medRxiv preprint doi: https://doi.org/10.1101/2020.02.20.20025882; this version posted February 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Evaluating the impact of international airline suspensions on COVID-19 direct importation risk

Aniruddha Adiga, Srini Venkatramanan*, Akhil Peddireddy, Alex Telionis, Allan Dickerman, Amanda Wilson, Andrei Bura, Andrew Warren, Anil Vullikanti, Brian D Klahn, Chunhong Mao, Dawen Xie, Dustin Machi, Erin Raymond, Fanchao Meng, Golda Barrow, Hannah Baek, Henning Mortveit, James Schlitt, Jiangzhuo Chen, Jim Walke, Joshua Goldstein, Mark Orr, Przemyslaw Porebski, Richard Beckman, Ron Kenyon, Samarth Swarup, Stefan Hoops, Stephen Eubank, Bryan Lewis*, Madhav Marathe*, Chris Barrett**

Biocomplexity Institute & Initiative, University of Virginia

**{aniruddha, srini, brylew, marathe, ChrisBarrett}@virginia.edu*

ABSTRACT

Global airline networks play a key role in the global importation of emerging infectious diseases. Detailed information on air traffic between international airports has been demonstrated to be useful in retrospectively validating and prospectively predicting case emergence in other countries. In this paper, we use a well-established metric known as effective distance on the global air traffic data from IATA to predict COVID-19 times of arrival (ToA) for different countries as a consequence of direct importation from China. Using this model trained on official first reports from WHO, we provide estimated ToA for all other countries. By combining effective distance with a measure for the country's vulnerability (Infectious Disease Vulnerability Index (IDVI)), we propose a metric to rank vulnerable countries at immediate risk of case emergence. We then incorporate data on airline suspensions to recompute the effective distance and assess the effect of such cancellations in delaying the estimated arrival time for all other countries.

KEYWORDS

Infectious diseases; COVID-19; International airline network; Effective distance; IDVI; Airline suspensions;

OVERVIEW

For the ongoing COVID-19 epidemic, 24 countries[1] have officially reported cases. All of the first reports in these countries have had travel history to China (mostly to Wuhan city or Hubei province). In this work, we assess the global importation risk of novel coronavirus disease (COVID-19) for other countries based on travel to China. We employ the concept of *effective distance* [3] which has been retrospectively validated for the SARS and H1N1 cases, to estimate the time-of-arrival (ToA) for all countries. We apply this metric on global traffic flows out of China obtained from the International Air Travel Association (IATA) [5] for the month of February 2019. We provide a risk assessment for countries based on two factors: the connectivity of the country to China (determined by the effective distance) and its vulnerability to disease outbreaks (determined by Infectious Disease Vulnerability Index (IDVI)).

We discuss the computation of the effective distance and the network construction in the subsequent sections. We show that with the constructed network, a moderately high linear relationship (coefficient of determination, $R^2 = 0.78$) holds for ToA of COVID-19 cases as reported by World Health Organization (WHO) (Figure 1). We then employ the linear estimator to compute estimated ToAs for countries across the globe. By plotting the estimated ToAs against IDVI, a measure of vulnerability, we observe that of the countries that have reported cases, most are developed economies with typically high IDVI (low vulnerability (Figure 2). This is likely due to (a) developed economies having strong air traffic and connectivity especially to China, and (b) their ability to detect and report imported cases fairly quickly. Our

[1] As February 10, 2020, excluding Hong Kong, Macau, and Taiwan

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.20.20025882; this version posted February 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

results indicate multiple countries (many with low IDVI, hence highly vulnerable) might see (or already have seen) case emergence in the month of February. We propose IDVI-weighted effective distance as a metric to compare two countries accounting for vulnerability and risk of importation.

