EARLY EPIDEMIC SPREAD, PERCOLATION AND COVID-19

GONÇALO OLIVEIRA

ABSTRACT. Human to human transmissible infectious diseases spread in a population using human interactions as its transmission vector. The early stages of such an outbreak can be modeled by a graph whose edges encode these interactions between individuals, the vertices. This article attempts to account for the case when each individual entails in different kinds of interactions which have therefore different probabilities of transmitting the disease. The majority of these results can be also stated in the language of percolation theory.

The main contributions of the article are: (1) Extend to this setting some results which were previously known in the case when each individual has only one kind of interactions. (2) Find an explicit formula for the basic reproduction number R_0 which depends only on the probabilities of transmitting the disease along the different edges and the first two moments of the degree distributions of the associated graphs. (3) Motivated by the recent Covid-19 pandemic, we use the framework developed to compute the R_0 of a model disease spreading in populations whose trees and degree distributions are adjusted to several different countries. In this setting, we shall also compute the probability that the outbreak will not lead to an epidemic. In all cases we find such probability to be very low if no interventions are put in place.

CONTENTS

1. INTRODUCTION

The vector by which several infectious diseases propagate in a population is the human to human interaction. It is therefore natural to model their spread using such interactions as

Date: April 2020.

GONÇALO OLIVEIRA

the "basic mechanism". In general, it gives rise to a dynamic process which evolves in time with its early stages being reasonably well approximated by considering the patient zero as the root (apex) of a tree whose branches encode the interactions through which the disease can propagate. In such epidemiological and percolation problems it is commonly assumed that each individual has an equal probability to transmit the disease to any of its contacts. However, this is an over-simplification of the actual situation as the same person can have several classes of interactions. Namely, it is conceivable that it is more likely that an infected individual will transmit the disease to someone with which it maintains a close familiar relation than to someone which it sporadically meets. In this setting we shall encode these different probabilities of transmitting the disease by distinct trees $\mathcal{T}_1, \ldots, \mathcal{T}_n$ all with the same root which is the patient zero. To each tree $\mathcal{T}_i$ we associated a probability T_i of transmitting the disease along an interaction modeled by the corresponding tree. In the case when the T_i are all the same for each vertex (individual) it is known that the basic reproduction number R_0 controls the possibility of almost surely avoiding an epidemic, see [Br, Ha, Ja, KA, Sc] and [Ca] which phrases these results in terms of the percolation interpretation. When there are different T_i a similar framework can be used to prove the following result.

Theorem 1. *Suppose the basic reproduction number $R_0 < 1$, then the outbreak will almost surely, i.e. with probability one, not lead to an epidemic. On the other hand, if $R_0 > 1$ there is still a nonzero probability $\mathbf{P} \in (0,1)$ that the outbreak will be contained.*

In section 2 we shall develop the framework of multivariate generating functions on which this work will be based. Section 3 will show how to use this framework to effectively compute R_0 and finally we will prove the main abstract results in section 4. We will then exemplify the theory with a few examples. These are instructive in order to unravel an explicit formula for R_0 which does not depend on any "abstract" generating function. Such a formula is deduced in section 6 where we show the following result.

Theorem 2. *For each $j \in \{1, \ldots, n\}$ let $\mathbf{E}(k_j)$ denote the average degree of the tree $\mathcal{T}_j$ and $\sigma(k_j)$ the standard deviation of the associated degree distribution. Suppose that a fraction Q of all infected individuals is completely isolated and does not transmit the disease to anyone. Then, if the probability of transmitting the disease along an arm of the tree $\mathcal{T}_j$ is $T_j \in [0,1]$, the basic reproduction number is*

$$R_0 = (1-Q) \sum_{j=1}^{n} T_j \mathbf{E}(k_j) \left(1 - \frac{1}{\sum_{l=1}^{n} \mathbf{E}(k_l)} \right) + (1-Q) \sum_{j=1}^{n} T_j \sigma(k_j) \frac{\sigma(k_j)}{\sum_{l=1}^{n} \mathbf{E}(k_l)},$$

or

$$R_0 = (1-Q) \sum_{j=1}^{n} T_j \left(\mathbf{E}(k_j) \left(1 - \frac{1}{\sum_{l=1}^{n} \mathbf{E}(k_l)} \right) + \frac{\sigma(k_j)^2}{\sum_{l=1}^{n} \mathbf{E}(k_l)} \right).$$

A somewhat interesting feature of the previous formula is that it together with the transmission probabilities $\{T_i\}_{i=1}^n$, it depends solely on the first two moments of the degree distribution of the trees encoding the interactions.

As a final application of the theory, and motivated by the recent outbreak of Covid-19, we reserve the last section to do some specific country analysis. We have attempted to make the parameters of the theory be somewhat adequate to model the initial spread of Covid-19 but the results should be regarded as an "academic" toy example. A more robust analysis using our framework is possible, but would require detailed knowledge on the habits, family ties/interactions, attendace of public gatherings and other features of the analyzed populations, which are not uniform in each country. That last section computes the relevant values of R_0 for the countries considered and the probabilities that the outbreak will be contained. As we shall see, these are very small and in order to increase it we will investigate the effect of quarantining part of the infected individuals.

2. Generating functions

2.1. **From the generating vertex.** Let $\mathcal{T}$ be a graph having a tree structure and whose edges are divided into n groups. Each of groups gives rise to subgraphs $\mathcal{T}_i \subset \mathcal{T}$ having the same vertices and the edges of the corresponding group. We will also assume that $\mathcal{T}_i \subset \mathcal{T}$ has a tree structure. For each of these trees $\mathcal{T}_i$, for $i = 1, \ldots, n$, we shall denote by $\{P_i(k)\}_{k \in \mathbb{N}_0}$ the corresponding degree distribution, i.e. for a randomly chosen vertex its degree is k with probability $P_i(k)$. Using these we can construct the corresponding generating functions

$$G_i(x) = \sum_{k=0}^{+\infty} P_i(k) x^k,$$

and the joint generating function

$$G(x_1, \ldots, x_n) = \sum_{k_1=0}^{+\infty} \cdots \sum_{k_n=0}^{+\infty} \prod_{i=1}^n P_i(k_i) x_i^{k_i} = \prod_{i=1}^n G_i(x_i).$$

Remark 1. *One can readily check that $G_i(0) = 0$, $G_i(1) = 1$ and $G_i(x)$ converges for $|x| \leq 1$. Furthermore, we compute that $G_i'(x) = \sum_{k=1}^{+\infty} k P_i(k) x^{k-1}$ and thus the average degree $\mathbf{E}(k_i)$ of the tree $\mathcal{T}_i$ is*

$$\mathbf{E}(k_i) = G_i'(1).$$

Similarly, we find that from $G(x, \ldots, x)$ the total average degree is

$$\mathbf{E}(k) = \sum_{i=1}^{n} G_i'(1) = \frac{d}{dx} G(x, \ldots, x)\Big|_{x=1}.$$

Each of these trees corresponds to different classes of contacts which have unequal probability of transmitting the disease. For instance, people that live in the same house are more likely to transmit the disease to each other than those which occasionally meet on public transport. Thus, to each tree, i.e., to each class of interactions, we associate a probability of transmission T_i and assume with no loss of generality that $T_1 > \ldots > T_n$. Fix $m_1, \ldots, m_n$, then the probability that a randomly picked first infected individual transmits the disease to m_i other individuals along the tree $\mathcal{T}_i$ is

$$\sum_{k_1=m_1}^{+\infty} \cdots \sum_{k_n=m_n}^{+\infty} \prod_{i=1}^{n} P_i(k_i) \binom{k_i}{m_i} T_i^{m_i} (1 - T_i)^{k_i - m_i}.$$

Further suppose there is a probability Q that an infected individual is detected and quarantined in complete isolation. Then, if $m_1 + \ldots + m_n \neq 0$, the probability above must be multiplied by a factor of $1 - Q$ which accounts for the possibility that it is not quarantined. At this point, it is convenient to define the generating function

$$
\begin{aligned}
Z(x) &= Q + (1-Q) \sum_{k_1=0}^{+\infty} \cdots \sum_{k_n=0}^{+\infty} \prod_{i=1}^{n} P_i(k_i) \sum_{m_1=0}^{k_1} \cdots \sum_{m_n=0}^{k_n} \binom{k_i}{m_i} (xT_i)^{m_i} (1 - T_i)^{k_i - m_i} \\
&= Q + (1-Q) \sum_{k_1=0}^{+\infty} \cdots \sum_{k_n=0}^{+\infty} \left(\prod_{i=1}^{n} P_i(k_i) \right) \left(\prod_{i=1}^{n} (xT_i + 1 - T_i)^{k_i} \right) \\
&= Q + (1-Q) \sum_{k_1=0}^{+\infty} \cdots \sum_{k_n=0}^{+\infty} \prod_{i=1}^{n} P_i(k_i)(xT_i + 1 - T_i)^{k_i} \\
&= Q + (1-Q) G(1 - T_1(1 - x), \ldots, 1 - T_n(1 - x)).
\end{aligned}
$$

