Dynamical processes on transportation networks : Effects on infectious disease spreading

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Mansi Budamagunta

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

July, 2021

Supervisor: Prof. M. S. Santhanam

Certificate

This is to certify that this dissertation entitled Dynamical processes on transportation networks : Effects on infectious disease spreading towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Mansi Budamaguntaat Indian Institute of Science Education and Research under the supervision of Prof. M. S. Santhanam, Professor, Department of Physics, during the academic year 2020-2021.

Santhanam

Prof. M. S. Santhanam

30.7.2021

Committee:

Prof. M. S. Santhanam

Dr. Sachin Jain

Declaration

I hereby declare that the matter embodied in the report entitled Dynamical processes on transportation networks : Effects on infectious disease spreading are the results of the work carried out by me at the Department of Physics, Indian Institute of Science Education and Research, Pune, under the supervision of Prof. M. S. Santhanam and the same has not been submitted elsewhere for any other degree.

Mansi

Mansi Budamagunta

Acknowledgments

Firstly, I would like to thank Prof. M. S. Santhanam for his kind and patient guidance throughout the year.

Next, I would like to thank my friend and colleague, Onkar, who worked on a huge chunk of the Static Hazard Map discussed in this thesis. He has been supportive and encouraging throughout the year and has helped me through many hurdles, academic and otherwise.

I would like to thank Dr. Sachin Jain and Dr. G. J. Sreejith for many lively and informative discussions over the course of this project.

My time at IISER Pune has been an incredible journey. I made many memories to cherish with my friends - Mohan, Adithya, Arya, Koustav, Onkar, Adarsh, Sanjana, Deva, Shriya, Snigdha, Aagam, Nishant, Rounak, Jhelam, Aasim, Aarti, Sriram, Abhishek, and Sirisha. I would also like to thank Dr. Collins Assisi, Dr. Suhita Nadkarni, and all present and past members of the Computational Neurobio lab at IISER Pune for a fun and fulfilling lab experience. The birthday pranks will remain a highlight of my IISER life.

My friends back home have ensured that I could always seek refuge at home. All our time together has been a blessing, especially the last year spent at my hometown in the midst of a pandemic. I thank Palak, Palak, Isha, Saanketh, Chinmay, Tanya and Nancy.

Lastly, I would like to thank my parents for being exceptionally understanding and supportive. They've been my rock.

I would like to acknowledge the INSPIRE scholarship by DST as well.

Abstract

In this thesis, we study the spread of an infectious disease in the context of India. For this, we construct a network of 446 cities having a population of over 1 Lakh and study the infection spread using an SIR metapopulation model. We build the traffic matrix for this network based on real-life data about three modes of transport - air, rail, and road. Given an outbreak location and some information about the disease parameters, we can predict the *hazard* at all other cities in the network. This is called a Hazard Map. In this thesis, we discuss two variations of it - Static Hazard Map and Dynamic Hazard Map. In Dynamic Hazard Map, the difference in time scales associated with the different modes of transport is taken into account. This is not the case for the former.

Contents

List of tables

List of figures

Chapter 1

Introduction

It is a truth universally acknowledged that pandemics are bad.

Humanity has found itself crippled by pandemics many times over the last few centuries. The most obvious example is the spread of the recent SARS-CoV-2 virus, first reported from Wuhan in China in December 2019. As of August 24, 2021, the SARS-CoV-2 strain has infected more than 19 Crore (190 million) people and led to the deaths of over 42 Lakh (4.2 million) people worldwide [1]. In India, the two waves of the pandemic have ravaged the country. Starting from a few cases reported in February 2020, the infection has spread to over 3 crores (30 million) cases, as of August 24, 2021[2].

To effectively deal with the infection spread, minimizing the harm caused by it in society, we must understand two broad aspects of the disease.

Firstly, we require an understanding of the biological details of the disease and the related required medical infrastructure. This would include studying the virus causing it, symptoms caused by it, the mode of transmission (airborne, human to human transmission, etc.), effective PPE equipment, medicines to manage symptoms, developing vaccines, etc. All of the above is outside the scope of this project.

Secondly, we need to understand the patterns of spreading of the infection. We could take measures to prevent the disease from spreading to many regions and allocate the limited medical resources appropriately to provide relief to high-risk areas. It is known that irrespective of the intrinsic properties of the virus to cause disease, the spread of the infection from one geographical region to another is mainly influenced by the mobility patterns of people [3, 4, 5, 6]. Long-distance travel has made the world smaller. It has been shown in many studies that this has led to the spread of any infection to remote corners of the world at short time-scales [7, 8, 9]. The spread of the SARS-CoV-2 in India, similar to the case of previous worldwide infections, is influenced by the movement of people along multiple modes of transport [10, 11, 12, 13].

In India, most of the modelling studies have focused on evolution of the disease inside a city/town and the prediction of caseloads at that level [14, 15, 16, 17, 18, 19]. There have been some studies attempting to understand the spread of the infection from one geographical region to another owing to the mobility of people via transportation networks [20, 21, 22, 23]. The approach taken in this project is aligned with the efforts mentioned in the latter group of studies.

Specifically, we are considering a network of 446 cities in the country having a population above 1 lakh. We seek to answer this: Given an outbreak location and some details about the spread of an infection, how long does it take for the disease to spread to another city/town?

In this project, we aim to construct a real-time infectious diseases hazard map for India, somewhat similar to the seismic hazard maps produced for earthquakes [24]. Here, the hazard value would represent the risk faced by a city/town in the network from the infection given an outbreak location. This hazard map considers real-time rail and air transportation updates to provide a changing picture of the risk as the situation changes.

This hazard map has been built in the context of the spread of the SARS-CoV-2 in India but can be modified to use on any future epidemic if the model's basic assumptions remain valid.

The main features of this hazard map are long-range travel which takes into account the time scales associated with the modes of transport involved, an extensive network that covers most of the country, and real-life evolving transportation data about the three primary modes of transport, namely - Air, Rail, and Road.

The plan for the thesis is as follows: Chapter 2 describes the properties of the network and the data collection process, Chapter 3 discusses the Static Hazard Map, Chapter 4 discusses the Dynamic Hazard Map, Chapter 5 has conclusion and future directions.

Chapter 2

Properties of the Network and Data Collection

To study the spread of an infectious disease over the country, we have constructed a bidirectional network with cities as nodes whose edge weights are determined by real-life transportation data. In this chapter, we describe the properties of the networks used for the Static Hazard Map. Key features of the network and the data collection process remain the same for the Dynamic Hazard Map. Differences that occur have been described in Chapter 4.

In total, 446 cities having population over 1 Lakh [25] have been taken into consideration. These cities are connected by at least one of the three modes of transport - air, rail, or road. In the network, the weight of an edge AB(directional) is determined by the number of passengers traveling from A to B on a typical day . 4 networks have been constructed, one corresponding to each and one with all three modes of transport combined. These networks are called $\boldsymbol{F}$ matrices. The data collection was the most challenging and time-taking part of this project. Official data sources that have been used were available, and when they weren't, assumptions have been made, and algorithms were employed to construct passenger data. Details of the data collection, construction of the networks, and resulting properties have been described in the following sections. Properties of the networks have been summarised in Table 2.1

2.1 Air

Air data was collected in the form of monthly travel statistics from a private website [26] for the year 2019. The DGCA [27] collects information about the number of passengers traveling between every pair of airports every month. This information was averaged over for the year 2019, then converted to number of passengers per day. We have 85 cities in our datasets that are connected by air transport. The resulting transport matrix is symmetric.

2.2 Rail

For rail transport, information about the number of passengers traveling between cities is not available in the public domain. So, we collected all information about the train timetable from the official railway website [28] in July 2020. This contained information about the number of trains traveling between stations as well as the route and train type of each train. We know the capacity of a train, given its type. In total, we collected information about trains traveling between 827 stations associated with 435 cities connected by rail transport (many big cities contained multiple stations). It is common knowledge that train routes are generally in pairs going to and fro between a pair of stations. When we found a route in our dataset with its reverse route missing, this was assumed to be an error, and the reverse route was artificially constructed.

