

Noname manuscript No.  
(will be inserted by the editor)

# A Model-Based Approach To Assess Epidemic Risk

Hugo Dolan · Riccardo Rastelli

Received: date / Accepted: date

**Abstract** We study how international flights can facilitate the spread of an epidemic to a worldwide scale. We combine an infrastructure network of flight connections with a population density dataset to derive the mobility network, and then define an epidemic framework to model the spread of disease. Our approach combines a compartmental SEIRS model with a graph diffusion model to capture the clusteredness of the distribution of the population. The resulting model is characterised by the dynamics of a metapopulation SEIRS, with amplification or reduction of the infection rate which is determined by the mobility of individuals. Then, we use simulations to characterise and study a variety of realistic scenarios that resemble the recent spread of COVID-19. Crucially, we define a formal framework that can be used to design epidemic mitigation strategies: we propose an optimisation approach based on genetic algorithms that can be used to identify an optimal airport closure strategy, and that can be employed to aid decision making for the mitigation of the epidemic, in a timely manner.

**Keywords** COVID-19 · SEIRS compartmental model · Genetic Algorithm · Network Analysis · Human Mobility

## 1 Introduction

In recent years, the extensive development of the transportation infrastructure has radically changed how connected our world is. International flights allow individuals to travel around the globe in just a few hours or days. This has

---

Hugo Dolan  
University College Dublin  
E-mail: hugo.dolan@ucdconnect.ie

Riccardo Rastelli  
University College Dublin  
E-mail: riccardo.rastelli@ucd.ie

important negative implications on the potential spread of pandemic diseases, whereby epidemics can reach a worldwide scale before effective responses are set in place. The recent COVID-19 outbreak [17] has clearly raised and emphasised this problem. As a response to the emergency, many countries have taken drastic measures to contain and slow down the spread of the virus by imposing lockdowns and airport closures [5]. While these measures have been successful in confining the epidemic, the immediate and chaotic response has blurred the actual role played by the topology of the infrastructure network on the spread of the virus.

The main goal of this paper is to create a model-based framework that can inform decision making regarding airport closures as a means to slowing down a potential pandemic without causing excessive economic damage. In particular, we introduce a new framework to study networks of international flights as potential vehicles for the spread of pandemic diseases.

In this paper, we first propose an in-depth analysis of the Open Flights network dataset [13], which describes a large number of flight connections between more than 3,000 airports. We calculate a number of descriptive statistics from the data, in order to study the underlying topology of this infrastructure network, and essentially to understand how individuals can move between distant locations. We use various centrality measures to identify key airports, and we test the resilience of the network when these key airports are removed.

Then, we fit a Stochastic BlockModel (SBM) to partition the airports into homogeneous groups. The SBM, originally introduced and studied by [20], is a fundamental model and tool for statistical network analysis, since it can highlight groups of nodes that exhibit similar connectivity patterns. Inference for the SBM can be performed using both classical and Bayesian approaches [12]. One fundamental aspect of this model is that it can be interpreted as a finite mixture model for networks [6], and thus it borrows many concepts and tools from this related research area. In this context, a useful by-product of the SBM framework is that it allows us to compare the partition on the airports with their actual geographical location.

After the exploratory data analysis, we use the infrastructure network to create a model for the simulation of epidemics. An essential aspect of this task is the development of a statistical network model that can combine these flight routes data with the geographical distribution of the population. Our aim is to give a model-based quantification of the pandemic risk which is amplified by the travelling of individuals, and to possibly identify effective interventions that can mitigate this risk. [5] propose an approach similar to ours, in that they combine the infrastructure network with the gridded population data to study the effects of the airport closure interventions that were actioned at the beginning of this year. [5] use a tool called GLEAM [4, 3] which can combine data from different sources to predict the behaviour of the epidemic using an individual-based compartmental SEIR model. While they focus on the effects following the actual airport closures, in this paper we aim at defining a framework to take new decisions that can lead to optimal airport closures, or potential future airport reopenings.

Our approach relies on epidemic compartmental models [8], and in particular on a SEIRS model. This framework postulates that the population is divided in 4 ordered compartments (susceptible, exposed, infectious, recovered) and that different rates determine the flow of individuals from one compartment into the next, and eventually back into the first compartment. This family of models has been largely employed in various research fields to model the evolution of epidemics, and it has been also successfully used within the context of COVID-19 [9, 5, 15, 16, 10].

One fundamental aspect of our epidemic model is that, similarly to [5, 19, 2], we consider a metapopulation where each subpopulation is centred at an airport location, and whereby the local epidemic is determined by a SEIRS model. We use a graph diffusion process to describe the flows between the various subpopulations, which in turn affects the local dynamics. Not only this allows us to observe the epidemic locally for each subpopulation, and globally, but it also allows us to appreciate the temporal-geographical progression of the virus.

We calibrate the parameters of our model using the recent COVID-19 pandemic as an example and as a meter of comparison. We do not claim that our results are specific to the COVID-19 epidemic nor that they should be used within this context; rather, we provide a general methodology that could be employed in any epidemic setting, and, for example purposes, we recreate realistic situations that resemble the recent COVID-19 epidemic. In fact, we test the sensitivity of our model by running a number of simulations that encompass a variety of possible real scenarios.

The fundamental contribution of this work regards the study of optimal epidemic mitigation strategies. Once we possess a model which is calibrated to a realistic setting, we explore an optimisation approach to identify what could be the optimal airport closure strategies that should be implemented. We use the predictions from our epidemic model to construct an objective function that takes into account measures for the spread of the disease as well as economic losses. We perform the optimisation using Genetic Algorithms (GAs) [18]. GAs are heuristic stochastic optimisation algorithms that explore new candidate solutions by selecting and transforming a set of current solutions using some basic principles of evolution and natural selection. In our context, GAs are especially convenient since the problem that we address is a combinatorial one, where we want to find the optimal subset of airports that should be closed to minimise our objective function.

The software that we have used in this paper is maintained by one of the authors and is available from the GitHub repository: [1].

## 2 Network Topology

In this section we propose an exploratory data analysis and basic statistical modelling of the Open Flights dataset [13]. The Open Flights dataset contains information on 3,425 airports globally, including a database of 37,594 commer-

cial routes between these airports collected in 2014. The dataset is transformed into an adjacency matrix with nodes representing airports and directed edges representing whether there exists a direct route between any two airports. In Figure 1 we present the network visually, and on initial inspection it is clear that the network exhibits extremely high degree of connectivity, with the plot of degree distribution indicating that over 20% exhibit a degree greater than 10.

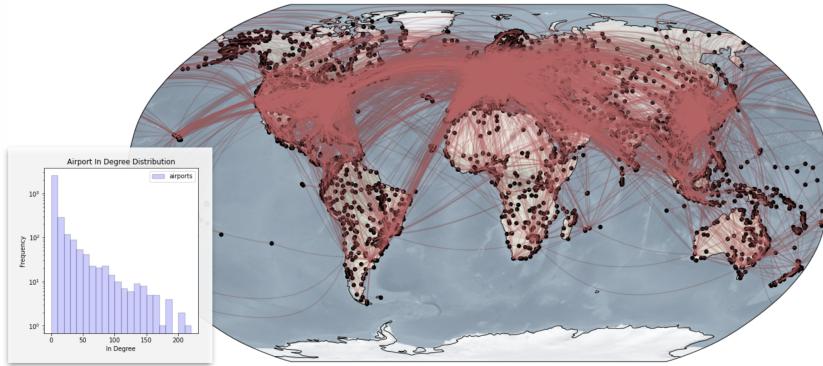


Fig. 1: Open Flights Network Visualisation and its in-degree distribution.

Identifying airports which are important to the overall connectivity of the network is crucial in gaining a better understanding of the network’s topology. We consider several metrics to measure the importance of nodes. These include the Page-Rank centrality, betweenness centrality, coreness ranks, as well as the in-degrees and out-degrees of nodes [11]. We present a table of the 20 most important airports according to Page-Rank in Table 1. Airports with high Page-Rank are clearly major international destinations, and they form an extremely well connected subnetwork with a coreness of over 60, meaning every airport in this subnetwork has in-degree + out-degree higher than 60, or, equivalently, since the graph is approximately symmetric (the majority of air routes run return flights), we can say that every airport in this subnetwork has an out-degree close to or exceeding 30. It is interesting to note that airports with high betweenness centrality (Charles De Gaulle, Dubai, Beijing, Amsterdam, Los Angeles, Toronto, Frankfurt) are also major connecting flight hubs, exhibiting out-degrees of over 200.

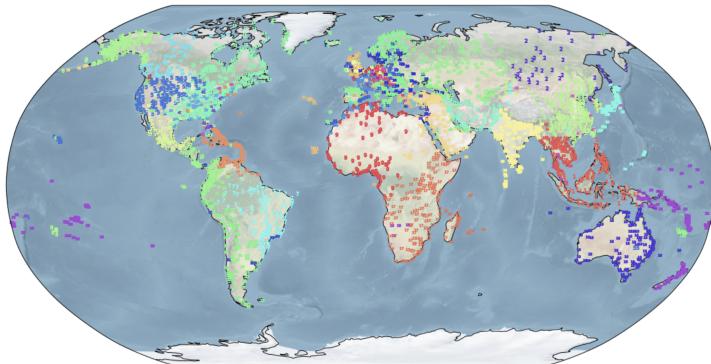
We identify homogeneous subgroups of airports within the network by employing a SBM framework: we utilise a Python implementation of an efficient Markov chain Monte Carlo method, which is suitable for inferring SBMs in large networks, as described by [14]. The optimal SBM partition and the corresponding block matrix are shown in Figures 2a and 2b, respectively. It is clear that the communities found are very strongly associated with the geograph-

Table 1: Summary Statistics for the top 20 airports, sorted according to Page-Rank centrality values.

