IV. MEAN CORRELATION VALUES FOR EIGENVECTOR INDICATOR (EV) VS INFECTED PEOPLE (I)

Mean	EV vs I Actual	EV vs I 15 days	EV vs I 30 days
Lineal	0.591	0.394	0.307
Spearman	0.922	0.635	0.505
Kendall Tau	0.939	0.541	0.407

After analysing the three main indicators, we can observe that the Imported Risk indicator provides better results to explain how the COVID-19 was spread through the air traffic network both 15 and 30 days later. In addition, the Page Rank indicator for the whole air transport network weighted with the Imported Risk is the best to determine the risk of the actual day. Eigenvector indicator was discarded as results showed poorer performances.

On the other hand, we can observe how the correlations decreased with the number of infected in the long term. This could be due to the effectiveness of the local measures implemented in the different countries to prevent the expansion of the COVID-19.

Putting the focus on the indicator with the best results, the Imported Risk indicator, we can also observe in Figure 2 that the box plots with the dispersion of the lineal correlation values is expanded when comparing with the number of infected people in longer periods. This wider dispersion, specially 30 days later, could be due to the differences in the local measures implemented in each country. Different measures with different effectiveness in the contention of the SARS-CoV-2 could be one of the factors producing this effect.

B. Use Case – Spain

We have analysed the evolution of the COVID-19 epidemic in one specific country to better understand if our indicators could have been used to predict and prevent the spread. In this section, the evolution of the pandemic in Spain was analysed.

Spain had a lockdown period, starting on March 14th, 2020, where almost all flights were modified or cancelled.

As in the previous sections, the period analysed comprise from 22/01/2020 to 31/03/2020. In Figure 4, it can be seen the evolution of the Imported Risk indicator (in blue) and the COVID-19 confirmed cases (in red) for Spain. In this country, the correlation between the Imported Risk indicator and the amount of people infected was around 0.8.

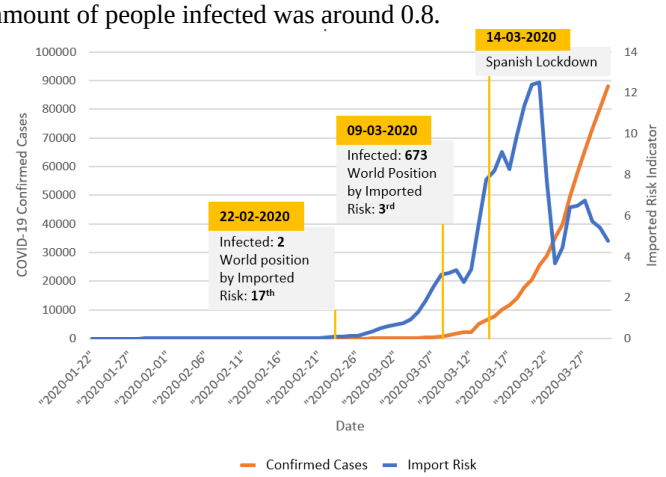

Figure 4. Evolution of the Imported Risk Indicator vs COVID-19 confirmed cases in Spain.

We can observe that, in the beginning stage of the pandemic, the Imported Risk was alerting at least 15 days before the confirmed cases started rising dramatically. By 22-02-2020, Spain only had 2 infected people, but at a worldwide level it was placed in the 17th position of the countries by the highest Imported Risk. After 15 days, the number of infected people were 673 in Spain, being the 3rd country in the world with the higher risk of importing passengers with COVID-19.

These results highlight that the Imported Risk indicator was in fact alerting of the risk of importing passengers infected around 15 days before the amount of infected people started increasing,

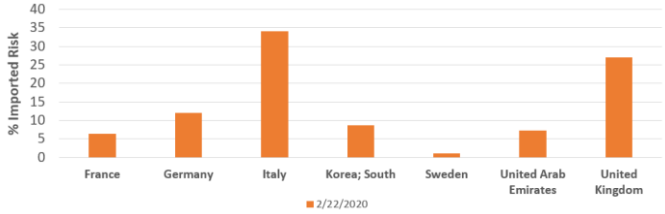

Figure 5 Main risk-exporter countries for Spain (22-02-2020).