We used an algorithm to estimate the number of passengers based on this information. The details of this algorithm are published in the supplementary material of the arxiv paper [29]. The key assumptions and the resulting expression for the number of people traveling between two cities along a route are described here, along with a toy example in Figure 2.1

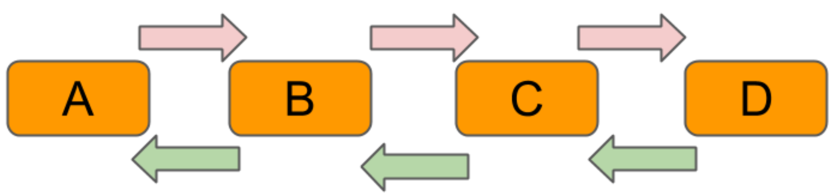
The two key assumptions are:

1. The train is always full.
2. The number of passengers traveling between any two cities on a route (not just the adjacent ones) is proportional to their population.

Given the above assumptions, we can construct the number of passengers traveling between cities on any given route. On a route consisting of k stops, we get the number of people going from city a to city b ($2 \leq a < b \leq k$) to be,

$$
\begin{aligned}
F_a^b &= \left[C \frac{N_b}{\sum_{j=a+1}^{k} N_j} \right], && \text{for,} \quad 1 = a < b \leq k, \\
F_a^b &= \left[C \, \frac{N_a}{\sum_{j=a}^{k} N_j} \right] \left[\frac{N_b}{\sum_{j=a+1}^{k} N_j} \right], && \text{for,} \quad 2 \leq a < b \leq k
\end{aligned}
\tag{2.1}
$$

where F_a^b is the number of passengers traveling from city a to city b on the given route, N_j is the population of city j, and C is the capacity of the train. The passenger data generated by this algorithm is illustrated by an example in Figure 2.1 for a route $A \rightarrow B \rightarrow C \rightarrow D$ and the corresponding reverse route $D \rightarrow C \rightarrow B \rightarrow A$. In this example, it is assumed that the population of all cities on the route is the same. The capacity of the train is taken to be 120.



Capacity=120

	A	B	C	D
A	0	40	40	40
B	60	0	20	20
C	20	20	0	60
D	40	40	40	0

Figure 2.1: Example of passenger traffic generated along a route using the rail algorithm.

The total number of passengers traveling between two cities on the rail network is a sum of the numbers obtained by considering all routes that connect the two cities. In Figure 2.1, we see that the network generated by one pair of routes is not symmetric, but the total influx for a city is equal to its total outflux. These properties remain true of the resulting complete railway network as well.

2.3 Road

Unlike air and rail transport, there is no source information of information that chronicles the number of people traveling from one city to another by road for a network as extensive as ours. We have used an algorithm based on some assumptions to construct the road network in the absence of the desired information. The details of this algorithm are published in the supplementary material of the arxiv paper [29]. The key assumptions and parameters used are described here, along with a toy example in Figure 2.3

The key assumptions are:

1. The total number of people from a city commuting by road is proportional to the population of the city

2. People only use road transport for traveling over short distances

3. The resulting road network is symmetric

Based on the above assumptions, we can choose the value of two parameters - desired mobility and the radius of road travel to construct the passenger traffic for the road network, where mobility is defined as the ratio of the number of people from a city traveling by road to the population of the city.

We use the python library *geopy* to obtain the latitude and longitude info for the cities in our network. Using this information, we compute the distance between pairs of cities. It should be noted that the distance calculated in this manner is the aerial distance, and in general, the actual distance of the road between the cities may be longer. However, we are not taking that into account for now. We construct the adjacency matrix for the road network by creating a circle of a desired radius around each city and establishing links for that city with all cities inside this circle. In Figure 2.2, we see what area the circles occupy for a few cities when a radius of 200 km is chosen; we also see a zoomed-in image of the same for the city of Pune.

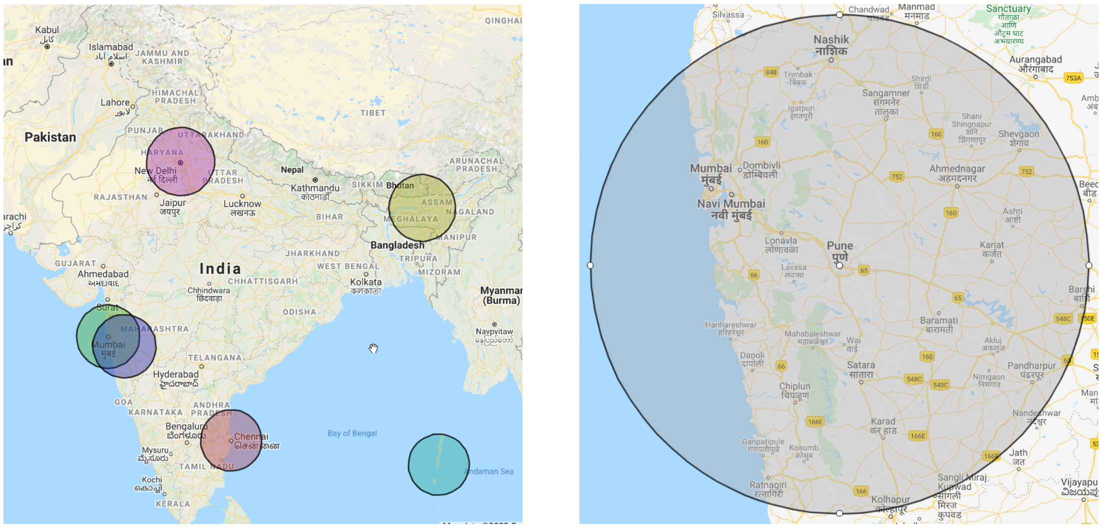


Figure 2.2: On the left: Circles drawn over a radius of 200km centered at Mumbai, Pune, Delhi, Chennai, Guwahati, and Port-Blair. On the right: Zoomed in image of the same for the city of Pune

The algorithm used to construct the passenger traffic is described using an example illustrated in Figure 2.3. Consider a network of 5 cities - A, B, C, D, and E. Assume that their geographical locations are such that the links AD, DA, BC, and CB are missing. Their populations are in the following ratio - A : B : C : D : E :: 10 : 25 : 50 : 75 : 100. Further, we assume the population of city A to be 1 Lakh and choose the value of desired mobility to be 0.0175. Given these conditions, the adjacency matrix is constructed as follows:

1. For each city, given population and mobility, we can compute the desired number of people traveling by road. This number is given in the last column of the table.

2. The cities are sorted in order of increasing population

3. We begin with the first row. The desired number of passengers are reported in the last column are distributed among the cities that A is connected to proportional to their respective populations. Then, symmetric entries are made in the first column. See all blue entries.

4. Moving to the second row, the already generated value is subtracted from the desired traffic. The rest of the passengers are distributed among the remaining links proportional

to the respective populations. Symmetric entries are generated as before. See all red entries.

5. This process is repeated for all cities, one by one. In some cases, the number of passengers can exceed the desired traffic for any city due to previously generated symmetric entries. In that case, the desired mobility for that city is increased by a small amount, and then the remaining steps are carried out.

6. The actual number of people traveling by road for each city generated by the algorithm is reported in the column ”Final”. We see that it is close to the desired number for most cities.

	A	B	C	D	E	Final	Desired
A	0	250	500	0	1000	1750	1750
B	250	0	0	1768	2357	4375	4375
C	500	0	0	3536	4714	8750	8750
D	0	1768	3536	0	7821	13125	13125
E	1000	2357	4714	7821	0	15892	17500

Figure 2.3: An example illustrating the road algorithm

For the road network constructed for the Static Hazard Map, the following parameters were used:

1. Desired mobility $\gamma = 0.015$

2. Radius of the circle of proximity = 200 Kms.

3. The desired mobility of a city is increased by $0.5\times$ desired mobility if the previously generated symmetric entries exceed the original desired traffic.