We then model the impact of flight suspensions to and from China that is being witnessed globally, while ensuring that the effective distances with and without interventions (i.e., suspensions) are comparable through normalization. The reduced flow from China results in a modified network with reduced edge weights and almost all countries see an increase in their effective distance from China, with some countries having higher increase than others. We also observe a change in rank of countries as per the IDVI-weighted effective distance (Figures 3 and 4).

Related Work

There are multiple ongoing attempts to use airline traffic data to quantify global risk posed by COVID-19. In [1], the authors employ air travel volume obtained through IATA from ten major cities across China to rank various countries along with the IDVI to convey their vulnerability. [12] consider the task of forecasting international and domestic spread of COVID-19 and employ Official Airline Group (OAG) data for determining air traffic to various countries. [6] fit a generalized linear model for observed number of cases in various countries as a function of air traffic volume obtained from OAG data (authors refer to [12]) to determine countries with potential risk of under-detection. [7] provide Africa-specific case-study of vulnerability and preparedness using data from Civil Aviation Administration of China. In order to determine vulnerability of nations to COVID-19, the popular metric has been IDVI [1, 7] with [7] also employing WHO International Health Regulation Monitoring and Evaluation Framework to determine preparedness. For the current COVID-19 outbreak, [2] provides the *relative import risk* to various countries estimated using the worldwide air transportation network consisting of 4000 airports, and also an interactive visualization for arbitrary origin airports.

Some key differences: (a) We use IATA data which contains actual flow volume between origin-destination airports (including transit points) instead of OAG data (used by [3]) which contains total number of seats available for a given segment. We believe that origin-destination flows by IATA provides a better estimate of population level exposure, and the likelihood of detecting the first case in a country (rather than at the transit airports), thus serves as a better network on which to compute effective distance (cf. Appendix A). While [1, 4, 8, 12] have used the IATA data for estimating relative risk for COVID-19 spread, this is the first work to show the relationship between the emergence times of COVID-19 using effective distance with actual arrival times for 24 countries. We provide estimated ToA for countries that have not officially reported COVID-19, and provide a way to quantify the effect of ongoing airline suspensions.

METHODS

Data

IATA [5] provides comprehensive air travel data covering 12000 airports across the world. Given a source and destination airport, it provides the number of passengers traveling between the two airports as reported by the airlines ("reported PAX") and the IATA estimated ("reported + estimated PAX") per month. The travel details include both direct and multi-hop connections between the airports. In addition, it provides the different carriers serving the two airports, which allow for simulating airline cancellations using historical data. For this analysis, we used data from the month of February 2019. **Infectious Disease Vulnerability Index** (IDVI) [9] is a metric developed by the RAND Corporation to identify countries potentially most vulnerable to poorly controlled infectious disease outbreaks because of confluence of factors (including political, economic, public health, medical, demographics), and disease dynamics. **Arrival times** are curated from the official situation reports [10] published by World Health Organization. Each country is associated with the first date of confirmed case, with China being assigned December 31st, 2019. **Airline suspensions** information until February 10th 2020, by airline and routes was obtained and curated from Bloomberg [11].

medRxiv preprint doi: https://doi.org/10.1101/2020.02.20.20025882; this version posted February 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Network construction

We use international air traffic data from February 2019, which is the most recent data that coincides with the Lunar New Year period, known to involve a large movement of people in and out of China. As mentioned previously, IATA data provides number of passengers between two airports. We aggregate the flows out of all the airports in China to all the airports within the country of interest. We treat non-stop routes and routes with transit equivalently, since we are primarily interested in passenger volumes traveling out of China disembarking at various destination countries.

Effective distance and estimated ToA

In order to model the risk due to direct importation, we consider the number of passengers whose trip originated in airport m and ended in airport n (through multiple paths) as obtained from IATA database. This data extracted from IATA is different from the number of available seats between m and n (i.e., the link capacity which could be used by passengers using $m \rightarrow n$ as a transit) considered in [3]. Hence, the computed effective distances would be different and the difference is explained in Appendix A. We compute the fraction of flows from China to any given country-of-interest and determine its effective distance. We develop a linear estimator for COVID-19 ToA at various countries using the effective distance to China. This allows us to "predict" when COVID-19 is most likely to be reported by various countries.

