Modeling an Epidemic

Sahel Mohammad Iqbal
1711112

November 30, 2020

Table of Contents

Why model an epidemic?

- Imagine you are responsible for deciding which cities to put under lockdown during Covid-19. How do you make your decision?
- You can go ahead and put any current hotspots under lockdown, but this fails to stop the spread.
- To effectively reign in the spread of an epidemic, we need a certain degree of predictive power, to find where the virus is likely to spread from the current locations.

Table of Contents

Overview of the meta-population model

- Meta-population models consider the existence of small, discrete geographical units within which people may intermingle freely, with multiple such units connected by some mode of transportation.
- To model a pandemic like Covid-19, we choose cities or urban centres with airports as our discrete sub-populations, among which people may travel via flights.

Overview of the meta-population model

- These models need a lot of data.
- The International Air Transport Association (IATA) database has information on 3100 airports and 17182 connections between them representing more than 99% of worldwide commercial traffic by plane [1].
- Also rely on census data to get population and population density information.

Overview of the meta-population model

- We can now write down a high-level equation for the rate of change of I_i which represents the number of infected people in a city i [1].

$$\frac{\partial I_i}{\partial t} = K_i(t) + \Omega_i(t) \tag{1}$$

- $K_i(t)$ represents the variation of infected individuals due to the *infection dynamics inside the city*
- $\Omega_i(t)$ corresponds to the net balance of infectious individuals *traveling in and out of the city*.

Overview of the meta-population model

- Ω_i, which we call the transport operator, depends on the probability p_{ij} that an infected individual will go from city i to city j, and may be written as

$$\Omega_i = \sum_{j \in V(i)} (p_{ji} I_j - p_{ij} I_I) \tag{2}$$

- This represents the number of infected people coming into the city minus those going out to all connected cities.
- It acts as a coupling term among the evolution of the epidemics in the various cities.

Disease Evolution in a Closed Population

- We'll now look at how we can model the function $K_i(t)$, which models the disease evolution within a closed population.

Compartmentalized models

- The population is divided into compartments such as susceptible (S), infected (I), permanently recovered (R), diseased (D), etc.
- These compartments are mutually exclusive and exhuastive.