2.4 Combined

The combined network was constructed by summing up the three $\boldsymbol{F}$ matrices constructed for each mode of transport. So, this network contains the average number of passengers per day computed for each edge for all 3 modes of transport combined. The properties of all 4 networks constructed are given in Table 2.1. All networks are bidirectional. The number of edges for the combined network is not a simple sum of the same for the three individual networks because some edges are common to two or more individual networks. The average degree reported is the out-degree (the same as the in-degree) for all networks. The route symmetry index is computed as follows:

$$\text{Route Symmetry Index} = 1 - \left\langle \frac{\max_j\{|F_i^j - F_j^i|\}}{N_i} \right\rangle_i . \tag{2.2}$$

where F_i^j is the traffic on the edge from city i to city j and N_i is the population of city i. In the context of a pair of edges, if the traffic on an edge AB is the same as the traffic on its symmetric pair BA, it would contribute positively to the route symmetry index. The index is equal to 1 when the aforementioned property is true for all edges in the network. Locality of mobility tells us whether the local mobility is the same or different from the global mobility of the network. Here, local mobility for a city is the ratio of total outflux to the population of the city, and global mobility is the ratio of the total number of people in transit (sum of all values in the F matrix) to the total number of people in the network (sum of the population of all cities).

Property	Airway	Railway	Roadway	Combined
Number of Nodes	85	435	446	446
Number of Edges	1182	41594	9128	46448
Average Degree	13	95	20	104
Route symmetry index	1	0.9875	1	0.9878
Locality of Mobility	Same	Different	Same	Different
Passengers/day	7.5×10^5	8.8×10^6	2.5×10^6	1.2×10^7
Fraction of total	0.06	0.73	0.21	1.0

Table 2.1: Properties of the networks used for the Static Hazard Map

Chapter 3

Static Hazard Map

As mentioned in the previous chapter (Chapter 2), we have a network of a 446 cities across which we study the spread of an infection. If there is an outbreak at one of these cities, The Hazard Map predicts the risk of infection at the other 445 cities. The version of the model described in this chapter is called the static hazard map because the transport matrix is constant in time. In this chapter, we'll be describing the static hazard map in detail. It is divided into two sections - model and results. The model is an extension of the classical SIR model. The details are explained in Section 3.1. The results of the Static Hazard Map have been published in arxiv paper [29]. In Section 3.2, we briefly summarise the results published there.

3.1 Model

The SIR compartmental model is a staple of epidemiology, and many variants of it have been used to study the spread of infections to a reasonable degree of success. In this section, we describe the standard SIR model, followed by the modification used in our model.

3.1.1 Classical SIR Model

The classical SIR model studies the spread of an infectious disease in a well-mixed population [30, 31]. At every time step, each individual of the population belongs to one of three compartments - Susceptible (S), Infected (I), or Recovered (R). The susceptible population converts to infected at a rate α known as the infection rate. An infected individual can recover from the disease or die. In either case, they would no longer participate in the dynamics. In the rest of this document, we use the terms Recovered and Removed interchangeably. The transition from infected to recovered/removed occurs at a rate β, known as the recovery rate. The ratio α/β is known as the reproduction number R_0; it is a measure of how many people does one infected person infects on average. The classical SIR model is governed by the following set of differential equations:

$$\begin{aligned} \partial_t s &= -\alpha s i \\ \partial_t i &= \alpha s i - \beta i \\ \partial_t r &= \beta i \end{aligned}$$

where, s, i, r are fractions, α is the per capita infection rate and β is the per capita recovery rate.

3.1.2 SIR Metapopulation Model

The assumption of a well-mixed population is unreasonable for a country as large as India with a spread-out distribution of population density. To address this, we have considered a metapopulation SIR model incorporating transport dynamics over the network [10, 32]. We consider a network of cities - the population of each city is taken from the 2011 Census of India [25]. It is assumed that the population within a city is well-mixed, and the disease spread within a city is governed by the SIR dynamics. On top of this, we allow people to move from one city to another. This movement is governed by the properties of the network described in Chapter 2. It is assumed that there is no transmission of infection during transit, i. e., with respect to transit from a city A to a city B, if an individual is Susceptible (Infected/Recovered) while leaving city A, they would be Susceptible (Infected/Recovered) when arriving at city B. The movement kinetics can be described using the following set of equations:

$$\partial_t X_n = \sum_m \left[\frac{F_m^n}{N_n} X_m - \frac{F_n^m}{N_n} X_n \right]$$

where,
X could be any of the compartments S, I or R
S_n, I_n, R_n are the number of Susceptible, Infected and Recovered people in the city n respectively
$\boldsymbol{F}$ is the traffic matrix
F_m^n: Number of people going from city m to city n in a unit of time
N_n: Population of node n

Incoporating the movement over the network with the SIR model, we get the equations for our model governing the Static Hazard Map as follows:

$$\begin{aligned}
\partial_t S_n(t) &= -\alpha \frac{S_n(t) I_n(t)}{N_n} + \sum_m \left[\frac{F_m^n}{N_m} S_m(t) - \frac{F_n^m}{N_n} S_n(t) \right] \\
\partial_t I_n(t) &= +\alpha \frac{S_n(t) I_n(t)}{N_n} - \beta I_n(t) + \sum_m \left[\frac{F_m^n}{N_m} I_m(t) - \frac{F_n^m}{N_n} I_n(t) \right] \\
\partial_t R_n(t) &= +\beta I_n(t) + \sum_m \left[\frac{F_m^n}{N_m} R_m(t) - \frac{F_n^m}{N_n} R_n(t) \right] \qquad (3.1)
\end{aligned}$$

where (as previously defined):
S_n, I_n, R_n are the number of Susceptible, Infected and Recovered people in city n respectively
$\boldsymbol{F}$ is the traffic matrix
F_m^n: Number of people going from city m to city n in a unit of time
N_n: Population of node n
α: per capita infection rate
β: per capita recovery rate.

Here, we assume that α and β are the same across all cities. It should be noted that the population of each city N_n stays constant over time. The set of Equations 3.1 is solved using the Runge-Kutta-4 method using the python library *odespy*. The results obtained by running simulations according to Equations 3.1 for various conditions are discussed in the next section.

3.2 Results

We ran simulations with the initial condition as infected fraction = 0.0001 for the outbreak location and everyone else in the system being susceptible. Then we trace the spread of the infection as the simulation progresses. We analyze these results in the context of the time of arrival of the infection. The time of arrival for a city n is defined as the time at which the number of active cases in the city first crosses a certain threshold. This is a measure of the time taken for the disease to reach a city. Another important measure is Effective distance D_{eff}, which was first introduced in an article in *Science* by Brockmann and Helbeing [10]. It is defined as follows:
For Two Adjacent cities m and n: $d_m^n = 1 - \log P_m^n$
For Two Non-Adjacent cities m and n: $D_m^n = \min\left(\sum_k d_{m'}^{n'}\right)$, where k is the number of cities on some path connecting m and n.
where, $P_m^n = \frac{F_m^n}{F_m}$, $F_m = \sum_n F_m^n$. Unless specified otherwise, F refers to the transport matrix of the combined network.

We found that given an outbreak location k, the time of arrival for a city n is proportional to the Effective Distance of the city from the outbreak location D_k^n. Owing to this relation, we have defined hazard index as $h = \exp(-D_{\text{eff}})$. A higher hazard index indicates that in terms of effective distance, a city is closer to outbreak location than a city with a lower hazard index. The infection reaches a city with a higher hazard index faster than it reaches a city with a lower hazard index. In the extreme case, $D_{\text{eff}} = 0$ by construction for the outbreak location corresponding to $h = 1$ which is the highest possible hazard value. In the other extreme case, we define $D_{\text{eff}} = \infty$ for a city not connected to the outbreak location. This corresponds to a hazard index of $h = 0$. This only occurs if there are isolated cities in the network. By construction, there are no such cities in our dataset. Given an outbreak location, you can compute hazard values for all other cities. We call this the Static Hazard Map for that outbreak location. In Figure 3.1, you can find static hazard maps for some cities. In this figure, the top ten highest risk cities with respect to the outbreak location have been plotted. The radius of the circle is proportional to the hazard index. You can find such hazard maps for all 446 cities in the combined network on the following website: `https://www.iiserpune.ac.in/~hazardmap/` [33]

