Continuous time microsimulation of COVID-19 epidemics

Marcin Bodych, Tyll Krueger, Agata Migalska and Tomasz Ozanski,
Wroclaw University of Science and Technology and MOCOS research group.

Keywords: *Continuous time microsimulations, epidemic models, COVID 19, non-pharmaceutical intervention measures*

1 Introduction

Microsimulations of agent based models form an increasingly important tool for long and short time predictions of the evolution of key socio-economic quantities. For decision making on all levels of governance the outcomes of such microsimulations might provide valuable additional information. Contrary to classical mean field models for macroscopic observables, agent based models can take into account the intrinsic stochasticity of individual behavior and provide as outcomes probability distributions over possible trajectories. The complexity, sophistication and specificities of an agent based model depends very much on the aims of the model and the available data. For predictions of slowly varying sociological factors like the demographic age profile of a population it is usually enough to consider a discrete time evolution on a yearly base. Predictions for labour market related observables are subject to monthly or even weekly changes. Forecasting epidemic outbreaks of rapidly spreading diseases like influenca or COVID 19 has to be done at least on a daily base which is also the time unit on which data are collected and available. Complex models formulated in continuous time form up to now a minority among agent based models. The reason for that is mainly traditional and the unjustified but widespread belief that continuous time models would add an additional degree of computational complexity and hence are not suitable for large scale simulations on country or regional levels which involve usually millions of individuals. On the other side continuous time microsimulations have the advantage that events are disjoint and the simulation can proceed in an event driven manner that is process on event after another and updating the systems state accordingly. Properly implemented one can argue, that the run time of a continuous time microsimulation for a time period T scales linear in the number of events taking place within time T. As long as events affect only a small number of individuals in the population and assuming that the number of events per unit time is at most linear in population size n one can achieve runtimes proportional to $const \cdot T \cdot n$.

2 Methods

2.1 Model description

We present here a complex continuous time agent based model for COVID 19 developed by the MOCOS group in spring 2020. We describe the epidemic dynamics as a directed first passage percolation process on a directed random graph G. The random graph encodes the infectious link structure and as such is a subgraph of the random graph of physical contacts between individuals. The time dependent state $\chi_t(i)$ of an individual i corresponds to the setting of an SIR epidemics , that is 0 corresponds to the susceptible class S , state 1 correspond to the class of active infected I and state 2 corresponds to the removal or recovered class R e.g. individuals which have been

infected in the past, are cured and immune to further infections. The population is the node set V of the random graph space and an initial state is specified by the three subsets S_0, I_0 and R_0 . Besides the epidemic state variable $\chi_t(i)$ we assign to each node i a feature vector $x_i \in F \subset \mathbb{R}^k$ characterizing properties of a node which are relevant for the link structure and the timing of the infection process. To account for differences between in-household and outside household contact structures we furthermore partition the node set V into disjoint households $\{H_k\}_k^M$ with M being the total number of households. Links between nodes represent potential infectious contacts and are defined for out-houshold contacts via a kernel function $\varkappa : F \times F \to \mathbb{R}^+$ such that the probability of an infected node i to have an infectious contact to node j (denoted by $i \to j$) is given by

$$\Pr\{i \to j\} = \min\left\{1, \frac{\varkappa(x_i, x_j)}{n}\right\}, \ i,j \text{ not in the same household} \tag{1}$$

. Links between nodes i, j living in the same household are defined via a household kernel $\varkappa_H$ such that $\Pr\{i \to j\} = \varkappa_H(x_i, x_j)$.For both types of contacts a directed link from node i to node j is interpreted such that node i if infected would infect during the time when node i is infectious the noninfected node j. If node j is already infected respectively recovered we can keep the edge or remove it since it has no effect of the further progression of the epidemic. To define the evolution in continuous time we assign to each edge a random "travel " time - in epidemics called the generation time or serial interval - τ_{ij} drawn from a distribution $\varphi_i(\tau)$ which usually depends on the features of node i. τ_{ij} specify the time gap between the time when node i gets infected and when node j gets infected given that node i infects node j. The travel times τ_{ij} are themselves the sum of the random incubation time $T_i^{(0)}$ and a random time a_i drawn from a uniform distribution $U\left[0, T_i^{(1)}\right]$ where $T_i^{(1)}$ is the duration of the time node i is infectious. The distributions of τ_{ij} maybe different for in- and out-household contacts since infected individuals eventually stay home depending on the severeness of the disease progression. For a directed path $\gamma_{ij} = (i = k_0, ...k_{l-1}, k_l = j)$ from i to j containing no nodes from R_0 we define the distance between i and j as $d_\gamma(i,j) = \sum_{l=0}^{l-1} \tau_{k_l k_{l+1}}$. The set of all such directed paths from i to j is denoted by Γ_{ij}.We define the time of infection $\tilde{T}_j(I_0, R_0)$ for a node $j \in S_0$ which is in the the forward connected susceptible component $F(I_0)$ of I_0 as

$$\tilde{T}_j(I_0, R_0) = \min\{d_\gamma(i,j) : i \in I_0 \text{ and } \gamma \in \Gamma_{ij}\} \tag{2}$$