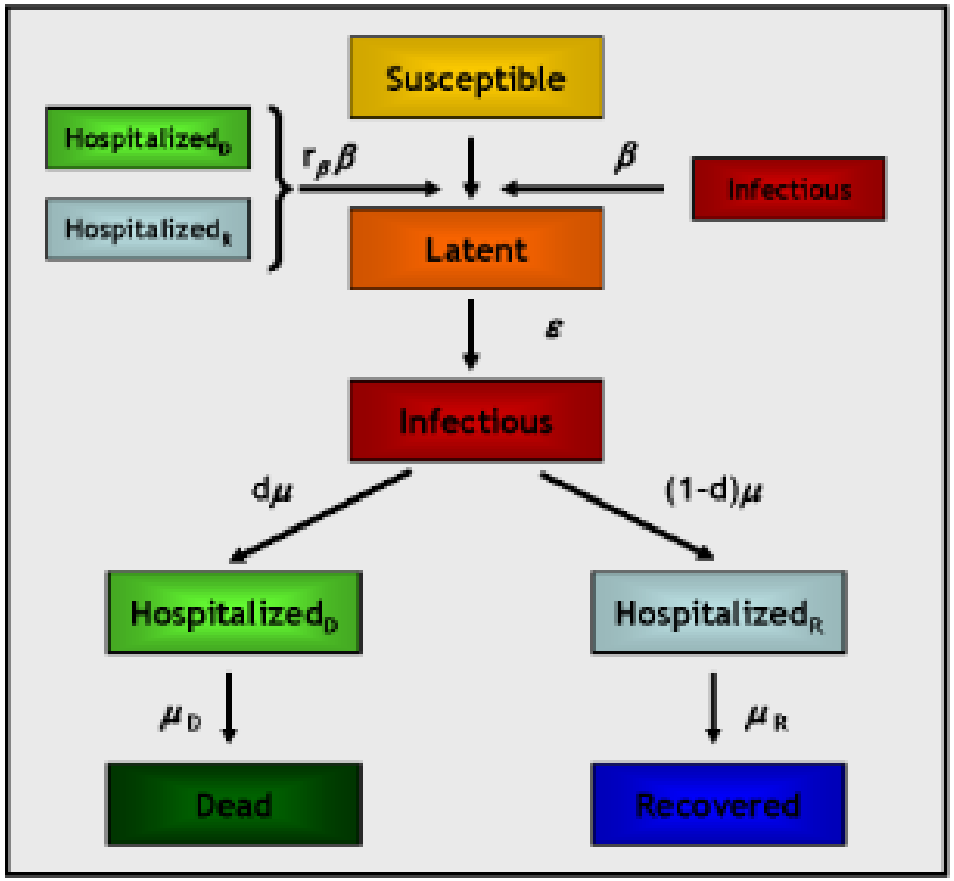

Figure: A possible compartmentalization of a population with various transition probabilities.

Compartmentalized models

- A person can jump from one compartment to the other with probabilities that depend on various parameters specific to the disease under consideration, such as
  1. The reproduction number R (the average number of secondary infections caused by an infected individual),
  2. The latency period (the time taken for a person who contracted the infection to show symptoms),
  3. The fatality and recovery rates

Some notation

- The total population of each city is N_j
- $X_j^{[m]}(t)$ is the number of individuals in the compartment $[m]$ in the city j at time t.

Compartmentalized models

- What are the various processes by which people change compartments?

Compartmentalized models

- The first class of processes refers to the spontaneous transition of one individual from one compartment $[m]$ to another compartment $[h]$. In that case, the following equation holds.

$$X^{[m]} \to X^{[m]} - 1, X^{[h]} \to X^{[h]} + 1 \tag{3}$$

- Example: A person recovering from the illness ($I \to R$), or starting to show symptoms ($L \to I$).

Compartmentalized models

- The variation in the number of individuals $X^{[m]}$ is given by

$$\sum_h \nu_h^m a_h^m X^{[h]}$$

- a_h^m is the transition rate from the class $[h]$ to the class $[m]$
- $\nu_h^m \in [-1, 0, 1]$ signifies whether people are added to or subtracted from $X^{[m]}$ due to this process [2]

Compartmentalized models

- The second class of processes refers to binary interaction among individuals such as when a susceptible person contracts the virus on interaction with an infectious one ($S + I \rightarrow 2I$).
- In this case, the rate of variation of $X^{[m]}$ is given by

$$\sum_{h,g} \nu_{h,g}^{m} a_{h,g}^{m} X^{[h]} X^{[g]} N^{-1}$$

where similar to the previous process, $a_{h,g}^{m}$ is the transition probability for the process and $\nu_{h,g}^{m} \in [-1, 0, 1]$.

Compartmentalized models

- Homogeneous approximation - The probability for each individual of class $[h]$ to interact with an individual of class $[g]$ is simply equal to the density $X^{[g]}/N$ of such individuals [2].

Compartmentalized models

- Combining both the processes, the full equation is

$$\partial_t X_j^{[m]} = \sum_h \nu_h^m a_h^m X_j^{[h]} + \sum_{h,g} \nu_{h,g}^m a_{h,g}^m X_j^{[h]} X_j^{[g]} N^{-1} \tag{4}$$

Sources of stochasticity

- In getting to the above equation, we made two assumptions.
  1. We know the transition rates exactly.
  2. The transition rates are constants for the whole population.
- Both these assumptions are wrong.

Sources of stochasticity

- We can only calculate the transition rates from the data that we observe, and our knowledge of them evolves with the pandemic.
- Also, there are large variations in different people's calculations of these transition rates [3].