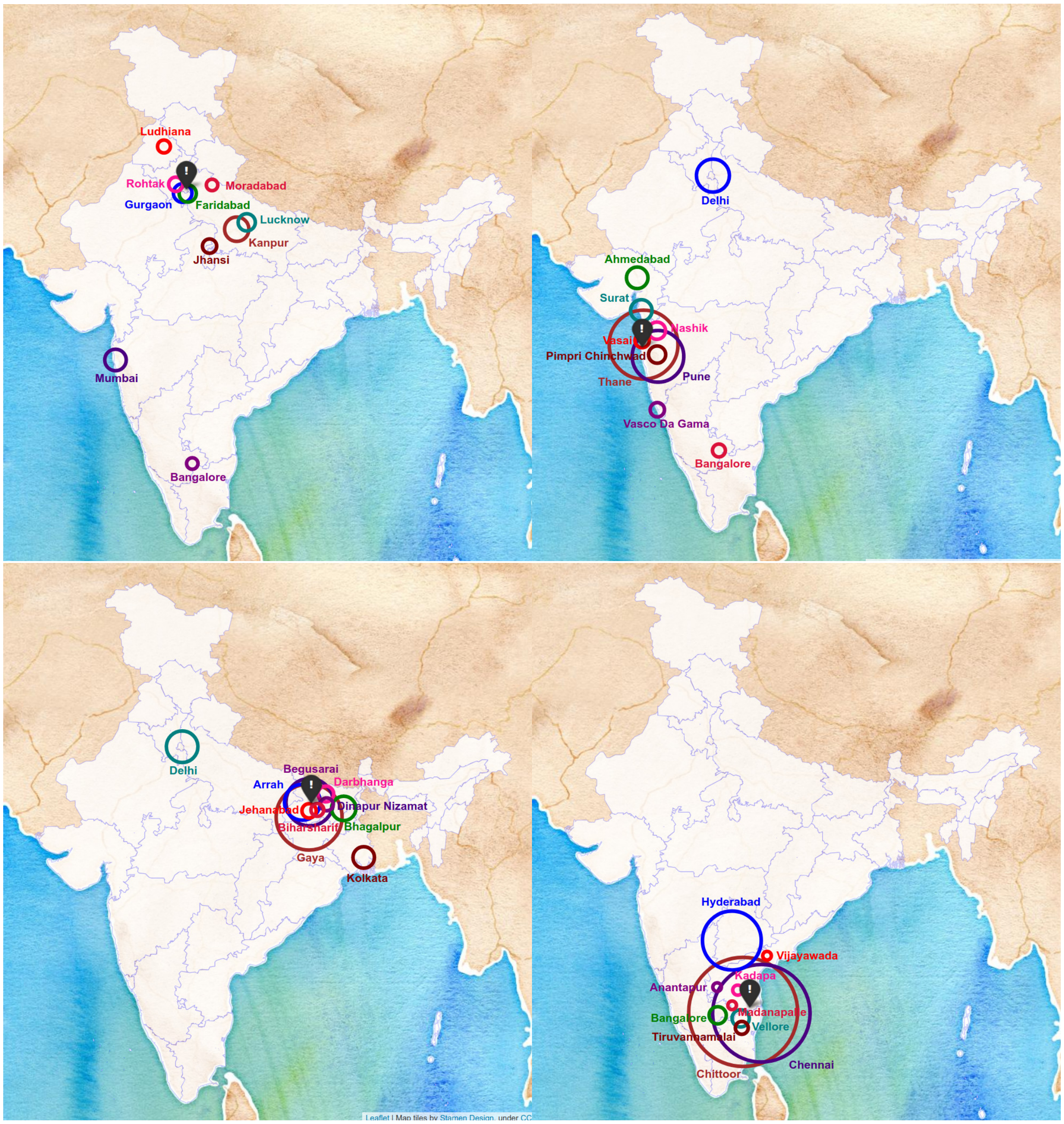


Figure 3.1: Static Hazard Maps for outbreak locations (in clockwise order) Delhi, Mumbai, Tirupathi, and Patna. The black location icon shows the outbreak location. The radius of the circle is proportional to the hazard value. The colours of the circles have no special meaning. These plots, along with the rest of the map plots in the paper, are made using Leaflet [34].

Similarly, we can construct D_{eff} and, consequently, the Static Hazard Maps with respect to a particular mode of transport. In Figures 3.2 and 3.3, you can find static hazard maps with respect to individual modes of transport - air, rail, road - and for the combined network. In the figures, the top ten highest risk cities with respect to the outbreak location have been plotted, and the radius of the circle is proportional to the hazard index.

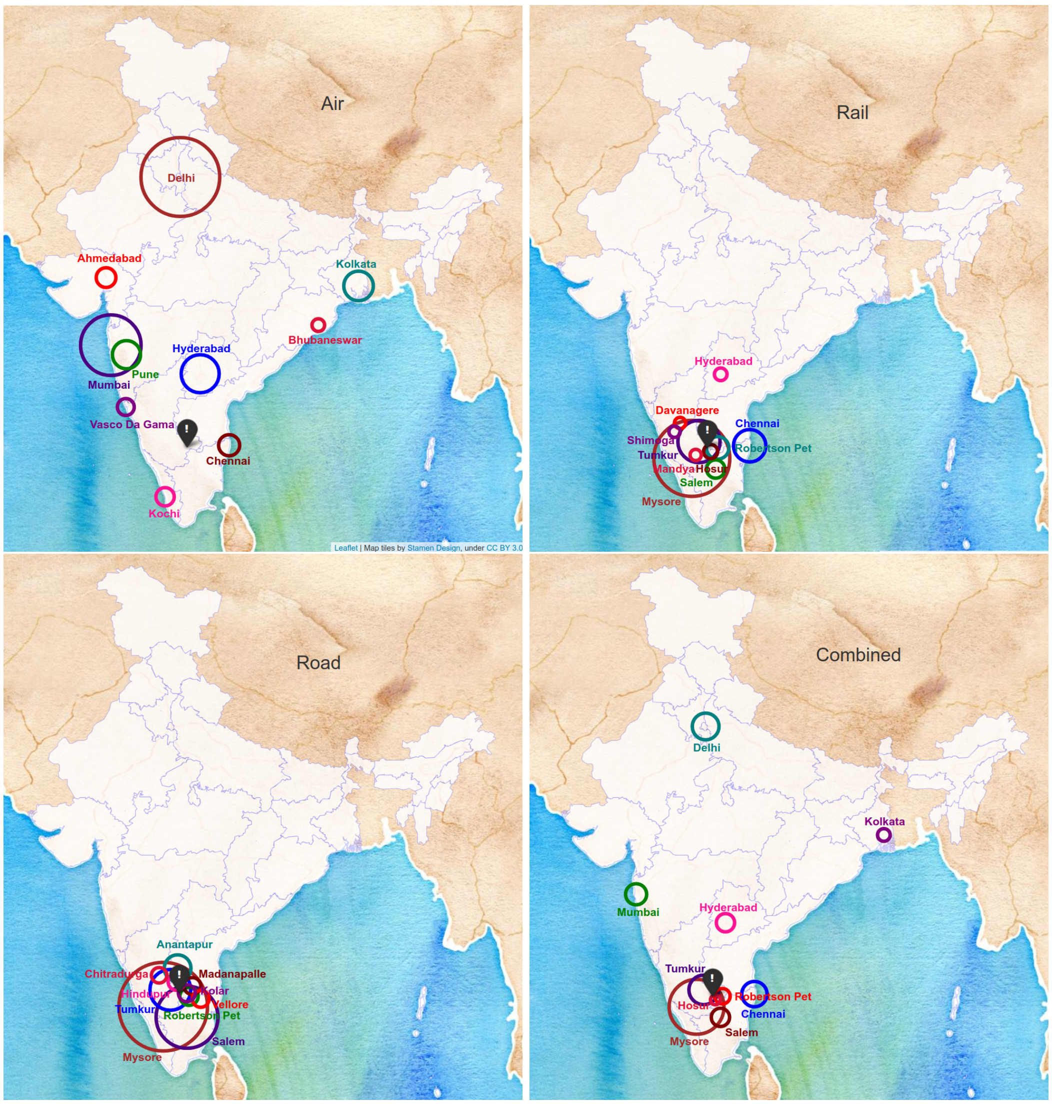


Figure 3.2: Static Hazard Maps for outbreak location Bangalore with respect to specific modes of transport (as labelled in the subfigures). The black location icon shows the outbreak location. The radius of the circle is proportional to the hazard value. The colours of the circles have no special meaning.