Note that we consider only paths avoiding the initial removal node set R_0. Since the progression from state 0 to states 1 and 2 depends only on x_i the epidemic process is well defined via the sets $A_t = I_t \cup R_t = \left\{i \in S_0 : \tilde{T}_i(I_0, R_0) \le t\right\} \cup I_0 \cup R_0$ and $R_t = \left\{i \in S_0 : \tilde{T}_i(I_0, R_0) + T_i^{(0)} + T_i^{(1)} \le t\right\} \cup \left\{i \in I_0 : T_i^{(0)} + T_i^{(1)} \le t\right\} \cup R_0$. Note that the set of active infected nodes I_t is given by $A_t - R_t$.

The feature space F may include socio-economic and health related features like age, gender, profession, social activity, comorbidities and household index of an individual as well as the timings defining the progression of the disease like incubation time and duration of the infectious period. Furthermore we assign to the feature space indicator variables wether an individual will be discovered via testing or contact tracing and put under quarantine which affects the duration of the infectious time for out-household contacts. We also assume that when a case is detected the whole household goes under quarantine. Contact tracing itself can be described as an additional tracing respectively search process B_t taking place on A_t and affecting the epidemic process

itself since traced individuals will be quarantined. B_t can be again viewed as a first passage process and is mainly specified by the distribution of the random time lag it takes to discover an infected offspring of an detected index case.
For clarity and simplicity of the exposition we have not given the description of the above epidemic process in the most general form. Natural extensions which are also implemented in the simulations include time dependent additional removal sets due to ongoing vaccinations and time dependence of the kernel functions due to changes in state imposed contact restrictions or process dependent changes in the behavior of the infected and susceptible population affecting the contact kernels. We also have not described additional state refinements like hospital stay, ICU treatment or death which are part of the actual simulation model.

2.2 Simulations and data

Simulations for realistic settings of the parameters and node features where made on regional and country level for Germany and Poland and included up to 80 million individuals. Data for the age structure and household composition where taken from German microcensus data and for Poland from census and recent registry data provided by GUS (Polish Statistical Office). Medical data for the progression of the disease like incubation time, time till hospitalization and mortality where taken from referenced in literature for Germany and from a large set of individual patient data for Poland [1] provided by the Polish National Institute for Public Health (NIPZ). The patient data set was also used to estimate the efficiency of contact tracing and testing for Poland and their changes over time. Computations for Poland with a population of 38 millions where made on the lower Silesia supercluster computation grid and runtimes for a single trajectory on a single core for a time span of two year took at most 10 minutes. The program code was written in JULIA and is public available on github. The simulation itself is taking place in an event driven manner. Due to the model definition the number of secondary out-household infections generated by an infected individuals i (the out- household outdegree with respect to the contact graph) is for large n Poisson distributed with parameter $\lambda_i = \int_F \varkappa(x_i, y)\, d\mu(y)$ where $\mu(y)$ is the empirical probability measure of the features over the sample population. The number of in - household infections is sampled from a multinomial distribution over the household member according to the corresponding household kernel $\varkappa_H$. Hence when a new infected node is born we first sample the number k of secondary cases from $Poiss(\lambda_i)$, than sample k individuals from V with respect to the induced measure $\frac{1}{\lambda_i}\varkappa(x_i, y)\, d\mu(y)$ and finally sample for each of those k individuals the generation times τ_{ij} and update the list of forthcoming events. Figure 1 shows key aspects of the algorithm in a schematic way.

