Supplementary Material for "An Agent Based Modeling of COVID-19: Validation, Analysis, and Recommendations"

Md. Salman Shamil, Farhanaz Farheen, Nabil Ibtehaz, Irtesam Mahmud Khan, M. Sohel Rahman

Department of CSE, BUET, ECE Building, West Palasi, Dhaka-1205

1. Details of the Agent Based Model

The set of states in our model is given by $S = \{H,N,A,S,D\}$ and any agent at any period of time will be in one of these five states. When an agent reaches state S, (s)he may or may not be hospitalized which, in our model, has been determined by assigning a probability to it. Once hospitalized, all activities previously performed by the patient are halted and the infected person stays at the hospital throughout the day.

The set of groups in our model is given by $G = \{F, T, W, E, H\}$ where the elements in the set represent the following -

F = Stay home
T = Commute
W = Work or attend school
E = Attend event
H = Stay at the hospital

Each agent at any period of time belongs to one of these five groups for an hour before being allocated to a new group based on supplementary Table 3 and Table 9.

For transportation, at first, our model calculates the total number of free seats by multiplying the number of transports with the passenger capacity. Then it assigns a free seat to a new agent. Every time an agent is given a free seat, the seat becomes occupied and the number of free seats in that transport is decremented.

2. Digital Herd Immunity

When most of the people in a population become immune to an infectious disease (via vaccines or mass infection), the spread of the disease can be controlled. This condition is called herd immunity.¹ This means that although people might come in contact with infected patients, they won't fall sick i.e. it will be defeated by their immune system. However, even without making the population biologically immune to an infectious disease, the population can be made immune to epidemic growth with the use of technology.² This condition is known as digital herd immunity. The concept of contact tracing is important to comprehend digital herd immunity. In order to contain the spread of infectious diseases like COVID-19 where no vaccines have been produced yet, Non-Pharmaceutical Interventions (NPI) such as lock-down, awareness and preventive measures are undertaken. Contact tracing is the act of identifying the people who came within close proximity of an infected person and subsequently quarantining them before they can infect others. Many diseases like COVID-19 have an omega period - which is the portion of the incubation period where the disease becomes contagious i.e. the period of asymptomatic transmission. Since new infections are only found due to onset of symptoms, it is likely that some people who came into contact with the infected person during the omega period have been infected as well. Contacts can be traced by manually interviewing the potential contacts of the patient but this is quite impracticable. Digital contact tracing uses

technology i.e. smartphones to store information on people in close proximity of infected patients. Bluetooth or GPS are two competing ways to trace contacts.³ By digital contact tracing and quarantining potentially infected people even before they show any symptoms during their latent period (the asymptomatic and not contagious stage), the spread of the disease can be significantly lowered, eventually leading to elimination of any further epidemic growth and, therefore, digital herd immunity. Our extensive experiments on different percentages of people being smartphone owners give a general idea about the time and total number of infections to reach digital herd immunity by contact tracing.

3. Supplementary Tables for Ford County, Kansas, USA:

Name	Value
min_age	1
max_age	77
min_name_length	3
max_name_length	7
min_family_size	1
max_family_size	5
n_workgroup	4000
n_transport	1000
transport_seat_limit	40
n_events	335
n_persons	33619
n_infected_init	2
awareness_start	1
quarantine_start	1
quarantined_person_ratio	0.4

Table 1: Part 1 of location-specific data for Ford county including total population, initial cases, average family size, life expectancy, lock-down declaration etc.

Name	Minimum age	Maximum age	Percentage	Total number
Student	4	25	0.32	10758

Service	18	65	0.488	16406
Doctor	30	65	0.005	168
Unemployed	10	77	0.187	6286

Table 2: Part 2 of location-specific data for Ford county including percentage of the total population belonging to different professions.

Task	Minimum start time	Maximum start time	Minimum duration	Maximum duration	Profession	Minimum probability	Maximum probability
Stay Home	0	0	7	8	Service	1	1
Stay Home	0	0	6	7	Driver	1	1
Stay Home	0	0	7	8	Doctor	1	1
Go to Work	7	8	1	2	Service	0.4	1
Go to Work	6	7	1	1	Driver	0.4	1
Go to Work	7	8	1	2	Doctor	0.8	1
Work	8	10	8	9	Service	0.4	1
Treat Patients	7	8	8	12	Driver	0.4	1
Treat Patients	8	9	8	10	Doctor	0.8	1
Returns Home	16	19	1	2	Service	0.4	1
Returns Home	15	20	1	1	Driver	0.4	1
Returns Home	16	19	1	2	Doctor	0.8	1
Stay Home	17	21	10	10	Service	1	1
Stay Home	16	21	10	10	Driver	1	1
Stay Home	17	21	10	10	Doctor	1	1
Attend Event	12	14	2	4	Service	0.3	0.9
Attend Event	12	14	2	4	Driver	0.3	0.6
Attend Event	12	14	2	3	Doctor	0.3	0.7
Stay Home	0	0	12	13	Unemployed	1	1
Stay Hospital	0	0	24	24	Hospitalized	1	1
Stay Home	0	0	7	8	Student	1	1