(2.1)

2.2. **By following a random infection.** Suppose we place ourselves at a vertex which is obtained from following a randomly chosen transmission. As an element of the tree $\mathcal{T}_i$, the ramification of this vertex is the number of remaining edges emanating from it. The probability that such a vertex has ramification $k_1, \ldots, k_n$ is therefore proportional to

$$(k_1+1)P_1(k_1+1) \prod_{i\neq 1} P_i(k_i) + (k_2+1)P_2(k_2+1) \prod_{i\neq 2} P_i(k_i) + \ldots + (k_n+1)P_n(k_n+1) \prod_{i\neq n} P_i(k_i),$$

with each term accounting from the probability of arriving at the chosen vertex via a given tree. Normalizing this we find that such probability $\tilde{P}(k_1, \ldots, k_n)$ is obtained from the previous formula by dividing by

$$\mathbf{E}(k) = \mathbf{E}(k_1) + \ldots + \mathbf{E}(k_n) = G_1'(1) + \ldots + G_n'(1),$$

i.e. the average total ramification. We have thus concluded that

$$\tilde{P}(k_1, \ldots, k_n) = \frac{\sum_{l=1}^{n}(k_l + 1)P_l(k_l + 1)\prod_{i \neq l} P_i(k_i)}{\mathbf{E}(k)}.$$

Associated with this we define the generating function

$$(2.2) \qquad \tilde{G}(x_1, \ldots, x_n) := \sum_{k_1, \ldots, k_n} \tilde{P}(k_1, \ldots, k_n)x_1^{k_1} \ldots x_n^{k_n}.$$

For future reference, it is convenient to have this written in terms of the simpler generating functions $G_i(x_i)$ for $i = 1, \ldots, n$. For this, we insert the formula for $\tilde{P}(k_1, \ldots, k_n)$ previously obtained. This yields

$$\tilde{G}(x_1, \ldots, x_n) = \sum_{k_1, \ldots, k_n} \tilde{P}(k_1, \ldots, k_n)x_1^{k_1} \ldots x_n^{k_n}$$

$$= \frac{1}{\mathbf{E}(k)} \sum_{k_1, \ldots, k_n} \left(\sum_{l=1}^{n}(k_l + 1)P_l(k_l + 1)\prod_{i \neq l} P_i(k_i) \right) x_1^{k_1} \ldots x_n^{k_n}$$

$$= \frac{1}{\mathbf{E}(k)} \sum_{k_1, \ldots, k_n} \left(\sum_{l=1}^{n}(k_l + 1)P_l(k_l + 1)x_l^{k_l}\prod_{i \neq l} P_i(k_i)x_i^{k_i} \right)$$

$$= \frac{1}{\mathbf{E}(k)} \left(\sum_{l=1}^{n} G_l'(x_l)\prod_{i \neq l} G_i(x_i) \right),$$

which may be written in the following more explicit form

$$(2.3) \qquad \tilde{G}(x_1, \ldots, x_n) = \frac{\sum_{l=1}^{n} G_l'(x_l)\prod_{i \neq l} G_i(x_i)}{\sum_{l=1}^{n} G_l'(1)}.$$

Remark 2. *Consider of a sole tree $\mathcal{T}_i$, the corresponding ramification distribution is*

$$\tilde{P}_i(k) = \frac{(k+1)P_i(k+1)}{\mathbf{E}(k_i)} = \frac{(k+1)P_i(k+1)}{G_i'(1)}.$$

which the individual generating function

$$\tilde{G}_i(x) = \sum_{k=0}^{+\infty} \tilde{P}_i(k)x^k = \frac{\sum_{k=0}^{+\infty}(k+1)P_i(k+1)x^k}{G_i'(1)} = \frac{G_i'(x)}{G_i'(1)}.$$

Then, we have $G_i'(x) = G_i'(1)\tilde{G}_i(x)$ which upon inserting in equation 2.3 yields

$$\tilde{G}(x_1, \ldots, x_n) = \frac{\sum_{l=1}^{n} \tilde{G}_l(x_l)G_l'(1)\prod_{i \neq l} G_i(x_i)}{\sum_{l=1}^{n} G_l'(1)}.$$

Let $m_1, \ldots, m_n \in \mathbb{N}_0$ with $m_1 + \ldots + m_n > 0$. Using the distribution for the ramification, we conclude that by following the contacts of the trees $\mathcal{T}_1$ up to $\mathcal{T}_n$ a randomly infected

individual infects m_1 up to m_n other ones is

$$(1-Q) \sum_{k_1,\ldots,k_n} \tilde{P}(k_1,\ldots,k_n) \prod_{i=1}^{n} \binom{k_i}{m_i} T_i^{m_i} (1-T_i)^{k_i-m_i},$$

if any of the m_i is nonzero. Based on this, we define the generating function

$$
\begin{aligned}
\tilde{Z}(x) &= Q + (1-Q) \sum_{k_1,\ldots,k_n} \tilde{P}(k_1,\ldots,k_n) \prod_{i=1}^{n} \sum_{m_i=0}^{k_i} \binom{k_i}{m_i} (xT_i)^{m_i} (1-T_i)^{k_i-m_i} \\
&= Q + (1-Q) \sum_{k_1,\ldots,k_n} \tilde{P}(k_1,\ldots,k_n) \prod_{i=1}^{n} (xT_i + 1 - T_i)^{k_i} \\
&= Q + (1-Q)\tilde{G}(1 + T_1(x-1), \ldots, 1 + T_n(x-1)).
\end{aligned}
$$

(2.4)

3. The basic reproduction number

The basic reproduction number, usually denoted by R_0, is defined as the average number of individuals which are infected by each previously infected one. In our setting this can be immediately computed as follows. First, suppose we stand at a randomly chosen individual which has ramification $(k_1,\ldots,k_n)$. The average number of individuals it infects is

$$
\begin{aligned}
a_{k_1\ldots k_n} &= (1-Q) \sum_{m_1=0}^{k_1} \ldots \sum_{m_n=0}^{k_n} (m_1 + \ldots + m_n) \prod_{i=1}^{n} \binom{k_i}{m_i} (T_i)^{m_i} (1-T_i)^{k_i-m_i} \\
&= (1-Q) \sum_{l=1}^{n} \sum_{m_1=0}^{k_1} \ldots \sum_{m_n=0}^{k_n} m_l \prod_{i=1}^{n} \binom{k_i}{m_i} (T_i)^{m_i} (1-T_i)^{k_i-m_i} \\
&= (1-Q) \sum_{i=1}^{n} \sum_{m_i=1}^{k_i} m_i \binom{k_i}{m_i} (T_i)^{m_i} (1-T_i)^{k_i-m_i},
\end{aligned}
$$

where we have used the binomial formula to deduce $\sum_{m_i=0}^{n} \binom{k_i}{m_i}(T_i)^{m_i}(1-T_i)^{k_i-m_i} = 1$. For the average vertex we must weight this with the ramification distribution, i.e.

$$R_0 := \sum_{k_1\ldots k_n} \tilde{P}(k_1,\ldots,k_n) a_{k_1\ldots k_n}$$

Having in mind the formula for $\binom{k_i}{m_i}$ we find

$$m_i \binom{k_i}{m_i} = \frac{k_i!}{(k_i-m_i)!(m_i-1)!} = k_i \frac{(k_i-1)!}{((k_i-1)-(m_i-1))!(m_i-1)!} = k_i \binom{k_i-1}{m_i-1}.$$

Inserting into the above equation for R_0 and using again the binomial formula gives

$$R_0 = (1-Q) \sum_{k_1...k_n} \tilde{P}(k_1,\ldots,k_n) \sum_{i=1}^{n} \sum_{m_i=1}^{k_i} k_i \binom{k_i-1}{m_i-1} (T_i)^{m_i}(1-T_i)^{k_i-m_i}$$

$$= (1-Q) \sum_{k_1...k_n} \tilde{P}(k_1,\ldots,k_n) \sum_{i=1}^{n} k_i T_i \sum_{m_i=1}^{k_i} \binom{k_i-1}{m_i-1} (T_i)^{m_i-1}(1-T_i)^{(k_i-1)-(m_i-1)}$$

$$= (1-Q) \sum_{k_1...k_n} \tilde{P}(k_1,\ldots,k_n) \sum_{i=1}^{n} k_i T_i.$$

Comparing this with the formula for $\tilde{Z}(x)$ we conclude the following result.