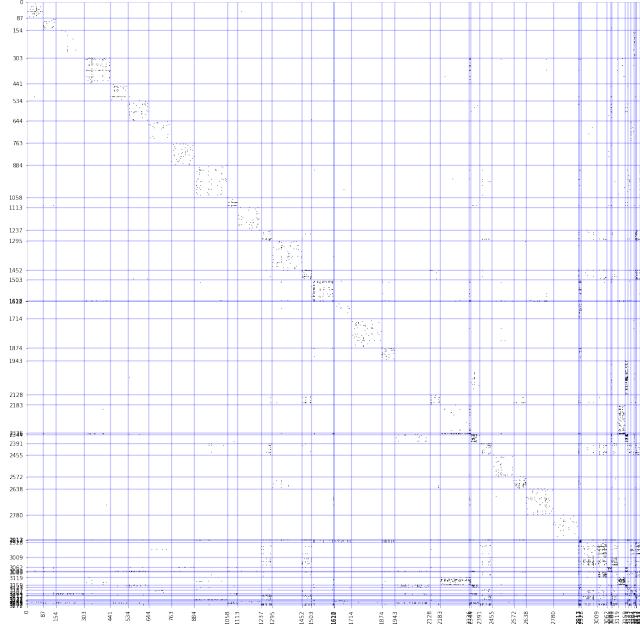
Airport	PageRank	Betweeness	Coreness	deg(+)	deg(-)	Country
ATL	0.0047	0.0294	60	216	<b>217</b>	United States
IST	0.0044	<b>0.0412</b>	62	230	<b>227</b>	Turkey
ORD	0.0043	<b>0.0474</b>	60	203	<b>206</b>	United States
DEN	0.0043	0.0262	60	168	169	United States
DFW	0.0042	0.0251	60	185	187	United States
DME	0.0041	0.0294	62	189	189	Russia
CDG	0.0039	<b>0.0617</b>	62	233	<b>237</b>	France
FRA	0.0038	<b>0.051</b>	62	238	<b>239</b>	Germany
PEK	0.0038	<b>0.0492</b>	60	206	<b>206</b>	China
DXB	0.0036	<b>0.0594</b>	62	182	188	United Arab Emirates
AMS	0.0036	<b>0.0427</b>	62	231	<b>232</b>	Netherlands
IAH	0.0035	0.023	60	168	169	United States
LAX	0.0033	<b>0.0662</b>	60	148	149	United States
SYD	0.0032	0.0326	45	83	85	Australia
YYZ	0.003	<b>0.0425</b>	60	146	147	Canada
JFK	0.003	0.0258	62	160	162	United States
BOG	0.0029	0.0248	40	74	74	Colombia
PVG	0.0029	0.0221	56	153	152	China

ical location of the airports, and with their region or province. This is quite surprising as this information is not encoded explicitly in the data provided to the algorithm. This would strongly suggest a high degree of connectivity of airports not only globally but also within regions or geographical areas. We also note from the block matrix in Figure 2b that the majority of connections are within relatively large communities representing the geographic clustering observed in Figure 2a, but also towards the lower right corner there is significant disassortative behaviour, likely these nodes are large international hubs such as the small community of London, Frankfurt, Amsterdam, Charles De Gaulle which share connections to many cities across the world.

We continue our exploratory analysis by studying the percolation properties of the network [11]. We percolate the network by sequentially removing the nodes and their connections, and observing how the connectivity of the graph changes. We remove the nodes following both a random order, and following a decreasing order of the out-degree and other centrality measures. The results are shown in Figure 3. The network is highly resilient to random attacks, since the removal of almost all nodes is required in order to disrupt network connectivity. However, we note that the network is moderately more vulnerable to targeted (degree based) attacks, yet this would still require more than half of all airports to be removed for the single giant component to disappear. Sim-



(a) Map of the SBM results, with colours indicating the cluster memberships.



(b) Block matrix visualisation of the adjacency matrix, where the black dots indicate routes between airports and the blue boxes indicate the clusters.

Fig. 2: Fitted SBM results.

ilarly, the network is moderately vulnerable to targeted attacks according to other ranking factors (Page-Rank, betweenness, coreness). We note that these procedures are also averaged over many trials to account for the removal of vertices of equal rankings in different orders, however, we find these results very quite similar to percolation by degree.

In conclusion, the Open Flights network summary statistics show that airports which are large regional destinations or hubs for connecting flights tend

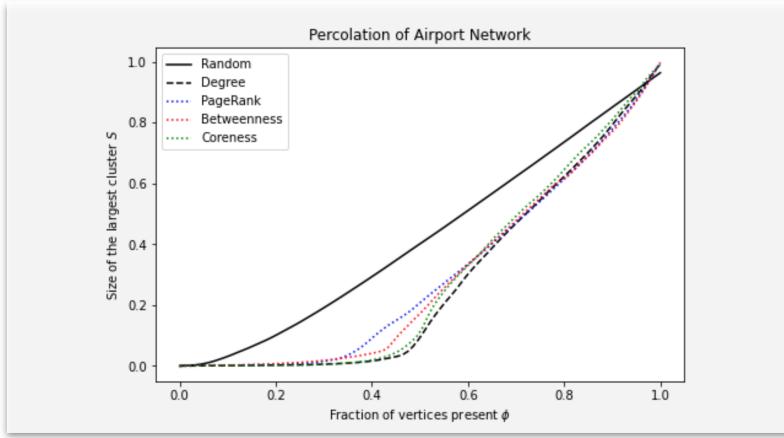


Fig. 3: Percolation on the Open Flights dataset via a variety of ranking criteria (results are averaged over 100 independent simulations).

to have high importance to network connectivity. Furthermore it is observed that some nodes in the network are extremely well connected, both at regional and global level, with significant geographical community structure. The network is also highly resilient to random or deterministic damage.

In the context of epidemics, these initial findings provide a very solid evidence that the flight connections can sadly be a very efficient vehicle to facilitate the spread of diseases, and, more importantly, that substantial network “damage” (e.g. airport closures) is required to ensure that the disease does not spread to a worldwide scale. This evidence motivates our work, in that we aim at finding optimal mitigation strategies that can reduce the pace of the epidemic to a much smaller scale, without causing excessive disruption to economies and to this particular infrastructure network.

### 3 Model Specification

#### 3.1 Theoretical Underpinnings

Before we develop the main model of this paper, we must first introduce two existing models which can be found in the domains of applied mathematics and epidemiology [11]. Firstly we specify the graph diffusion model, which describes the flow of a fluid across a network, driven by pressure differences between adjacent nodes. This can be expressed as a vector of differential equations denoting changes of fluid volumes at each node and time step:

$$\frac{d\psi}{dt} = c(A - D)\psi$$

We use the notation  $\psi$  to represent the vector of fluid volumes at every node,  $\mathbf{A}$  to denote the adjacency matrix of the network and  $\mathbf{D}$  to denote a diagonal matrix of containing the degrees of every node. A full derivation of this can be found in [11].

Additionally, we introduce the SEIRS compartmental epidemiology model. Each letter of the model name denotes a compartment of the system (Susceptible ( $\mathbf{S}$ ), Exposed ( $\mathbf{E}$ ), Infectious ( $\mathbf{I}$ ) and Recovered ( $\mathbf{R}$ )), in which some number of individuals from the total population ( $\mathbf{M}$ ) reside. Figure 4 illustrates the direction of progression from state to state, whilst Equation 2 indicates the exact rates at which the population in each compartment changes. The

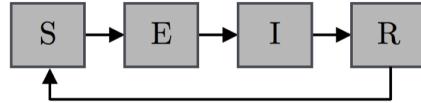


Fig. 4: SEIRS compartmental flow diagram for the equation set:  $\frac{dS}{dt} = \delta R - \frac{S\beta I}{M}$ ,  $\frac{dE}{dt} = \frac{S\beta I}{M} - \epsilon E$ ,  $\frac{dI}{dt} = \epsilon E - \gamma I$ ,  $\frac{dR}{dt} = \gamma I - \delta R$

greek letters are parameters of the system which can be fitted to match the characteristics of some observed epidemic. Whilst the system cannot be solved analytically, we can find a numerical approximation to the solution, which is sufficient for our simulation purposes.

### 3.2 Model Definition

In order to model the transmission of disease through the international flights network, we opt to use the SEIRS model combined with a graph diffusion model as described in the previous section. We will refer to the airports' adjacency matrix as  $\mathbf{A}$  and denote airport nodes as  $v_j ; j = 1, \dots, N$ . The total number of nodes in the network is  $N$  and the associated population at each node is  $M_j$ . Let us define the local epidemic state vector as  $\theta_j(t) = (S_j \ E_j \ I_j \ R_j)^T$ , which represent the compartments of the SEIRS model for airport population at any given time. A condition of the SEIRS model constrains the total population of all compartments to equal the total population:

$$S_j(t) + E_j(t) + I_j(t) + R_j(t) = M_j$$

We assume that the local population is fully mixed (i.e. everyone has equal chance of being infected), as this is a standard assumption of compartmental epidemic models, however we assume this only to be true at the individual airport level, and not for the entire global system. Additionally we introduce  $\alpha_j$  as the proportion of the population which can travel, and  $c$  as the probability that an individual departs from an airport on any given day. The proportions

allow us to define an additional variable  $\psi_j(t) = \alpha_j \theta_j(t)$  which corresponds to the mobile epidemic state.