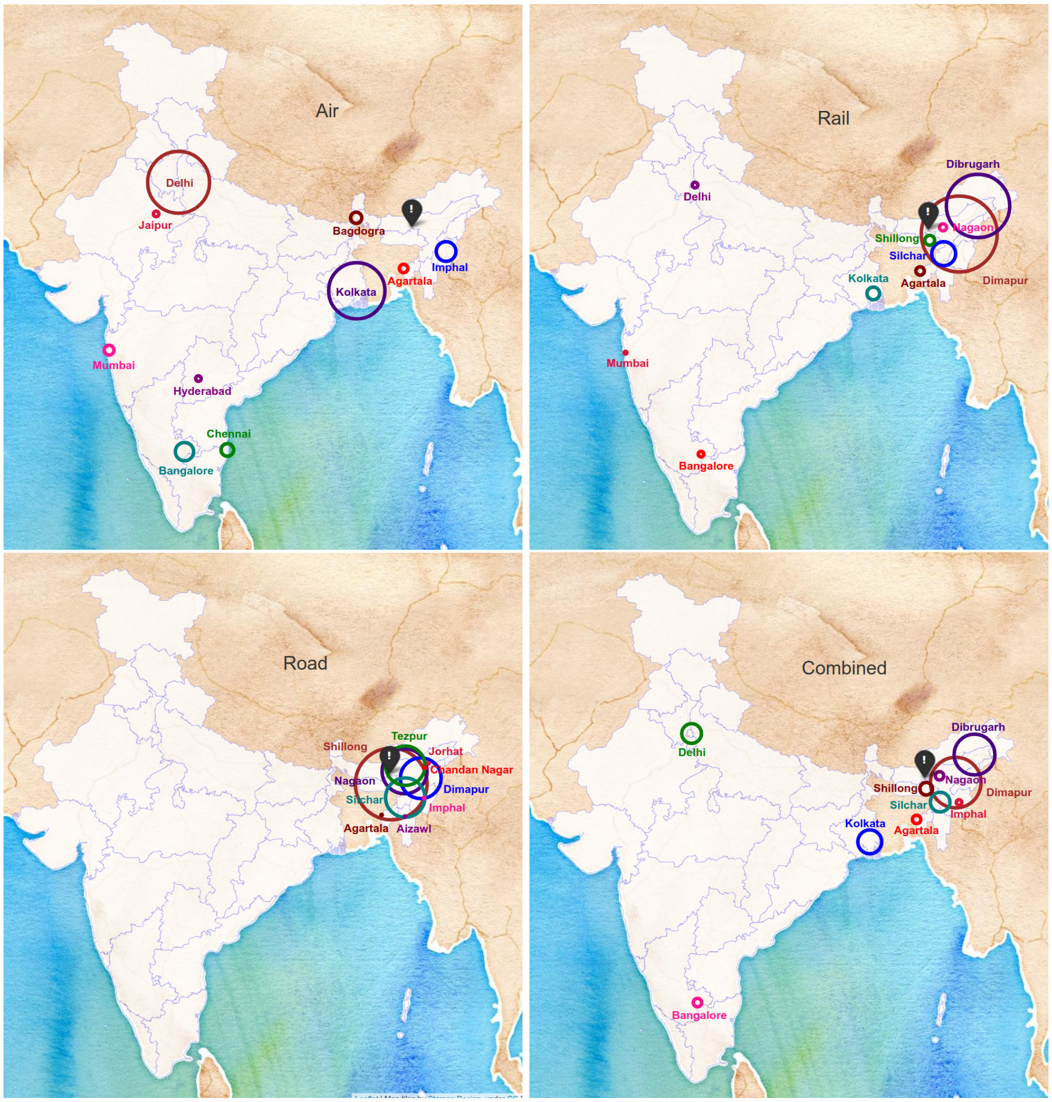


Figure 3.3: Static Hazard Maps for outbreak location Guwahati with respect to specific modes of transport (as labelled in the subfigures). The black location icon shows the outbreak location. The radius of the circle is proportional to the hazard value. The colours of the circles have no special meaning.

From the above figures, we can see that cities geographically closer to the outbreak location are at high risk for the rail and road networks, and the farther cities are virtually safe. On the other hand, for the air network, major cities of the country (having a high number of connections) are at high risk irrespective of the geographical distance from the outbreak location. For the combined network, we see the contribution of both effects. The cities at highest risk include some major cities and some which are geographically close. The Static Hazard Map for the combined network more closely resembles those of the rail and road networks as these contribute to a majority of the passenger traffic (See Table 2.1).

The brief discussion above is a summary of the results published in the arxiv paper [29].

Chapter 4

Dynamic Hazard Map

The Dynamic Hazard Map is built on the same principle as that of the Static Hazard Map. The key difference is that it takes into account the difference in speeds of the different modes of transport, where for the Static Hazard Map, just the average number of passengers per day was taken into account for all modes of transport. This chapter describes the data used for the Dynamic Hazard, the equations governing the model, and some results.

4.1 Data

In this section, we first describe the format of the data required to construct the Dynamic Hazard Map. It is illustrated with a few examples. Then we describe the specific parameters used to construct the dataset, which are used for the simulations whose results are discussed in Section 4.3

The data input for the Dynamic Hazard Map is somewhat similar to the traffic matrices described in Chapter 2. For the Static Hazard Map, we used passenger traffic in terms of the average number of passengers traveling between a pair of cities in one day. The key difference here is that the time taken to travel between said pair of cities is also taken into account. The temporal precision is now half an hour as one timestep. To input this information into the Dynamic Hazard Map, we define two time-dependent matrices F^{dep} and T^{arr}. These matrices are 3 dimensional (time, source and destination). ${F_{mn}^{dep}}(t)$ is the

number of people who left city m at time t to reach city n and $T_{mn}^{arr}(t)$ is the time at which people who left city m at time t will reach city n.
For example, suppose x people leave city m at time t_m (t_{dep}) to reach city n at time t_n (t_{arr}), then the corresponding entries for this movement in the matrices F^{dep} and T^{arr} would be:
$F_{mn}^{dep}(t_m) = x$ and $T_{mn}^{arr}(t_m) = t_n$

We can understand the F^{dep} and T^{arr} matrices better using the example of toy network. Consider a network of 3 cities: A, B, and C. For a hypothetical mode of transport, consider a simple route: 10 people leave from city A at t_0 and reach city B at t_1, 20 people leave from city B at t_1 and reach city C at t_2, finally, 30 people leave from city C at t_2 and reach city A at t_3. This is illustrated below:

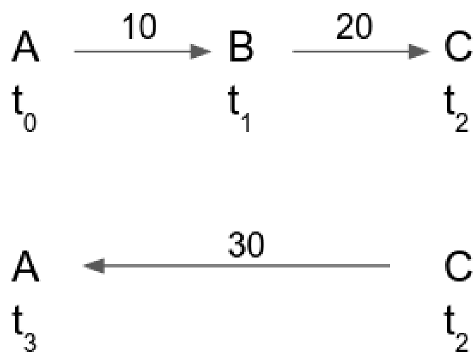


Figure 4.1: A simple route on a toy network. The number above the arrow is the number of passengers moving between the pair of cities connected by the arrow. The time index below the city is the corresponding t_{dep} or t_{arr} as applicable.

This route can be expressed in F^{dep} and T^{arr} matrices as shown in Figure 4.2. Each arrow corresponds to one entry (shown in blue). The rest of the entries are zero/empty. We use sparse matrices for the same.

F^{dep}

$F^{dep}(t_0)$

	A	B	C
A	0	10	0
B	0	0	0
C	0	0	0

$F^{dep}(t_1)$

	A	B	C
A	0	0	0
B	0	0	20
C	0	0	0

$F^{dep}(t_2)$

	A	B	C
A	0	0	0
B	0	0	0
C	30	0	0

T^{arr}

$T^{arr}(t_0)$

	A	B	C
A	-	t_1	-
B	-	-	-
C	-	-	-

$T^{arr}(t_1)$

	A	B	C
A	-	-	-
B	-	-	t_2
C	-	-	-

$T^{arr}(t_2)$

	A	B	C
A	-	-	-
B	-	-	-
C	t_3	-	-

Figure 4.2: F^{dep} and T^{arr} matrices for the route shown in Figure 4.1. The relevant entries are shown in blue.