In addition, Figure 5 shows the main risk-exporter countries to Spain on the 22-02-2020 expressed in percentages. Italy was the top risk-exporter (representing 34% of Spain's risk). More in detail, the top three airports adding risk from Italy were: LIRF (22.8%), LIMC (18.4%) and LIME (14.4%). By taking the measure of restricting flights from Italy, Spain reduces his Import Risk and changes the position in the world ranking from the place 17th to the 21st.

C. *Platform interface*

This section describes the main functionalities of the web-based interface of the platform[1], which can be visited in the following links with data in map format and table format, respectively.

http://213.229.174.138:8866/voila/render/stap_map.ipynb

http://213.229.174.138:8866/voila/render/stap_itable.ipynb

The interface shows the daily evolution of the imported risk indicators during the expansion of the SARS-CoV-2 at the beginning of 2020. This early prototyping allows assessing the evolution of the COVID-19 pandemic from the perspective of the future users of the platform, the stakeholders supporting the decision-making process. In addition, it helps to identify the user requirements for the adaptation of the platform to future epidemics.

Figure 6 shows the main window of the interface with the display in map format. The user can select one of the three imported risk indicators – Imported risk, Eigenvector or Page Rank – and the day. An additional filter is available to consider the imported risk originated by domestic or by international flights. This filter allows distinguishing between the overall imported risk in the catchment area of an airport which is originated by international flights exclusively, or by all flights arriving to the airport.

Figure 6 shows the Imported Risk on the 22nd of January 2020 for all airports in the world. The size of the bubbles represents the Imported Risk of each airport, and the darker colours are the countries with higher number of infected individuals on the selected date. We can see that the size of the bubbles is bigger in airports highly connected with China, such as those which are in the vicinity – three airports in Thailand, South Korea and Japan have the highest risk –. The size of the bubbles in European airports is not that big, although there are some airports like Istanbul which is the 8th airport with the highest risk in the world, and at the same time, it has a lot of connections with other European airports. Although our Risk Indicator is becoming high in some airports, the platform shows that, out of China, only 7 infected individuals in 5 countries were officially reported by the Johns Hopkins database in that day.

The prototype is updated automatically as soon as new data of infected people and flights at world scale is available. This process allows calculating the new imported risk indicators as soon as a new day has gone. Figure 7 shows the Imported Risk on the 29th of February 2020. After one month, the user can see that the Imported Risk increases in most of the airports in the world. Airports of Southeast Asia are 6 of the top-ten airports with the highest Imported Risk, but we can now identify Frankfurt airport and London Heathrow in this top-ten. In general, the risk of most of the European airports is very high.

The interface can also show the total imported risk of a country by taking into consideration the risk of all its airports. For the 29th of February 2020, four European countries are included in the top-ten: United Kingdom – 3rd position–,

[1] This is an initial prototype that was developed for validation purposes. Thus, computation time could take longer than expected in an operational platform. For support, please report to jpbarbero@e-crida.enaire.es.

Germany, Spain, and France. This risk is not necessary aligned with the number or infected people or the number of deceases due to SARS-CoV-2; for instance, in the case of United Kingdom, 61 infected and 0 deceases were reported by that date.

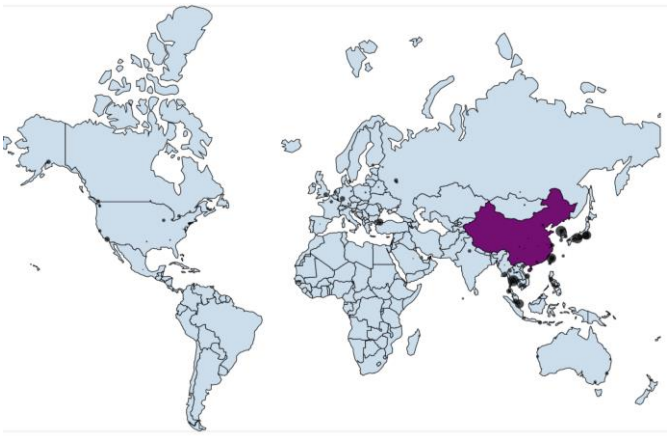

Figure 6 Imported Risk Map on January 22nd, 2020.