Further, two countries with similar effective distances do not necessarily rank the same in terms of risk of sustained, undetected or uncontrolled epidemic outbreak. In order to capture this aspect, we consider the product of effective distance of a country to its IDVI to rank the countries by risk to COVID-19.

Effect of airline suspensions

Finally, using airline suspensions data from [11], we alter the flow volumes in the original air traffic network. To ensure that the effective distances on the air traffic network are comparable with and without interventions, we scale the flow volume in the latter with total outflow from origin in the former. Assuming that the total outflow from a country without airline suspensions is reflective of the country population size, normalizing the reduced flows with it could provide better estimates for P_{mn}. Formally, if G is the original weighted flow network on which effective distance is computed from source i_0, and G' is the flow network derived by adjusting flow volumes based on airline suspensions, then the edge weights between nodes m and n on G' are $\bar{F}'_{mn} = F'_{mn}/\sum_n F_{mn}$, where F'_{mn} denotes the reduced flow between the nodes. We use IDVI-weighted effective distance on both networks, and the estimated ToA (using the same linear estimator as before) to measure the effect of airline suspensions of altered COVID-19 importation risk.

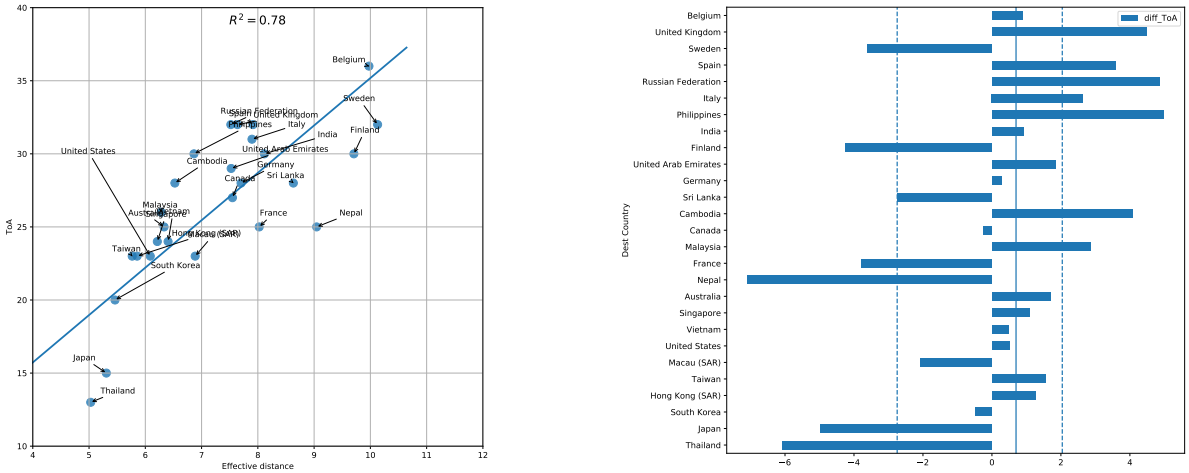
RESULTS

Importation risk prior to airline suspensions

Using a linear model for the reported time of arrival against the effective distance computed on IATA data, we observe a moderately high coefficient of determination $R^2 = 0.78$. A scatter plot of COVID-19 ToA for countries with reported cases with respect to their effective distance from China is shown in Figure 1a. It should be noted that in the regression analysis, China has effective distance and time of arrival set to zero. The difference between reported ToA and the estimated ToA is provided in Figure 1b. The median error in estimation is less than a day with the estimated ToA typically being earlier than the reported ToA. Note that there are a few outliers which have much earlier reported ToA than estimated ToA (e.g., Nepal, Japan, Thailand). For these countries, airline traffic may not be representative of the total connectivity with China due to their spatial proximity (land or sea). In order to assess the risk of direct importation of COVID-19 from China to various countries we show a plot of the estimated ToA at each country to their vulnerability (IDVI) in Figure 2. The Combatant Commands provide a rough continent-level grouping of countries. As demonstrated in Figure 1a, the countries with reported cases (shown with plus marker) are mostly to the left. Interestingly, most countries with reported cases also have higher IDVI, which may be due to a combination of (a) high connectivity and air traffic to China, (b) better detection and reporting capabilities for imported cases. In general, the