We can define at high level our simulation procedure:

1. At each time-step we must first initiate community spread of the disease through each airport's local population. This is represented by the compartments of its SEIRS model denoted by the local epidemic state vector  $\theta_j$ . Using a fourth-order Runge-Kutta approximation technique we update each compartments value in accordance with the dynamics of the SEIRS model. The new state of the compartments after community spread is denoted  $\theta^*$  as opposed to the previous state  $\theta$ , notice that we drop the airport subscript  $j$  as this process generalises to a matrix  $\theta$  whose columns are vectors  $\theta_j$ .
2. Now take this new local epidemic state  $\theta^*$  and split it into two types of population members: a base population denoted  $\theta_B$ , who are permanent residents to the local area; and a second group  $\theta_T$ , i.e. the transient population who are temporary visitors, who have arrived at the airport on business or holidays and will return home after a short period. This differentiation is necessary to ensure that the system does not arrive at an unrealistic equilibrium, where airports have drastically different populations relative to their initial values (we assume no permanent migration in our model).
3. Next, we compute the proportions of  $\theta_B$  and  $\theta_T$  who can fly (in each compartment), this is simply achieved by multiplying the vectors by  $\alpha_j$  which was defined at population level  $\theta_j$ , such that  $\psi_+ = \alpha \theta_B$  describes the number of outbound passengers and  $\psi_- = \alpha \theta_T$  describes the number of returning travellers.
4. Using our diffusion model compute the changes to  $\theta_B$  and  $\theta_T$  at each airport. The exact value of these changes is weighted based on several factors including: the differences between outbound and returning passengers at every connected airport (corresponding to the “fluid”, in the interpretation of Section 3.1), the relative importance of the airports in the network and the node's degree (controlling how the travellers get distributed across routes).
5. Recombine updated values of  $\theta_B$  and  $\theta_T$  into the aggregate populations  $\theta$  and loop for as many iterations of simulation as required.

For completeness we include both a diagrammatic form of the algorithm (Figure 5) as well as a the full algorithm (Table 2) which reflects the high level overview above. A full derivation is provided in Appendix 7.1. The algorithm described above has been vectorised so that  $\theta$  is now a  $4 \times N$  matrix containing the states for all airports  $\theta(t) = (\theta_1(t) \dots \theta_N(t))$ , and similarly for  $\psi_+$ ,  $\psi_-$  which represent the states of outbound and returning travellers respectively. The matrix  $1/D$  is a diagonal matrix with entries that are  $1/\deg(v_j)$ , this normalises outflows from airports preventing more passengers from leaving a node when there are not individuals in that location. The matrix  $B$  is an operator encoding the differential equations of the SEIRS model for vectorised

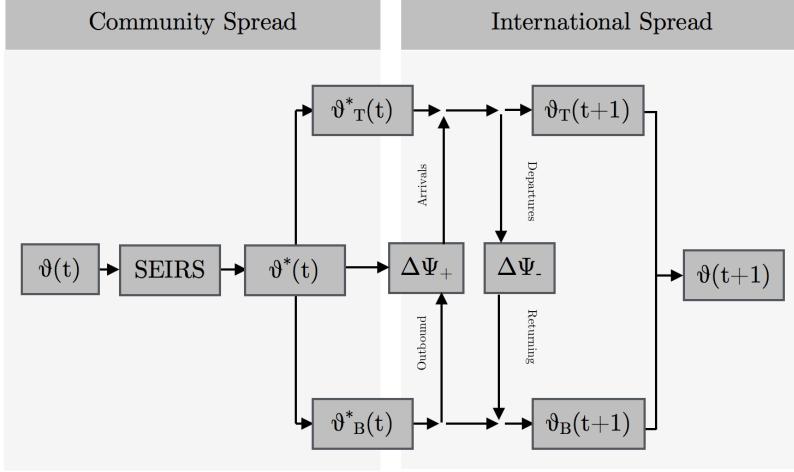


Fig. 5: Diagram of the epidemic diffusion model steps.

Table 2: Simulation pseudocode.

**Algorithm: Flow & Degree Corrected Epidemic Diffusion Model**

<b>for</b> $t = 1, \dots, T:$	
(5)	$\vartheta(t-1) = \vartheta_B(t-1) + \vartheta_T(t-1)$
(1)	$\vartheta^*(t-1) = \vartheta(t-1) + \mathbf{B}\tilde{\boldsymbol{\theta}}$
(2)	$\vartheta^*_B(t-1) = ( \vartheta_B / \vartheta ) * \vartheta^*(t-1)$ $\vartheta^*_T(t-1) = ( \vartheta_T / \vartheta ) * \vartheta^*(t-1)$
(3)	$\Psi^*_{+}(t-1) = \vartheta^*_B(t-1) \alpha_+$ $\Psi^*_{-}(t-1) = \vartheta^*_T(t-1) \alpha_-$
(4)	$\Delta\Psi_{+}(t) = -c_{+}\Psi^*_{+}(t-1) (1/D) (\mathbf{I} - \mathbf{C}^T)$ $\Delta\Psi_{-}(t) = -c_{-}\Psi^*_{-}(t-1) (1/D) (\mathbf{I} - \mathbf{C}^T)$ $\vartheta_B(t) = \vartheta^*_B(t-1) + \text{Min}(\Delta\Psi_{+}(t), 0) + \text{Max}(\Delta\Psi_{-}(t), 0)$ $\vartheta_T(t) = \vartheta^*_T(t-1) + \text{Max}(\Delta\Psi_{+}(t), 0) + \text{Min}(\Delta\Psi_{-}(t), 0)$

application to many airports simultaneously, whilst  $\tilde{\boldsymbol{\theta}}$  is a modified version of  $\boldsymbol{\theta}$  to enable  $\mathbf{B}$  to be applied as a linear operator. Finally  $\mathbf{C}$  is a weighted version of adjacency matrix  $\mathbf{A}$ , so that the outflows from vertices reflect the relative importance of adjacent airports. The above should provide an intuition, however we reiterate that there is a far more comprehensive derivation of the mathematics required in Appendix 7.1.

### 3.3 Data Sources, Parameters, Estimation

In this section we provide further information on the parameters of our model, on their interpretation, and on how we have used different data sources to infer or calibrate these parameters. Our framework requires information on airports,

routes, demographics, wealth, which is difficult to obtain and use. A summary of the model's parameters is provided in Table 3, for convenience.

Table 3: Model parameters [\*] necessary conditions for the escalation of the epidemic.

	Parameter	Description	Dimension	Range
Data Based	$\theta(0)$	Initial Populations (S,I,E,R) for each airport.	$\mathbf{R}^{4 \times N}$	(N/A) Gridded Population of the World (GPW), SEDAC
	$C$	Relative Centrality	$\mathbf{R}^{N \times N}$	(N/A) Open Flights Data
	$\alpha_+$	% Base Population who can fly	$\mathbf{R}^{N \times N}$ (Diagonal)	(N/A) World Bank Data
Model Assumption	$\alpha_-$	% Transient Population who can fly	$\mathbf{R}^{N \times N}$ (Diagonal)	$\alpha_- = I$ , assuming no permanent migration.
	$c_-$	% of people who return at the end of the week	$\mathbf{R}$	$c_- = 1$ , assuming no permanent migration
	$c_+$	% of people who depart on any given day	$\mathbf{R}$	$c_+ \in (0,1)$
Heuristic Based	$\beta$	Rate of infection / contact (Daily)	$\mathbf{R}$	$\beta \in (0, \infty) ; \beta > \epsilon > \gamma$ [*]
	$\epsilon$	Rate of transfer from exposed to infectious state (Daily)	$\mathbf{R}$	$\epsilon \in (0, \infty)$
	$\gamma$	Rate of transfer from infectious to recovered state (Daily)	$\mathbf{R}$	$\gamma \in (0, \infty)$
	$\delta$	Rate of loss of immunity (Daily)	$\mathbf{R}$	$\delta \in (0, \infty)$

One fundamental transformation that we apply is the following: let  $\mathbf{P}$  be the vector of Page-Rank values obtained from earlier analysis and  $\mathbf{A}$  be the adjacency matrix of flight connections. Then, we define and work with the matrix  $\mathbf{C}$ , which is the relative centrality matrix with elements defined by  $C_{ij} = \frac{A_{ij}P_j}{\sum_k A_{ik}P_k}$ . The interpretation of this new matrix is that its values are identical to those of  $\mathbf{A}$ , except now the edges have been assigned weights based on the relative importance of adjacent nodes. This transformation ensures that the flows of individuals between airports are scaled so that they reflect higher traffic to major airports and less to airports of little significance in the network thus preventing the system from obtaining unrealistic configurations as the simulation progresses.

With regard to the metapopulation, we must define the number of individuals that have access to, and are served by, each of the airports. This corresponds to estimating the initial population parameters  $\theta$ . For this task, we utilise the gridded population of the world dataset [7], and we assign a value for total population at each airport location. In order to perform this assignment there is an immediate problem as many airports are often in close proximity to each other (for example some cities are served by multiple airports). We assume that the maximal distance anyone will travel to reach an

airport is 240km ( $60 \text{ kph} * 4 \text{ hours} = 240\text{km}$ ). Using this assumption, for each cell of the grid of the population dataset, we search all the airports that are within this radius. Then, we assign a population contribution to each, proportionally to their Page-Rank values. This metric is chosen as it provides a proxy for the likelihood that the airport will be preferred by travellers in the local region. Note that this approach will automatically exclude population grid cells which are not within 240km of any airport. These populations are excluded from the simulation as they are unlikely to be flying regularly, thus they will not play a relevant role in the spread of the disease.

Finally, we estimate the percentage of each population who can fly  $\alpha_+$ . It is obvious that this percentage will vary across countries, depending on a number of factors, primarily wealth. To find a reasonable value for this model parameter, we acquire passenger estimates by country as supplied by the World Bank and divide these through by the total airport populations for the given country. We then assume the proportion by airport is the same as at country level<sup>1</sup>.

The SEIRS model, as well as the graph diffusion model, rely on several strong assumptions that inevitably impact the results. For clarity, we provide a list of our modelling assumptions here below, to highlight the specific features that our model will exhibit.

1. **Fully Mixed Local Populations:** Within any given node every member of the population has equal chance of contact and thus passing on disease.
2. **Fully Mixed Wealth:** The proportion of population which may fly is distributed the same at node / airport level as at country level.
3. **Maximal Travel Distance:** We assume the maximum distance someone will travel is 240km to get to an airport, and thus anyone who exists outside of all airport radius is assumed to be isolated and may be excluded from the model<sup>2</sup>.
4. **Air Transit Only:** We assume the only way for the disease to spread between nodes is via air routes and that spread via other means eg. boat & road links are negligible.
5. **No Permanent Immigration:** Assume all individuals will return to their home country by the end of the business week.
6. **No Seasonality:** We assume that all parameters of the model are constant throughout the duration of simulation.
7. **Universal Rates:** We assume that the parameters of the SEIRS model are universal and do not vary significantly between countries.

