Multi-type branching and graph product theory of infectious disease outbreaks

Alexei Vazquez[1, 2, *]

[1] *Cancer Research UK Beatson Institute, Glasgow, United Kingdom*
[2] *Institute for Cancer Sciences, University of Glasgow, Glasgow, United Kingdom*

(Dated: October 9, 2020)

The heterogeneity of human populations is a major challenge to mathematical descriptions of infectious disease outbreaks. Numerical simulations are therefore deployed to account for the many factors influencing the disease spreading dynamics. Yet, the results from numerical simulations are often as complicated as the reality, leaving us with a sense of confusion about how the different factors account for the simulation results. Here, using a multi-type branching together with a graph tensor product approach, I derive a single equation for the effective reproductive number of an infectious disease outbreak. Using this equation I deconvolute the impact of crowd management, contact heterogeneity, testing, vaccination, mask use and smartphone tracing app use. This equation can be used to gain a basic understanding of infectious disease outbreaks and their simulations.

Infectious disease outbreaks take place in the background of a heterogenous population of susceptible individuals. A key source of heterogeneity is the variability in the number of potentially transmitting contacts. For airborne diseases like the ongoing COVID-19 outbreak, the relevant contacts are close physical proximity for a certain amount of time. Simulations of people movement within a city has shown that the number of proximity contacts within a day has a broad distribution across individuals [1]. Another source of variability are age groups, which dictate the mixing patterns between children, adults and the elderly [2]. Finally, the pattern of adherence to intervention guidelines to manage the disease introduces heterogeneity in the susceptibility of individuals to be infected [3].

These heterogeneities may sound too complex to be handle by means of analytical descriptions, leaving us with the choice of numerical simulations. Numerical simulations are indeed the right context to introduce all kinds of parametrizations of spreading processes on heterogenous populations [3–5]. Yet, we would also like to have a basic understanding of the problem under consideration, albeit sacrificing numerical precision. Here I demonstrate that a combination of multi-type branching process theory and graph tensor products allow us to disentangle the contributions of different factors and containment strategies to the outbreak dynamics.

The susceptible, infected and recovered (SIR) model is a good representation of infectious disease outbreaks when the recovery from the disease confers immunity. In the case of COVID-19 it is not clear how long a person remains immune to the disease after infection, but it is expected to be at least of the order of months. In the SIR model the individuals of a population can be susceptible to acquire the disease, infected and as a consequence infectious (with some probability) or dead/recovered from the disease. Infected individuals can transmit the disease to susceptible individuals when they are in contact. In the case of COVID-19, contact means physical proximity for a certain amount of time. In the case of HIV contact means sexual intercourse, syringe-needle sharing or mother giving birth baby. Here is were the population heterogeneity starts to kick in.

Some individuals may visit crowded places during a day, getting in contact with several people. Other individuals may work at home and get in contact mostly with their house mates. With relevance to HIV or other sexually transmitted diseases, there is a broad distribution in the number of sexual partners of individuals across a population [6]. I will call this *contact heterogeneity*. The number of physical proximity contacts in a day, or the number of sexual partners within a year, can vary from zero to 100s and it is better represented by a probability distribution.

Individuals are also different regarding their perception about the effectiveness of containment strategies enforced or suggested by the relevant authorities. In the context of COVID-19, face masks and smartphone tracing apps are not in use by all individuals. In the context of HIV and other sexually transmitted diseases there are several sources of heterogeneity, including sexual orientation, condom use, drug use, among other factors. I will call this *type heterogeneity*, where a type can be any property taking values over a discrete set of small size that can have an impact on the infectious disease dynamics. The types are characterized by their frequency in the population and the mixing patterns between individuals according to type.