Go to Work	7	8	1	1	Student	0.4	1
Work	8	9	6	7	Student	0.4	1
Returns Home	14	16	1	1	Student	0.4	1
Stay Home	15	17	10	10	Student	1	1
Attend Event	12	14	2	4	Student	0.4	0.9
Go to Work	12	13	1	1	Unemployed	0.3	0.8
Attend Event	13	14	2	4	Unemployed	0.35	0.9
Returns Home	15	18	1	1	Unemployed	0.3	0.8
Stay Home	16	17	10	10	Unemployed	1	1
Stay Home	0	0	24	24	No Outing Allowed	1	1

Table 3: The lower and upper bounds of duration, probability of occurrence, starting and ending times for different tasks performed by an agent in Ford county.

Name	Value
action_occurring_threshold	0.6
action_affecting_threshold	0.6
action_infect_threshold	0.7
infection_threshold	0.65

Table 4: Values of different thresholds for Ford county.

Day	Real Data (R _i)	Model Output (M _i)	Difference (R _i - M _i)
1	2	2	0
2	2	2	0
3	3	2	1
4	5	2	3
5	13	20	7
6	16	28	12

7	18	37	19
8	32	46	14
9	51	46	5
10	85	68	17
11	107	88	19
12	127	108	19
13	181	134	47
14	188	173	15
15	209	211	2
16	289	262	27
17	350	316	34
18	420	374	46
19	459	426	33
20	516	492	24
21	544	539	5
22	601	584	17
23	675	627	48
24	702	676	26
25	770	729	41
26	815	775	40
27	832	829	3
28	869	872	3
29	880	916	36
30	933	981	48
31	1028	1025	3
32	1064	1076	12
33	1102	1121	19
			-

34	1135	1166	31
35	1135	1217	82
36	1233	1250	17
37	1233	1303	70
38	1299	1363	64
39	1299	1398	99
40	1299	1437	138
41	1403	1479	76
42	1404	1503	99
43	1440	1542	102
44	1440	1572	132
45	1558	1615	57
46	1558	1633	75
47	1558	1650	92
48	1607	1665	58
49	1607	1686	79
50	1628	1703	75
51	1628	1718	90
52	1738	1734	4
53	1738	1747	9
54	1738	1756	18
55	1790	1770	20
56	1790	1785	5
57	1804	1799	5
58	1804	1809	5
59	1823	1819	4
60	1823	1825	2

Table 5: Data points for daily confirmed cases of Ford county along with values obtained from our simulation and their absolute difference.

Name	Value
min_age	1
max_age	81
min_name_length	3
max_name_length	7
min_family_size	1
max_family_size	6
n_workgroup	600
n_transport	2500
transport_seat_limit	60
n_events	100
n_persons	10000
n_infected_init	1
awareness_start	7
quarantine_start	27
quarantined_person_ratio	0.5

4. Supplementary Tables for New York City, USA:

Table 6: Part 1 of location-specific data for New York city including total population, initial cases, average family size, life expectancy, lock-down declaration etc. The values have been scaled to accommodate for a population of 10000 from 8.3 million.

Name	Minimum age	Maximum age	Percentage	Total number (Per 10000 population)
Student	4	25	0.22	2200

Service	18	62	0.741	7410
Doctor	25	70	0.019	190
Unemployed	10	81	0.02	200

Table 7: Part 2 of location-specific data for New York city including percentage of the total population belonging to different professions.

Name	Value
action_occurring_threshold	0.55
action_affecting_threshold	0.55
action_infect_threshold	0.45
infection_threshold	0.55

Table 8: Values of different thresholds for New York city.