The format of the matrices discussed above is not based on any particular mode of transport or any assumptions made about the passenger data. These matrices can be constructed from any dataset for any mode of transport if the passenger information is available at the required precision. The final aim of this project is to incorporate *real-time* updates into the Dynamic Hazard Map.

However, collecting real-life transportation data at this level of detail is a significant challenge. This part of the project is a work in progress. For now, we have converted the F matrices used for the Static Hazard Map (See Chapter 2) into the required format based on some assumptions. We construct F^{dep} and T^{arr} matrices for all three modes of transport independently.

It should be noted that when constructed in the following manner, the traffic doesn't change from one day to another.

As mentioned in Chapter 2, for each mode of transport (air, rail, and road), we have a corresponding F matrix. Here, we first describe the process of converting this matrix into the required format for air and rail transport (in Section 4.1.1). The process for rail transport is slightly different and is discussed in Section 4.1.2. Using the process described below, we obtain F^{dep} and T^{arr} matrices over the timescale of one day. In the absence of real-time data, these matrices are repeated as required over the timescale of the simulation.

4.1.1 Air and Rail

For air and rail networks, each entry in the F matrix is converted to one or more entries in the F^{dep} and T^{arr} matrices based on the following parameters:

1. Speed of the mode of transport
2. Average capacity of the vehicle

Let C be the average capacity of the vehicle and v be the speed of the mode of transport. An entry F_i^j of the static matrix is converted to n entries each for the corresponding F^{dep} and T^{arr} matrices, where $n = F_i^j/C$. Each constructed entry is of the form $F_{ij}^{dep}(t_{dep}) = C$ and $T_{ij}^{arr}(t_{dep}) = t_{arr}$, where t_{dep} is chosen n times randomly from the numbers 0-47 (Each day corresponds to 48 time points, since we have chosen the temporal precision of half an hour.), and $t_{arr} = t_{dep} + t_{travel}$ where t_{travel} is the time taken to travel from i to j. $t_{travel} = d_i^j/v$ where d_i^j is the aerial distance between cities i and j calculated from their latitude and longitude information.

The speed of the mode of transport and capacity of vehicle assumed are given below.

For air:

1. Speed = 600km/h
2. Average number of people on a flight = 200

For rail:

1. Speed = 50km/h
2. Average number of people on a train = 1500

4.1.2 Road

For road transport, the traffic in a day (and corresponding t_{dep}) is distributed evenly from 6 am to midnight instead of assuming a capacity of the vehicle and picking t_{dep} randomly. The rest of the process is the same as described in Section 4.1.1. The speed was assumed to be 50km/h.

4.1.3 Combined

Unlike for the Static Hazard Map, we cannot perform an element-wise sum of the F^{dep} and T^{arr} matrices for the individual modes of transport to obtain a combined network. Instead of this, transport over the three modes of transport is integrated when the simulation is run. Details are given in Section 4.2.1.

4.2 Model

In principle, the model for the Dynamic Hazard Map is very similar to that of the Static Hazard Map. The key difference is in how the movement kinetics are now described. The details about the movement kinetics are described in the following section.

4.2.1 Movement Kinetics

We can describe the evolution of the number of passengers belonging to a certain compartment X of a city n as follows:

$$\partial_t X_n = X_n^{arr}(t) - X_n^{dep}(t) \tag{4.1}$$

where,
X is could be any of the compartments S, I or R
S_n, I_n, R_n are the number of Susceptible, Infected and Recovered people in city n respectively
$X_n^{arr}(t)$ is the number of people of compartment X that are arriving at city n at time t
$X_n^{dep}(t)$ is the number of people of compartment X that are leaving city n at time t

In the context of the Dynamic Hazard Map, there are two variables associated with the population of a city n. We call these N'_n and $N_n(t)$, where N'_n is the city population as taken from the census and $N_n(t)$ is the instantaneous population of the city. Henceforth, "population" refers to N'_n and "instantaneous population" refers to $N_n(t)$.

$X_n^{dep}(t)$ is computed as follows:
$X_n^{dep}(t)$ = number of people of compartment X that are leaving city n at time t
$X_n^{dep}(t)$ = (fractional number of people belonging to compartment X of city n at time t)$\times$
(number of people of leaving n at time t)
Therefore,

$$X_n^{dep}(t) = \frac{X_n(t)}{N_n(t)} \sum_m F_{nm}^{dep}(t) \tag{4.2}$$

$X_n^{arr}(t)$ is computed iteratively. We maintain vectors $X^{arr}(t)$ for all timesteps t for the duration of the simulation. These vectors are updated at every timestep. Consider an entry $X_n^{arr}(t)$, it is the number of people of compartment X that are arriving at city n at time t. It is computed as follows:
at time t_0, $X_n^{arr}(t) =$ all people of compartment X that leave any city $m \neq n$ at time t_0
and arrive at city n at time t.
$X_n^{arr}(t) = \sum_m$ (no. of people of compartment X going from m to n in time t_0 to t)

$X_n^{arr}(t)$ = $\sum_m$ [(fractional number of people belonging to compartment X of city m at time t_0)$\times$(number of people going from m to n in time t_0 to t)]

Therefore, at time t_0

$$X_n^{arr}(t) = \sum_m \left[\frac{X_m(t_0)}{N_m(t_0)} F_{mn}^{dep}(t_0) \right] \qquad \forall \quad T_{mn}^{arr}(t_0) = t$$

Similarily, at time t_1, we would add an additional term as shown:

$$\Delta X_n^{arr}(t) = \sum_m \left[\frac{X_m(t_1)}{N_m(t_1)} F_{mn}^{dep}(t_1) \right] \qquad \forall \quad T_{mn}^{arr}(t_1) = t$$

Therefore, at time t_1

$$X_n^{arr}(t) = \sum_m \left[\frac{X_m(t_0)}{N_m(t_0)} F_{mn}^{dep}(t_0) \right] + \sum_m \left[\frac{X_m(t_1)}{N_m(t_1)} F_{mn}^{dep}(t_1) \right] \qquad \forall \quad T_{mn}^{arr}(t_0) = t, T_{mn}^{arr}(t_1) = t$$

We keep adding terms as shown above for all time steps before t.

Finally, we can write,

$$X_n^{arr}(t) = \sum_{k=t_0}^{t} \sum_m \left[\frac{X_m(k)}{N_m(k)} F_{mn}^{dep}(k) \right] \qquad \forall \quad T_{mn}^{arr}(k) = t \tag{4.3}$$

$X^{dep}(t)$ and $X^{arr}(t)$ are computed as described above for all three modes of transport individually and then summed over at every timestep. The transport of passengers over the *combined* network is thus described completely by Equations 4.1, 4.2, and 4.3.

4.2.2 SIR Metapopulation Model

Similar to the model for the Static Hazard Map (See Equation 3.1), the model for the Dynamic Hazard Map employs an SIR metapopulation model with transport dynamics over the network. It is assumed that the population within a city is well-mixed, and the disease spread within a city is governed by the SIR dynamics. It is also assumed that there is no transmission of infection during transit i. e., with respect to transit from a city A to a city

B, if an individual is Susceptible (Infected/Recovered) while leaving city A, they would be Susceptible (Infected/Recovered) when arriving at city B.

The equations governing the Dynamic Hazard Map are as follows:

$$\begin{aligned}
\partial_t S_n(t) &= -\alpha \frac{S_n(t) I_n(t)}{N_n(t)} + S_n^{arr}(t) - S_n^{dep}(t) \\
\partial_t I_n(t) &= +\alpha \frac{S_n(t) I_n(t)}{N_n(t)} - \beta I_n(t) + I_n^{arr}(t) - I_n^{dep}(t) \\
\partial_t R_n(t) &= +\beta I_n(t) + R_n^{arr}(t) - R_n^{dep}(t)
\end{aligned} \tag{4.4}$$

where

$$N_n(t) = S_n(t) + I_n(t) + R_n(t)$$

where all variables are as previously defined.