Another potential source of variability is the disease dynamics within a given individual, from the time of receiving the infection to recovery. This dynamics could be, in general, correlated with the contact or type heterogeneities. Here I focus on the contact and type heterogeneity and assume that the disease dynamics within individuals is uncorrelated from the contact and type heterogeneity. In this case the disease transmission dynamics from an infected to a susceptible individual is characterized by the generating time, denoted by τ, defined as the interval from the time of infection of an individual to the time it transmit the disease to a susceptible

* avazque1@protonmail.com

individual. I will denote by $g(\tau)$ the probability density function of the generating time.

Using well stablished mathematics from the theory of multi-type branching processes, I have previously calculated the expected number of infected individuals of infectious disease outbreaks on heterogeneous populations [7]. In a nutshell, the multi-type formalism replaces the average reproductive number, an scalar, by a matrix of reproductive numbers, making an distinction between patient zero an any other infected individual. The average reproductive number matrix for patient zero has elements

$$R_{ab} = \frac{\langle\beta\rangle}{\gamma} r e_{ab} \tag{1}$$

where a and b are indexes over the types, $\langle\beta\rangle$ is the average contact rate in the population, γ is the recovery rate, r is the probability of infection transmission and e_{ab} is the probability that an individual of type a reaches a type b individual upon contact. For infected cases other than patient zero one needs to take into account that the disease spreading biases the disease transmission to individuals with a higher contact rate. The patient zero can be thought as an individual selected at random from the population. Any other infected individual will not be selected at random, but it will be found with a probability proportional to its contact rate: $\beta/N\langle\beta\rangle$, where N is the population size. Once infected, the individual found by contact will engage in new contacts at a rate β. Therefore, the average reproductive number matrix for patients other than patient zero has elements

$$\tilde{R}_{ab} = \frac{\langle\beta^2\rangle}{\langle\beta\rangle\gamma} r e_{ab} \tag{2}$$

R gives the average number of infectious at the first generation, those generated by patient zero. $R\tilde{R}$ gives the average number of infections at the second generation and $R\tilde{R}^{d-1}$ gives the average number of infections at the d generation. The actual time when an infected case at generation d becomes infected equals the sum of d generation times and it has a probability density function $g^{\star d}(t)$, where the symbol $\star$ denotes convolution ($g^{\star d} = \int_0^t g^{\star(d-1)}(\tau)g(t-\tau)d\tau$). Therefore, the average number of new infected individuals at time t is given by (equation (36) in Ref. [7])

$$I(t) = \sum_{ab} N_a \sum_{d=1}^{D} \left(R\tilde{R}^{d-1}\right)_{ab} g^{\star d}(t) \tag{3}$$

where N_a is the number of patients zero of type a and D is the maximum generation, when the disease transmission ends.

If infected individuals become infectious at a rate α and get recovered at rate γ, then the probability of disease transmission upon contact between an infected and a susceptible individual is given by [7]

$$r = \frac{\alpha}{\alpha+\gamma} \tag{4}$$

and the probability density function of the disease transmission times is exponential

$$g(\tau) = \frac{1}{\alpha+\gamma} e^{-(\alpha+\gamma)\tau} \tag{5}$$

Notice that, when the rate of becoming infectious is very high ($\alpha \gg \gamma$), disease transmission is almost certain and most disease transmissions will take place in a time scale of the order of $1/\alpha$.

Under the assumptions made from (1) to (5), equation (3) has two different limiting behaviours depending on the parameter

$$\theta = \frac{D-1}{\rho} \tag{6}$$

where ρ is the largest eigenvalue of $\tilde{R}$ [7]. When $\rho > 1$ and $\theta \gg 1$, then for $(\lambda+\mu)t \ll \theta$ the number of new infectious grows exponentially according to

$$I(t) \sim e^{(\rho-1)(\alpha+\gamma)t} \tag{7}$$

In contrast, when $\theta \ll 1$, then for $(\lambda+\mu)t \gg \theta$ the number of new infectious grows as a power law with an exponential cutoff

$$I(t) \sim t^{D-1} e^{-(\alpha+\gamma)t} \tag{8}$$

These results tell us that the outbreaks dynamics is mostly determined by the largest eigenvalue of the reproductive number matrix $\tilde{R}$ and the maximum number of generations D the outbreak goes through. I have already made the point that a lockdown is translated into a small value of D, leading to the manifestation of the power law growth with an exponential truncation [8, 9]. According to equation (8) this behaviour persist regardless of the type heterogeneity, provided $\theta \ll 1$. Here I will focus on the impact of heterogeneities on the largest eigenvalue ρ.