medRxiv preprint doi: https://doi.org/10.1101/2020.02.20.20025882; this version posted February 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

lower left region (low IDVI and low estimated ToA) indicates a regime of high risk, which is relatively empty for this outbreak. However, we notice that several low IDVI countries (especially those in AFRICOM) are estimated to have times of arrival in the first three weeks of February. Equally concerning are countries to the left with circle markers (some of which are highlighted), which have an early estimated time of arrival but haven't officially reported cases yet.



(a) a **(b)** b

Figure 1. (a) Validating effective distance to the arrival times with first reports (Source: WHO Situation reports). Time of arrival is indexed starting Dec 31st, 2019 as day 0. (b) The difference between reported ToA and estimated ToA (with median error marked with solid vertical line and dashed lines indicating the interquartile range).

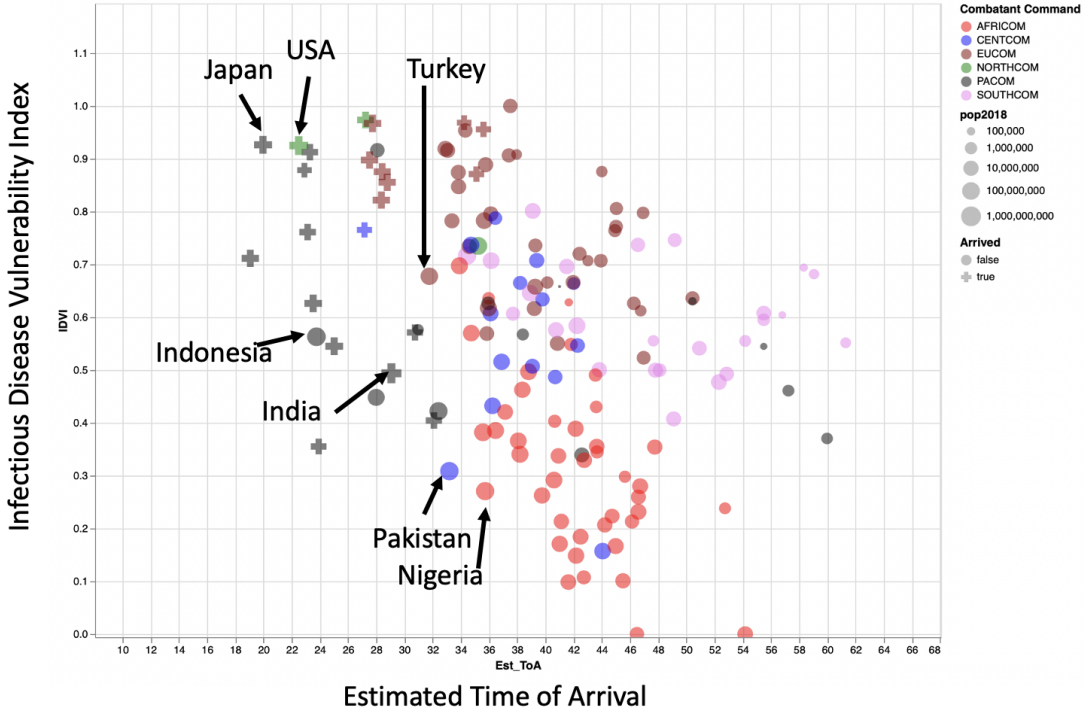
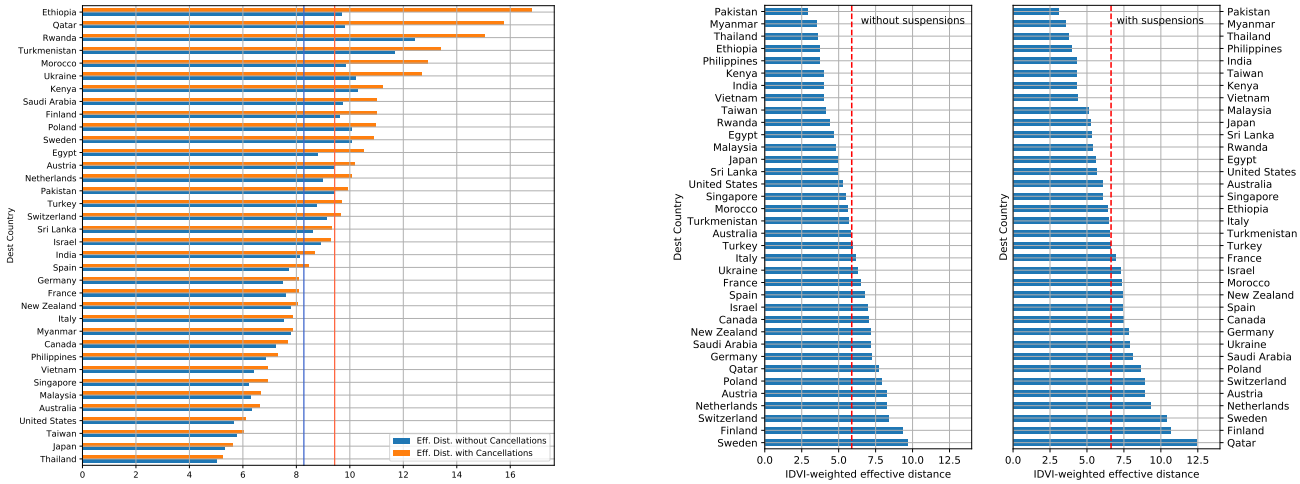