While it is obvious that each of these assumptions can have a non-negligible effect, we do argue that our simulations can show a remarkable resemblance with a realistic epidemic, such as the recent COVID-19 case.

<sup>1</sup> Sometimes this proportion is greater than 1 particularly for major hubs, where annual traffic through the airport may exceed the size of the local population assigned. In these cases we replace our estimate by the average global proportion, clearly this is imperfect but is likely as accurate as we will be able to obtain given this information is not readily available.

<sup>2</sup> The percentage of total population excluded by the model is less than 3%

## 4 Simulations

### 4.1 Visualisation of Unmitigated Spread

Now that we have outlined the theory and processes to develop our model, we proceed to simulate and visualise an unmitigated epidemic (i.e. no measures to decrease the infection rate). We set the SEIRS model parameters as  $\epsilon = 0.14$ ,  $\delta = \frac{1}{730}$ ,  $\gamma = 0.048$  and  $\beta = 0.4$ . These values were obtained from a study by [9], with the additional parameter  $\delta$  which is set conservatively as we do not know the duration of which people will remain immune to COVID-19. We also setup the model such that the first cases occur in Wuhan for results with similarity to the COVID-19 outbreak [17].

We proceed to aggregate the time series data for the  $S$ ,  $E$ ,  $I$ ,  $R$  compartments to community levels (here, a community is a geographic region which we have labelled according to SBM resulting partition), and we consider the progression of the epidemic with regards to the most relevant compartments: Exposed ( $E$ ) and Infected ( $I$ ). The community labelling choices are somewhat arbitrary but the main purpose of these plots in Figure 6 is to illustrate the network effect results in a somewhat staggered epidemic across different geographic communities. There is evidence of a gradual dispersion of the virus

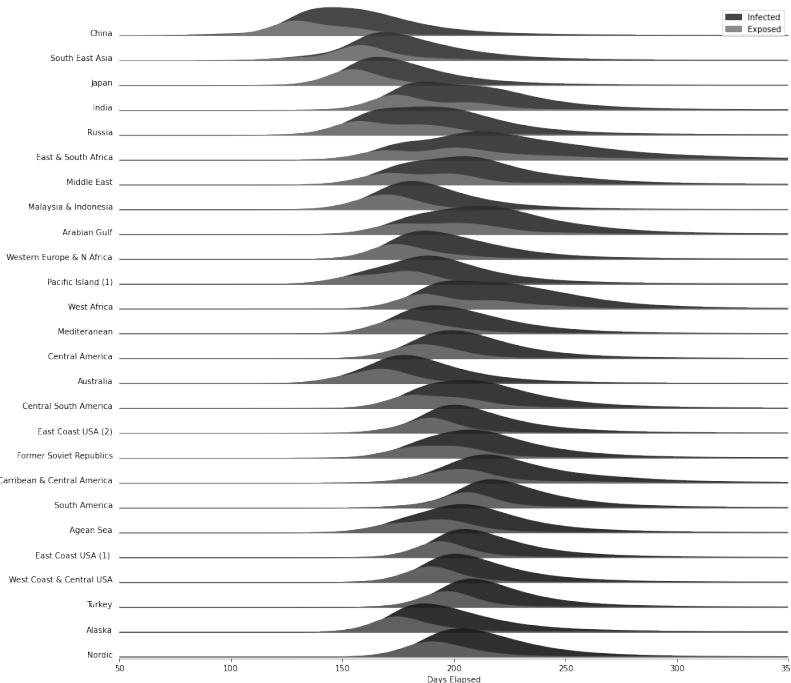


Fig. 6: Exposed and Infectious local epidemic states for key communities derived from SBM. (Y-axis scales are not comparable)

across the world starting in China, moving onward into other areas including South East Asia, Japan, Russia, India, South Africa, Middle East. Some of the last places to be infected are the Americas, Nordic Countries, Alaska and Turkey. Figure 7 is an illustrative sketch of one possible spread through the network which could be derived from the spread sequence in Figure 6.

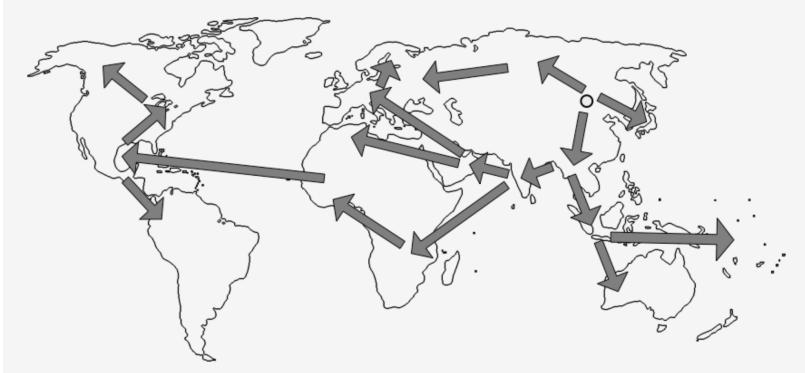


Fig. 7: Illustrative sketch of the epidemic based on the sequencing of the epidemic spread provided by Figure 6.

#### 4.2 Sensitivity Analysis of Key Parameters

Where possible, we have utilised data to calibrate the model parameters, and where not available, we have elected to make some simplifying assumptions regarding travel behaviour. However, we have not yet addressed the estimation or calibration of the SEIRS parameters in a more formal way. In the previous section, we have obtained the results for the unmitigated scenario using the values estimated by [9]. In this section, in order to obtain a better understanding of the impact of these SEIRS parameters on our results, we proceed to conduct a sensitivity analysis.

Whilst we can be reasonably confident that reported rates of recovery are reliable, this is perhaps less true for infection and exposure rates. Thus we let  $\beta = k\gamma$  (infection rate) and  $\epsilon = s\gamma$  (exposed to infectious rate) where we assume  $\gamma$  (recovery rate) is known. It is clear that  $k > 1$  otherwise  $\beta < \gamma$  and the epidemic quickly vanishes, similarly let  $k > s > 1$  to ensure valid parametrisation of the model. We will not consider varying  $\delta$  (loss of immunity rate) as we will primarily focus our analysis exclusively to the first wave of the epidemic. Let  $k \in \{1.1, 1.5, 2, 2.5, 3, 4, 5\}$ , and  $s \in \{1.05, 1.25, 1.75, 2, 2.5, 3\}$  with  $\gamma = \frac{1}{16}$ . We choose a smaller range for  $s$  as medical research suggests that people remain in the exposed state for up to 14 days for COVID-19. We see from Figure 8 that the model behaves

Dubai International Airport										Atlanta International Airport														
s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	
Time to $I(t) > 10K$	1.05	-	-	443	335	276	213	179	1.05	+	+	+	+	458	350	291	1.05	-	-	385	316	243	203	
Max $I(t)$ (infectious)	1.25	-	-	410	310	255	197	165	1.25	-	+	+	+	425	323	270	1.25	-	-	474	356	293	225	
Time to $I(t) > 1000$	1.75	-	-	359	270	222	171	143	1.75	-	-	+	450	368	280	232	1.75	-	-	414	310	255	195	
Max $I(t)$ (infectious)	2.0	-	-	-	257	211	163	136	2.0	-	-	-	428	350	265	220	2.0	-	-	-	295	242	185	
Time to $I(t) > 10K$	2.5	-	-	-	-	195	150	125	2.5	-	-	-	-	322	245	201	2.5	-	-	-	224	170	147	
Max $I(t)$ (infectious)	3.0	-	-	-	-	140	117	-	3.0	-	-	-	-	-	228	188	3.0	-	-	-	-	-	160	
Time to $I(t) > 1000$	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	133	
Max $I(t)$ (infectious)	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	245	
Time to $I(t) > 10K$	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	202	
Max $I(t)$ (infectious)	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	
London Heathrow Airport										JFK International														
s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	
Time to $I(t) > 10K$	1.05	-	-	455	344	283	219	183	1.05	+	+	+	+	455	345	287	1.05	-	-	475	357	295	227	
Max $I(t)$ (infectious)	1.25	-	-	422	318	262	202	169	1.25	-	+	+	+	420	320	265	1.25	-	-	440	331	273	210	
Time to $I(t) > 1000$	1.75	-	-	369	277	228	175	147	1.75	-	-	+	445	365	276	230	1.75	-	-	384	289	237	182	
Max $I(t)$ (infectious)	2.0	-	-	-	264	217	166	139	2.0	-	-	-	425	345	262	217	2.0	-	-	-	275	225	173	
Time to $I(t) > 10K$	2.5	-	-	-	-	200	153	128	2.5	-	-	-	-	320	240	200	2.5	-	-	-	208	159	133	
Max $I(t)$ (infectious)	3.0	-	-	-	-	144	120	-	3.0	-	-	-	-	-	225	186	3.0	-	-	-	-	-	150	
Time to $I(t) > 1000$	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	124	
Max $I(t)$ (infectious)	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	
Hong Kong International										Benchmark No Network														
s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	
Time to $I(t) > 10K$	1.05	-	-	335	249	204	155	129	1.05	+	+	+	+	477	389	294	244	1.05	-	-	484	367	293	227
Max $I(t)$ (infectious)	1.25	-	-	308	229	187	143	119	1.25	-	+	+	+	439	358	271	225	1.25	-	-	440	331	273	210
Time to $I(t) > 1000$	1.75	-	-	266	197	161	123	102	1.75	-	-	+	379	309	234	193	1.75	-	-	384	289	237	182	
Max $I(t)$ (infectious)	2.0	-	-	-	187	152	116	96	2.0	-	-	-	361	293	221	182	2.0	-	-	426	471	519	541	
Time to $I(t) > 10K$	2.5	-	-	-	-	140	106	88	2.5	-	-	-	-	269	202	167	2.5	-	-	-	320	243	201	
Max $I(t)$ (infectious)	3.0	-	-	-	-	-	99	82	3.0	-	-	-	-	-	189	155	3.0	-	-	-	-	-	536	
Time to $I(t) > 1000$	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	563	
Max $I(t)$ (infectious)	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	228	
Time to $I(t) > 10K$	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	
Benchmark No Network										s/k 1.1 1.5 2.0 2.5 3.0 4.0 5.0														
s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	s/k	1.1	1.5	2.0	2.5	3.0	4.0	5.0	
Time to $I(t) > 10K$	1.05	-	-	484	367	293	227	193	1.05	+	+	+	+	435	325	266	203	1.05	-	-	484	367	293	227
Max $I(t)$ (infectious)	1.25	-	-	440	324	242	175	140	1.25	-	+	+	+	400	299	244	186	1.25	-	-	440	321	267	203
Time to $I(t) > 1000$	1.75	-	-	205	148	118	88	72	1.75	-	-	+	446	365	278	231	1.75	-	-	345	257	210	159	
Max $I(t)$ (infectious)	2.0	-	-	-	139	111	82	67	2.0	-	-	-	-	344	299	244	186	2.0	-	-	-	345	347	265
Time to $I(t) > 10K$	2.5	-	-	-	-	100	74	61	2.5	-	-	-	-	-	320	243	201	2.5	-	-	-	-	-	187
Max $I(t)$ (infectious)	3.0	-	-	-	-	-	69	56	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	128	
Time to $I(t) > 1000$	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	
Max $I(t)$ (infectious)	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	-	3.0	-	-	-	-	-	-	

