Social Mixing Matrix Literature Review and an Intuitive Approach for the $\mathcal{R}_0$

GitHub: `https://github.com/epidemic-research-team/`

April 29, 2020

1 Introduction

Standard epidemic models, such as SIR, SIS and their extensions SIRS, MSEIRS, SEIR are well established mathematical methods that simulate the spread of an infectious decease [4, 8, 7]. One limitation of these models is that they base their prediction on the a-priori assumption that the rate of contact in a population is constant across environments and age groups. As a result, these methods neglect the vital component of how populations interact. Therefore, researchers who solely rely on the SIR/SIS family of models often formulate contact rates (σ) and reproduction numbers ($\mathcal{R}_0$) based on assumptions and non-data driven hypotheses. As a reminder, $\mathcal{R}_0$ is the *"basic reproduction number $\mathcal{R}_0$, which is defined as the average number of secondary infections produced when one infected individual is introduced into a host population where everyone is susceptible"* [8, p.21]. In general, $\mathcal{R}_0 > \sigma$ but in many models $\mathcal{R}_0 = \sigma$.

Social Mixing Matrices (SMMs) are considered important tools for formulating realistic $\mathcal{R}$, and σ. SMMs capture social distancing patterns based on direct observations and surveys and as a result they provide crucial information to both scientists and public officials on how a directly transmitted infectious decease is spreading. Because of the granularity of the data that SMMs offer, their use has the potential to increase model robustness and prediction accuracy compared to models that rely solely on indirect data [9, 1].

2 Literature Review

On their study, [2] explored the construction of the called next generation matrices (NGMs). NGMs are a vital part of the SMMs construction and utilisation process. Many papers that have used SMMs, one way or the other, they had to transform their initial population interaction matrix to be NGM compliant in order to adequately compute the integral $\mathcal{R}_0$ and σ [9, 3, 5]. In other words, NGMs show the number of novel infections from each category (age groups, gender groups, etc) in consecutive generations [2, p.874]. From the NGM, we can map the $\mathcal{R}_0$ to the dominant eigenvalue of the matrix. The ODE models that are used in epidemiology are linearised and form a Jacobian matrix [4, Section 2.1.2], then [2] showed how to decompose this matrix into a transition component Σ (change in state) and a transmission part T (new infections).

A landmark SMM based study was POLYMOD [9] that includes $n = 7,290$ participants from 8 European countries between years 2005-2006. This is the first study of such inter-state scale and since then it has been widely used amongst publications in the field. POLYMOD, contains information for $97,904$ contacts grouped by age, gender, type of contact (physical, conversational), place and frequency. The study expanded the pool of potential information that researchers for epidemiological deceases could access. In their results, the authors showed some interesting cultural outcomes. For example, using countries as dummy variables do seem to make a difference in the social mixing matrix. For instance, German participants showed per daily average $\mu = 8$ fewer number of contacts compared to their Italian counterparts $\mu = 20$.

One limitation with the POLYMOD study is the population bias that exhibits on children and young adult ages as the focus of the study has been set deliberately on these age groups. The reason is that the infectious decease that the authors studied (influenza) were most likely to be transmitted through young ages. For example, 38% of the participants are under 20 years old of age and 12% were over 60 years old. At the time of survey, the European average for these 8 countries for 20 y and under was 17% and the same average for 60 y and over was 21%. As a reference, the EU average for 20 y and under was 16% and 23% respectively. There is also a small change over the years, in 2018 (the same year as the BBC Pandemic survey in UK) the population in the 8 countries for 20 y and under was 15% and for 60y and over was 25%. The inclusion of developing Eastern European countries

has not managed to reverse the ageing population that developed countries suffer from. The subsequent rapid development of the new European Union countries has affected them with the same problem of an ageing population [1]. Another potential issue with the POLYMOD data is the time that has passed since the study was contacted. Since then, social behaviours and patterns might have changed dramatically.