Figure 2. Estimated arrival times based on effective distance to all countries. Countries are colored by the US combatant command. Countries with official confirmed cases as of February 10th are shown with a plus marker. The full interactive visualization is available at `https://nssac.github.io/covid-19/import_risk.html`

Impact of airline suspensions

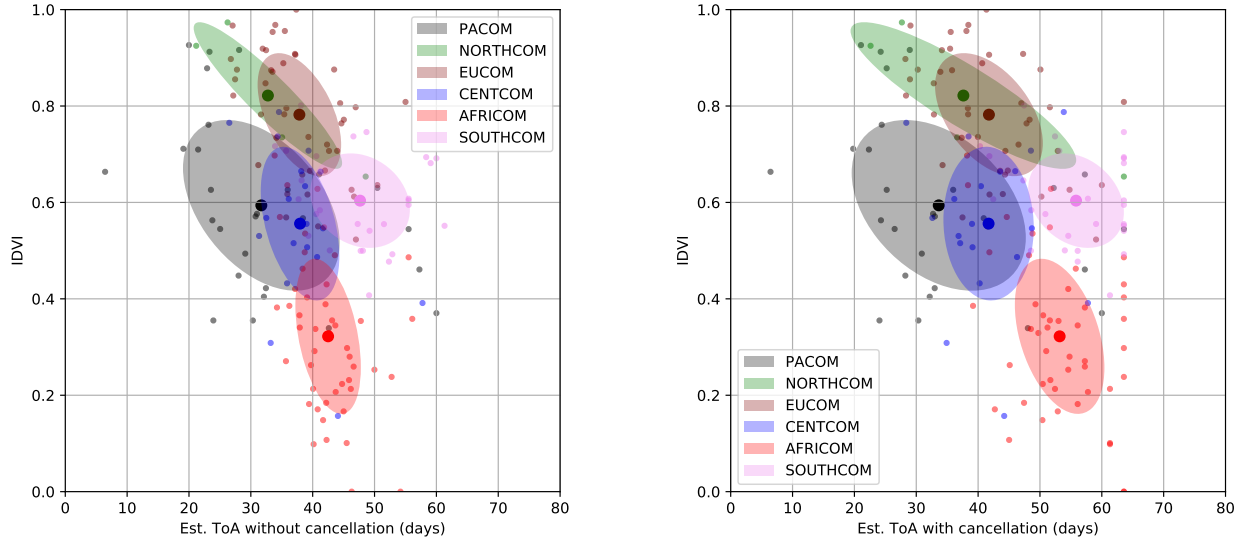
For a select set of countries, we compare and report their effective distance with and without airline suspensions and present the results in Figure 3. It is observed that for countries like Singapore, Thailand, United States, etc. which typ-