Fig. 8: Sensitivity analysis conducted on 3 major airports: (–) indicates  $\beta < \gamma$  (+) indicates time exceeded 500 days.

as expected. As  $s$  increases, the maximum number of infections decreases and the time until peak infections increases. This makes sense in the context of Figure 4, because, when  $s$  approaches  $k$ , the rate of change in the exposure compartment approaches zero. Similarly, as  $k$  increases, the maximum number of infections increases, and the time to reach peak infections decreases, which makes sense in the context of the converse of the previous argument. It is also interesting to note that the time for the epidemic to return below 1000 infections is often greater than 500 days, except when  $k$  is very large. Similarly, a larger value of  $s$  will also prolong the epidemic due to the slower rate of change in the exposed compartment. It is clear from Figure 8. that the values of  $s$  and  $k$  can vary the peak number of infections quite drastically, often on the order of millions of cases, even with only a 0.25 change in parameter values. Thus, we suggest that in the absence of reliable estimates of  $\beta$ ,  $\gamma$ ,  $\epsilon$  the

reader should treat any numerical conclusions presented as stylised versions of reality, which convey general trends but not precise predictions.

Finally we have included a benchmark case in which the world average city population is located in a single airport - essentially a global SEIRS model. The benchmark is provided for comparison but also to highlight how this global approximation is grossly inadequate for modelling a metapopulation which exhibits strong geographical clusteredness. The benchmark clearly underestimates the time until peak infections and the peak number of infections.

## 5 Mitigation Strategies

This section examines potential mitigation strategies in the context of epidemics on international flight networks. In order to evaluate these strategies we must first select some performance metrics. We decide to select metrics which are easily interpretable by policy makers and the general public, whilst also being useful in the context of managing hospital intensive care unit capacity and overall impact of the epidemic. Specifically we will measure the peak number of infections and the total number of cases of the disease, as these are transparent and can easily be measured from our simulations.

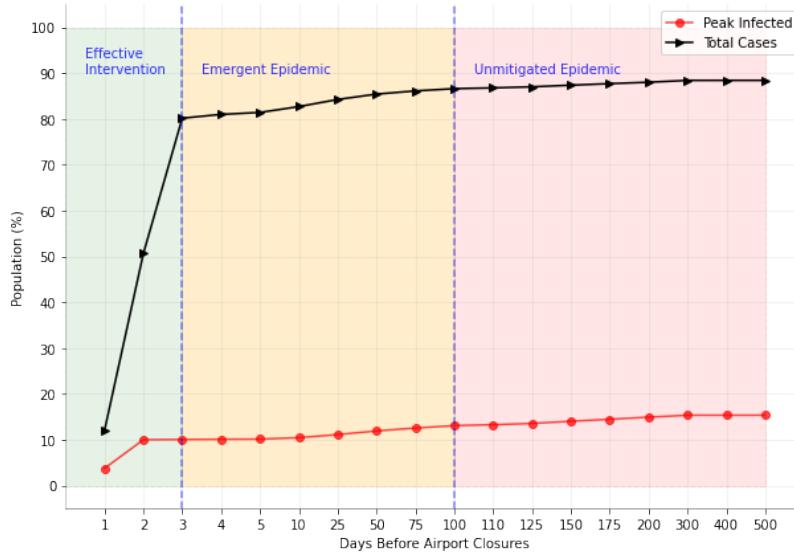


Fig. 9: Impact of worldwide permanent airport closures from nth day since first infection.

### 5.1 Nth Day Rule

The first mitigation strategy that we consider is defined by the permanent closure of air routes from the Nth day after the initial outbreak (with  $s = 2$  and  $k = 5$ ). Our simulations in Figure 9 demonstrate that closing routes earlier reduces the peak number of infections, but does not significantly reduce the total number of cases, unless all airports are closed by the end of day **2**. This demonstrates the high level of connectivity within the network with cases of infections proliferated across every continent within the first **3** days of the outbreak. It is quite remarkable that the impact of these infections only becomes noticeable after around **50** days, which would explain how the recent COVID-19 could spread across the world unnoticed for months before a pandemic was declared. We note that closing airports immediately after day 1 would have a significant impact, reducing the total number of cases by almost **80%**. Our interpretation of these results is that the graph diffusion model spreads tiny portions of the epidemic in each of the seed's neighbours, thus seeding also these new locations. Then, since the SEIRS parameters guarantee a local escalation of the epidemic, we do observe an increasing number of cases even a long time after the airport closures. This is in fact a very positive message, since in these cases the epidemic escalates at a very slow pace, thus permitting the introduction of new measure that can control it.

Extending on the previous result we consider whether the impact of social distancing, mask wearing and other measures which collectively aid in reducing the rate of infection can have a joint effect on our metrics when employed alongside the earlier airport closures. Since we already observed in Figure 8 that reducing the rate of infection reduces the speed at which the disease spreads a joint effect on our metrics seem plausible. After running the same experiment as previous, but with different values for the infection rate multiplier ( $k = 5$  and  $k = 3$ ), we find that the reductions in both peak and total cases due to the earlier airport closures is 4 - 5% larger when the infection rate is lower ( $k = 3$ ) as seen in Figure. 10. However the joint reductions due to

		% Reduction on 500 Day Closure Scenario *					% Population
	k	1	2	3	4	5	(500)
Peak Infections	5	75.69	34.86	34.55	34.26	33.99	15.33
	3	75.91	39.01	38.76	38.53	38.31	9.94
	Interaction Effect **	0.22	<b>4.15</b>	<b>4.21</b>	<b>4.27</b>	<b>4.32</b>	-
Total Cases / Recoveries	5	86.44	42.76	9.3	8.4	7.92	88.40
	3	86.12	42.89	14.0	13.09	12.56	79.87
	Interaction Effect **	-0.32	<b>0.13</b>	<b>4.70</b>	<b>4.69</b>	<b>4.64</b>	-

Fig. 10: Comparison of  $k = 5$  and  $k = 3$  on relative reduction. \* Cell value is computed as:  $1 - \frac{\max(\text{Infections} - \text{Day } N \text{ Closure})}{\max(\text{Infections} - \text{Day 500 Closure})}$  and similarly for total cases / recoveries \*\* Interaction Effect =  $k_3 \text{Row} - k_5 \text{Row}$

the combination both measures becomes trivial if we close airports swiftly in Day 1 or Day 2, suggesting that early closures will simply stop the virus from spreading in the first place. Additionally an interaction may still be present at days greater than Day 5, however it is not useful to consider this as the absolute reductions in cases by any measure after this period is small.

### 5.2 Threshold Infected Rule

A further modification to the previous results involves dynamically closing airports whenever the total number of cases exceeds **1** in every **10 $X$**  people within the local population. This differs from the previous method in which we implemented blanket global closures. The results of this new experiment, which are shown in Figure 11, indicate that this ‘wait and see’ strategy is totally ineffective under our modelling framework. Even considering a highly

		% Reduction on No Closure Scenario *					% Population	
		k	$10^7$	$10^6$	$10^5$	$10^3$	$10^2$	(1)
Peak Infections	5	8.07	6.07	6.19	4.2	2.87	15.33	
	3	7.92	5.65	6.05	4.2	2.9	9.94	
	Interaction Effect	-0.15	-0.42	-0.14	0.00	0.03	-	
Total Cases / Recoveries	5	0.4	0.25	0.26	0.12	0.03	88.40	
	3	31.5	26.77	8.35	0.81	0.03	79.87	
	Interaction Effect	31.10	26.52	8.09	0.69	0.00	-	

Fig. 11: Figure 14 Comparison of  $k = 5$  versus  $k = 3$  on relative reduction.

\* Cell value is computed as:  $1 - \frac{\max(\text{Infections} - 1 \text{ in } 10N \text{ threshold})}{\max(\text{Infections} - 1 \text{ in } 100 \text{ threshold})}$

unrealistic version of this strategy in where we suppose it is possible to detect cases up to a fineness of **1** in every **10** million people, it is already too late to close airports, providing little more than a **8%** reduction in peak infections and almost no change in the total number of cases. Again, we motivate this result by emphasising that most of the observed cases are due to the escalation of the local epidemic, and are not travel-related.