Sources of stochasticity

- The more erroneous assumption is that the transition rates are constants for the whole population.
- As an illustrative example, let's consider the recovery rates of individuals. The strenght of the immune system varies wildly person to person.
- So from a statistical point of view, the recovery rate of a particular individual in the population can only be described as random (within a given range).

Sources of stochasticity

- In addition, decisions taken by the government or by the individuals themselves may also affect these values. For example, a person who chooses to go out less frequently during the time of an epidemic will have a far less chance of contracting the virus than someone who chooses to do the opposite.

Sources of stochasticity

- Hence there's a lot of randomness in the various transition rates, which we did not account for in the previous equation.
- So to add this stochasticity to our model is the obvious progression.

The epidemic Langevin equation

- Under the assumption of large populations, to each reaction term we associate a noise term with amplitude proportional to the square root of the reaction term [4].
- The Langevin equation can then be written as [2]

$$\partial_t X_j^{[m]} = \sum_h \nu_h^m a_h^m X_j^{[h]} + \sum_{h,g} \nu_{h,g}^m a_{h,g}^m X_j^{[h]} X_j^{[g]} N^{-1} + \sum_h \nu_h^m \sqrt{a_h^m X_j^{[h]}} \cdot \eta_h + \sum_{h,g} \nu_{h,g}^m \sqrt{a_{h,g}^m X_j^{[h]} X_j^{[g]} N^{-1}} \cdot \eta_{h,g} \tag{5}$$

The epidemic Langevin equation

- The full equation that models the epidemic will include the transport operator added to the above equation.
- Extremely large number of equations to solve.
- Can be solved using numerical techniques, such as the Euler method.

Table of Contents

How do we check the accuracy of the model?

- We create multiple realizations of the epidemic, and check the level of similarity between various pairs.
- A realization is an evolution of the epidemic from the same set of initial conditions but different values noise values.
- The logic is that if a large number of realizations are fairly similar, then the model should have decent accuracy because we're tried out a lot of different possible values for noise.

The similarity function

- Consider the vector $\vec{\pi}(t)$ whose components are $\pi_j(t) = I_j(t)/\sum I_l$ which is the normalized probability that an infected individual is in city j.
- Such a measure of similarity $sim(\vec{\pi}^A(t), \vec{\pi}^B(t))$ is given by a function called the Hellinger affinity, which is given by [2]

$$sim(\vec{\pi}^A(t), \vec{\pi}^B(t)) = \sum_j \sqrt{\pi_j^A \pi_j^B} \tag{6}$$

The overlap function

- We need to consider the possible differences in the total percentage of infected people worldwide given by $i = \sum_j I_j/P$ where P is the total global population.
- This is done using the similarity function $sim(\vec{i}^A(t), (\vec{i}^B(t))$ where $\vec{i}(t) = (i, 1-i)$. So the total overlap function measuring the similarity between two different outbreak realizations is defined by [2]

$$\Theta(t) = sim(\vec{i}^A, \vec{i}^B) \cdot sim(\vec{\pi}^A, \vec{\pi}^B) \tag{7}$$

The overlap function

- The overlap is maximal when the very same cities have the very same number of infectious individuals in both realizations, and $\Theta(t) = 0$ if the two realizations do not have any common infected cities at time t.
- A large overlap corresponds to a predictable evolution, providing a direct measure of the reliability of the epidemic forecast. An example of this sort of analysis is given in figure 32.