Proposition 1. *The basic reproduction number can be obtained from the generating function $\tilde{Z}(x)$ for the distribution of individuals infected by following a randomly chosen infected individual, via*

$$R_0 = \tilde{Z}'(1),$$

which may also be written as

$$R_0 = (1-Q) \sum_{j=1}^{n} T_j \frac{\partial \tilde{G}}{\partial x_j}\Big|_{x_1=...=x_n=1}.$$

4. Containing an outbreak

In this section we shall compute the probability that the infection as it propagates eventually dies out. In the case of a unique tree, i.e. assuming all interactions have the same probability of transmitting the disease, such a computation have been carried out in [Br]. See also [Ha, Ja, KA, Sc] for related results and [Ca] for the same setup in the context of the theory of percolation.

Consider an individual which has been infected by another one, i.e. a vertex of the tree $\mathcal{T}$ which is not its root and consider the probability that the infections generated by that vertex disappear within N generations. Denote such a probability by p_N, then

$$\mathbf{P}_N = Q + (1-Q) \sum_{m_1,\ldots,m_n} \mathbf{P}_{N-1}^{m_1+\ldots+m_n} \left(\sum_{k_1 \geq m_1,\ldots,k_n \geq m_n} \tilde{P}(k_1,\ldots,k_n) \prod_{i=1}^{n} \binom{k_i}{m_i} T_i^{m_i}(1-T_i)^{k_i-m_i} \right)$$

$$= Q + (1-Q) \sum_{m_1,\ldots,m_n} \sum_{k_1 \geq m_1} \cdots \sum_{k_n \geq m_n} \tilde{P}(k_1,\ldots,k_n) \prod_{i=1}^{n} \binom{k_i}{m_i} \mathbf{P}_{N-1}^{m_i} T_i^{m_i}(1-T_i)^{k_i-m_i}$$

$$= Q + (1-Q) \sum_{k_1,\ldots,k_n} \sum_{m_1=0}^{k_1} \cdots \sum_{m_1=0}^{k_1} \tilde{P}(k_1,\ldots,k_n) \prod_{i=1}^{n} \binom{k_i}{m_i} (\mathbf{P}_{N-1}T_i)^{m_i}(1-T_i)^{k_i-m_i},$$

which upon comparing with the definition of $\tilde{Z}(x)$ in equation 2.4 can equally be read as

$$(4.1) \qquad \mathbf{P}_N = \tilde{Z}(\mathbf{P}_{N-1}).$$

By construction we must have

$$\mathbf{P}_0 = Q + (1 - Q) \sum_{k_1 \geq 0, \ldots, k_n \geq 0} \tilde{P}(k_1, \ldots, k_n) \prod_{i=1}^{n}(1 - T_i)^{k_i},$$

with equation 4.1 yielding all the following iterations. From inspection we find that $\mathbf{P}_1 := \tilde{Z}(\mathbf{P}_0) > \mathbf{P}_0$ and as $\tilde{Z}$ is increasing we find assuming $\mathbf{P}_{N-1} > \mathbf{P}_{N-2}$ that

$$\mathbf{P}_N - \mathbf{P}_{N-1} = \tilde{Z}(\mathbf{P}_{N-1}) - \tilde{G}(\mathbf{P}_{N-2}) > 0,$$

which inductively proves that the sequence $\{\mathbf{P}_N\}_{N \in \mathbb{N}_0}$ is increasing. Hence, the number

$$\mathbf{P}_\infty = \lim_{N \to +\infty} \mathbf{P}_N \in [0, 1],$$

is well defined and encodes the probability that the infection starting from any such individual eventually dies out. We can then conclude the following.

Proposition 2. *The probability $\mathbf{P}_\infty$ that the chain of infections generated from a randomly infected individual eventually disappears in a finite number of generations satisfies*

$$\mathbf{P}_\infty = \tilde{Z}(\mathbf{P}_\infty).$$

Furthermore, there ate most two fixed points of $\tilde{Z}$ in $[0, 1]$.

Proof. The fact that $\mathbf{P}_\infty$ satisfies $\mathbf{P}_\infty = \tilde{Z}(\mathbf{P}_\infty)$ follows immediately from the preceding discussion. Hence, the only remaining item to be shown is that there are at most two fixed points of $\tilde{Z}$ in the interval $[0, 1]$. The fact that there is at least one is obvious as $\tilde{Z}(1) = 1$. We must now show that there is at most one other.
We argue by contradiction and assume there are at least two other different fixed points $\mathbf{P}_* < \mathbf{P}_{**}$ both in $(0, 1)$.[1] First, notice that both $\tilde{Z}'$, $\tilde{Z}''$ are positive in $(0, 1)$. Secondly, consider the function $f(x) = x - \tilde{Z}(x)$ whose zeros correspond to the fixed points of $\tilde{Z}$, we have

$$f(0) > 0, \quad f(P_*) = 0, \quad f(P_{*}*) = 0, \quad f(1) = 0.$$

Hence, by the intermediate value theorem there must be two critical points $c_* \in (\mathbf{P}_*, \mathbf{P}_{**})$ and $c_{**} \in (\mathbf{P}_{**}, 1)$ of f. As in $(0, 1)$ we have $f''(x) = -\tilde{Z}''(x) < 0$ we have that each of these must a maximum. Again, by the intermediate value theorem, between the two maxima must be a minimum $c_{***} \in (c_*, c_{**})$ contradicting $f''(c_{***}) < 0$. $\square$

---

[1] We assume with no loss of generality that $\mathbf{P}_*, \mathbf{P}_{**}$ are positive as $\tilde{Z}(0) = \mathbf{P}_0 \neq 0$ and so 0 can never be a fixed point.

Placing ourselves at the tip of the tree which originated the infection chain, the so called patient zero, the probability that the infection eventually dies out is

$$\mathbf{P} = Q + (1-Q) \sum_{k_1=0}^{+\infty} \cdots \sum_{k_n=0}^{+\infty} \sum_{m_1=0}^{k_1} \cdots \sum_{m_n=0}^{k_n} \left(\prod_{i=1}^{n} P_i(k_i) \binom{k_i}{m_i} T_i^{m_i} (1-T_i)^{k_i-m_i} \right) p_\infty^{m_1+\ldots+m_n}$$

$$= Q + (1-Q) \sum_{k_1=0}^{+\infty} \cdots \sum_{k_n=0}^{+\infty} \sum_{m_1=0}^{k_1} \cdots \sum_{m_n=0}^{k_n} \prod_{i=1}^{n} P_i(k_i) \binom{k_i}{m_i} (p_\infty T_i)^{m_i} (1-T_i)^{k_i-m_i}$$

$$= Z(\mathbf{P}_\infty),$$

where the last equality follows from comparison with the formula 2.1 for the generating function for infections starting from the patient zero.

4.1. **The case when $R_0 < 1$.** We shall now prove that when the basic reproduction number is smaller than one, the chain of infections will almost surely extinguish.

Theorem 3. *If $R_0 < 1$, then both $\mathbf{P}$ and $\mathbf{P}_\infty$ equal 1.*

Proof. By proposition 2 we know that $\mathbf{P}_\infty$ it must be a fixed point of $\tilde{Z}$. Furthermore, we know that $\tilde{Z}(1) = 1$ and so the statement follows if we can show that there is no other fixed point of $\tilde{Z}$ in the interval $[0, 1]$. In that direction we shall show that under the hypothesis that $R_0 < 1$ the map

$$\tilde{Z} : [0, 1] \to [0, 1]$$

is a contraction and so has a unique fixed point which must therefore the 1. This follows immediately from realizing that $\tilde{Z}''$ is nonnegative and so $\tilde{Z}'(x) \leq \tilde{Z}'(1) = R_0$ by 1. Hence, for $x, y \in [0, 1]$ we have

$$|\tilde{Z}(x) - \tilde{Z}(y)| \leq \left(\sup_{t \in [0,1]} \tilde{Z}'(t) \right) |x - y| \leq R_0 |x - y|,$$

which shows that $\tilde{Z}$ is a contraction if $R_0 < 1$.

Finally, the fact that also $\mathbf{P} = 1$ is then a consequence of $\mathbf{P} = Z(\mathbf{P}_\infty) = Z(1) = 1$. □

Remark 3 (The minimum required quarantined). *At the beginning of an outbreak the question arises of what is the minimum number of infected individuals that must be detected and subsequently quarantined in order to contain the possible epidemic outbreak.*

If the disease is already well known, such as flu, measles or any other standard disease, not Covid-19, then its "free" basic reproduction number R_0^{free} is known. Of course, this may depend on local conditions of where the outbreak takes place. By "free" we intend to emphasize that this is the basic reproduction number when the disease is free to propagate without taking in account any non-pharmaceutical intervention directed to slow its spread.