### 5.3 Limited Nth Day Rule

In the previous section we realise that it is far more effective to close airports preemptively than it is to wait on some threshold level on infections to be attained within the local population. However, one could argue that it is impractical to close all airports globally, both from a economic and political point of view. Thus, we proceed to examine what performance we can achieve by only closing a subset of key airports, which we rank by several metrics: population,

Page-Rank and betweenness. It is quite interesting to see that even with only the top 1% of airports closed we still obtained significantly greater reductions than the threshold rule (Figure 12a). This further emphasises the point that early intervention is far more important in the network than attempting to detect when certain infection thresholds have been breached. We also note that the fall-off in the strategy's performance is quite non linear both across the rows or columns. In Figure 12b, we present the same experiment but using the Page-Rank ranking of the airports, and we notice that the reductions in peak infections are drastically superior to that obtained when ranking the airports by populations, suggesting that prioritising the closure of airports with high centrality is very important in terms of epidemic mitigation. Finally, we apply the same methodology using a ranking based on betweenness. In Figure 12c, we see the results are somewhat mixed, with the strategy performing better than with previous metrics at the 5% level, and marginally worse at other levels.

#### 5.4 Optimisation Approach

In section 4 we explored several mitigation strategies, ultimately finding greatest flexibility and mitigation in the limited Nth day rule. This strongly suggests that we could find a very effective airport closure strategy that will achieve excellent mitigation results, without totally disrupting the network connectivity. In this section, we employ a genetic algorithm (GA) to search for the optimal combination of airport closures that maximises the utility of the Nth Day Rule. GAs, in their simplest form, operate on binary strings called chromosomes which are an encoding of the parameters of interest, commonly referred to as genes. Any particular instance of a chromosome has a genotype which refers to a specific string of bits each with values 1 or 0 representing a particular gene's allele. Once the problem can be formulated within this framework the procedure followed by genetic algorithms is the following:

1. Define a fitness function  $F(X)$  which evaluates the optimality of a given genotype.
2. Initialise a population of chromosomes with randomly assigned genotypes.
3. Evaluate the fitness of all members of the population. Individuals will then be selected for breeding at a frequency proportional to their fitness (this emulates the “survival of the fittest” mechanism).
4. In a process known as *crossover*, pairs of *selected* genotypes are split uniformly at some locus along the chromosome and then *recombined*, to form new chromosomes.
5. *Mutations* are then applied at random to alleles of the recombined chromosomes (simply by bit flipping) in order to prevent irrevocable loss of any characteristic.
6. The process (3 - 5) is then repeated so that the fittest member of the population are selected.

% Airports Closed (Ranked by Population)	% Reduction on 500 Day Closure Scenario *					% Population (500)
	1	2	3	4	5	
Peak Infections	1%	31.47	30.44	29.3	27.5	25.75
	5%	64.77	28.48	25.49	23.85	22.75
	10%	75.4	24.82	24.72	24.6	24.46
	25%	75.69	33.64	33.36	33.1	32.85
	100%	75.69	34.86	34.55	34.26	33.99
Total Cases / Recoveries	1%	24.74	9.97	9.37	8.4	7.79
	5%	51.89	21.13	8.28	7.52	7.06
	10%	66.01	28.23	8.32	7.69	7.29
	25%	86.44	40.55	8.71	8.0	7.58
	100%	86.44	42.76	9.33	8.4	7.92

(a) Relative reductions: airports sorted by **population**.

% Airports Closed (Ranked by PageRank)	% Reduction on 500 Day Closure Scenario *					% Population (500)
	1	2	3	4	5	
Peak Infections	1%	25.63	23.35	21.64	20.65	20.02
	5%	68.68	30.77	30.36	29.93	29.52
	10%	74.9	35.29	34.85	34.49	34.15
	25%	75.68	34.66	34.35	34.07	33.81
	100%	75.69	34.86	34.55	34.26	33.99
Total Cases / Recoveries	1%	5.7	2.28	2.2	2.13	2.06
	5%	42.16	12.21	5.24	4.64	4.28
	10%	82.23	30.92	6.96	6.31	5.93
	25%	86.38	41.93	8.22	7.34	6.91
	100%	86.44	42.76	9.33	8.4	7.92

(b) Relative reductions: airports sorted by **Page-Rank**.

% Airports Closed (Ranked by Betweenness)	% Reduction on 500 Day Closure Scenario *					% Population (500)
	1	2	3	4	5	
Peak Infections	1%	24.69	22.85	20.93	19.67	18.84
	5%	68.26	34.1	34.01	33.87	33.7
	10%	70.57	39.43	38.91	38.46	38.04
	25%	75.44	36.53	36.1	35.72	35.38
	100%	75.69	34.86	34.55	34.26	33.99
Total Cases / Recoveries	1%	6.26	2.35	2.22	2.1	1.99
	5%	46.49	19.12	8.08	7.32	6.85
	10%	77.89	29.89	8.47	7.71	7.27
	25%	84.58	40.16	9.37	8.46	7.97
	100%	86.44	42.76	9.33	8.4	7.92

(c) Relative reductions: airports sorted by **betweenness**.

Fig. 12: Relative reductions on peak infections and total cases using different percentages of closures and ranking factors.

For a more detailed explanation of the entire process see [18]. We opt to use a genetic algorithm for this problem since the solutions of our combinatorial problem can be easily represented as a binary strings. By contrast, problems of this nature are ill suited to classical optimisation methods such as Gradient Descent, as it is not possible to analytically compute gradients and our search space is too large for approximation methods. Additionally a fitness function can be easily defined from the metrics we have described previously. In the Table 13 we identify the key information required to formulate our problem within the genetic framework.

Parameter	Definition
<b>Chromosome</b>	<p>Instead of optimising for <i>which airports to close</i>, we choose to optimise for <i>which countries should close their airports</i>. This reduces the GA search space by 15 times which is highly desirable for convergence to global optimum. This reframing is also sensible from a policy perspective, as if some airports were to remain open, then travellers to the closed airports would simply be rebook to the remaining open airports (completing the remainder of their journey by road) defeating the point of the closure (which was to reduce international spread of the disease).</p> <p>We now define our chromosome as a 195 bit string where 1 in the <math>i^{\text{th}}</math> indicates that the <math>i^{\text{th}}</math> country's airports are closed and 0 indicates the <math>i^{\text{th}}</math> country's airports are open.</p>
<b>Fitness Function</b>	<p>In designing the fitness function we seek to balance the economic impacts of airport closures along with the reduction in Peak &amp; Total Cases of infections, relative to an unmitigated epidemic. Since we have no prior biases towards Peak, Total Case or economic impacts we will weight their contributions equally in the fitness function:</p> $\begin{aligned} \text{Let } T &= \% \text{ Reduction in Total Infections} \\ \text{Let } P &= \% \text{ Reduction in Peak Infections} \\ \text{Let } A &= \% \text{ Airports Remaining Open} \end{aligned}$ $F(X) = T(X) * P(X) * \sin(0.5\pi * A(X))$ <p>The sin term in the fitness function encodes our prior knowledge that very few airport closures are associated with increasingly worse performance on our key metrics. This non linear shaping of the fitness function should help our algorithm to converge faster, avoiding exploration of poor solutions.</p>

Fig. 13: Defining key components of our problem in the GA framework.

In order to evaluate the fitness function, we must compute the values of  $T$  and  $P$  (computing  $A$  is trivial). This clearly involves inputting the parameters encoded in the chromosomes genotype into our simulation developed in previous sections. We perform this by first using a lookup table to convert between the 195 bit country closures string specified in the genotype to a **3,425** bit string airport closures vector required by our model. Next we set to zero the

rows and columns of the closed airports within the adjacency matrix at the appropriate time steps (to disconnect an airport from the network). Finally we proceed to run the algorithm for sufficient iterations as for the first wave of the epidemic to be completed.

### 5.5 Optimisation Results and Comparisons

The results of the algorithm in Figure 14 are presented in the usual format for consistency, however each row now represents a different GA optimised for metrics on of days **1** to **5** respectively, with the corresponding quantities optimised bolded in the table.

% Airports Closed (Ranked by Genetic Algo)	% Reduction on 500 Day Closure Scenario *					% Population (500)
	1	2	3	4	5	
Peak Infections	18%	<b>78.21</b>	-1.88	-1.55	-1.31	-1.12
	40%	60.71	<b>48.12</b>	45.47	43.57	42.04
	33%	51.12	50.21	<b>45.57</b>	42.42	40.05
	25%	42.4	42.36	42.29	<b>42.19</b>	42.06
	32%	52.49	50.5	46.52	43.81	<b>41.75</b>
Total Cases / Recoveries	18%	<b>86.55</b>	5.2	1.38	1.14	1.01
	40%	60.92	<b>34.34</b>	7.16	5.69	5.19
	33%	50.2	29.28	<b>8.16</b>	6.68	6.12
	25%	43.55	26.51	8.47	<b>7.27</b>	6.71
	32%	51.3	30.07	7.75	6.27	<b>5.73</b>

Fig. 14: Relative reductions on **500** day closure scenario for percentages of airport closures. (airports sorted by **Genetic Algorithm**), **bolded** number indicate is the day for which the row was optimised by GA.

The results are quite remarkable, but better visualised in Figure 15 for the day **3** closure scenario. Not only does the GA outperform our previous ranking

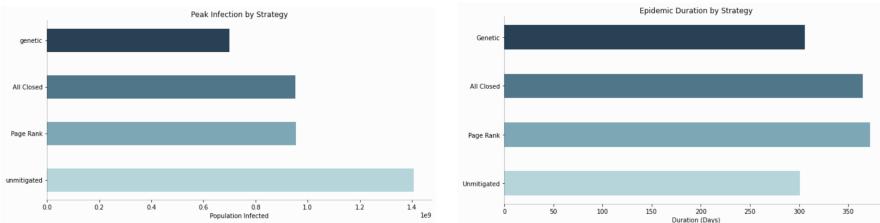


Fig. 15: Day 3 comparison of strategy performance

methods (with an equivalent % of airport closures), it also improves on the

close all airports strategy by **15–20%**. Whilst this is counter-intuitive, the GA leverages the hidden structure within the network to flatten the curve over **50** days earlier, while also reducing peak and total infections when compared with Page-Rank, all-closed strategy and unmitigated strategy, as seen in Figure 16.