3 Results

Besides forecast on medium and short time which are provided at a weekly base to the public German - Polish Forecast Hub (https://kitmetricslab.github.io/forecasthub/forecast) and to the Polish Ministry of Health, our main focus is on the in depth understanding of the impact and efficiency of nonpharmaceutical intervention measures (NIP) on the progression of the epidemics . This requires multiple large scale simulation over large parts of the parameter space specifying the NIP measures like contact reduction, contact tracing success probability , testing rates and the probability to get tested and corresponding delay times. We also include the effect of contact tracing via the COVID 10 smartphone tracing app which is widely used in Germany. Results are

presented in the form of heatmaps over the parameter space for the main observables of prevalence (at the end of the epidemics or after a year) , peak height of the epidemics and duration of the epidemics. In Fig 2, 3 and 4 we show some examples of the simulation outcomes. One of the key onbservations is the steepness of the phase transition when passing the critical line of reproduction number $R = 1$ which makes it difficult to get lasting and stable control of the epidemic without strong NPI measures (see also [2])
Besides the numerical simulation outcomes we also obtained theoretical results for the prevalence and peak time in the limit of $n \to \infty$. The theoretical results extend known theoretical work on phase transitions, size of the giant component and first passage percolation for heterogenous random graphs (see [3] and [4]) to our setting.

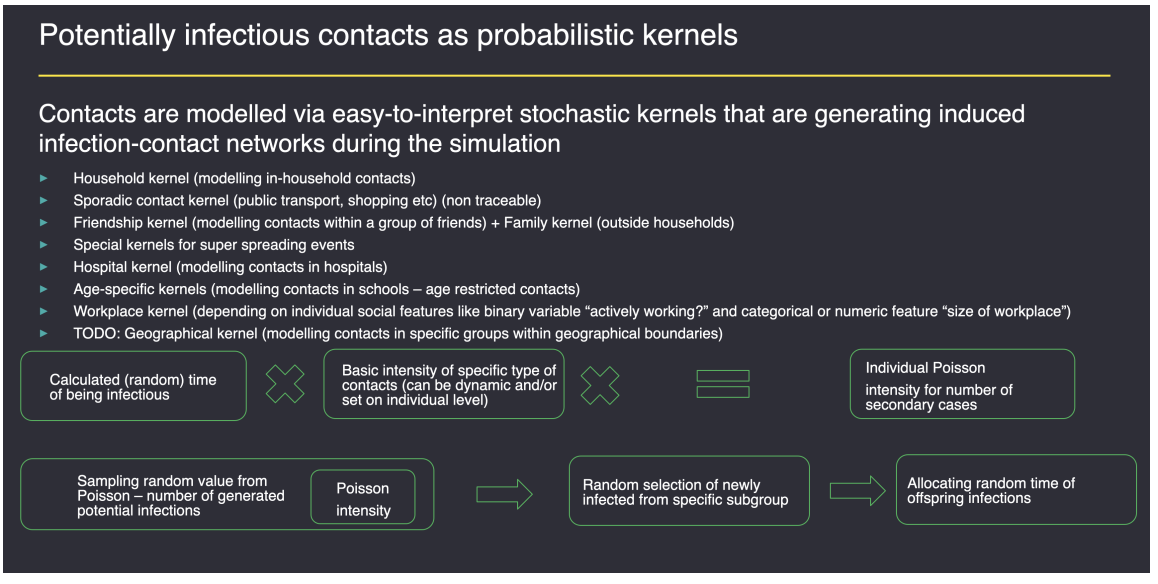

Figure 1: Schematic representation of the simulation algorithm

References

[1] Barbara Adamik, Marek Bawiec, Viktor Bezborodov, Przemyslaw Biecek, Wolfgang Bock, Marcin Bodych, Jan Pablo Burgard, Tyll Krueger, Agata Migalska, Tomasz Ożański, et al. Estimation of the severeness rate, death rate, household attack rate and the total number of covid-19 cases based on 16115 polish surveillance records. *medRxiv*, 2020. URL https://www.medrxiv.org/content/10.1101/2020.10.29.20222513v1. DOI: 10.1101/2020.10.29.20222513.

[2] Wolfgang Bock, Barbara Adamik, Marek Bawiec, Viktor Bezborodov, Marcin Bodych, Jan Pablo Burgard, Thomas Goetz, Tyll Krueger, Agata Migalska, Barbara Pabjan, et al. Mitigation and herd immunity strategy for covid-19 is likely to fail. *medRxiv*, 2020. URL https://www.medrxiv.org/content/10.1101/2020.03.25.20043109v2. DOI: 10.1101/2020.03.25.20043109.

[3] Shankar Bhamidi, Remco van der Hofstad, Gerard Hooghiemstra, et al. Universality for first passage percolation on sparse random graphs. *Annals of Probability*, 45 (4):2568–2630, 2017. URL https://projecteuclid.org/euclid.aop/1502438435. DOI: 10.1214/16-AOP1120.

[4] Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007. DOI: https://doi.org/10.1002/rsa.20168.