Now, suppose an aggressive testing capacity can be put in place in order to detect those which

have been infected. We would like to know the minimal fraction Q of infected individuals which must be completely isolated so that the outbreak is almost surely controlled without having to take any other measures. The answer, as we shall now see is that $Q > 1 - \frac{1}{R_0^{free}}$. The generating functions $\tilde{Z}^{free}$ and $\tilde{Z}$ can both be written as in equation 2.4 with the exception that $Q = 0$ for $\tilde{Z}^{free}$. Hence, $\tilde{Z} = Q + (1 - Q)\tilde{Z}^{free}$ and by Proposition 1 we find

$$R_0 = \tilde{Z}'(1) = (1 - Q)(\tilde{Z}^{free})'(1) = (1 - Q)R_0^{free}.$$

Then, the condition $R_0 < 1$ required to apply Theorem 3 turns into

$$Q > 1 - \frac{1}{R_0^{free}},$$

which corroborates the common intuition behind the basic reproduction number. For example, suppose there is an outbreak of disease for which each infected individual is expected to infect 3 others if nothing is done to prevent it, i.e. R_0^{free}, then we expect that in order to cut the chain of transmission less than a third of the infections can be allowed to transmit the disease. Indeed, from the above computation, at least $Q > 2/3$ of the whole infected must be detected and isolated.

4.2. **The case when $R_0 > 1$.** Finally, in the case when $R_0 > 1$ we shall now prove that there is still a positive, but not certain, probability that the infection disappears.

Theorem 4. *If $R_0 > 1$, then $\mathbf{P}_\infty \in (0, 1)$ and $\mathbf{P} = Z(\mathbf{P}_\infty) \in (0, 1)$.*

Proof. In this setup we consider the function $f(x) = x - \tilde{Z}(x)$ used in the proof of Proposition 2. This satisfies $f(0) = -\tilde{Z}(0) < 0$, $f(1) = 0$ and by Proposition 1 $f'(1) = 1 - R_0 < 0$ and so $f(1 - \varepsilon) > 0$ for sufficiently small nonzero $\varepsilon \ll 1$. Thus, again the intermediate value theorem shows the existence of a zero of f which we shall denote by $x_* \in (0, 1 - \varepsilon)$. Recalling that zeros of f correspond to fixed points of $\tilde{Z}$ which by Proposition 2 has only x_* and 1 as fixed points. Thus, in this case we can also have $\mathbf{P}_\infty = x_*$. $\square$

Remark 4 (Lower bounds for $\mathbf{P}$ and $\mathbf{P}_\infty$). *From the fixed point equation $\mathbf{P}_\infty = \tilde{Z}(\mathbf{P}_\infty)$ and writing the generating function $\tilde{Z}$ as in Remark 3, i.e. $\tilde{Z}(x) = Q + (1 - Q)\tilde{Z}^{free}(x)$, we find that*

$$\mathbf{P}_\infty = Q + (1 - Q)\tilde{Z}^{free}(\mathbf{P}_\infty),$$

and so $\mathbf{P}_\infty - Q = (1 - Q)\tilde{Z}^{free}(\mathbf{P}_\infty) > 0$ and so

$$\mathbf{P}_\infty > Q.$$

Furthermore, we can equally write $\mathbf{P} = Z(\mathbf{P}_\infty) = Q + (1 - Q)\tilde{Z}(\mathbf{P}_\infty)$. Said in other words, we find that the probability of the infection eventually dying out is at least the fraction of infected individuals which are completely isolated.

4.3. The case when $R_0 = 1$. We shall now consider the case when $R_0 = 1$. We go back to the setup in the proof of Proposition 2 and Theorem 3, namely we consider the function $f(x) = x - \tilde{Z}(x)$ whose zeros correspond to the fixed points of the map $\tilde{Z} : [0,1] \to [0,1]$. We have seen that $f''(x) < 0$ in $(0,1)$ and $f(1) = 0$. Under the hypothesis that $R_0 = 1$ we have $f'(1) = 1 - R_0 = 0$ and as $f''(x) < 0$ for $x < 1$ we find that f is negative immediately before $x = 1$. Hence, if there was another zero $x_* \in (0,1)$ of f, between $(x_*, 1)$ the function f would have a minimum which contradicts $f'' < 0$ in $(0,1)$. We then conclude that also in this case

$$\mathbf{P} = \mathbf{P}_\infty = 1.$$

4.4. Lower bounds on P. In this section we shall elaborate on the question raised in Remark 4, namely: Whether it is possible to find lower bounds on the probability that the chain of infections eventually dies out not leading to an epidemic.

To answer the question raised we proceed by direct inspection of the fixed point equation in Proposition 2. Start by noticing that all the terms in the Taylor series for $\tilde{Z}$ are positive as one can check from its definition in equation 2.4. Thus, as $\mathbf{P}_\infty = \tilde{Z}(\mathbf{P}_\infty)$ we find that $\mathbf{P}_\infty$ is larger than the zeroth order term of $\tilde{Z}$, i.e.

$$\mathbf{P}_\infty > \tilde{Z}(0),$$

and from the monotonicity of $Z(x)$ we then have

$$\mathbf{P} > Z(\tilde{Z}(0)),$$

which is itself grater than $Z(0)$.

5. Examples

In the simplest nontrivial example we can consider a population in which individuals have to kinds of interactions: a close and continuous interaction with their family and friends, and a a more distant sporadic interaction with not so close friends and other people which cross their path, by chance, in their daily lives as they commute to work and so on. Of course, we expect the probability of transmitting the disease to be larger in the first case and so assign to a it a larger transmissibility T_1 than to the second interactions T_2, i.e. $T_1 > T_2$.

5.1. Delta and Poisson. In this first example we assume for simplicity that all person have the same number, $N_1 \geq 1$, of close contacts and their sporadic contacts follow a Poisson distribution with intensity N_2. In formulas, we have

$$P_1(k_1) = \delta_{N_1 k_1}, \text{ and } P_2(k_2) = \frac{N_2^{k_2}}{k_2!} e^{-N_2}.$$

Then, we find the generating function for the degree distribution

$$G_1(x_1) = x_1^{N_1}, \text{ and } G_2(x_2) = e^{-N_2(1-x_2)},$$

from which we compute $G(x_1, x_2) = x_1^{N_1} e^{-N_2(1-x_2)}$ and so

$$Z(x) = Q + (1-Q)G(1 - T_1(1-x), 1 - T_2(1-x))$$
$$= Q + (1-Q)(1 - T_1(1-x))^{N_1} e^{-N_2 T_2(1-x)}.$$

In order to compute $\tilde{Z}(x)$ we may first find the generating function $\tilde{G}(x_1, x_2)$ for the ramification distribution. This can be done using equation 2.3 which yields

$$\tilde{G}(x_1, x_2) = \frac{G_1(x_1)G_2'(x_2) + G_1'(x_1)G_2(x_2)}{G_1'(1) + G_2'(1)}$$
$$= \frac{x_1^N N_2 e^{-N_2(1-x_2)} + N_1 x_1^{N_1-1} e^{-N_2(1-x_2)}}{N_1 + N_2}$$
$$= \frac{N_1 + N_2 x_1}{N_1 + N_2} x_1^{N_1-1} e^{-N_2(1-x_2)}.$$

We can finally use this to compute the generating function $\tilde{Z}$ using the formula 2.4. This gives

$$\tilde{Z}(x) = Q + (1-Q)\tilde{G}(1 - T_1(1-x), 1 - T_2(1-x))$$
$$= Q + (1-Q)\frac{N_1 + N_2(1 - T_1(1-x))}{N_1 + N_2}(1 - T_1(1-x))^{N_1-1} e^{-N_2(1-1+T_2(1-x))}$$
$$= Q + (1-Q)\frac{N_1 + N_2 - N_2 T_1(1-x)}{N_1 + N_2}(1 - T_1(1-x))^{N-1} e^{-N_2 T_2(1-x)}$$
$$= Q + (1-Q)\left(1 - \frac{N_2}{N_2 + N_1} T_1(1-x)\right)(1 - T_1(1-x))^{N_1-1} e^{-N_2 T_2(1-x)},$$

and using it we can compute the basic reproduction number R_0, which by Proposition 1 is

$$R_0 = (1-Q)\frac{N_2 N_1}{N_2 + N_1}\left(\frac{N_1 - 1}{N_2}T_1 + \frac{N_2}{N_1}T_2 + (T_1 + T_2)\right),$$

or perhaps in a somewhat more suggestive manner

$$R_0 = (1-Q)\left(N_2 T_2 + \left(N_1 - \frac{N_1}{N_2 + N_1}\right)T_1\right).$$

Notice in particular that R_0 scales homogeneously with degree 1 as a function of N_1, N_2, T_1, T_2.