medRxiv preprint doi: https://doi.org/10.1101/2020.02.20.20025882; this version posted February 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



(a) (b)

Figure 3. (a) Effective distances to various countries from China with and without airline suspensions (vertical lines indicate the mean effective distance of the displayed countries). (b) IDVI-weighted effective distance index of various countries before and after the airline suspensions.



(a) (b)

Figure 4. Estimated arrival times vs IDVI (a) without cancellation (b) with cancellation incorporated into the model. The mean-denoted ellipses indicate the standard deviation for the point clouds corresponding to the respective combatant commands.

ically have large flow of traffic during normal operations, the change in effective distance after reduced connectivity is small. This could be attributed to the fact that current data on airline suspensions mainly include carriers from respective countries, and do not have official records for Chinese carriers serving these countries. In the current model, we have retained flight traffic contributed by Chinese carriers (which constitute a large share of the market) at full capacity. We rank (from high risk to low risk) the countries based on difference in the effective distance and compare their relative standing before and after airline suspensions in Figure 3b. It is observed that countries such as Qatar and Ethiopia see large increase in their effective distance, when weighted by IDVI Qatar's risk rank drops whereas Ethiopia remains high (due to low IDVI).

We select the countries for which there have been no imported cases, and compute their new estimated time of arrival using the updated effective distance. For this process, we use the regression coefficients that were computed using known arrival times. Figure 4 shows the scatter plots of estimated ToA vs. IDVI for the countries before and after the

medRxiv preprint doi: https://doi.org/10.1101/2020.02.20.20025882; this version posted February 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

imposition of airline suspensions. The ellipses indicate the σ confidence interval of the point clouds corresponding to the respective combatant commands. The increase in ellipse widths and mean shift to the right (due to constant IDVI) between Figure 4a and Figure 4b indicate the effect of flight cancellations on estimated ToA. On an average, countries in AFRICOM see the largest average increase in estimated ToA of nearly 11 days, followed by countries in SOUTHCOM showing an average increase of 8 days.

DISCUSSION

Our work is an initial attempt at quantifying the impact of airline suspensions on COVID-19 direct importation risk. Some of the limitations of the current work can be overcome with more timely data availability and improved model assumptions. Firstly, the current observations are based on first official reports which in some cases can be quite different from first importations. Also, due to the evolving nature of the outbreak and limited observations, the linear estimator coefficients could change with new reports and altered travel conditions.

While air traffic data from IATA allows us to quantify population exposure, (a) it is dated and may not be reflective of current conditions; (b) may not be representative of all human mobility between the countries. As we observed, some countries that are geographically closer to China (e.g., Nepal, Thailand and Japan) have very early arrival times (in relation to estimates based on effective distance). This highlights the need to account for multi-modal transport networks for quantifying the risk of global importations. However, this also raises concerns about other countries and regions (such as Pakistan, Myanmar and Northeastern India) which are geographically adjacent to China but haven't reported any cases yet.