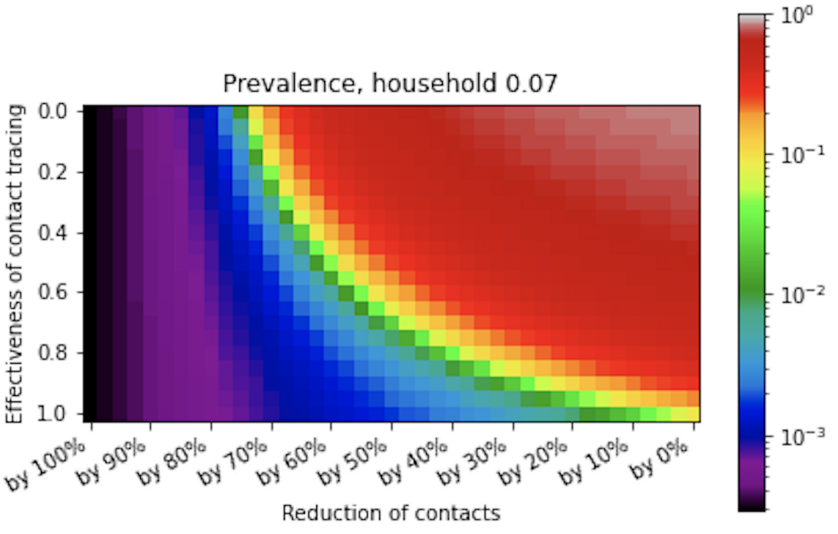

Figure 2: Prevalence as a function of contact reduction and tracing efficiency for two different in - household attack rates (detection prob. via testing is assumed to be 0.6 and time lag for testing and contact tracing 2 days)

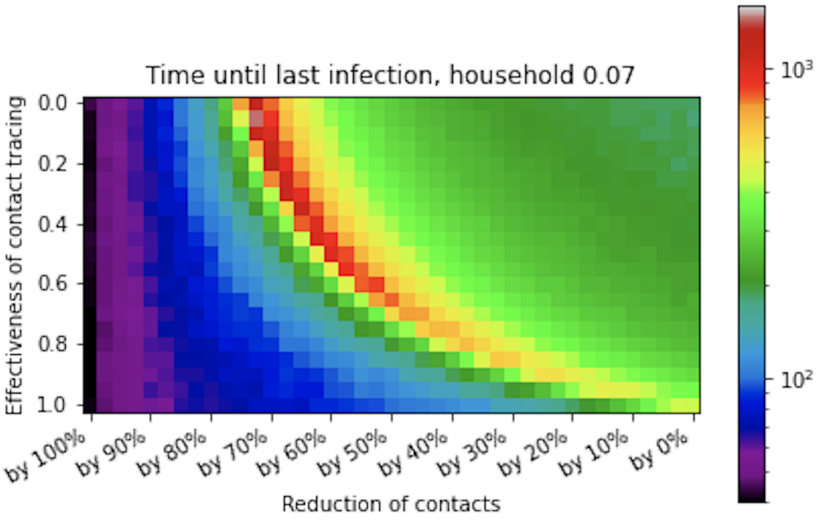

Figure 3: Duration till zero cases for the same setting as in picture 2. Simulations for Poland with mean over 100 runs per parameter combinations

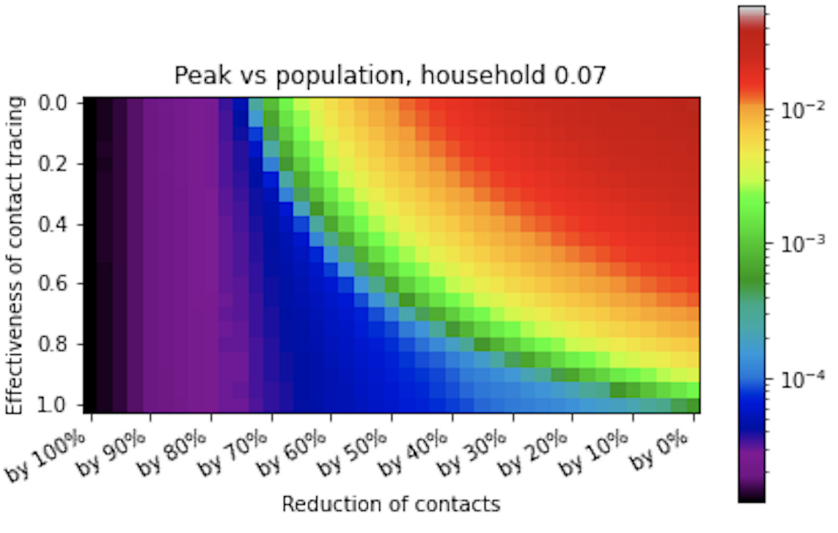

Figure 4: Maximal number of daily incidence as fraction of the population.