The interface has two 'what-if' functionalities to assess the impact of potential mitigation solutions and to assess future scenarios depending on the evolution of the epidemic. The 1st functionality provides information about the impact of implementing local solutions in an airport, route or for specific airlines which are operating in areas at high risk. The functionality allows removing all flights of one or several countries, specific airports in the countries or even single routes of one airport. After filtering those elements, we can obtain how the imported risk indicators vary. As an example, the decision of closing the routes with China on the 22nd of January 2020 would have implied the overall reduction of the imported risk in Europe – we can see the Imported Risk in Madrid-Barajas airport in Figure 7–.

The 2nd "what-if" functionality assesses the impact on the overall network of changes in the evolution of the epidemic in one or several countries. This functionality would have allowed anticipating the consequences at a world scale of an exponential growth of the COVID-19 epidemic in China. The user can change the number of Susceptible, Infected and Recovered (SIR) individuals in each country, together with the percentage of flight occupancy per country.

VI. CONCLUSIONS AND NEXT STEPS

A. *Conclusions*

We designed a new indicator, the Imported Risk, to predict how the diseases are propagated through the air transport network, and two complementary indicators which shows the centrality of this indicator in a complex network, the Eigenvector, and the Page Rank.

We developed a complex network model based on Tagged Graphs to calculate our indicators by using real data of SARS-CoV-2 – mainly Flight Radar 24 data for the air traffic and Johns Hopkins University for the COVID-19 evolution –.

We analysed the statistical correlation of the proposed indicators with the actual infected people on the day, the accumulated number of infected 15 days later, and the accumulated number of infected 30 days later for all days from 22/01/2020 to 31/03/2020. Three correlation coefficients were used: Pearson, Spearman and Kendall Tau.

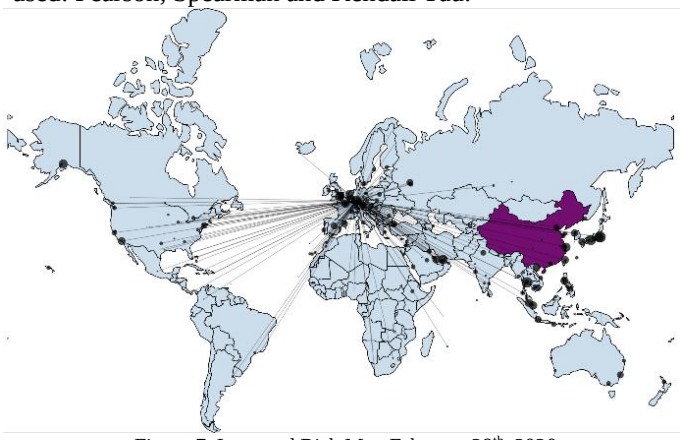

Figure 7. Imported Risk Map February 29th, 2020.

In conclusion, our main indicator, the Imported Risk, provided better results to explain how the COVID-19 was spread through the air traffic network both 15 and 30 days later. The statistical analysis showed strong correlations between this indicator and the subsequent evolution of the number of infected, recovered, and susceptible people in the area of influence of each airport in the world. Countries with higher values of the Imported Risk for a specific day have higher infected individuals both 15 and 30 days later, reporting an increase in the number of cases.

We also observed that correlations were slightly reduced when comparing with the accumulated number of infected people 30 days later. The proposed Imported Risk indicator provided better results to predict the evolution of the epidemic in the time horizon of 15 days later. This behaviour could be due to the local measures implemented in the different countries to prevent the expansion of the COVID-19 such as closing commercial centres and restaurants, promoting the use of the masks or imposing lockdown periods in the whole country, among others. These measures are probably avoiding an exponential increase in the number of cases in the long-term. The effect of these local measures was not included in this 1st model, and it is identified as one of the next steps - Item 5 of section B -.