Example

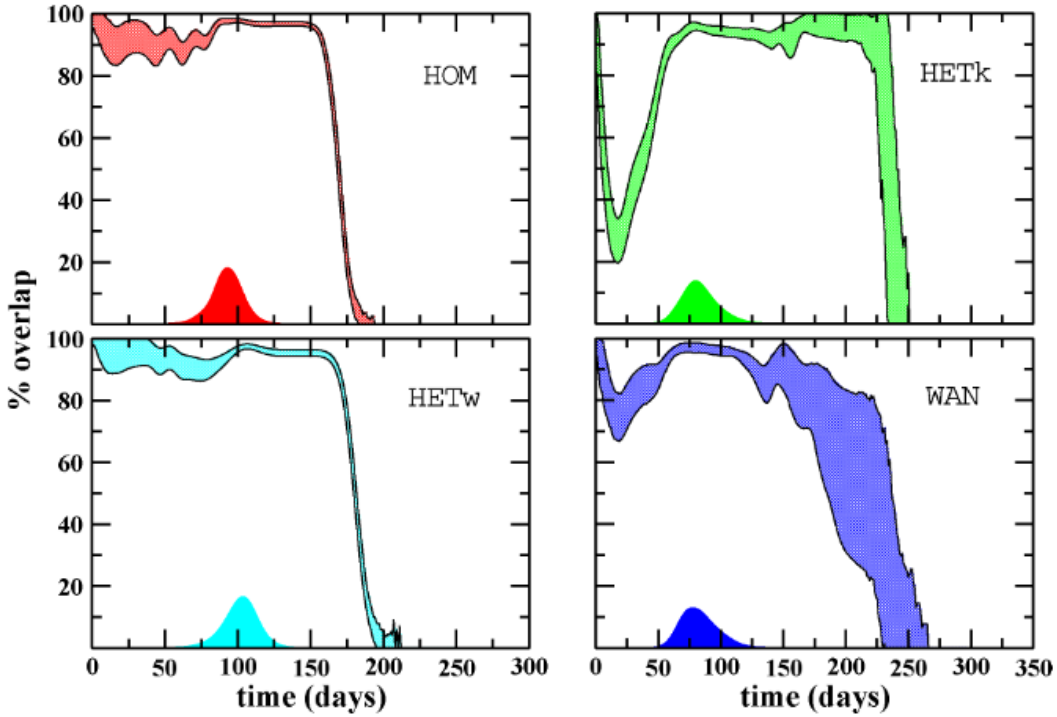

Figure: Overlap functions plotted for 5000 pairs of realizations of an epidemic [1]. The four plots are for four different air-trasportation datasets of which WAN (bottom-right) is the complete one. The shaded areas denote standard deviations. The small bumps at the bottom of each of the four graphs are prevalence profiles of the epidemic (with different units on the y-axis).

Takeaways from the graph

- At time 0, the overlap function is 100%
- This is followed by a sharp dip, which shows that the early days of the epidemic are very prone to stochastic fluctuations.
- The overlap function subsequently rises to above 90% at around the same location as the peak of the prevalence function.
- This means that when the epidemic is at its most widespread, the models have great predictability.
- Then the overlap function quickly dips to zero as the epidemic dies down.

Why are different realizations similar?

- Heterogeneity in the airline transportation network. ie, some routes (and locations) are frequented more by travellers (consider Kochi-Dubai and Kochi-Bhubaneswar as an example).
- Models can hence pick out places that are especially vulnerable to large-scale infections, and that is what makes them very accurate, particularly during times when the epidemic/pandemic is at its height.