Finally, we have considered China as a single origin, while there exists case counts by province. The current analysis can be extended by considering weighting the effective distance from multiple origins with their relative levels of infection. Further, while we have tried to incorporate data on airline suspensions, this may not be complete (for instance Chinese airlines are not listed), and will not capture the actual reductions in flow volumes due to travel advisories, government restrictions and behavioral adaptations. There are also a number of screening procedures in place at international airports, which could potentially delay case importations.

While providing a preliminary analysis using COVID-19 official reports and airline suspensions, this work also lays out a framework for rapid risk assessment for an emerging and ongoing outbreak. We believe that using near real-time multi-modal mobility datasets and detailed disease surveillance with qualitative and quantitative inputs on ongoing interventions and preparedness efforts will aid in swift and efficient global response to such outbreaks.

REFERENCES

1. Isaac I Bogoch, Alexander Watts, Andrea Thomas-Bachli, Carmen Huber, Moritz UG Kraemer, and Kamran Khan. Potential for global spread of a novel coronavirus from china. *Journal of travel medicine*, 2020.

2. Dirk Brockmann. *Event Horizon – COVID-19*, 2020. URL: `http://rocs.hu-berlin.de/corona/`.

3. Dirk Brockmann and Dirk Helbing. The hidden geometry of complex, network-driven contagion phenomena. *science*, 342(6164):1337–1342, 2013.

4. Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Luca Rossi, Kaiyuan Sun, Cécile Viboud, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (2019-nCoV) outbreak. *medRxiv*, 2020.

5. Market IS data. International air transport association (IATA). URL: `https://www.iata.org`.

6. P.M. De Salazar, R. Niehus, A. Taylor, C.O. Buckee, and M. Lipsitch. Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases. *medRxiv*, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.20.20025882; this version posted February 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

7. M. Gilbert, G. Pullano, F. Pinotti, and Poletto C. and Boelle P.Y. and D'Ortenzio E. Valdano, E., and Eholie S.P. and Altmann M. Yazdanpanah, Y., and B. Gutierrez. Preparedness and vulnerability of african countries against introductions of 2019-nCoV. *medRxiv*, 2020.

8. Shengjie Lai, Isaac Bogoch, Nick Ruktanonchai, Alexander Watts, Yu Li, Jianzing Yu, Xin Lv, Weizhong Yang, Hongjie Yu, Kamran Khan, et al. Assessing spread risk of wuhan novel coronavirus within and beyond china, january-april 2020: a travel network-based modelling study. *medRxiv*, 2020.

9. Melinda Moore, Bill Gelfeld, and Christopher Paul Adeyemi Okunogbe. Infectious Disease Vulnerability Index. *Rand Health Quarterly*, 6(3), 2017.

10. World Health Organization. 2019 novel coronavirus (COVID-19) situation reports, 2020. URL: `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports`.

11. Srini Venkatramanan. Flight cancellations related to 2019-nCoV (COVID-19). *University of Virginia Dataverse*, 2020. `doi:10.18130/V3/Z6524P`.

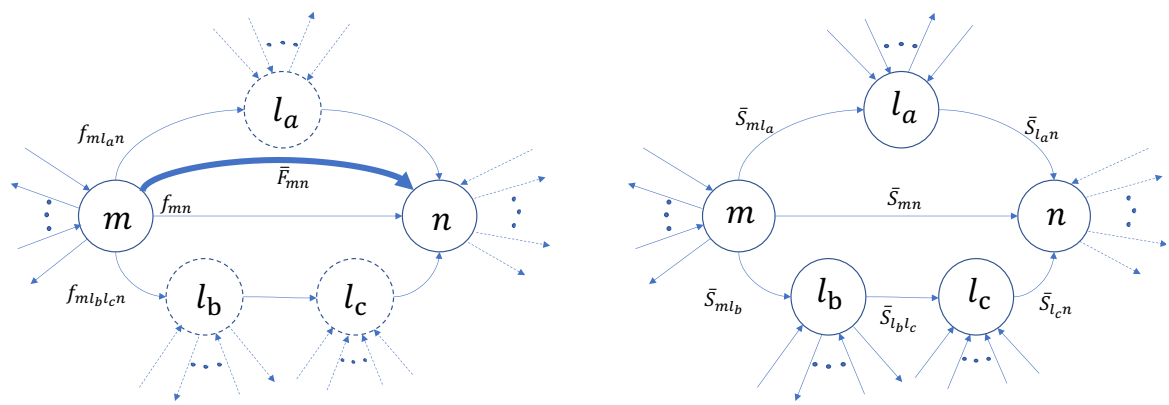
12. J.T. Wu, K. Leung, and G.M. Leung. Forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in wuhan, china: A modelling study. *The Lancet*, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.20.20025882; this version posted February 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Appendix