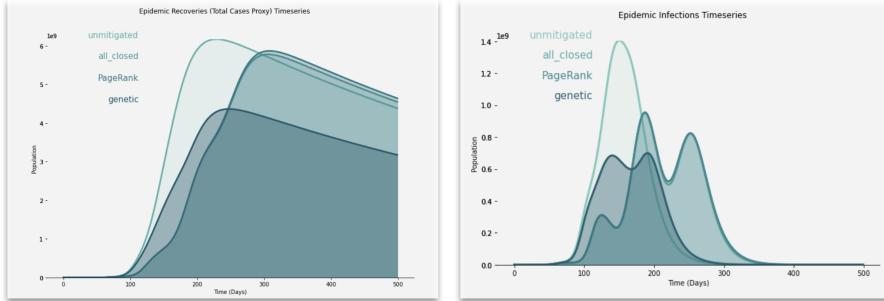


Fig. 16: Day 3 comparison of strategy performance

This suggests that the GA learns to leverage the network structure via closures in such a way as to accelerate the initial infection rate, but achieving a lower point of equilibrium. To gain further intuition into the GA behaviour, it is best to look at Figure 17 which exhibits the evolution of the GA strategy as we alter the day at which closures occur. What we observe is that it is initially optimal to close China and certain other countries such as France which are very well connected to China via air routes. However as governments delay closures up to day **5**, we find that the GA shifts focus away from China and starts to establish “firebreaks” in other surrounding countries such as India, Kazakhstan and Russia, whilst also selective targeting certain several African and South American countries.

Returning our focus back to the day 3 strategy learned by the GA in Figure 18, we examine the percentage change in peak infections and total cases under the GA strategy compared with the all-closed strategy and unmitigated strategy. Despite only **33%** of airports being closed under the GA strategy **73%** of countries see a reduction in peak infections relative to an unmitigated case, whilst **60%** see a reduction in peak infections relative to the all closed case. Similarly, we see that **67%** of countries see a reduction in total cases relative to the unmitigated case and **69%** see a reduction in total cases relative to the all-closed scenario. This provides overwhelming evidence of the effectiveness of the genetic algorithm strategy over naive propositions of worldwide closures of airports, and highlights how a model-based approach to solving such a problem can provide significant added value over simplistic intuitive strategies, such as closing by population or network centrality.

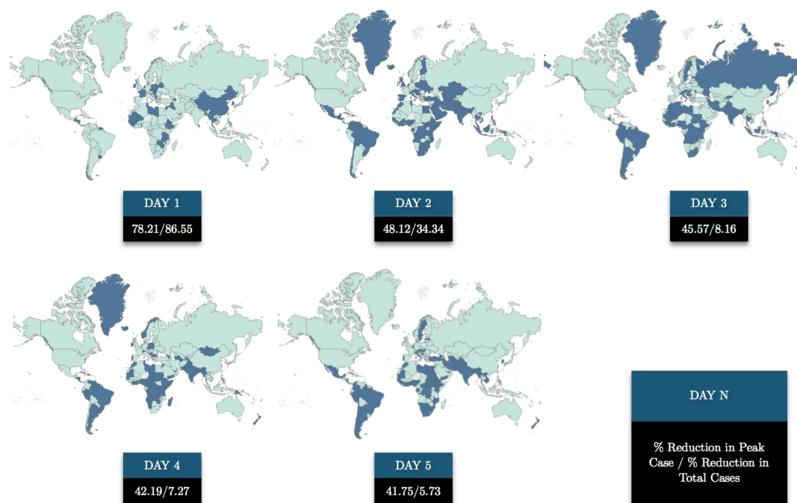


Fig. 17: GA: the optimal countries to close starting from nth day (dark blue indicates closed).

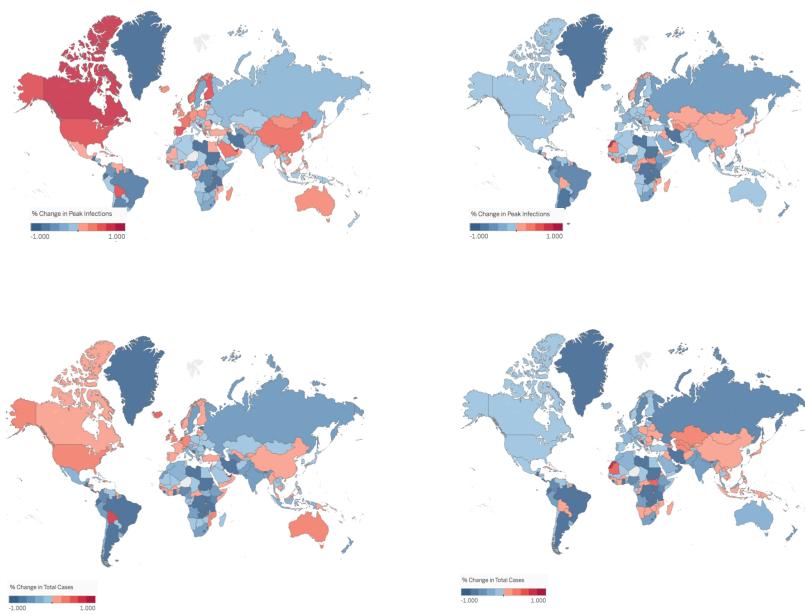


Fig. 18: A comparison of GA strategy versus naive strategies and total inaction.

## 6 Conclusion

In this paper, we have shown that human mobility infrastructure networks are highly complex and highly resilient to node removals. This has important consequences in epidemiology, since the high connectivity facilitates the spread of pandemic diseases to a worldwide scale. These observations motivated us to seek and test non-trivial airport closure strategies that could maintain a good network connectivity, while slowing down the spread of a potential disease.

Our framework is based on a metapopulation SEIRS model, where the sub-populations are located in the airports' locations, and they are composed of the individuals living in those nearby areas. Our simulations allowed us to explore and study a variety of realistic scenarios to understand the dynamics of the spread of the disease. We considered several different airport closure strategies, and we compared them using measures extracted from our simulations.

One main message that arises from our results is that the disease is seeded in many locations worldwide with impressive speed. If conditions are met for isolated local escalations of the epidemic, then most countries will be hit by the seeded disease after a variable delay, regardless of any late interventions on travelling restrictions.

Our findings suggest that the first week of dispersion of the disease through the network is a critical time period for effective intervention, however interventions in the network, such as airport closures, still provide some reductions to peak infections and total cases up to 3 months into the simulation. Furthermore, we show that policies which reduce community spread can be combined with our proposed airport closure strategies to provide greater benefits than if either policy had been used separately.

Finally, we explored the application of an optimisation approach to identify optimal airport closures within the critical first week of disease spread, in order to reduce the global impact of the epidemic while keeping as many airports open as possible. This optimisation approach, based on a genetic algorithm, improved upon all of the other methods, hence providing a new model-based perspective on the decision making process that leads to the travel restrictions. One very interesting aspect of our results is that the algorithm leverages the complex structure of the network to place strategic "fire breaks", which drastically reduce peak infections and total cases.

## 7 Appendix

### 7.1 Algorithm Derivation

In our review of theoretical underpinnings (Section 3.1) we discussed the graph diffusion model. Where there is an amount of fluid  $\psi_j$  at nodes  $j = 1 \dots N$ . Thus fluid flows from node  $i$  to node  $j$  at a rate proportional to the difference in the amount of fluid at each node  $c(\psi_i - \psi_j)$  where  $c$  is the constant of proportionality or more commonly referred to as the *diffusion constant*. We note however that fluid can only flow between nodes if they are adjacent in graph  $G$  with adjacency matrix  $A$ . Thus the total instantaneous change in fluid volume at node  $j$  is given by the following equation.

$$\frac{d\psi_j}{dt} = c \sum_{i=1}^N (\psi_i - \psi_j) A_{ij}$$

Following from this we can easily proceed to vectorise this equation for all vertices as follows, where  $\psi = (\psi_1 \dots \psi_N)^T$  and  $D = \text{diag}(\deg(v_1), \dots, \deg(v_N))$ .

$$\begin{aligned} \frac{d\psi_j}{dt} &= c \sum_{i=1}^N \psi_i A_{ij} - c \psi_j \deg(v_j) \\ \frac{d\psi}{dt} &= c(A - D)\psi \\ \frac{d\psi}{dt} &= -c(D - A)\psi \end{aligned}$$

In the basic implementation described above we consider there to be only a single fluid, however we will require 4 different fluids for the 4 states of our SEIRS model. Thus we let  $\theta_j = (S_j \ E_j \ I_j \ R_j)^T$  represent the number of people in the airport / node population which are in the Susceptible, Exposed, Infectious and Recovered states and  $\psi_j = \alpha_j \theta_j$  be the proportion of people available for travel. Hence  $\psi \in \mathbb{R}^{4 \times N}$  is now a matrix and thus we must modify the ordering of our equation to correct for this.

$$\frac{d\psi}{dt} = -c\psi(D - A^T) \in \mathbb{R}^{4 \times N}$$

Breaking this down, the components of the original definition are reconfigured such that  $\sum_{i=1}^N \psi_i A_{ij}$  becomes  $\psi A^T = (\sum_{j=1}^N A_{1j} \psi_j \dots \sum_{j=1}^N A_{Nj} \psi_j) \in \mathbb{R}^{4 \times N}$  and  $\psi_j \deg(v_j)$  becomes  $\psi D \in \mathbb{R}^{4 \times N}$ . As a final simplification we assume that  $A = A^T$  which is appropriate in the case of flight routes which are almost always operated in both directions.

$$\frac{d\psi}{dt} = -c\psi(D - A) \in \mathbb{R}^{4 \times N}$$

This equation provides a simple model for the diffusion / travel of people through the international flight network. Thus we can now define a simple

update algorithm which enables travel through the network and updating of the sizes of the local populations.