An additional effect to be considered is that countries implemented different measures which had different effectiveness in the contention of the pandemic. This is one of the factors which could imply that correlations are better for some airports and countries than for others.

As an example of how this mathematical approach could be put into operation, we developed a platform with a web-based interface to exploit the results by decision makers. We assessed if the platform could have been used for the early identification

of the risk of virus spread in those countries that were impacted at the beginning of the pandemic – no local transmission – and also if it was possible to identify air traffic-related measures to prevent the expansion. We developed "what-if" functionalities that would allow the user assessing the impact of one element of the network such as one single airport or one single route in the world. However, the multiple and diverse changes that could be implemented in the air transportation network make it difficult for the user to determine the more effective and less penalising mitigation measure.

We identified that all alternatives, such as cancelling one single route or reducing the flights occupancy of airports operating in one country, influence our Imported Risk indicator, but it is difficult to determine which option is the most appropriate. The development of optimization criteria to select the best options, probably based on multi-objective performance framework, was not part of this 1st model, and it is identified as one of the next steps – Item 6 of section B -.

B. Next steps

We have identified the following areas of improvement to take advantage of all the potential of the proposed approach:

1. Full characterization of the air transport network considering daily data of connecting passengers. This will allow identifying the imported risk derived from passengers which come from airports or countries with high incidence but are not flying directly from the affected zones.
2. Characterisation and modelling of the airports according to their main passengers' flows and layout. The risk of infection at airports will be taken on board as an input to the complex network model, similarly to what was already done with the risk of infection during the flight.
3. Parameter setting of essential characteristics of epidemics that determines its expansion. This characterization of epidemics will allow using the platform for the prediction of the impact of future pandemics.
4. Integration of the platform with sanitary models that are being developed for the early detection of local epidemics. As we saw in the SARS-CoV2, the lack of knowledge of the virus at early phases makes difficult to determine the number of infected individuals and their characteristics. Our complex network model could benefit from on-going initiatives on syndromic surveillance systems for the early identification of epidemic in areas at high risk of disease outbreaks with similar signs and symptoms. Whilst these surveillance tools will focus on the early detection of new epidemics and the potential number of cases, our platform could use this initial data to predict the risk of expansion at world scale.
5. Inclusion of the impact of external mitigation measures which are not part of the air traffic-related measures, but they influence on how the emergency is spread. A good example is the obligation of using masks to enter a country. An initial prospection was done in this area with the inclusion in the model of the stringency index [28] which quantifies how restrictive the measures implemented in each country are.
6. Identification of pre-defined air traffic-related measures with decision-makers to develop capabilities for the quantification of the impact of each action. A multi-objective performance framework will allow analysing each measure from diverse perspectives. On one hand, emergency-related metrics will quantify the effectiveness and resilience of each measure to changes in the evolution of the epidemic around the world. On the other hand, ATM-related metrics will quantify the impact on capacity, efficiency or environment Key Performance Areas (KPAs). Both perspectives will be combined into cost-related metrics which will determine the overall costs derived from the implementation of each measure, and in particular the economic implications in the air transport sector. This approach will facilitate joint decision-making among airlines, airports, regions, and countries, taking decisions more efficient and less aggressive for the air transport sector than simply closing airports or the entire airspace.
7. Calibration of the model to ensure the alignment of the values of the Imported Risk indicator with real potential entrees of infected individuals through each air transport route.
8. Validation of the model by using verified data of how past epidemics, e.g., original SARS-CoV2 and/or the recently developed new viral lineages, entered into regions or countries. This will allow assessing the sensitivity of the model with real data. The most reasonable approach seems to be using phylogenetic models. Following this approach, we will infer the SARS-CoV2 population dynamics using publicly available viral genetic sequences. Subsequently, we will correlate the population dynamics parameters inferred such as migration rates between geographic locations and population size changes at specific locations with the predictions obtained with the developed complex network model.

ACKNOWLEDGMENT

This research was promoted by ENAIRE and CRIDA within the initiative 'Open Innovation', which has the aim of identifying proposals which contribute to the fight against coronavirus. In addition to the entities of the authors and project MTM2017-86875-C3-3-R, we would thank for the support of the Technical School of Aeronautical and Space Engineering (Polytechnic University of Madrid) and the Technical School of Telecommunications and Systems Engineering (Autonomous University of Barcelona).