The type mixing matrix, with elements e_{ab}, can be represented by a directed weighted graph with loops. A directed edge (arc) will be drawn from type a to type b whenever $e_{ab} > 0$. Loops account for the fact that infected individuals of a given type could infect susceptible individuals of the same type. The arcs have weights e_{ab}, indicating the probability that, starting from a type a, there is a type b at the other end. Figure 1 provides examples of the type graphs associated with vaccination, mask use or smartphone use. In each case there are two types: vaccinated or not, wears mask or does not, smartphone tracing app user or not. The associated mixing matrices are 2×2 matrices and it is straightforward to calculate the largest eigenvalue. The challenge begins when we consider a combinations of those or other population stratifications at once. We would have to include several types and deal with matrices of largest dimension, making an analytical description cumbersome and prompting calculation errors.

Yet, if the different stratifications are not correlated the problem can be tackle with the use of graph tensor products. Under the assumption of independence, the type graph taking into account n independent population stratifications can be represented by the graph tensor product of each independent stratification

$$G = G^{(1)} \times G^{(2)} \times \cdots \times G^{(n)} \tag{9}$$

An example is shown in Fig. 1. In turn, the type mixing matrix of graph G can be written as a Kronecker product of the type mixing matrices of graphs G_i,

$$e = e^{(1)} \times e^{(2)} \times \cdots \times e^{(n)} \tag{10}$$

Here comes the trick. The eigenvalues of the Kronecker product of two matrixes are given by the pairwise product of the eigenvalues of each matrix (Theorem 13.12, [10]). An obvious corollary of this theorem if that the largest eigenvalue of e is equal to the products of the largest eigenvalues of the $e^{(i)}$,

$$\Lambda = \Lambda_1 \Lambda_2 \cdots \Lambda_n \tag{11}$$

where Λ_i denotes the largest eigenvalue of $e^{(i)}$. Finally, the largest eigenvalue of $\tilde{R}$ in equation (2) is given by

$$\rho = \frac{\langle \beta^2 \rangle}{\langle \beta \rangle \gamma} r \Lambda_1 \Lambda_2 \cdots \Lambda_n \tag{12}$$

We can use equation (12) to estimate the effectiveness of mixed strategies to contain an infectious disease outbreak. To illustrate how it is done, let us consider the case of a population where crowd management, testing, vaccination, mask use and smartphone tracing apps have been deployed. Crowd management alters the distribution of β across the population and its effect can be represented by the trasnformation

$$\frac{\langle \beta^2 \rangle}{\langle \beta \rangle \gamma} \to c \frac{\langle \beta^2 \rangle}{\langle \beta \rangle \gamma} \tag{13}$$

where $0 \leq c \leq 1$ is the reduction in contact heterogeneity due to the crowd management. Testing will increase the rate at with infected individuals are removed from the disease transmission chain. Assuming that testing is done at rate ξ (test reports per person per unit of time), then the probability of disease transmission upon contact should be updated to

$$r = \frac{\alpha}{\alpha + \gamma + \xi} \tag{14}$$

Vaccination can be modelled by the type graph G_1 in Fig. 1 and the associated type mixing matrix

$$e^{(1)} = \begin{bmatrix} 0 & 1-v \\ 0 & 1-v \end{bmatrix} \tag{15}$$

where v is the fraction of vaccinated individuals in the population. In this case