Remark 5. *Notice that in the case $N_1 = 0$, i.e. if only the sporadic contacts occur, we have $R_0 = (1-Q)N_2 T_2$, while if $N_2 = 0$ meaning that all sporadic contacts are cut out, then $R_0 = (1-Q)(N_1 - 1)T_1$.*

Also, by evaluating the generating function at $x = 0$ we find

$$\tilde{Z}(0) = Q + (1-Q)\left(1 - \frac{N_2}{N_2 + N_1}T_1\right)(1 - T_1)^{N_1-1} e^{-N_2 T_2}.$$

It then follows from the discussion in section 4.4 that

$$\mathbf{P}_\infty > Q + (1 - Q)\left(1 - \frac{N_2}{N_2 + N_1}T_1\right)(1 - T_1)^{N_1 - 1}e^{-N_2 T_2},$$

and

$$\mathbf{P} > Z(\tilde{Z}(0)) > Z(0) = Q + (1 - Q)(1 - T_1)^{N_1}e^{-N_2 T_2}.$$

5.2. **Polynomial and Poisson I.** In this second model we will elaborate slightly on the first example, in the sense that we still assume everyone to establishes sporadic contacts following a Poisson distribution with intensity N_2. On the other hand, we shall encode the distribution of close contacts by a polynomial of degree $N \geq 1$. Furthermore, let N_1 be the average degree of the distribution of close contacts, i.e. $N_1 := G_1'(1)$. In formulas, we have

$$P_1(k_1) = \sum_{i=1}^{N}\delta_{ik_1}p_{k_1}, \text{ and } P_2(k_2) = \frac{N_2^{k_2}}{k_2!}e^{-N_2}.$$

As in the previous case, we can now find the generating function for these degree distributions

$$G_1(x_1) = p_0 + p_1 x + \ldots + p_{N_1}x_1^{N_1}, \text{ and } G_2(x_2) = e^{-N_2(1-x_2)},$$

and compute $G(x_1, x_2) = G_1(x_1)e^{-N_2(1-x_2)}$. Then,

$$Z(x) = Q + (1 - Q)G_1(1 - T_1(1 - x))e^{-N_2 T_2(1-x)},$$

and to compute $\tilde{Z}(x)$ we start by obtaining the generating function $\tilde{G}(x_1, x_2)$ for the ramification distribution.Equation 2.3 yields

$$\tilde{G}(x_1, x_2) = \frac{(\log G_1(x_1))' + N_2}{N_1 + N_2}G_1(x_1)e^{-N_2(1-x_2)}.$$

Finally using equation 2.4 as before, we find

$$\tilde{Z}(x) = Q + (1 - Q)\frac{G_1'(1 - T_1(1 - x)) + N_2 G_1(1 - T_1(1 - x))}{N_1 + N_2}e^{-N_2 T_2(1-x)},$$

from which we can compute R_0 using Proposition 1, i.e. using the formula $\tilde{Z}'(1)$. This requires computing the derivative of $\tilde{Z}$ which reads

$$\tilde{Z}'(x) = (1 - Q)T_1\frac{G_1''(1 - T_1(1 - x)) + N_2 G_1'(1 - T_1(1 - x))}{N_1 + N_2}e^{-N_2 T_2(1-x)}$$
$$+ (1 - Q)N_2 T_2\frac{G_1'(1 - T_1(1 - x)) + N_2 G_1(1 - T_1(1 - x))}{N_1 + N_2}e^{-N_2 T_2(1-x)},$$

and so

$$R_0 = (1 - Q)T_1\frac{G_1''(1) + N_2 N_1}{N_1 + N_2} + (1 - Q)N_2 T_2.$$

Remark 6. *In order to evaluate R_0 more clearly, we must understand what is $G_1''(1)$. This can be computed for any distribution $\{P(k)\}_{k \in \mathbb{N}_0}$ with generating function $G(x) = \sum_{k=0}^{+\infty}P(k)x^k$.*

Indeed, from direct differentiation

$$G''(1) = \sum_{k=2}^{+\infty} k(k-1)P(k) = \sum_{k=0}^{+\infty} k^2 P(k) - \sum_{k=0}^{+\infty} kP(k) = \mathbf{E}(k^2) - \mathbf{E}(k),$$

as the $k = 0$ terms vanish an the $k = 1$ terms cancel.

Inserting this into the previous formula for R_0 yields

$$R_0 = (1 - Q)\left(\frac{\mathbf{E}(k_1^2) + N_1(N_2 - 1)}{N_2 + N_1}T_1 + N_2 T_2\right).$$

Notice in particular that R_0 scales homogeneously with degree 1 as a function of N_1, N_2, T_1, T_2.

We turn now to compute lower bounds on $\mathbf{P}$ and $\mathbf{P}_\infty$ following the strategy of section 4.4, from which we infer that

$$\mathbf{P}_\infty > \tilde{Z}(0) = Q + (1 - Q)\frac{G_1'(1 - T_1) + N_2 G_1(1 - T_1)}{N_1 + N_2}e^{-N_2 T_2},$$

while

$$\mathbf{P} > Z(\tilde{Z}(0)) > Z(0) = Q + (1 - Q)G_1(1 - T_1)e^{-N_2 T_2}.$$

5.3. **Polynomial and Poisson II.** We shall now continue with a population organized as in the previous example but we assume that a fraction f of the population decides to social isolate and cut its sporadic contacts. One can imagine this can be done by not going into public transport, reducing contact with unknown people and working from home. To implement this we modify the previous example by writing

$$P_1(k_1) = \sum_{i=1}^{N} \delta_{ik_1}p_{k_1}, \text{ and } P_2(k_2) = f\delta_{0k_2} + (1 - f)\frac{N_2^{k_2}}{k_2!}e^{-N_2}.$$

As in the previous case, we can now find the generating function for these degree distributions

$$G_1(x_1) = p_0 + p_1 x + \ldots + p_{N_1}x_1^{N_1}, \text{ and } G_2(x_2) = f + (1 - f)e^{-N_2(1-x_2)},$$

and compute $G(x_1, x_2) = G_1(x_1)(f + (1 - f)e^{-N_2(1-x_2)})$. Then,

$$Z(x) = Q + (1 - Q)G_1(1 - T_1(1 - x))(f + (1 - f)e^{-N_2 T_2(1-x)}),$$

and

$$\tilde{G}(x_1, x_2) = f\frac{G_1'(x_1)}{N_1 + N_2} + (1 - f)\frac{G_1'(x_1) + N_2 G_1(x_1)}{N_1 + N_2}e^{-N_2(1-x_2)},$$

from which we can compute

$$\tilde{Z}(x) = Q + (1 - Q)f\frac{G_1'(1 - T_1(1 - x))}{N_1 + N_2}$$
$$+ (1 - Q)(1 - f)\frac{G_1'(1 - T_1(1 - x)) + N_2 G_1(1 - T_1(1 - x))}{N_1 + N_2}e^{-N_2 T_2(1-x)},$$

from which we can compute R_0 using Proposition 1, i.e. using the formula $\tilde{Z}'(1)$. This requires computing the derivative of $\tilde{Z}$ which reads

$$\tilde{Z}'(x) = Q + (1-Q)fT_1 \frac{G_1''(1-T_1(1-x))}{N_1+N_2}$$
$$+ (1-Q)(1-f)T_1 \frac{G_1''(1-T_1(1-x)) + N_2 G_1'(1-T_1(1-x))}{N_1+N_2} e^{-N_2 T_2(1-x)}$$
$$+ (1-Q)(1-f)N_2 T_2 \frac{G_1'(1-T_1(1-x)) + N_2 G_1(1-T_1(1-x))}{N_1+N_2} e^{-N_2 T_2(1-x)}$$

and evaluating this at $x = 1$ yields

$$R_0 = Q + (1-Q)fT_1 \frac{G_1''(1)}{N_1+N_2}$$
$$+ (1-Q)(1-f)\left(T_1 \frac{G_1''(1) + N_2 G_1'(1)}{N_1+N_2} + N_2 T_2 \frac{G_1'(1) + N_2 G_1(1)}{N_1+N_2}\right)$$
$$= Q + (1-Q)fT_1 \frac{G_1''(1)}{N_1+N_2} + (1-Q)(1-f)\left(T_1 \frac{G_1''(1) + N_1 N_2}{N_1+N_2} + N_2 T_2 \frac{N_1+N_2}{N_1+N_2}\right).$$