A EFFECTIVE DISTANCE COMPUTATION OVER IATA-BASED AND OAG-BASED NETWORKS

Consider the two networks presented in Figure 5. Here m, ℓ_a, ℓ_b, ℓ_c, and n represent the nodes and the effective distance is computed between node m and n. Figure 5a represents the network construction using IATA data for computing the effective distance between node m and n. In this construction, f_{mn} represents the number of passengers (passenger flow) on the path $[(m,n)]$. IATA data consists of all the paths (along with the flow) that were used by passengers in the travel between source m and destination n. In the example network (Figure 5a), the paths are $\Gamma = \{[(m,n)], [(m,\ell_a),(\ell_a,n)], [(m,\ell_b),(\ell_b,\ell_c),(\ell_c,n)]\}$ with respective flows as f_{mn}, $f_{m\ell_a n}$, $f_{m\ell_b\ell_c n}$. The total flow between nodes m and n is the sum of the flows through all the paths connecting them and we denote it as $F_{mn} = f_{m\ell_a n} + f_{mn} + f_{m\ell_b\ell_c n}$. With $\sum_k F_{mk}$ as the total outflow from node m, we define $\bar{F}_{mn} = F_{mn}/\sum_k F_{mk}$, as the probability of a traveller exiting node m has destination node n. The effective distance between nodes m and n is defined as $D^{\text{IATA}}_{mn} = 1 - \log \bar{F}_{mn}$.



(a) (b)

Figure 5. A toy example that compares networks derived from IATA and OAG data. Nodes n and m are considered as source and destination, respectively. (a) Network constructed using IATA data. $\bar{F}_{mn} = f_{mn}/\sum_k F_{mk}$, $\bar{F}_{m\ell_a n} = f_{m\ell_a n}/\sum_k F_{mk}$, and $\bar{F}_{m\ell_b\ell_c n} = f_{m\ell_b\ell_c n}/\sum_k F_{mk}$, the bold arrow is used to represents $\bar{F}_{mn} = \bar{f}_{m\ell_a n} + \bar{f}_{mn} + \bar{f}_{m\ell_b\ell_c n}$, the total fraction of flow from origin m to destination n. (b) Network constructed using the OAG data. The dashed lines are used to represent nodes and outflows that do not contribute to the computation of the edge weights in the graph.

As for network construction using OAG data (cf. Figure 5b) we denote S_{mn} as the total number of seats between airports m and n. (Mention random walk over the graph). $\bar{S}_{mn} = S_{mn}/\sum_k S_{mk}$ denotes the probability of a person exiting node m (as either origin or transit) has the next stop at node n and the effective distance is defined as $d_{mn} = 1 - \log \bar{S}_{mn}$. In [3], the authors consider all the paths that can be constructed between the two nodes which in this case is the set Γ and define the effective distance as the shortest path between the two nodes: $D^{OAG}_{mn} = \min_{\gamma\in\Gamma} \lambda(\gamma)$, where $\lambda(\gamma) = \sum_{(i,j)\in\gamma} d_{ij}$ is the length of path γ.