$$\Lambda_1 = 1 - v \tag{16}$$

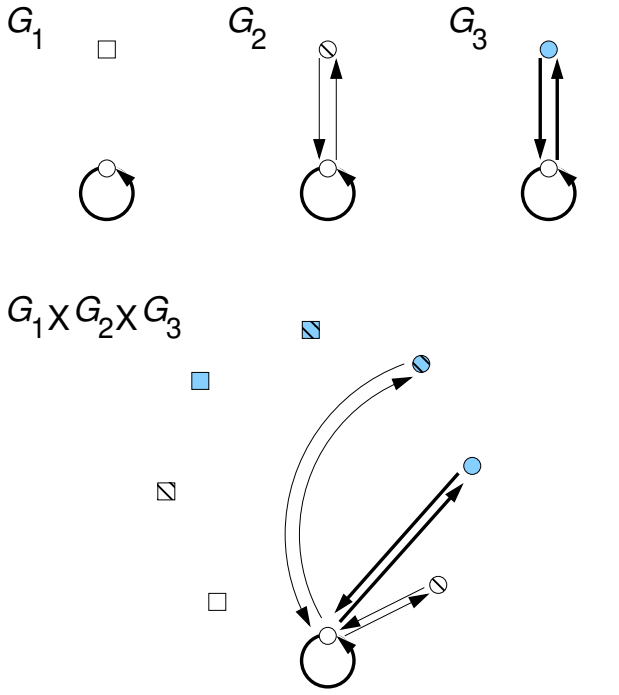

FIG. 1. Type graphs for vaccination (G_1), mask use (G_2), smartphone tracing app use (G_3) and their graph tensor product ($G_1 \times G_2 \times G_3$). In every case open circles represent individuals that are not covered by the containment strategy. Square are vaccinated individuals, symbols with a diagonal line are individuals that wear mask and solid symbols are individuals that use the smartphone tracing app

Mask use is modelled by the type graph G_2 in Fig. 1 and the associated type mixing matrix

$$e^{(2)} = \begin{bmatrix} 0 & a_1(1-m) \\ a_0 m & 1-m \end{bmatrix} \tag{17}$$

where m is the fraction of individuals that wear mask, $0 \leq a_1 < 1$ is the attenuation of the disease transmission from a mask user to a non-user, and $0 \leq a_0 < 1$ is the attenuation of the disease transmission from a non-mask user to a user. Here we have assumed that there is no disease transmission between mask users. In this case the largest eigenvalue equals to

$$\Lambda_2 = \frac{1 - m + \sqrt{(1-m)^2 + 4m(1-m)a_0 a_1}}{2} \tag{18}$$

Finally, smartphone tracing app use is modelled by the type graph G_3 in Fig. 1 and the associated type mixing matrix

$$e^{(3)} = \begin{bmatrix} 0 & 1-u \\ u & 1-u \end{bmatrix} \tag{19}$$

where u is the fraction of individuals that use the smartphone tracing app. Here we have assumed that there is no disease transmission between smartphone app users. This is because, once a smartphone app user test positive for the disease, any forward transmission to other

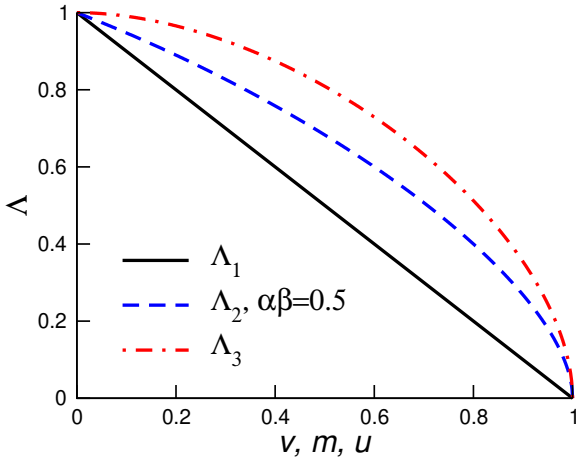

FIG. 2. Largest eigenvalue as a function of the relevant parameter for the listed containment strategies.