1. Repeat for  $t = 1 \dots T$

- (a) Compute  $\psi = \theta\alpha$  where  $\alpha = \text{diag}(\alpha_1, \dots, \alpha_N)$ .
- (b) Update the local populations due international travel  $\theta = \theta + \frac{d\psi}{dt}$ .

However there are some obvious issues with this simplistic model, namely that  $\theta_i$  typically converge to some equilibrium value  $\theta^{(0)} \neq \theta^{(eq)}$ , which does not represent the reality that people travelling on holidays or business will typically return to their home country. Secondly there are some technical issues which arise from implementation, such as for vertices of large degree *simultaneous* outflows may exceed the total population at the node (this is an artefact of vectorisation). Finally the flows to other vertices are not influenced by how central the airport is to the network, which is obviously important as travellers are more likely to visit major tourist / business destinations and this information is not necessarily reflected in the population (eg. Ethiopia has a very large population but it is not as important in the network as France or the UK).

The first of these 3 issues is quite easy to address, we can simply split the population  $\theta$  into a base population  $\theta_B$  who live permanently in the local area and  $\theta_T$  being the transient population in a given node on holidays / business. Refactoring our algorithm to incorporate this change is relatively simple.

1. Repeat for  $t = 1 \dots T$

- (a) Compute  $\psi_+ = \theta_B\alpha_+$  and  $\psi_- = \theta_T\alpha_-$  where  $\psi_+$  represents out-bound travellers from the local base population and  $\psi_-$  represents returning transient travellers to their home countries.
- (b) We then compute  $\frac{d\psi_+}{dt} = -c_+\psi_+(D - A)$  and  $\frac{d\psi_-}{dt} = -c_-\psi_-(D - A)$  respectively
- (c) Finally we now have two update equations for international travel
  - i.  $\theta_B = \theta_B + \min(\frac{d\psi_+}{dt}, 0) + \max(\frac{d\psi_-}{dt}, 0)$  where we subtract the *outbound* base populations and add the *arriving* travellers.
  - ii.  $\theta_T = \theta_T + \max(\frac{d\psi_+}{dt}, 0) + \min(\frac{d\psi_-}{dt}, 0)$ , where we add the *returning* base populations and subtract the *departing* travellers.

Note that we have chosen the wording *outbound*, *arriving*, *departing*, *returning* carefully to reflect the sequence of steps all travellers pass through in order. This is crucial to ensure logical behaviour of travellers and also to prevent leakage of fluid / people from the network.

To address the second issue we can simply divide  $\psi_i$  by the degree of the node which ensures that the outflows computed to every other node will be at most  $\psi_i$ . This modification we will denote by  $\frac{\psi_i}{D}$  but it is understood that this represents  $\frac{\psi_i}{D_{ii}}$  element wise and will be implemented by Numpy broadcasting

operation in Python.

To Address the final issue we replace the adjacency matrix  $\mathbf{A}$  with a weighted matrix  $\mathbf{C}$  to encourage flows to more central airports, where  $\mathbf{C}$  is defined as follows, where  $P_j$  is the page-rank centrality of node  $v_j$ .

$$\frac{\mathbf{A}_{ij}P_j}{\sum_{j=1}^N \mathbf{A}_{ij}P_j}$$

We then translate this back into our original formulation, with  $\mathbf{A}$  replaced by  $\mathbf{C}$  and  $\mathbf{D}$  replaced by  $\mathbf{I}$ , the identity matrix (it is trivial to check by a similar derivation to previously that since  $\mathbf{C}$  is row stochastic  $\mathbf{D}$  may be replaced by the identity matrix). Similarly this generalises to the more complex models addressing the other issues with the original graph diffusion model.

$$\frac{d\psi}{dt} = -c\psi(\mathbf{I} - \mathbf{C}) \in \mathbb{R}^{4 \times N}$$

Next we addition another the SEIRS model which models the changes in  $\theta$  due to community spread of the disease (rather than due to international travel). As in Section 3.1 we define the following differential equations which define the SEIRS model.

$$\begin{aligned} \frac{dS}{dt} &= \delta R - \frac{S\beta I}{M} \\ \frac{dE}{dt} &= \frac{S\beta I}{M} - \epsilon E \\ \frac{dI}{dt} &= \epsilon E - \gamma I \\ \frac{dR}{dt} &= \gamma I - \delta R \end{aligned}$$

These changes represent the spread of disease within a single community, but for efficiency we will define an operator  $\mathbf{B}$  which acts on a modified  $\theta$  which enables application of the linear transformation to obtain the updated community state.

$$\mathbf{B} = \begin{pmatrix} -\beta & 0 & 0 & \delta \\ \beta & -\epsilon & 0 & 0 \\ 0 & \epsilon & -\gamma & 0 \\ 0 & 0 & \gamma & -\delta \end{pmatrix}$$

$$\tilde{\theta}_j = (\frac{S_j I_j}{M_j}, E_j, I_j, R_j)$$

Note that the community spread occurs prior to the splitting of the population  $\theta$  into  $\theta_B$  and  $\theta_T$  as it is assumed that travellers and the local community are fully mixed. Thus the final implementation of the algorithm

may be described in table 4, where we include one final step of splitting  $\theta$  into  $\theta_B$  and  $\theta_T$  according to the prior populations  $\|\theta_B\|$  and  $\|\theta_T\|$  respectively.

Table 4: Simulation Pseudocode.

Algorithm: Flow & Degree Corrected Epidemic Diffusion Model	
<b>for</b> $t = 1, \dots, T$ :	
(5) $\vartheta(t-1) = \vartheta_B(t-1) + \vartheta_T(t-1)$	
(1) $\vartheta^*(t-1) = \vartheta(t-1) + B\bar{\theta}$	
(2) $\vartheta^*_B(t-1) = ( \vartheta_B / \vartheta ) * \vartheta^*(t-1)$	Community Spread
$\vartheta^*_T(t-1) = ( \vartheta_T / \vartheta ) * \vartheta^*(t-1)$	(SEIRS)
(3) $\Psi^*_+(t-1) = \vartheta^*_B(t-1) \alpha_+$	
$\Psi^*_.(t-1) = \vartheta^*_T(t-1) \alpha_-$	
(4) $\Delta\Psi_+(t) = -c_{+}\Psi^*_+(t-1) (I/D) (I - C^T)$	International Spread
$\Delta\Psi_-(t) = -c_{-}\Psi^*_.(t-1) (I/D) (I - C^T)$	
$\vartheta_B(t) = \vartheta^*_B(t-1) + \text{Min}(\Delta\Psi_+(t), 0) + \text{Max}(\Delta\Psi_-(t), 0)$	(Diffusion)
$\vartheta_T(t) = \vartheta^*_T(t-1) + \text{Max}(\Delta\Psi_+(t), 0) + \text{Min}(\Delta\Psi_-(t), 0)$	

## References

- [1] “A-Model-Based-Approach-To-Assess-Epidemic-Risk”. In: (Aug. 2020). URL: <https://github.com/hugo1005/A-Model-Based-Approach-To-Assess-Epidemic-Risk>.
- [2] A. Apolloni et al. “Metapopulation epidemic models with heterogeneous mixing and travel behaviour”. In: *Theoretical Biology and Medical Modelling* 11.1 (2014), p. 3.
- [3] D. Balcan et al. “Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model”. In: *Journal of computational science* 1.3 (2010), pp. 132–145.
- [4] D. Balcan et al. “Multiscale mobility networks and the spatial spreading of infectious diseases”. In: *Proceedings of the National Academy of Sciences* 106.51 (2009), pp. 21484–21489.
- [5] M. Chinazzi et al. “The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak”. In: *Science* 368.6489 (2020), pp. 395–400.
- [6] J-J Daudin, Franck Picard, and Stéphane Robin. “A mixture model for random graphs”. In: *Statistics and computing* 18.2 (2008), pp. 173–183.
- [7] “Gridded Population of the World (GPW), v4”. In: (Jan. 2020). URL: <https://sedac.ciesin.columbia.edu/data/set/gpw-v4-admin-unit-center-points-population-estimates-rev11>.
- [8] H. W. Hethcote. “The mathematics of infectious diseases”. In: *SIAM review* 42.4 (2000), pp. 599–653.
- [9] Can Hou et al. “The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): A well-mixed SEIR model analysis”. In: *Journal of medical virology* (2020).

- [10] L. López and X. Rodo. “A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics”. In: *Available at SSRN 3576802* (2020).
- [11] Mark Newman. *Networks*. OUP Oxford, Mar. 2010. doi: 9780191500701.
- [12] K. Nowicki and T. A. B. Snijders. “Estimation and prediction for stochastic blockstructures”. In: *Journal of the American statistical association* 96.455 (2001), pp. 1077–1087.
- [13] “OpenFlights: Airport And Airline Data”. In: (Jan. 2017). URL: <https://openflights.org/data.html>.
- [14] Tiago P. Peixoto. “Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models”. In: *Physical Review E* 89.1 (Jan. 2014). ISSN: 1550-2376. doi: 10.1103/physreve.89.012804. URL: <http://dx.doi.org/10.1103/PhysRevE.89.012804>.
- [15] L. Peng et al. “Epidemic analysis of COVID-19 in China by dynamical modeling”. In: *arXiv preprint arXiv:2002.06563* (2020).
- [16] A. Radulescu and K. Cavanagh. “Management strategies in a SEIR model of COVID 19 community spread”. In: *arXiv preprint arXiv:2003.11150* (2020).
- [17] Muhammad Adnan Shereen et al. “COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses”. In: *Journal of Advanced Research* (2020).
- [18] J. C. Spall. *Introduction to stochastic search and optimization: estimation, simulation, and control*. Vol. 65. John Wiley & Sons, 2005.
- [19] M. Tizzoni et al. “On the use of human mobility proxies for modeling epidemics”. In: *PLoS Comput Biol* 10.7 (2014), e1003716.
- [20] Y. J. Wang and G. Y. Wong. “Stochastic blockmodels for directed graphs”. In: *Journal of the American Statistical Association* 82.397 (1987), pp. 8–19.