

# 1 Methods

## 1.1 Introducing the Model

In our model, all nodes have a so-called disease level. Possible levels include "susceptible", or capable of receiving disease from an infected neighbor with some probability; "infected", or capable of transmitting disease to neighbor nodes; and "recovered", or incapable of transmitting or receiving disease, due to immunity. In addition, a so-called immunity state determines the degree of susceptibility. Hence each susceptible node has a certain immunity which is represented with its immunity state.

At the beginning of every simulated flu season, all nodes are in a susceptible state. In a naïve network that has never experienced disease, corresponding to season  $s = 0$ , all susceptibles are in the disease state  $\gamma = 0$ , representing the fact that no immunity exists within the population. The probability of transmission ("transmissibility") between an infected node and a susceptible neighbor is given by

$$T_{naive} = \frac{R_0(s=0)}{T_{critical}} = R_0(s=0) \frac{\sum_{k=0}^{\infty} k(k-1)p_k}{\sum_{k=0}^{\infty} kp_k} \quad (1)$$

where  $R_0$  is the expected number of secondary infections caused by each infection, and  $T_{critical}$  is the threshold transmissibility, below which only small outbreaks can occur. The critical transmissibility only depends on the network structure and is the reciprocal of the average excess degree.

The state of susceptibility (or conversely, immunity) for a given node is dependent only on the number of seasons  $\gamma = \Delta s$  since the node was most recently infected. Therefore susceptibility for  $\gamma > 0$  is modeled as

$$T_v(\gamma) = T_{naive} e^{\frac{-D}{\gamma_v}} \quad (2)$$

where  $T_v$  is the incoming transmissibility for node  $v$ , giving the probability that an infected neighbor will transmit the infection to  $v$ . The immunity state of  $v$  is  $\gamma_v$ , representing the number of seasons since node  $v$  was most recently infected, and  $D$  is a non-negative immune decay parameter that determines how quickly a node loses its immunity. If a node has never experienced infection, then  $\gamma_v$  is 0 by definition and the incoming transmissibility is simply  $T_{naive}$ .

[ I don't know if this fits very well, maybe there is a better place for this part In practice, the distinction between "outbreak" and "epidemic" can be difficult for small networks ( $10^1$  -  $10^2$  individuals), but is generally straightforward for larger networks. This is because the expected absolute outbreak size is constant as network size varies, while for epidemics, it is instead the expected relative size that remains constant (assuming constant degree distribution).]

## 1.2 Analytical Methods

Analytical calculations allow us to estimate the possible epidemic sizes and probabilities over the course of a multi-seasonal infection. Assuming an infinite network, it is possible to determine the probability of being in a certain immunity state for each node and season using a Markov chain model. The probability of being in a certain state can be obtained using

$$\rho_{\gamma}^{(k)}(s+1) = \sum_{\gamma'=0}^s \omega_{\gamma',\gamma}^{(k)}(s) \rho_{\gamma'}^{(k)}(s) - \omega_{\gamma,\gamma'}^{(k)}(s) \rho_n^{(k)}(s) \quad (3)$$

This master-equation defines every immunity state probability  $\rho$  for each node with degree  $k$ , only using the the transfer probabilities  $\omega$  to change an immunity state, and the knowledge that at season 0 all nodes where in state 0

$$\rho_{\gamma}^{(k)}(s=0) = \begin{cases} 1 & \text{if } \gamma = 1 \\ 0 & \text{else} \end{cases} \quad (4)$$

$\gamma_{max}$  equals in our case the total number of observed seasons.

A node stays in the immunity state 0 as long as it has never been infected before. It changes into state 1, if it has been infected in the last season, and it moves up one state per season, if it has not been infected in the last season but it already had some immunity, hence its original state was greater than state 0. Using this, it is possible to derive the transfer probabilities. The probability of getting not infected gives

$$\begin{aligned} \omega_{0,0}^{(k)}(s+1) &= (1 + (u(s) - 1)T(0))^k \\ \omega_{\gamma,\gamma+1}^{(k)}(s+1) &= (1 + (u(s) - 1)T(\gamma))^k \end{aligned} \quad (5)$$

with  $\gamma > 0$  and  $u(s)$  as the probability that the disease could be transmitted but the neighbor on the other side of the edge is not infected. The other possible non-zero transfer probabilities are given by infections. It is for all possible  $\gamma$

$$\omega_{\gamma,1}^{(k)}(s+1) = 1 - (1 + (u(s) - 1)T(\gamma))^k \quad (6)$$

All other transfer probabilities are zero.