Considering the aforementioned limitations that the POLYMOD data suffer from, [5], with the assistance of BBC Pandemic project, conducted a large scale population and social behaviour survey in UK, 2017-2018. The study included $n = 38,000$ volunteers between 13-90+ by contact type (physical, conversational) and by social environment (work, home, etc) . The survey, groups the individuals in 5 year age segments, up to 75+ for whom there is a single category. To construct the social mixing matrix [5] start from the *raw contact matrix* (M), where its elements are $m_{ij} = \frac{t_{ij}}{n_j}$. For t_{ij} is the total number of contacts, within 24h, between the reference age group j and people in age group i. Then, we compute the average number of contacts per surveyed person j by $m_{ij} = \frac{t_{ij}}{n_j}$. Next, the authors create a reciprocal matrix, or *population contact matrix* C, which is a biased version of the matrix M. The matrix C, uses ONS population data for 2018 w, for each age group, as a "prior" element that adds to the m_{ij} in order to get more realistic results, $c_{ij} = \frac{1}{2}(m_{ij} + m_{ij}\frac{w_i}{w_j})$.

The BBC Pandemic survey from [5] covers most of the limitations that POLYMOD study [2] has but BBC Pandemic has its own drawbacks too. For instance, because the BBC app allowed users from 13 y and more the survey does not contain information for ages below 13 y. For the study of infectious deceases that are transmitted more easily amongst children there is a gap in the survey. There is also an element of population bias considering the fact that the sample has not been selected randomly from the overall geographical population. It might be that population from big cities dominate the study. Also, the survey is not gender specific, whereas POLYMOD is.

A study on the magnitude of [5] is comprehensive and extensive but also costly, time consuming and logistically complicated. As alternative, researchers used statistical methods to map the social mixing characteristics from the 8 countries of [2] into other geographic areas and from there to infer results. These methods, could potentially by utilised in order to expand research on local authority level in UK or to other countries. For

---

[1]Information sourced from Eurostat data [demo_pjanbroad]

instance, [3] used census data from 28 European countries (household, work, education etc) in order to construct synthetic NGM. In general, they assume a uniform distribution to depict the contacts amongst individuals and then use empirical evidence from past research as a prior to modify these contact rates according to the social setting (work, household, school, etc) before they construct the average frequency of contacts matrix M that leads to the NGM matrix.

Other papers that have evaluated and expanded the POLYMOD survey using statistical methods include [10], where the authors sourced socio-demographic data from 152 countries in order to project the POLYMOD from 8 to 144 countries, using a hierarchical Bayesian model and MCMC to sample the posterior parameters.

3   Computation of $\mathcal{R}_0$ from NGM

3.1   Next Generation Matrices: SEIR

In their paper, [2, p.875] use a SEIR model with 2 exposed categories as an example (E_1, E_2).

$$\dot{S} = \mu N - \beta \frac{SI}{N} + \mu S$$

$$\dot{E}_1 = \rho \beta \frac{SI}{N} - (\nu_1 + \mu)\dot{E}_1$$

$$\dot{E}_2 = (1 - \rho)\beta \frac{SI}{N} - (\nu_2 + \mu)\dot{E}_2$$

$$\dot{I} = \nu_1 \dot{E}_1 +$$

$$nu_2\dot{E}_2 - (\gamma + \mu)I$$

$$\dot{R} = \gamma I - \mu R$$

where μ per capita birth and death rates, β the transmission rate, $\nu_i \in \{1, 2\}$ the rates of leaving the respective latency states E_i, γ the rate of leaving the infection state and $N = S + E_1 + E_2 + I + R$. Let $\dot{x} = (E_1, E_2, I)^T$, and we want to decompose the system to have the form

$$\dot{x} = (T + \Sigma)x \tag{1}$$

4

where matrix T corresponds to transmissions and matrix Σ to transitions.