The set of Equations 4.4 is solved using the Euler method.

4.3 Results

Given an outbreak location and an initial infected fraction, we can solve the set of Equations 4.4 to compute $S(t), I(t)$ and $R(t)$ for all cities in the network. In this section, we discuss in detail the results obtained for the Outbreak Location Mumbai. A similar analysis can be carried out, and Dynamic Hazard Maps constructed with respect to any of the 446 cities as an outbreak location.

Let city M be Mumbai. The initial conditions are taken to be:
initial infected fraction, $i_0 = 0.001$
$S_M(t_0) = N'_n - I_M(t_0), \qquad I_M(t_0) = i_0 N'_n, \quad R_M(t_0) = 0$
$S_n(t_0) = N'_n, \qquad I_n(t_0) = 0, \qquad R_n(t_0) = 0 \qquad \forall\, n \neq M$
per capita infection rate, $\alpha = 1.5$
per capita recovery rate, $\beta = 1.0$
Time duration of simulation = 60 days

An important quantity to monitor is the infected fraction $i_n(t)$ for a city n. It is defined as $i_n(t) = I_n(t)/N'_n$. Higher the $i_n(t)$ for a city, the greater *hazard* that it faces. In Figure 4.3, the infected fraction $i_n(t)$ for all cities in the network is plotted as a function of time. The first red curve corresponds to $i_M(t)$, i. e., the evolution in infected fraction for the outbreak location Mumbai. The other coloured curves correspond to $i(t)$ for the other cities in the network. The colours have no specific meaning. We see that the infected fraction rises initially for all cities and then falls as infected people start getting recovered. We see that the peak of all the curves is at a similar height (around 0.06). This implies that at the height of the infection, about 6% of the city population is infected (active cases). This peak occurs at different times for different cities as the infection spreads through the network. The vertical dashed lines are addressed below.

Based on the above information, we can define a **hazard index**, h to predict the risk of infection at a city at a given time. We define:

$$h_n(t) = \frac{i_n(t)}{i_{max}} \qquad \text{where, } i_{max} = \max_{n,t} i_n(t) \tag{4.5}$$

The hazard index h lies between 0 and 1, with $h = 0$ signifying the lowest risk of infection and $h = 1$ the highest. Drawing from the earlier discussion, we can say that for all cities, h starts with $h = 0$ at $t = t_0$, rises closes to $h = 1$ at the peak of infection, then falls again. At a certain time t, the cities in the network will have varying values of h (Consider a vertical slice from Figure 4.3. For example, one of the black dashed lines). So, at each time step t, we can construct a ranked list of cities based on h. This is what we refer to as the Dynamic Hazard Map.

In Table 4.1, we have reported the top ten highest risk cities at 4 time points - 15, 25, 35, and 45 days after the outbreak. These time points are shown using dashed black lines in Figure 4.3. We see that the highest risk cities are initially in and around Maharashtra (close to Mumbai, outbreak location). As time progress, these cities recover from the infection, and the highest risk cities are farther away.

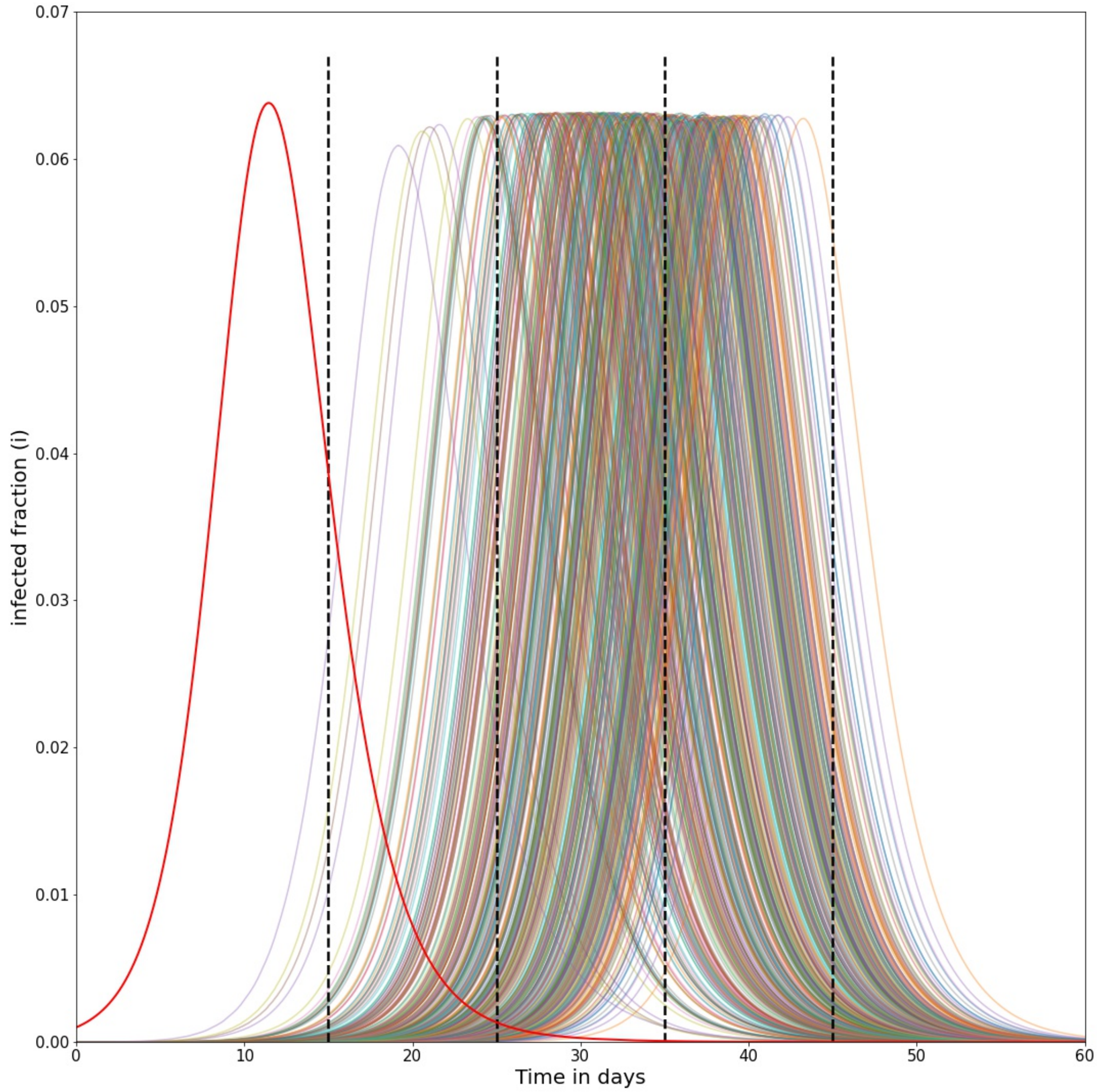


Figure 4.3: Infected fraction plotted against time for all cities in the network with respect to outbreak location Mumbai. The first red curve corresponds to Mumbai. For the other curves, the colour has no meaning. The black dashed vertical lines indicate the time points for which the highest risk cities as listed in Table 4.1.

t = 15 days		
Rank	City	h
1	Mumbai	0.6055
2	Vasco Da Gama	0.4334
3	Valsad	0.2509
4	Satara	0.2089
5	Thane	0.1625
6	Pune	0.0742
7	Vasai	0.0562
8	Nashik	0.0495
9	Mangalore	0.0477
10	Pimpri Chinchwad	0.0457

t = 25 days		
Rank	City	h
1	Bhusawal	0.9833
2	Kochi	0.9831
3	Udupi	0.9828
4	Bhiwandi	0.9801
5	Ahmadnagar	0.9761
6	Kozhikode	0.9728
7	Mira-Bhayandar	0.9718
8	Badlapur	0.9694
9	Pimpri Chinchwad	0.9632
10	Ulhas Nagar	0.9626

t = 35 days		
Rank	City	h
1	Karimnagar	0.9872
2	Kasganj	0.9871
3	Bettiah	0.9870
4	Durgapur	0.9865
5	Dhanbad	0.9862
6	Raiganj	0.9853
7	Katihar	0.9850
8	Gurgaon	0.9847
9	Kanchipuram	0.9839
10	Hassan	0.9837

t = 45 days		
Rank	City	h
1	Ganganagar	0.8589
2	Pilibhit	0.7310
3	Munger	0.6517
4	Proddatur	0.6409
5	Shimla	0.5853
6	Kulti	0.5296
7	Gangtok	0.5287
8	Darjeeling	0.5009
9	Bhind	0.4742
10	Gangawati	0.4680

Table 4.1: Top ten highest risk cities 15, 25, 35, and 45 days after disease outbreak at Mumbai. The hazard index h is computed as described in Equation 4.5

Chapter 5

Conclusion and Future Directions

In this thesis, we discussed in detail the construction of an infectious diseases hazard map for India. Two versions of the same were discussed - the Static Hazard Map and the Dynamic Hazard Map. Both were constructed by using an SIR metapopulation model incorporating real-life data for a transportation network spanning 446 cities and connected by three modes of transport.