To solve the transfer probabilities, it is necessary to know the probability  $u(s)$ . A way to think about it is as the probability that the edge connecting a node and one of its neighbors may be able to transmit the disease, but that the neighbor is not infective. Hence it is like following the edge to the neighbor and determining the probability that this neighbor is not infected. This probability clearly depends on the immunity state of the neighbor. We need to differentiate between the "incoming transmissibility"  $T_{in}$ , which only relies a node's own immunity, and the "outgoing transmissibility"  $T_{out}$ , which relies on the immunity of a node's neighbors. Knowing this, it is

clear that the neighbor's probability of not causing an infection needs to be calculated with the neighbor's incoming transmissibility and the probability of having infected neighbors itself. This leads to a recursive relation for  $u(s)$

$$u(s) = \frac{\sum_{k=0}^{\infty} k p_k \left(1 + (u(s) - 1) T_{in}^{(k)}(s)\right)^{k-1}}{\sum_{k=0}^{\infty} k p_k} \quad (7)$$

with the incoming transmissibility

$$T_{in}^{(k)}(s) = \sum_{\gamma=0}^s T(\gamma) \rho_{\gamma}^{(k)}(s) \quad (8)$$

The outgoing transmissibility depends on the states of a node's neighbors, which's states also depends on their neighbors, and so on. Every nodes state depends on all other states, which is impossible to derive in an infinite network. An good approximation is to include only the first order of neighbors, because the strength of this dependency is scales with  $\langle k \rangle^{-1}$  for each neighbor's order and the error for setting  $\mathcal{O}(n) \sim \langle k \rangle^{-n} \approx 0$  for  $n = 2, 3, 4, \dots, \infty$  is sufficiently small and approaches 0 fast for  $n \rightarrow \infty$ . The probability to reach a degree  $k'$  neighbor following one of the node's  $k$  edges is

$$q_{k'} = \frac{k' p_{k'}}{\sum_{k=0}^{\infty} k p_k} \quad (9)$$

and the outgoing transmissibility of a node is the incoming transmissibility of its neighbor. This leads to

$$T_{out}^{(k)}(s) = \sum_{k'=0}^{\infty} q_{k'} T_{in}^{(k')}(s) = \sum_{k'=0}^{\infty} \sum_{\gamma=0}^s q_{k'} T(\gamma) \rho_{\gamma}^{(k')}(s) \quad (10)$$

The outgoing transmissibility needs to be used to calculate the probability of an epidemic

$$p_{epi}(s) = 1 - \sum_{k=0}^{\infty} k p_k \left(1 + (u(s) - 1) T_{out}^{(k)}(s)\right)^k \quad (11)$$

which is the one minus the probability to not infect a neighbor. But it is also the fraction of all not-infected degree  $k$  nodes within a network, if an epidemic occurs. Therefore  $p_{epi}(s)$  equals also the expected epidemic size.

The average number of secondary infections  $R_0(s)$  can be derived using the average transmissibility for each season. Using the state probabilities

$\rho(\gamma)(s)$  and the corresponding transmissibilities, it is

$$\begin{aligned}
R_0(s) &= T_{critical} T_{average} \\
&= \frac{\sum_{k=0}^{\infty} k p_k \sum_{k=0}^{\infty} p_k \sum_{\gamma=0}^s T(\gamma) \rho_{\gamma}^{(k)}(s)}{\sum_{k=0}^{\infty} k(k-1)p_k}
\end{aligned} \tag{12}$$

for the corresponding expected epidemic sizes and probabilities.

### 1.3 Computational Methods

#### 1.3.1 Network Generation

Networks with  $10^6$  nodes were constructed using the urban degree distribution (unless otherwise noted) and the "modified matching" algorithm described by Milo et al. (2004) to connect the nodes and remove self-loops and multi-edges. As noted by Milo et al. (2004), this algorithm is fast but can introduce a subtle bias: feed-forward loops are slightly more common than one would expect in a randomly connected graph. This bias is not a concern in the research presented here, however, because on average 186.0 ( $n = 1000$ ) edges are self-loops or multi-edges in a random urban network, regardless of network size. In urban networks with  $10^6$  nodes, fewer than 0.012% of edges are expected to be self-loops or multi-edges.

The urban degree distribution is bimodal with a mean of 16.75 and is based on census data obtained for Vancouver, Canada.

#### 1.3.2 Percolation Simulation

We use the following percolation simulation to model epidemics: [...]

#### 1.3.3 Software

Analytical results were calculated using Python; network generation algorithms and epidemic simulations were implemented using an object-oriented C++ applications programming interface. Figures were generated using R.