According to [2], the role of T and Σ is similar to the sufficient statistics, as all information that is required is included in the decomposition. For the transition matrix T, we create a matrix with infected states, using indices $i, j \in 1, 2, 3$ and the record T_{ij} is the rate of infection term; individuals from state j infects individuals in i. If the element in T_{ij} is 0, then there is no chance of transmission. Therefore,

$$T = \begin{pmatrix} 0 & 0 & \rho\beta \\ 0 & 0 & c(1-\rho)\beta \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} -(\nu_1 + \mu) & 0 & 0 \\ 0 & -(\nu_2 + \mu) & 0 \\ \nu_1 & \nu_2 & -(\gamma + \mu) \end{pmatrix}$$

hence we solve for the NGM with large domain (K_L)

$$K_L = -T\Sigma^{-1} = \begin{pmatrix} 0 & 0 & \rho\beta \\ 0 & 0 & c(1-\rho)\beta \\ 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} -(\nu_1 + \mu) & 0 & 0 \\ 0 & -(\nu_2 + \mu) & 0 \\ \nu_1 & \nu_2 & -(\gamma + \mu) \end{pmatrix}$$

the dominant eigenvalue of this matrix K_L is the $\mathcal{R}$,

$$\mathcal{R}_0 = \left(\frac{\rho\nu_1}{\nu_1 + \mu} + \frac{(1-\rho)\nu_2}{\nu_2 + \mu} \right) \frac{\beta}{(\gamma + \mu)} \tag{2}$$

or similarly, by using the equation of NGM

$$K = E^T K_L E = -E^T T \Sigma^{-1} E \tag{3}$$

with E being a matrix that its columns contain unit vectors the non-zero rows of matrix T, which leads

$$\mathcal{R}_0 = \rho(K) = \frac{1}{2} \left(trace(K) + \sqrt{trace(K)^2 + 4det(K)} \right) \tag{4}$$

3.2 Next Generation Matrices: BBC Pandemic Survey

The NGM of [2] can be linked with [5] matrix C by

$$NGM = \frac{\mathcal{R}_0}{\rho(C)}C \tag{5}$$

Analogous to the 4, the $\rho(C)$ is the largest eigenvalue of C. As in PCA, the eigenvector associated with $\rho(C)$ shows which age group dominates in the social interactions and explain more the overall data variance. Therefore, following [2] examples we find the NGM and we already have $\rho(C)$ and C and we extract $\mathcal{R}_0$.

$$\mathcal{R}_0 = NGM \times C\rho(C)^{-1} \tag{6}$$

3.3 Next Generation Matrices: SIR

We saw in sub-Section 3.1 how to use [2] method to compute the NGM and subsequently the $\mathcal{R}_0$. Next, [3, p.4] provided an example of how to link [2] NGM method to a SIR model. Specifically, let the force of infection be λ_i of a SIR model to be a linear combination of the average frequency matrix M_{ij}, where it shows the frequency of contacts between age groups i and age groups j, and a transmissibility coefficient β_i, which accounts the probability of infection per contact. The matrix M_{ij} is similar to BBC Pandemic [5] matrix C_{ij} (C is a population biased M, see sub-Section 3.2).

The equations of the SIR model are

$$\dot{S} = \mu N - \beta M \frac{SI}{N} + \mu S$$
$$\dot{I} = \beta M \frac{SI}{N} - (\gamma + \mu)I$$
$$\dot{R} = \gamma I - \mu R$$

unlike the examples in [2], the authors do not use a population constant; nevertheless, we add them for consistency reasons, μ. Then, we calculate the $\mathcal{R}_0$ as the spectral radius of the NGM

$$\mathcal{R}_0 = \rho(\beta\gamma^{-1}M) \tag{7}$$

this is slightly different from [5] approach to compute $\mathcal{R}_0$ (Equation 6).