BIOGRAPHIES

Mr. **Javier García** is an ATM Research & Development Engineer at CRIDA. He received his M.Sc. in Air Transport Systems from UPM in 2017 in Madrid, and he obtained his B.Sc. in Aeronautical Engineering from UPC, in Barcelona in 2015.

Dr. **Ramón Lorenzo-Redondo** is a research assistant professor in the Department of Medicine, Division of Infectious Diseases, of Northwestern University Feinberg School of Medicine, Chicago, IL, USA. Dr. Lorenzo-Redondo has led published studies on the evolution of RNA viruses and contributed greatly to the field of host-virus genetics. He received a Ph.D. in Molecular Biology from the Universidad Autónoma de Madrid in 2011.

Mr. Prof. **Alfonso Mateos** is a full professor of Statistics and Operations Research in the Artificial Intelligence of the Universidad Politécnica de Madrid (Spain). He has directed the research group: Decision Analysis and Statistics Group.

Mr. **Javier Poveda** is a Principal R&D at CRIDA. Javier received a M.Sc. degree (2010) in Computer Science from the Technical School of Computer Engineers of the Universidad Politécnica de Madrid (Spain).

Mr. **Pablo Sánchez-Escalonilla** is technical manager at CRIDA. He is Aeronautical Engineer with a degree from the Polytechnic University of Madrid (Spain).

Dr. **Eloy Vicente** received a PhD in Artificial Intelligence from the Universidad Politécnica de Madrid (Spain). He is responsible for Artificial Intelligence and Data Science at "Grupo Preving".

Mr. **Óscar Villasante** is a R&D engineer at CRIDA He received a M.Sc. in Computer Science from the Technical School of Computer Engineers of the Universidad Politécnica de Madrid (Spain)

REFERENCES

[1] Q. Wu, T. Hadzibeganovic. "Pair quenched mean-field approach to epidemic spreading in multiplex networks," Appl. Math. Model, vol. 60, pp. 244-254, 2018.

[2] M. Chinazzi, J.T. Davis, M. Ajelli, C. Gionnini, M. Litvionova, S. Merler, A. Pastor, "The effect of travel restriction on the spread of the 2019 novel coronavirus (COVID-19) outbreak," Science, vol. 24, pp. 395-400, 2020.

[3] I. Bogoch, A. Watts, A. Thomas-Bachli, C. Huber, M.U.G. Kraemer, K. Khan, "Potential for global spread of a novel coronavirus from China," The Lancet, vo.l 27, pp. 871-877, 2020.

[4] M. Gilbert, G. Pullano, F. Pinotti, E. Valdano, C. Poletto, P.Y. Boelle, E. D'Ortenzio, Y. Yazdanpanah, S.P. Eholie, M. Altmann, B. Gutierrez, M.U.G. Kraemer, V. Colizza, "Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study," the Lancet, vol. 395, pp. 871-877, 2020.

[5] S. Lai, I.I. Bogoch, A. Watts, K. Khan, Z. Li, A. Tatem, "Preliminary risk analysis of 2019 novel coronavrus spread within and beyond China," University of Southhampton, pp. 1-20, 2020.

[6] N. Haider, A. Yavlinsky, D. Simons, A. Osman, F. Ntoumi, A. Zumla, R. Kock, "Passengers' destinations from China: Low risk of novel coronavirus 82019-nCoV) transmission into Africa and South America," J. Epid. And Infect., 1148, 2020.

[7] D. Zhao, L. Wang, Z. Wang, G. Xiao, "Virus propagation and patch distribution in multipex networks: modeling analysis and optimal allocation," IEEE Trans. Forensic Secur., vol. 14 (7), pp. 1755-1767, 2019.

[8] L. Zhu, W. Liu, Z. Zhang, "Delay differential equations modeling of rumor propagation in both homogeneous and heterogeneous networks with a forced silence function," Appl. Math. Comput, vol. 370, pp. 124925, 2020.