As before, if $R_0 > 1$, we find the lower bounds

$$\mathbf{P}_\infty > \tilde{Z}(0) = Q + (1-Q)\left(f\frac{G_1'(1-T_1)}{N_1+N_2} + (1-f)\frac{G_1'(1-T_1) + N_2 G_1(1-T_1)}{N_1+N_2}e^{-N_2 T_2}\right),$$

and

$$\mathbf{P} > Z(\tilde{Z}(0)) > Z(0) = Q + (1-Q)G_1(1-T_1)(f + (1-f)e^{-N_2 T_2}).$$

5.4. **Poisson and Poisson.** In this final example we split the contacts again in two groups, the familiar close interactions and distant sporadic ones. In contrast to the previous examples we shall assume a Poisson distribution for both of these classes of contacts having intensity N_1 and N_2 respectively. As in example 5.3 we will be assuming that a fractions of the population are isolating by cutting their contacts. To work with some generality we will assume that fraction f_1 of the population cuts its close contacts and another fraction f_2[2] cuts its sporadic contacts. Then, the degree distributions of the relevant trees $\mathcal{T}_1$ and $\mathcal{T}_2$ are

$$P_1(k_1) = f_1 \delta_{0k_1} + (1-f_1)\frac{N_1^{k_1}}{k_1!}e^{-N_1}, \text{ and } P_2(k_2) = f_2 \delta_{0k_2} + (1-f_2)\frac{N_2^{k_2}}{k_2!}e^{-N_2},$$

with the corresponding generating functions being

$$G_1(x_1) = f_1 + (1-f_1)e^{-N_1(1-x_1)}, \text{ and } G_2(x_2) = f_2 + (1-f_2)e^{-N_2(1-x_2)},$$

and compute $G(x_1, x_2) = (f_1 + (1-f_1)e^{-N_1(1-x_1)})(f_2 + (1-f_2)e^{-N_2(1-x_2)})$. Then,

$$Z(x) = Q + (1-Q)(f_1 + (1-f_1)e^{-N_1 T_1(1-x)})(f_2 + (1-f_2)e^{-N_2 T_2(1-x)}),$$

---

[2]It is probably reasonable to assume that $f_2 \geq f_1$.

while

$$\tilde{Z}(x) = Q + (1-Q) \sum_{i=1}^{2} \frac{N_i(1-f_i)f_j e^{-N_i T_i(1-x)}}{N_1(1-f_1) + N_2(1-f_2)}$$

$$+ (1-Q)(1-f_1)(1-f_2) \frac{N_1+N_2}{N_1(1-f_1)+N_2(1-f_2)} e^{-N_1 T_1(1-x) - N_2 T_2(1-x)},$$

where in the first sum $j \neq i$. We now compute R_0 using Proposition 1 which requires computing $\tilde{Z}'(1)$. This yields

$$R_0 = (1-Q)\left((1-f_1)N_1 T_1 + (1-f_2)N_2 T_2\right)$$

$$+ (1-Q)\frac{(1-f_1)f_1 N_1^2 T_1 + (1-f_2)f_2 N_2^2 T_2}{(1-f_1)N_1 + (1-f_2)N_2}.$$

If $R_0 > 1$, we have the lower bounds

$$\mathbf{P}_\infty > \tilde{Z}(0) = Q + (1-Q) \sum_{i=1}^{2} \frac{N_i(1-f_i)f_j e^{-N_i T_i}}{N_1(1-f_1)+N_2(1-f_2)}$$

$$+ (1-Q)(1-f_1)(1-f_2)\frac{N_1+N_2}{N_1(1-f_1)+N_2(1-f_2)} e^{-N_1 T_1 - N_2 T_2},$$

and

$$\mathbf{P} > Z(\tilde{Z}(0)) > Z(0) = Q + (1-Q)(f_1 + (1-f_1)e^{-N_1 T_1})(f_2 + (1-f_2)e^{-N_2 T_2}).$$

6. An alternative formula for R_0

This section is motivated by the examples in previous and in finding a more amenable general formula to compute R_0. We start with Proposition 1, namely the equation

$$R_0 = (1-Q) \sum_{j=1}^{n} T_j \frac{\partial \tilde{G}}{\partial x_j}\bigg|_{x_1=\ldots=x_n=1},$$

and the equation 2.3 which we rewrite here for simplicity

$$\tilde{G}(x_1,\ldots,x_n) = \frac{\sum_{l=1}^{n} G_l'(x_l) \prod_{i\neq l} G_i(x_i)}{\sum_{l=1}^{n} G_l'(1)}.$$

Using this, we compute

$$\frac{\partial \tilde{G}}{\partial x_j}\bigg|_{x_1=\ldots=x_n=1} = \frac{G_j''(x_j)\prod_{i\neq j}G_i(x_i) + \sum_{l\neq j}^{n} G_l'(x_l)G_j'(x_j)\prod_{i\neq l,j}G_i(x_i)}{\sum_{l=1}^{n}G_l'(1)}\bigg|_{x_1=\ldots=x_n=1}$$

$$= \frac{G_j''(1) + G_j'(1)\sum_{l\neq j}^{n}G_l'(1)}{\sum_{l=1}^{n}G_l'(1)}$$

and from the discussion in Remark 6 we further find

$$\frac{\partial \tilde{G}}{\partial x_j}\bigg|_{x_1=\ldots=x_n=1} = \frac{\mathbf{E}(k_j^2) - \mathbf{E}(k_j) + \mathbf{E}(k_j)\sum_{l\neq k}^n \mathbf{E}(k_l)}{\sum_{l=1}^n \mathbf{E}(k_l)}$$

$$= \mathbf{E}(k_j)\frac{\frac{\mathbf{E}(k_j^2)}{\mathbf{E}(k_j)} - 1 - \mathbf{E}(k_j) + \sum_{l=1}^n \mathbf{E}(k_l)}{\sum_{l=1}^n \mathbf{E}(k_l)}$$

$$= \mathbf{E}(k_j)\frac{\frac{\mathbf{E}(k_j^2)-\mathbf{E}(k_j)^2}{\mathbf{E}(k_j)} - 1 + \sum_{l=1}^n \mathbf{E}(k_l)}{\sum_{l=1}^n \mathbf{E}(k_l)}$$

$$= \frac{\mathbf{E}(k_j^2) - \mathbf{E}(k_j)^2}{\sum_{l=1}^n \mathbf{E}(k_l)} + \mathbf{E}(k_j)\left(1 - \frac{1}{\sum_{l=1}^n \mathbf{E}(k_l)}\right)$$

$$= \frac{\sigma(k_j)^2}{\sum_{l=1}^n \mathbf{E}(k_l)} + \mathbf{E}(k_j)\left(1 - \frac{1}{\sum_{l=1}^n \mathbf{E}(k_l)}\right),$$

where we have used $\sigma(k_j) = \sqrt{\mathbf{E}(k_j^2) - \mathbf{E}(k_j)^2}$ to denote the standard deviation of the degree distribution of the tree $\mathcal{T}_j$. Then, inserting this into the formula for R_0 yields the formulae in Theorem 2 which we shall restate here for convenience.

Theorem 5. *For each $j \in \{1,\ldots,n\}$ let $\mathbf{E}(k_j)$ denote the mean degree of the tree $\mathcal{T}_j$ and $\sigma(k_j)$ its standard deviation. Suppose that a fraction Q of all infected individuals is completely isolated and does not transmit the disease to anyone. Then, if the probability of transmitting the disease along an arm of the tree $\mathcal{T}_j$ is $T_j \in [0,1]$, the basic reproduction number is*

$$R_0 = (1-Q)\sum_{j=1}^n T_j\mathbf{E}(k_j)\left(1 - \frac{1}{\sum_{l=1}^n \mathbf{E}(k_l)}\right) + (1-Q)\sum_{j=1}^n T_j\sigma(k_j)\frac{\sigma(k_j)}{\sum_{l=1}^n \mathbf{E}(k_l)},$$

or

$$R_0 = (1-Q)\sum_{j=1}^n T_j\left(\mathbf{E}(k_j)\left(1 - \frac{1}{\sum_{l=1}^n \mathbf{E}(k_l)}\right) + \frac{\sigma(k_j)^2}{\sum_{l=1}^n \mathbf{E}(k_l)}\right),$$

7. APPLICATIONS

We shall now apply these results to a realistic scenario where a disease spreads through a population. Our goal is to investigate the possibility of an effective combination of isolation of infectious individuals, and practicing of social distancing, which together are capable of bringing R_0 below the threshold of 1. When that is not possible we will compute the probability of an epidemic developing.