3.4   Next Generation Matrices: SIR-COVID

The aim of this research is to employ the work of [2] and [3] in computing the $\mathcal{R}_0$ in the epidemiological model from [6]. The model will be adapted to COVID-19 cases and to include social mixing matrix data instead of the sexual contact matrix that the authors used. The compartmental transmission model has a form as shown in Figure 1 [6, p.5]. From the Figure 1, the surveyed population is divided into the following non-overlapping compartments: S is for individuals who are at risk of Covid-19 infection, I is for infected individuals, G is for infected who developed serious symptoms and required ICU admission, P is those who have recovered, and are seropositive and immune, N is for individuals who are recovered, immune but seronegative. The indices g, a and s indicate that every compartment is stratified by gender, age and level of social activity. Movements between compartments are occurring at per capita rates specified by the following parameters: λ is the **force of infection** dependent on the proportion of individuals in I, **PSC** is the probability of becoming seropositive, **WIP** is the Covid-19 incubation period, **DAI** is the duration of asymptomatic (i.e. without Covid-19) infection symptoms, **DWT** is the duration of treatment for Covid-19, and **DI** is the duration of immunity. Subscripts denote stratification of parameters: for example, $\mathbf{DWT}_g$ means that in our model this parameter is gender-dependent.

The system of ODEs is then formulated as follow
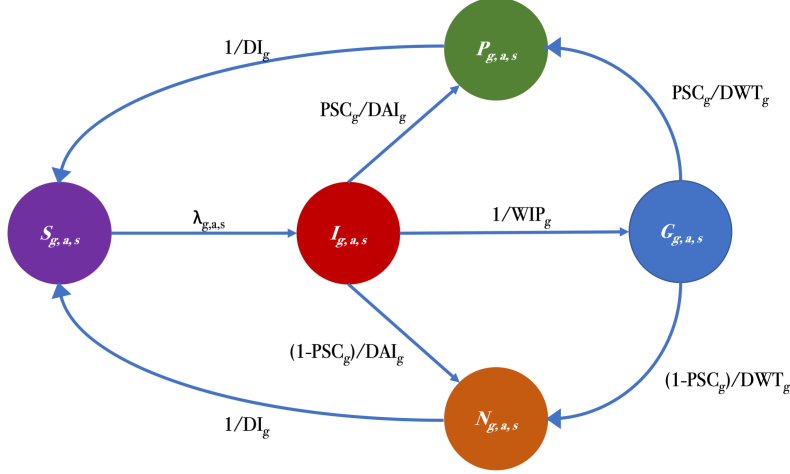
Figure 1: A compartmental Covid-19 transmission model.

$$\dot{S}_{g,s,a} = -\lambda_{g,s,a}(t)S_{g,s,a} + (P_{g,s,a} + N_{g,s,a})/DI_g + \frac{1}{r}S_{g,s,a-1} - \frac{1}{r}S_{g,s,a} +$$

$$\frac{1}{R}\sum_{g,s}(S_{g,s,20} + I_{g,s,20} + G_{g,s,20} + P_{g,s,20} + N_{g,s,20}) \times$$

$$\delta_1(a)(\pi_1\delta_1(s) + \pi_2\delta_2(s) + \pi_3\delta_3(s) + \pi_4\delta_4(s))$$

$$\dot{I}_{g,s,a} = \lambda_{g,s,a}(t)S_{g,s,a} - (1/WIP_g + 1/DAI_g)I_{g,s,a} + \frac{1}{r}I_{g,s,a-1} - \frac{1}{r}I_{g,s,a}$$

$$\dot{G}_{g,s,a} = I_{g,s,a}/WIP_g - G_{g,s,a}/DWT_g + \frac{1}{r}G_{g,s,a-1} - \frac{1}{r}G_{g,s,a}$$

$$\dot{P}_{g,s,a} = PSC_g(I_{g,s,a}/DAI_g + G_{g,s,a}/DWT_g) - P_{g,s,a}/DI_g + \frac{1}{r}P_{g,s,a-1} - \frac{1}{r}P_{g,s,a}$$

$$\dot{N}_{g,s,a} = (1 - PSC_g)(I_{g,s,a}/DAI_g + G_{g,s,a}/DWT_g) - N_{g,s,a}/DI_g + \frac{1}{r}N_{g,s,a-1} - \frac{1}{r}N_{g,s,a}$$

the stochastic compartmental system has in total 5 states, $T_{total\ population} = S + I + P + N + G$; of which two are infected states I and G and three uninfected states S, P and N. The total population remains constant by adding

an ageing term in the susceptible compartment, $\frac{1}{R}\sum(\dots) \times \delta\pi$ [6, p.7]. The ODE I is the state that contains the rate of infection term λ. The model is high dimensional since it involves the non-linear ODEs parameters that need to be calibrated $(\lambda, WIP, DAI, DWT, DI)$. At the infection free state, the number of susceptible equals the total population $S = Ttotal \quad population$.