smartphone tracing app users is halted. In this case the largest eigenvalue equals to

$$\Lambda_3 = \frac{1-u+\sqrt{(1-u)^2+4u(1-u)}}{2} \tag{20}$$

Figure (2) shows the largest eigenvalue of the different containment strategies as a function of the fraction of individuals subject to the intervention (vaccinated, mask user, smartphone tracing app user). It is evident that the largest eigenvalues associated with mask use and smartphone tracing app use are concave functions of the corresponding users fraction. The latter implies that for small user fractions there is not much reduction of the largest eigenvalue. These containment strategies requires that many individuals become users. For example, 50% of mask users will reduce the reproductive number by just 20%. Another important observation is that, based on the assumptions make here, mask use is more effective that smartphone tracing app use. This is because mask use reduces the probability of transmission between mask users and non-users, while the smartphone tracing app does not.

Now we proceed to combine the different containment strategies. Substituting equations (13)-(20) into equation (12) we finally obtain

$$\begin{aligned}\rho = c\frac{\langle\beta^2\rangle}{\langle\beta\rangle\gamma}\frac{\alpha}{\alpha+\gamma+\xi}(1-v) \\ \cdot \frac{1-m+\sqrt{(1-m)^2+4m(1-m)a_0a_1}}{2} \\ \cdot \frac{1-u+\sqrt{(1-u)^2+4u(1-u)}}{2}\end{aligned} \tag{21}$$

This master equation can be used as a starting point to obtain a comprehensive understanding of how intervention strategies impact the expected reproductive number. In the absence of contact heterogeneity ($\langle\beta^2\rangle = \langle\beta\rangle^2$), an instantaneous rate of becoming infectious ($\alpha \gg \gamma$) and no interventions ($c = 1$, $v = 0$, $m = 0$, $u = 0$, $\xi = 0$), we obtain $\rho = R_0 = \langle\beta\rangle/\gamma$, which is the basic reproductive number of the standard SIR model. Due the contact heterogeneity R_0 underestimates ρ in the absence of interventions. In the present of multiple containment strategies, we can use (21) to estimate the aggregate impact. For example, combining a 50% of mask users with a 50% of smartphone tracing app users will reduce the reproductive number by about a half. Add to that a 50% vaccination and it will reduce the reproductive number by about a third.

In conclusion, using a multi-type branching together with a graph tensor product approach I have derived an equation for the expected reproductive number of an infectious disease as a function of the parameters of the disease outbreaks and under the action of multiple containment strategies.

ACKNOWLEDGEMENTS

This work was supported by Cancer Research UK C596/A21140.

[1] S. Eubank, H. Guclu, V. S. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, Nature **429**, 180 (2004).
[2] J. Zhang and *et al*, Science **368**, 1481 (2020).
[3] A. Aleta, D. Martín-Corral, A. Pastore y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini Jr, S. Merler, A. Pentland, A. Vespignani, E. Moro, and Y. Moreno, Nature Human Behaviour **4**, 964 (2020).
[4] P. G. T. Walker and *et al*, Science **369**, 413 (2020).
[5] N. G. Davies and *et al*, The Lancet Public Health **5**, e375 (2020).
[6] F. Liljeros, C. R. Edling, L. A. Amaral, H. E. Stanley, and Y. Aberg, Nature **411**, 907 (2001).
[7] A. Vazquez, Phys. Rev. E **74**, 066114 (2006).
[8] A. Vazquez, Phys. Rev. Lett. **96**, 038702 (2006).
[9] A. Vazquez, medRxiv 2020.07.23.20160531.
[10] A. J. Laub, *Matrix Analysis For Scientists And Engineers* (Society for Industrial and Applied Mathematics, USA, 2004).