Task	Minimum start time	Maximum start time	Minimum duration	Maximum duration	Profession	Minimum probability	Maximum probability
Stay Home	0	0	7	8	Service	1	1
Stay Home	0	0	6	7	Driver	1	1
Stay Home	0	0	7	8	Doctor	1	1
Go to Work	7	8	1	2	Service	0.4	1
Go to Work	6	7	1	1	Driver	0.4	1
Go to Work	7	8	1	2	Doctor	0.8	1
Work	8	10	8	9	Service	0.4	1
Treat Patients	7	8	8	12	Driver	0.4	1
Treat Patients	8	9	8	10	Doctor	0.8	1
Returns Home	16	19	1	2	Service	0.4	1
Returns Home	15	20	1	1	Driver	0.4	1
Returns Home	16	19	1	2	Doctor	0.8	1

Stay Home	17	21	10	10	Service	1	1
Stay Home	16	21	10	10	Driver	1	1
Stay Home	17	21	10	10	Doctor	1	1
Attend Event	12	14	2	4	Service	0.3	0.9
Attend Event	12	14	2	4	Driver	0.3	0.6
Attend Event	12	14	2	3	Doctor	0.3	0.7
Stay Home	0	0	12	13	Unemploy ed	1	1
Stay Hospital	0	0	24	24	Hospitaliz ed	1	1
Stay Home	0	0	7	8	Student	1	1
Go to Work	7	8	1	1	Student	0.4	1
Work	8	9	6	7	Student	0.4	1
Returns Home	14	16	1	1	Student	0.4	1
Stay Home	15	17	10	10	Student	1	1
Attend Event	12	14	2	4	Student	0.4	0.9
Go to Work	12	13	1	1	Unemploy ed	0.3	0.55
Attend Event	13	14	2	4	Unemploy ed	0.35	0.6
Returns Home	15	18	1	1	Unemploy ed	0.3	0.55
Stay Home	16	17	10	10	Unemploy ed	1	1
Stay Home	0	0	24	24	No Outing Allowed	1	1

Table 9: The lower and upper bounds of duration, probability of occurrence, starting and ending times for different tasks performed by an agent in New York city.

5. Supplementary Tables for Physiological Data

Action	Min time gap	Max time gap	Min prob affect	Max prob affec t	task	Min prob	Max prob	Min effect other s	Max effect others	Min effect self	Max effect self
Sneeze	40	50	0.1	0.7	Work	0.1	0.605	0.1	0.705	0	0
Contaminate Thing	50	55	0.1	0.7	Work	0.1	0.605	0.1	0.705	0	0
Physical Contact	20	30	0.1	0.7	Work	0.1	0.605	0.1	0.705	0	0
Sneeze	40	50	0.1	0.7	Attend Event	0.1	0.65	0.1	0.8	0	0
Contaminate Thing	20	30	0.1	0.8	Attend Event	0.1	0.8	0.1	0.8	0	0
Physical Contact	20	30	0.2	0.8	Attend Event	0.2	0.8	0.3	0.8	0	0
Sneeze	40	50	0.1	0.7	Go to Work	0.1	0.7	0.1	0.705	0	0
Contaminate Thing	20	30	0.1	0.7	Go to Work	0.1	0.7	0.1	0.705	0	0
Physical Contact	20	30	0.1	0.7	Go to Work	0.1	0.7	0.1	0.705	0	0
Sneeze	40	50	0.1	0.7	Returns Home	0.1	0.7	0.1	0.705	0	0
Contaminate Thing	20	30	0.1	0.7	Returns Home	0.1	0.7	0.1	0.705	0	0
Physical Contact	20	30	0.1	0.7	Returns Home	0.1	0.7	0.1	0.705	0	0
Sneeze	40	50	0.1	0.7	Stay Hospital	0.1	0.7	0.1	0.705	0	0
Contaminate Thing	50	55	0.1	0.8	Stay Hospital	0.1	0.8	0.1	0.705	0	0
Physical Contact	20	30	0.1	0.8	Stay Hospital	0.1	0.8	0.1	0.705	0	0
Physical Contact	20	30	0.1	0.8	Treat Patients	0.1	0.8	0.1	0.705	0	0
Sleep	1000	1000	0	0	Stay Home	1	1	0	0	0	0
Wash Hands	30	40	0.5	1	Work	0.1	0.7	0	0	-0.75	-0.1
Wash Hands	30	40	0.5	1	Stay Home	0.2	0.7	0	0	-0.75	-0.1
Wash Hands	30	40	0.4	1	Attend Event	0.1	0.7	0	0	-0.75	-0.1
Wash Hands	20	40	0.4	1	Treat Patients	0.2	0.7	0	0	-0.75	-0.1
Contaminate Thing	50	55	0.1	0.7	Treat Patients	0.1	0.605	0.1	0.705	0	0
Sneeze	40	50	0.1	0.7	Treat Patients	0.1	0.605	0.1	0.705	0	0

Table 10: Lower and upper bounds of time interval between actions, probability of occurrence, effects on oneself and others etc. Here, min, max and prob refer to minimum, maximum and probability respectively.

6. Scaled down version of New York City

For conducting our experiments in case of New York city, we have chosen to run the simulations for 10,000 people. This involves scaling the location-specific input parameters for the smaller population. Table 7 shows that the proportion of people engaged in different professions are supplied as percentages to the model. However, some parameters are adjusted for the population of size = 10000. This can be understood from Table 11.