7.1. **Random contacts occurring with two different constant rates.** We shall use the setup of example 5.4 where the contacts established by the population in two groups. These are the close and distant contacts encoded in the trees $\mathcal{T}_1$ and $\mathcal{T}_2$ respectively. As in that example we assume both degree distributions to be Poisson, having intensities N_1 and N_2.

The fractions of the population which are cutting their close contacts is f_1 and that cutting its sporadic contacts f_2. We will also be assuming that $f_2 \geq f_1$ as it seems reasonable to assume that everyone which cuts its close contacts also cuts its sporadic ones.

As computed in example 5.4, the basic reproduction number is

$$R_0 = (1 - Q) \left((1 - f_1)N_1T_1 + (1 - f_2)N_2T_2 + \frac{(1 - f_1)f_1N_1^2T_1 + (1 - f_2)f_2N_2^2T_2}{(1 - f_1)N_1 + (1 - f_2)N_2} \right).$$

When no intervention is made all f_1, f_2 and Q vanish the disease is free to propagate and the corresponding basic reproduction number will be denoted by

$$R_0^{free} = N_1T_1 + N_2T_2.$$

Example 1. *Suppose for the sake of simplicity that $f_1 = f = f_2$. Then, R_0 can be reqritten in terms of R_0^{free} as follows*

$$R_0 = (1 - Q) \left((1 - f)(N_1T_1 + N_2T_2) + f\frac{N_1^2T_1 + N_2^2T_2}{N_1 + N_2} \right)$$

$$= (1 - Q) \left(R_0^{free} - f\frac{N_1N_2}{N_1 + N_2}(T_1 + T_2) \right).$$

Then, the condition that $R_0 < 1$ which is sufficient to contain the outbreak turns into

$$f > \frac{N_1 + N_2}{N_1N_2}\frac{1}{T_1 + T_2} \left(R_0^{free} - \frac{1}{1 - Q} \right).$$

For example, suppose that $Q = 1/10$, $T_1 = 1/2$, $T_2 = 1/50$ and $N_1 = 4$, $N_2 = 50$. Then, $R_0^{free} = 3$ which is actually a reasonable assumption a disease such as Covid-19 and the computation above yields $f > 51/52$ which seems extremely difficult to achieve. On the other hand, if $Q = 1/2$ while all other parameters remain the same $f > 27/52 \approx 0.52$ which seems a much more achievable goal.

The previous example assumes that one can also cut the close contacts and that is not a reasonable assumption in most situations. For instance, many people may leave in the same house in which way it is not possible to cut such contacts. In the next example we address this and suppose only the sporadic contacts are cut.

Example 2. *Only cutting the sporadic contacts corresponds to having $f_1 = 0$. Then, R_0 can be is*

$$R_0 = (1 - Q) \left(N_1T_1 + (1 - f_2)N_2T_2 + \frac{(1 - f_2)f_2N_2^2T_2}{N_1 + (1 - f_2)N_2} \right)$$

$$= (1 - Q) \left(R_0^{free} - f_2N_2T_2 + \frac{(1 - f_2)f_2N_2^2T_2}{N_1 + (1 - f_2)N_2} \right)$$

$$= (1 - Q) \left(R_0^{free} - f_2\frac{N_1N_2T_2}{N_1 + (1 - f_2)N_2} \right)$$

and the condition that $R_0 < 1$ becomes

$$f_2 > \frac{N_1 + N_2}{N_2} \frac{R_0^{free} - \frac{1}{1-Q}}{R_0^{free} - \frac{1}{1-Q} + N_1 T_2}.$$

We shall now use the same numerical values as in the previous example, $T_1 = 1/2$, $T_2 = 1/50$ and $N_1 = 4$, $N_2 = 50$. Then, we find from the previous computation that $f_2 > 1$ for any value of $Q \leq 1/2$. Hence, there is no way of surely containing the disease simply from cutting out the sporadic contacts and not increasing Q above $1/2$. Suppose then the borderline case when $Q = 1/2$ and consider the setting where also $f_2 = 1/2$ of the population is capable of cutting their sporadic contacts. Then, we have

$$Z(x) = \frac{1}{2} + \frac{1}{4}e^{-2(1-x)} + \frac{1}{4}e^{-2502(1-x)}$$

$$\tilde{Z}(x) = \frac{1}{2} + \frac{1}{29}e^{-2(1-x)} + \frac{27}{58}e^{-2502(1-x)},$$

from which we find

$$\mathbf{P}_\infty \approx 0.531, \quad and \quad \mathbf{P} \approx 0.594,$$

i.e. the probability of an epidemic forming, which is given by $1 - \mathbf{P}$, is approximately 0.406.

7.2. **Some country based analysis.** We shall now due some country based analysis using the household composition from the United Nations database [UN]. This analysis is motivated by the current pandemic of Coronavirus. However, given the existence of insufficient knowledge to estimate the T_i and the rude data from the beginning of the outbreak, this section should be regarded as within the realm of academic exercise. We shall use a simple model with two trees $\mathcal{T}_1$ and $\mathcal{T}_2$ with the first modeling the close interactions and given by the household composition of the respective country with the degree distribution being encoded in a generating function $G_1(x_1)$ which is then a polynomial. The second tree is intended to model the sporadic interactions in public gatherings. Assuming these to occur as a constant rate, the most suitable degree distribution is the Poisson of some intensity, say N_2, so that $G_2(x_2) = e^{-N_2(1-x_2)}$. Having this in mind, we shall use the model of sections 5.2 and 5.3 above. In all cases we shall consider, we have from those examples

$$G_1(x_1) = p_0 + p_1 x + \ldots + p_5 x_1^5, \text{ and } G_2(x_2) = f + (1 - f)e^{-N_2(1-x_2)},$$

so that writing $N_1 = G_1'(1)$

$$Z(x) = Q + (1 - Q)G_1(1 - T_1(1 - x))\left(f + (1 - f)e^{-N_2 T_2(1-x)}\right)$$

$$\tilde{Z}(x) = Q + (1 - Q)\frac{G_1'(1 - T_1(1 - x))\left(f + (1 - f)e^{-N_2 T_2(1-x)}\right) + N_2 G_1(1 - T_1(1 - x))e^{-N_2 T_2(1-x)}}{N_1 + N_2},$$

and

$$R_0 = Q + (1 - Q)fT_1\frac{G_1''(1)}{N_1 + N_2} + (1 - Q)(1 - f)\left(T_1\frac{G_1''(1) + N_1N_2}{N_1 + N_2} + N_2T_2\frac{N_1 + N_2}{N_1 + N_2}\right).$$

7.2.1. *Germany.* We start with the case of Germany, which according to the United Nations database [UN] has a household distribution in which 40% of the population live alone, 47% with one or two other individuals, 13% with three or four other and 1% with more than five other. The reader may note that we have not stated what is the exact percentage that live with only one or two other individuals. This is because such data is hard to locate for a large number of countries and we believe it will make little different in the qualitative, but also quantitative errors such as estimating the transmission probabilities T_1 and T_2. Indeed, there are more serious issues contributing to quantitative deviations. Thus, we will assume that half of the corresponding 47% live with one other individual and the other half with two other ones. Similar remarks hold for the 13% of the population that lives with three or four other individuals.

Remark 7. *Of course, if all households remain isolated there is no way for the disease to spread from one household to another. However, we know this is not a realistic situation as in general there are close family ties connecting individuals in living in different houses. Given the difficulty in quantifying this we shall simply use the household composition to model the degree distribution of the tree $\mathcal{T}_1$.*

Based on the data mentioned above we shall assume that

$$G_1(x_1) = 0.4 + 0.235x_1 + 0.235x_1^2 + 0.065x_1^3 + 0.065x_1^4 + 0.01x_1^5.$$