There are some key differences between the Static and Dynamic hazard maps. For the Static Map, the difference in time scales associated with the different modes of transport is not taken into account. The passenger data is aggregated over a length of time and an average number of passengers per day is computed and used for the Static Map. For the Dynamic Map, the model is set up so that real-time passenger data can be incorporated up to the precision of half an hour.

For each Map, we defined a **hazard index** that quantifies the risk from the infection for a city/town given an outbreak location. It implies that a city with a higher hazard index faces a higher risk than a city with a lower hazard index. This hazard index is computed differently for the two versions of the Map.

For the Static Map, the hazard index is dependent on the Time of Arrival of infection at a city, i.e., the time at which the number of active cases in the city crosses a certain threshold. Moreover, we know that this Time of Arrival for a city is directly proportional to the *effective distance* of the city from the outbreak location [29]. So, once the traffic matrix is fixed, the Static Hazard Map depends solely on the outbreak location.

For the Dynamic Map, the hazard index is dependent on the number of active cases in a city at a given time. In this context, all cities in the network have a high hazard value at a specific time. The Dynamic Hazard Map indicates the changing cluster of cities facing the brunt of the disease as the infection spreads through the country.

The future work for this project can be divided into two broad categories. Firstly, we have to set up a dataset with air and rail data that evolves in *real-time* to incorporate it into the Dynamic Hazard Map as described in Chapter 4. Secondly, we need to compare the results obtained from the Dynamic Hazard Map to the current real-life evolution of the COVID-19 pandemic in India.

Bibliography

[1] https://covid19.who.int/, https://coronavirus.jhu.edu/map.html.

[2] https://www.covid19india.org/.

[3] Belik, V., Geisel, T., Brockmann, D., Natural human mobility patterns and spatial spread of infectious diseases. Physical Review X **1**, 011001. (2011).

[4] Coltart, C.E., Behrens, R.H., The new health threats of exotic and global travel. British Journal of General Practice **62**, 512513. (2012).

[5] Helbing, D., Globally networked risks and how to respond. Nature **497**, 5159. (2013).

[6] Barbosa, H., Barthelemy, M., Ghoshal, G., James, C.R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J.J., Simini, F., Tomasini, M., 2018. Human mobility: Models and applications. Physics Reports **734**, 174. (2018).

[7] Gautreau, A., Barrat, A., Barthelemy, M., Global disease spread: statistics and estimation of arrival times. Journal of theoretical biology **251**, 509522. (2008).

[8] Barthlemy, M., Spatial networks. Physics Reports **499**, 1101. (2011).

[9] Feng, L., Zhao, Q., Zhou, C., Epidemic in networked population with recurrent mobility pattern. Chaos, Solitons & Fractals 139, 110016. (2020).

[10] Brockmann, D., Helbing, D., The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. Science **342**, 13371342. (2013).

[11] Arenas, A., Cota, W., Gmez-Gardees, J., Gmez, S., Granell, C., Matamalas, J.T., Soriano-Paos, D., Steinegger, B., Modeling the Spatiotemporal Epidemic Spreading of COVID-19 and the Impact of Mobility and Social Distancing Interventions. Physical Review X **10**, 041055. (2020).

[12] Althouse, B.M., Wenger, E.A., Miller, J.C., Scarpino, S.V., Allard, A., Hbert-Dufresne, L., Hu, H., Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2. arXiv:2005.13689. (2020).

[13] Garcia-Gasulla, D., Napagao, S.A., Li, I., Maruyama, H., Kanezashi, H., Perez-Arnal, R., Miyoshi, K., Ishii, E., Suzuki, K., Shiba, S., Global Data Science Project for COVID-19 Summary Report. arXiv:2006.05573. (2020).

[14] R. Gopal, V. K. Chandrasekar and M. Lakshmanan, Dynamical modelling and analysis of COVID-19 in India, Current Science **120** 1342 (2021).

[15] Bedi, P., Gole, P., Gupta, N., Jindal, V., Projections for COVID-19 spread in India and its worst affected five states using the Modified SEIRD and LSTM models. arXiv:2009.06457. (2020).

[16] Khajanchi, S., Sarkar, K., Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India. Chaos: An Interdisciplinary Journal of Nonlinear Science **30**, 071101. (2020).

[17] Das, A., Dhar, A., Goyal, S., Kundu, A., Pandey, S., COVID-19: Analytic results for a modified SEIR model and comparison of different intervention strategies. Chaos, Solitons & Fractals 110595. (2021).

[18] Jha, V., Forecasting the transmission of Covid-19 in India using a data driven SEIRD model. arXiv:2006.04464. (2020).

[19] Khajanchi, S., Sarkar, K., Mondal, J., Perc, M., Dynamics of the COVID-19 pandemic in India. arXiv:2005.06286. (2020).

[20] Mishra, R., Gupta, H.P., Dutta, T., Analysis, Modeling, and Representation of COVID-19 Spread: A Case Study on India. arXiv:2008.13116. (2020).

[21] Sarkar, K., Khajanchi, S., Nieto, J.J., Modeling and forecasting the COVID-19 pandemic in India. Chaos, Solitons & Fractals **139**, 110049. (2020).

[22] Pujari, B.S., Shekatkar, S., Multi-city modeling of epidemics using spatial networks: Application to 2019-nCov (COVID-19) coronavirus in India. medRxiv. https://doi.org/10.1101/2020.03.13.20035386 (2020).

[23] Gupta, S., Shah, S., Chaturvedi, S., Thakkar, P., Solanki, P., Dibyachintan, S., Roy, S., Sushma, M.B., Godbole, A., Jaseem, N. *et. al.*, An India-specific compartmental model for Covid-19: projections and intervention strategies by incorporating geographical, Infrastructural and response heterogeneity. arXiv:2007.14392. (2020).

[24] Earthquake hazard maps by the United States Geological Survey https://www.usgs.gov/natural-hazards/earthquake-hazards/maps.

[25] Census India (2011). https://www.censusindia.gov.in/2011-prov-results/paper2/data_files/India2/Table_2_PR_Cities_1Lakh_and_Above.pdf.

[26] `http://www.knowindia.net/aviation.html`.

[27] Website of the Directorate General of Civil Aviation `dgca.gov.in/`.

[28] `https://enquiry.indianrail.gov.in/mntes/`.

[29] O. Sadekar, M. Budamagunta, G. Sreejith, S. Jain, and M. Santhanam, "An infectious diseases hazard map for india based on mobility and transportation networks," *arXiv preprint arXiv:2105.15123*, 2021.

[30] Ronald, R., and Hilda, P., An application of the theory of probabilities to the study of a priori pathometry.–Part IIIProc. R. Soc. Lond. A **93**: 225240. (1917).

[31] Kermack, W. O. and McKendrick, A. G., A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. A **115**:700721. (1927).

[32] Gong, Y., Song, Y., Jiang, G., Epidemic spreading in metapopulation networks with heterogeneous infection rates, Physica A: Statistical Mechanics and its Applications, **416**, 208-218. (2014).

[33] For more information, visit `https://www.iiserpune.ac.in/~hazardmap/`.

[34] Leaflet — Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under CC BY SA.