[9] R. Pastor-Satorras, C. Castellano, P.V. Mieghem, A. Vespignani, "Epidemic processes in complex networks,", Rev. Mod. Phys., vol. 87 (3), pp. 925-979, 2015.

[10] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in spreading in scalefree networks," Phys. Rev. Lett., vol. 86 (14), pp. 3200-3203, 2001.

[11] Z. Wang, C. Xia, Z. Chen, G. Chen, "Epidemic propagation with positive and negativ preventive information in multipex networks," IEEE Trans. Cybern., 10.1109/TCYB.2019.290605.

[12] Z. Wang, Q. Guo, S. Sun, C. Xia, "The impact of awareness diffusion on SIR-like epidemics in multiplex networks," Appl. Math. Comput.", vol. 349, pp. 134-147, 2019.

[13] X. Wang, Z. Wang, H. Shen, Dynamical analysis of a discrete-time SIS epidemic model on complex networks," appl. Math. Lett., vol. 94, pp. 292-299, 2019.

[14] Z. Jin, G. Sun, H. Zhu, "Epidemic models for complex networks with demographics", Math. Biosci. Eng., vol. 11 (6), pp. 1295-1317, 2014.

[15] W. Pan, G. Sun, Z. Jin, "How domgorphy-driven evolving networks impact epidemic trnasmission between cities," J. Theor. Biol., vol. 382, pp.309-319, 2015.

[16] Y Yao, J. Zhang, "A two.strain epidemic model on complex networks with demographics," J. Biolo. Syst., vol. 24 (04), pp. 577-609, 2016.

[17] P. Nikolaou, L. Dimitrio, Identification of critial airports for controlling global infectious disease outbreaks. Stress-tests focusing in Europe," J. of Air Transp. Manag., vol. 85, 101819, 2020.

[18] E. Vicente, A. Mateos, "A COVID-19: analizamos el papel de los vuelos internacionales en su propagación, " The conversation, 2020.

[19] Y. Ling, S.B. Xu, Y.X. Lin, D. Tian et al., "Persistenci and clearance of viral RNA in 2019 novel coronavirus disease reahbilitation patients," Chines Med. J., vol. 133, pp. 1039-1043, 2020.

[20] F.P. Havers, C. Reed, T. Lim, J.M. Montgomery et al., Seroprvelanece of antibodies to SARS-COV-2 in 10 sites in the United States,", JAMA Inter Med.. 2020.

[21] A. Bilinski, R. Birger, S. Burn, M. Chitwood, et al., "Defining high-value information for COVID-19 decision-meaking," medRXiv, 2020.

[22] J.D. Murray, "Mathematical BiologyI. An Introduction, Springer:New York, 2004.

[23] D.K. Chu, E.A. Akl, S. Duda, K. Solo, S. Yaacoub, H.J. Schunemann, "Physical distancing, face masks, and eye protection, to prevent person-to-person trnsmission of SARS-Cov-2 and COVID-19: A siystematic review and meta-analysis," The Lancet, vol. 395, pp. 1973-1987, 2020.

[24] S. Hertzberg, H. Weiss, " Row rule for infectious disease trnasmission on aircraft," Annals of Global Health, vol. 82, pp. 819-823, 2017.

[25] V.S. Hetzberg, H. Weiss, W. Si, S.L. Norris, " Behaviors, movements, and trnsmission of droplet-mediated respiratory diseases during trnascontinental airline flights. In proceedings fo the National Academy of Sciences of the United States of America, Eds. (Burton H. Singer), University of Florida. Gainesville, FL, 2018.

[26] A. Barnett, K. Fleming, "Covid-19 risk among airline passengers: should the middle seat stay empty?, medRXiv, 2020.

[27] Guo, W., Toader, B., Feier, R. et al. Global air transport complex network: multi-scale analysis. SN Appl. Sci. 1, 680 (2019). https://doi.org/10.1007/s42452-019-0702-

[28] T. Hale, N. Angrist, R. Goldszmidt, B. Kira , A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, and H. Tatlow. (2021). "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)