Following [3, 1, 5], we will expand [6] by including in our research a social mixing matrix (POLYMOD or BBC Pandemic) to infer the $mathcalR_0$ and also we will consider the level of intervation and the impact it has on the social mixing patterns [10]. The current paper from [6] is using a sexual transmission matrix with dimensions $(2 \times 4 \times 4 \times 9 \times 9)$ for $g = 1|males$ and $g = 2|females$, four sexual activity groups $(s \in 1, \dots, 4)$ and a 9-bracket age group. In this initial stage of the model we will simplify the social distancing matrix by using only the age group as dimensions, we will move into more complex applications later.

The first step is to construct the NGM matrix [2] by using the survey data. In this example we will construct the NGM based on the BBC Pandemic data but following [3] approach for the SIR model. From [5], we compute the contact matrix M as a linear combination of age-specific contacts by context (home h, work w, school s, other r) and type (conversational Co or physical Ph) using the *population contact matrices* C. We also adjust the population contact matrices by a factor, based on temporal data collected from Google AI and Apple mobility reports [2].

$$M = H_{mob}C^h + W_{mob}C^w + S_{mob}C^s + R_{mob}C^r \tag{8}$$

where $H, W, S, R \quad mob$ are the mobility factors of home, work, school and other. The mobility factor matrices can be temporal, as a function of time similar to the ODEs or fixed. For the latter, we can use an average value of mobility reduction from the mobility reports. It is recommended though, because population movement restrictions could play a very significant role for the spread or not of the decease to utilise a temporal format. For instance, the mobility factor of H will have the approximate form:

---

[2]All data, including COVID, social distancing matrices, mobility reports, ONS and census UK specific data can be accessed in GitHub `https://bit.ly/2KNP06w`

$$H = \begin{pmatrix} 0.3 & 0 & \dots & 0 \\ 0 & 0.3 & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 0.3 \end{pmatrix}$$

References

[1] M. Baguelin, S. Flasche, A. Camacho, N. Demiris, E. Miller, and W. J. Edmunds. Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study. *PLoS Med*, 10(10):e1001527, 2013.

[2] O. Diekmann, J. A. P. Heesterbeek, and M. G. Roberts. The construction of next-generation matrices for compartmental epidemic models. *J R Soc Interface*, 7(47):873–885, 2010.

[3] L. Fumanelli, M. Ajelli, P. Manfredi, A. Vespignani, and S. Merler. Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLOS Computational Biology*, 8(9):1–10, 09 2012.

[4] M. Keeling, P. Rohani, and P. U. Press. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[5] P. Klepac, A. J. Kucharski, A. J. Conlan, S. Kissler, M. Tang, H. Fry, and J. R. Gog. Contacts in context: large-scale setting-specific social mixing matrices from the bbc pandemic project. *medRxiv*, 2020.

[6] I. A. Korostil, G. W. Peters, J. Cornebise, and D. G. Regan. Adaptive markov chain monte carlo forward projection for statistical analysis in epidemic modelling of human papillomavirus. *Statistics in Medicine*, 32(11):1917–1953, 2013.

[7] M. Li. *An Introduction to Mathematical Modeling of Infectious Diseases*. Mathematics of Planet Earth. Springer International Publishing, 2018.

[8] S. Ma and Y. Xia. *Mathematical Understanding of Infectious Disease Dynamics*. Lecture Notes Series, Institute for Mathematical Sciences. World Scientific, 2009.

[9] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikola-jczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, 5(3):1–1, 2008.

[10] K. Prem, A. R. Cook, and M. Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology*, 13:1–21, 09 2017.