Name of Parameter	Total population of NYC	Scaled down version of NYC
Population size	8,399,000	10,000
Number of events (approx)	84,000	100
Number of groups (approx)	504,000	600
Number of vehicles (approx)	2,000,000	2,500

Table 11: Scaling of parameters for New York city for a population of 10000.

Data pertaining to Table 11 has been collected from various sources.^{4,5,6,7} Moreover, we have considered each working group to contain 10-12 people approximately. In case of the gatherings, we have considered approximately 100 people or less to be present.

To compare the daily values of effective reproduction number (R_t) of our scaled down ABM model with an SIR model, we calculate the R_t values for each day using the following formula:

$$R_t(x) = \frac{\sum\limits_{i \in I(x)} d_i^{out}}{|I(x)|}$$

Here, $R_t(x)$ denotes the R_t value on day x. I(x) is the set of persons infected on day x. d_i^{out} is the number of secondary infections caused by person *i*. Venkatramanan *et al.* provided a formula for calculating the weekly values of R_t that has been adopted for determining daily values in the above equation.⁸ To remove noise generated by randomness of each day, the R_t curve was smoothened.

Methodology of SIR model:

We have used the well known SIR model,⁹ which divides the total population into three different compartments, namely *Susceptible*, *Infectious* and *Removed*. We have assumed the total population to be *Susceptible* initially. The rate of change from *Susceptible* to *Infectious* is defined as *Transmission Rate* (β). On the other hand, the rate of change from *Infectious* to *Removed* is termed as *Removal Rate* (γ). *Removal Rate* is assumed to be constant over the period of time, while the *Transmission Rate* is assumed to be time variant. We performed a grid search among the plausible values of *Removal Rate* and the value with maximum likelihood is used as *Removal Rate* henceforth. On the other hand, based on the work of *Kurchaski et al.*,¹⁰ we have modeled transmission (i.e., *Transmission Rate*) as a stochastic random walk process. We have used Sequential Monte Carlo simulation (i.e., Particle Filter),^{11,12} in order

to find *Transmission Rate* with time, and consequently \mathbf{R}_t as well. Sequential Monte Carlo simulation is run 100 times with bootstrap fits to deduce various confidence intervals of \mathbf{R}_t . Our model is fitted with the number of daily confirmed cases. While fitting the model, we have tried to maximize negative log likelihood.



Figure 1: R_t curves by ABM model with scaled down population (blue) and SIR model with full population (red) of NYC. The blue curve appears to be similar to the red curve with an RMSE value of 0.4626, although converging to zero sooner than the red curve because of the smaller size of population.

Figure 1 shows that the R_t curves obtained in the ABM and SIR models are consistent, thus supporting the reasoning behind choosing to scale down the parameters of New York city.

7. References

1 What is herd immunity; June 18, 2020.

https://www.jhsph.edu/covid-19/articles/achieving-herd-immunity-with-covid19.html

2 Bulchandani VB, Shivam S, Moudgalya S, Sondhi SL. Digital herd immunity and COVID-19. *arXiv preprint arXiv:2004.07237.* 2020 Apr 15.

- 3 Digital Contact Tracing; June 18, 2020. <u>https://en.wikipedia.org/wiki/Digital_contact_tracing</u>
- 4 New York City population; June 18, 2020. <u>https://en.wikipedia.org/wiki/New_York_City</u>
- 5 New York City cars; June 18, 2020.
- https://nyc.streetsblog.org/2018/10/03/car-ownership-continues-to-rise-under-mayor-de-blasio/#:~:text=Ov erall%2C%20there%20were%201%2C923%2C041%20cars.over%20the%20four%2Dyear%20period.

6 New York City buses; June 18, 2020. https://www.ny.com/transportation/buses/#:~:text=City%20Bus%20System-,The%20New%20York%20Cit y%20Bus%20System,West)%20routes%20and%20outlying%20areas.

7 New York City subways; June 18, 2020. <u>https://en.wikipedia.org/wiki/New_York_City_Subway</u>

8 Venkatramanan S, Lewis B, Chen J, Higdon D, Vullikanti A, Marathe M. Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics*. 2018 Mar 1;**22**:43-9.

9 Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character. 1927 Aug 1;115(772):700-21.

10 Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Eggo RM, Sun F, Jit M, Munday JD, Davies N. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The lancet infectious diseases*. 2020 Mar 11.

11 Liu JS, Chen R. Sequential Monte Carlo methods for dynamic systems. *Journal of the American statistical association*. 1998 Sep 1;93(443):1032-44.

12 Del Moral P. Nonlinear filtering: Interacting particle resolution. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*. 1997 Sep 1;**325**(6):653-8.