As for the transmission rates, we will set $T_1 = 2/3$, i.e. the probability of infecting a close family member is $2/3$. In a similar way, we shall assume that the probability of infecting one of the sporadic contacts is $T_2 = 1/50$ and the average number of sporadic contacts is $N_2 = 100$. This includes everyone that an infected person stays next to in public transport, markets, restaurants, work and other common areas. Of course, these numbers are debatable and we have chosen these simply to illustrate the theory. Using them, and assuming that both $Q = 0$ and $f = 0$ we compute that $N_1 = 1.21$ and

$$R_0 \approx 2.81.$$

Then, iterating $\tilde{Z}(\mathbf{P}_n)$ six times the sequence appears to stabilize around $\mathbf{P}_\infty \approx 0.0852$ which gives

$$\mathbf{P} \approx Z(\mathbf{P}_\infty) \approx 0.0855.$$

See figure 1 where the intersection point $\mathbf{P}_\infty$ can be visualized graphically. Hence, according to the model there is a very slight chance, of approximately 8.6% that the outbreak will not lead to an epidemic. Of course, this assumes that no non-pharmaceutical interventions have
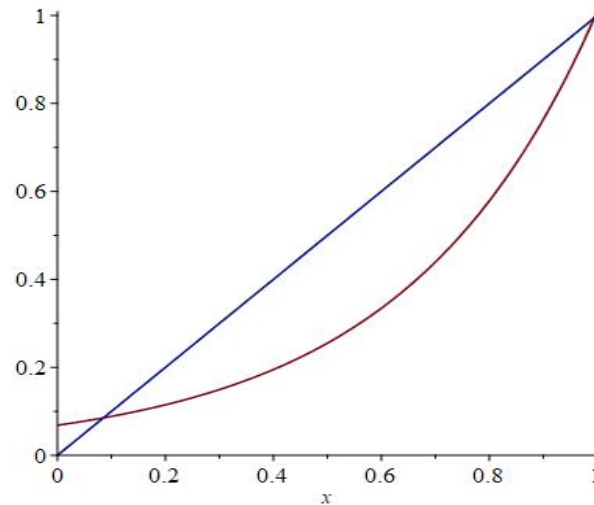
FIGURE 1. The function $\tilde{Z}(x)$ for Germany.

been put in place to control the outbreak. Suppose for instance that strict social distancing outside the household is imposed so that everyone adheres to it, which corresponds to $f = 1$. Then, a computation shows that $R_0 \approx 1.01 > 1$ and so is not yet enough to almost surely guarantee that the outbreak will not lead to an epidemic. Also, getting all the population to cut its sporadic contacts seems very difficult to achieve in practice. A more efficient and easier to achieve strategy seems to be that of identifying and isolating infected individuals and we shall now analyze it. For example, motivated by the case of Covid-19 let us consider a disease for which around 17% or 18% of individuals are asymptomatic. These estimates where obtained for Covid-19 in airport screening and the data from the Diamond Princess cruise ship, see [B], [S]. However, the validity of extrapolating thezse estimates to the remaining population is debatable as other studies seem to have quite disparate estimates for the fraction of asymptomatic carriers, see for instance [M]. Let us say then, that at least 0.18 of all infected individuals will not be detected, i.e $Q < 1 - 0.18$. We choose, for simplicity $Q = 0.7$ and $f = 0$ which yields $R_0 \approx 0.84$ and the outbreak will almost surely be contained. As another example, suppose that $Q = 0.6$ which is easier to achieve practically, then even with $f = 0.9$ we have $R_0 \approx 1.05 > 1$ and so it would be needed $f > 0.9$ which, again, is very difficult to achieve in practice. The conclusion is that, to contain the spread of the disease, it is much easier and effective to quarantine a sufficient fraction of infected individuals effectively than to simply cut the sporadic contacts of a large fraction of the entire population.

7.2.2. *Italy.* We shall now consider the Italian case, and based on [UN] we shall assume that

$$G_1(x_1) = 0.31 + 0.235x_1 + 0.235x_1^2 + 0.105x_1^3 + 0.105x_1^4 + 0.1x_1^5.$$

Then, using the same parameters as in the previous example we compute $R_0 \approx 3$ and $\mathbf{P}_\infty \approx 0.068$ so that

$$\mathbf{P} \approx Z(\mathbf{P}_\infty) \approx 0.068,$$

i.e. there is chance of 6.8% that an outbreak can be avoided without taking any precautions. Still, as a matter of comparison we find that if $Q = 0.7$ then $R_0 \approx 0.9$ which even though below 1 is visibly higher than that of the previous example.

7.2.3. *France.* For modeling France, we set

$$G_1(x_1) = 0.35 + 0.235x_1 + 0.235x_1^2 + 0.08x_1^3 + 0.08x_1^4 + 0.2x_1^5.$$

which with the same parameters of the previous two examples yields $R_0 \approx 3.09$ and $\mathbf{P}_\infty \approx 0.039$ which gives

$$\mathbf{P} \approx Z(\mathbf{P}_\infty) \approx 0.066,$$

yielding a probability of 6.6% to avoid an outbreak. When $Q = 0.7$ we compute $R_0 \approx 0.92$ which being below 1 is larger than that of the previous two examples.

7.2.4. *Portugal.* For Portugal, the United Nations database [UN] suggests using

$$G_1(x_1) = 0.19 + 0.28x_1 + 0.28x_1^2 + 0.115x_1^3 + 0.115x_1^4 + 0.2x_1^5.$$

Again, with the same parameters used in the previous examples we find $R_0 \approx 3.4$ and $\mathbf{P}_\infty \approx 0.04$. This gives

$$\mathbf{P} \approx Z(\mathbf{P}_\infty) \approx 0.04,$$

which yields a probability of 3.9% to avoid an outbreak. In this case, when $Q = 0.7$ we find $R_0 \approx 1.019$ which in contrast with the previous examples is already above 1. Thus, by Theorem 3 the probability $\mathbf{P}$ of avoiding an epidemic is below 1. Nevertheless, a computation shows that $\mathbf{P} \approx 0.984$ which is still quite high.

7.2.5. *Spain.* In the case of Spain we assume, using the same reference, that

$$G_1(x_1) = 0.19 + 0.265x_1 + 0.265x_1^2 + 0.13x_1^3 + 0.13x_1^4 + 0.3x_1^5.$$

Again, with the same parameters used in the previous examples we find $R_0 \approx 3.44$ and $\mathbf{P}_\infty \approx 0.039$. This gives

$$\mathbf{P} \approx Z(\mathbf{P}_\infty) \approx 0.039,$$

which yields a probability of 3.9% to avoid an outbreak. In this case, when $Q = 0.7$ we find $R_0 \approx 1.03$ as in the previous example. Indeed, the whole situation is very much parallel as that of the previous example.

7.2.6. *Brazil.* For Brazil we have

$$G_1(x_1) = 0.12 + 0.235x_1 + 0.235x_1^2 + 0.16x_1^3 + 0.16x_1^4 + 0.9x_1^5,$$

and with the same parameters of the previous examples we find $R_0 \approx 3.82$ which is the highest so far. Using this distributions we compute $\mathbf{P}_\infty \approx 0.026$ which has

$$\mathbf{P} \approx Z(\mathbf{P}_\infty) \approx 0.026,$$

which yields a very small probability of 2.6% to avoid an outbreak. In this case, even when $Q = 0.7$ we find $R_0 \approx 1.146$ which is still high. Indeed, the associated probability of avoiding an outbreak is $\mathbf{P} \approx 0.92$, i.e. there is a chance of 92% of containing the outbreak.

References

[AR] Anderson, Roy M. and May, Robert M. *Infectious diseases of humans: dynamics and control.* Oxford: Oxford University Press, (1991)

[Br] Brauer, Fred. *An introduction to networks in epidemic modeling.* Mathematical epidemiology. Springer, Berlin, Heidelberg, 2008. 133-146. 2, 7

[B] Quilty, Billy J., et al. *Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV).* Eurosurveillance 25.5 (2020): 2000080. 21

[Ha] Harris, T. E.: *The Theory of Branching Processes.* Springer, Berlin Heidelberg New York (1963) 2, 7

[Ja] Jagers, P.: *Branching Processes with Biological Applications.* Wiley, London (1975) 2, 7

[KA] Kimmel, M., Axelrod, D. E.: *Branching Processes in Biology.* Springer, Berlin Heidelberg New York (2002) 2, 7

[Sc] Schinazi, R. B.: Classical and Spatial Stochastic Processes. BirkhÂĺauser, Boston (1999) 2, 7

[C] Capasso, V. *Mathematical structures of epidemic systems.* Vol. 97. Springer Science and Business Media, (2008)

[M] Mizumoto, Kenji, et al. *Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020.* Eurosurveillance 25.10 (2020): 2000180. 21

[Ca] Callaway, Duncan S., et al. *Network robustness and fragility: Percolation on random graphs.* Physical review letters 85.25 (2000): 5468. 2, 7

[Re] Li, Ruiyun, et al. *Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2).* Science, 16 Mar (2020)

[S] Sun, W. W., et al. *Epidemiological characteristics of 2019 novel coronavirus family clustering in Zhejiang Province.* Zhonghua yu Fang yi xue za zhi [Chinese Journal of Preventive Medicine] 54 (2020): E027-E027. 21

[UN] United Nations, *Database on Household Size and Composition 2019.* Department of Economic and Social Affairs, Population Division (2019). 19, 20, 21, 22

Universidade Federal Fluminense

E-mail address: galato97@gmail.com