Verifying the model against SARS Outbreak

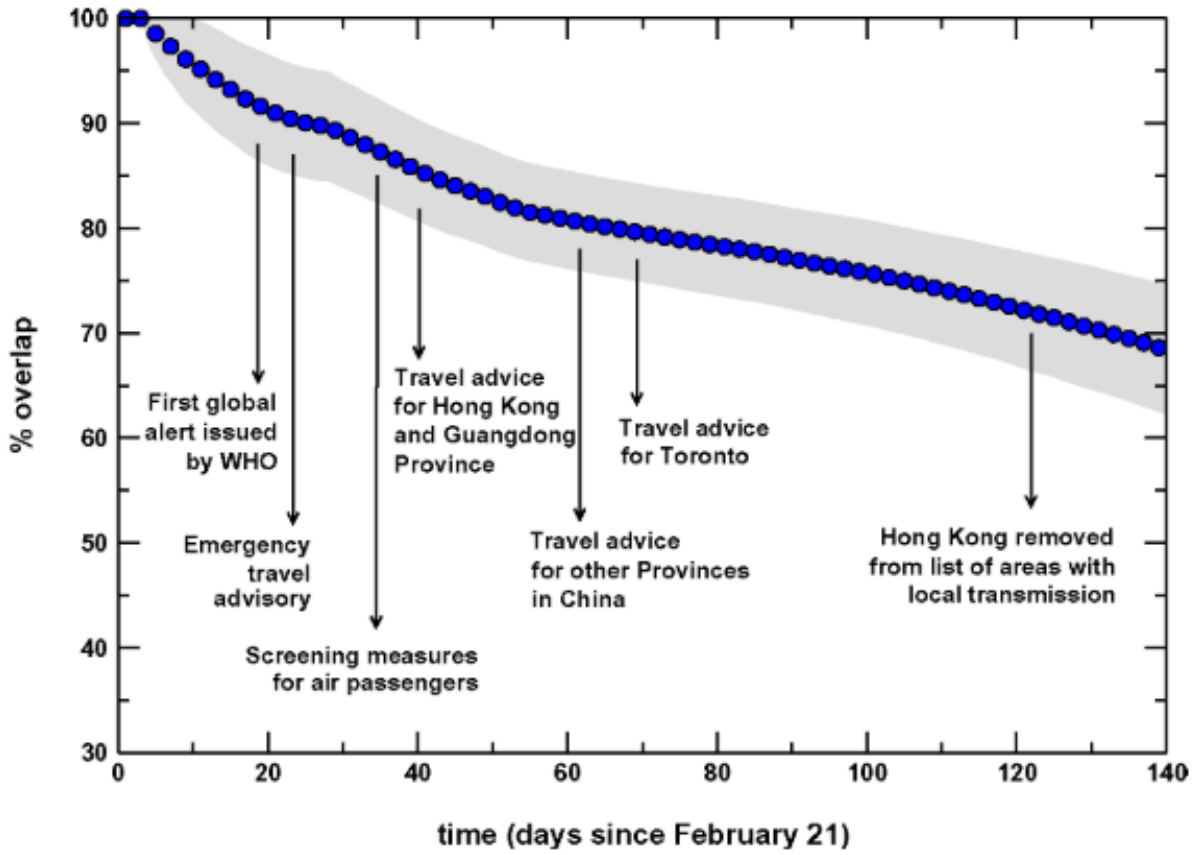

Figure: Overlap function for SARS pandemic [1].

Verifying the model against SARS Outbreak

- It was found that the overlap function takes values larger than 0.7 for a large portion of the time window investigated, which points to relatively strong computational reproducibility.
- This implies that the simulated disease followed a very similar evolution at each realization of the process.
- The origin of such reproducibility was attributed to the emergence of epidemic pathways, ie, preferential channels along which the epidemic will more likely spread [1].

Table of Contents

Conclusions and Applications

- The epidemic pathways that a model can resolve identifies for each country the possible origins of infection and provides a quantitative estimation of the probability of receiving the infection from each identified origin.
- Travel advisories or limitations and medical screenings at the ports of entry will strongly benefit from the analysis and identification of such epidemic pathways [1].

Conclusions and Applications

- It is also worth noting here that the models have the greatest predictive power when the epidemic is at its worst, which may also be the time when we need the most help.


Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani.
Epidemic predictions and predictability in complex environments.
In *Biophysical Reviews and Letters*, volume 3, pages 217–226, 2008.


Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani.
The modeling of global epidemics: Stochastic dynamics and predictability.
Bulletin of Mathematical Biology, 68(8):1893–1921, 2006.


Matthew Biggerstaff, Benjamin J. Cowling, Zulma M. Cucunubá, and et al.
Early insights from statistical and mathematical modeling of key epidemiologic parameters of covid-19.
Emerging Infectious Diseases, 26(11), November 2020.


Daniel T. Gillespie.
Chemical Langevin equation.
Journal of Chemical Physics, 113(1):297–306, 2000.