Optimizing transport frequency in multi-layered urban transportation networks for pandemic prevention

Calum MacRury[*], Nykyta Polituchyi[†], Paweł Prałat[‡],

Kinga Siuta[§], and Przemysław Szufel[¶]

Abstract

In this paper, we show how transport policy decisions can affect the pandemic dynamics in urban populations. Specifically, we develop a multi-agent simulation framework to model infection dynamics in complex networks. Our agents periodically commute between home and work via a combination of walking routes and public transit, and make decisions intelligently based upon their location, available routes, and expectations of public transport arrival times. Our infection scheme allows for different contagiousness levels, as a function of the virus's strain and where the agents interact (i.e., inside or outside). The results show that the pandemic's scale is heavily impacted by the network's structure, and the decision making of the agents. In particular, the progression of the pandemic greatly differs when agents primarily infect each other in a crowded urban transportation system, opposed to while walking. Additionally, the results show that local subgraph characteristics, including topology, structure, and statistics such as its degree distribution and density, affect the viruses' transmission rates. We also assess the effect of modifying the public transport's running frequency on the spread of two different virus strains (with different levels of contagiousness). In particular, lowering the running frequency can discourage agents from taking public transportation too often, especially for shorter distances. On the other hand, the low frequency contributes to more crowded streetcars or subway cars if the policy is not designed correctly, which is why such an analysis may prove valuable for finding "sweet spots" that optimize the system. The proposed approach has been validated on real world data, and a model of the transportation network of downtown Toronto. The framework used is flexible and can be easily adjusted to model other urban environments, and additional forms of transportation (such as carpooling, ride-share and more). This general approach can be used modeling of contiguous disease spread in an urban environments including influenza or various COVID-19 variants.

1 Introduction

The goal of the paper is to discuss how decisions made py public transport decision maker are affecting pandemic in urban population such as influenza or COVID-19 [1]. In th order to control

[*]Department of Computer Science, University of Toronto, Toronto, ON, Canada; e-mail:cmacrury@cs.toronto.edu

[†]SGH Warsaw School of Economics, Warsaw, Poland; e-mail:nickypolit2@gmail.com

[‡]Department of Mathematics, Ryerson University, Toronto, ON, Canada; e-mail:pralat@ryerson.ca.

[§]SGH Warsaw School of Economics, Warsaw, Poland; e-mail:kinga.siuta@gmail.com

[¶]SGH Warsaw School of Economics, Warsaw, Poland; e-mail:pszufe@sgh.waw.pl

the spread of the virus, a policy regulator may take several actions, which we refer to as policy designs. The development of the epidemic depends on various factors, including the frequency of social contact, the contagiousness of a particular virus strain, and the level of protection applied during those contacts. One area where individuals can transmit the disease is in the urban space of a large city. In such environments, people meet at various points of interest (POIs) or while travelling, including while on sidewalks and in public transportation. In this paper, we take an approach similar to [2] and model these processes using an agent-based model (AMB) to assess the effects of social distancing interventions in public transportation of a large urban area.

While the literature on pandemic evolution is incredibly rich and vast, our work can be classified as an agent based model (ABM) simulator that tracks infection rates in the population. As described in [3], ABMs are valuable tools for decision makers worldwide. For instance, Covasim [4] has informed policy decisions in the United States (US), Vietnam, the United Kingdom (UK), and Australia. Another ABM tool, OpenABM-Covid19 [5], assesses COVID-19 non-pharmaceutical interventions including contact tracing, and the default version of the model is designed for UK demographics (while it can reflect other countries through re-parametrization). On the other hand, [6] is an example of a tool measuring the efficacy of mitigation rules such as social distancing or mask-wearing in metropolises. Hoertel et al. [7] introduce a COVID-19 ABM that investigates health care considerations such as post-quarantine screening, COVID-19 treatments and their impact on ICU-bed occupancy. In this framework and the accompanying article, the focus is placed on the transportation network within the city. While there are multiple all-purpose simulators such as Covasim, there is less work that is transportation-oriented. In this paper, we first design a transportation restrictions simulator. We then explore and analyze its outcomes in order to provide qualitative guidance to policy makers on public transportation running frequency during an epidemic outbreak.

The adopted research methodology is as follows. We first build a simplified "toy model" of reality that only represents the key processes from the point of view of our research questions. We mathematically analyze this model and explore key measures such as the number of infected pedestrians in relation to the frequency of public transportation. Afterwards, we build a simulation model, and compare its numerical results with the toy model, where we verify that the results of both models agree. Next, we greatly expand upon the simulation model in order to better approximate real-world relations. We numerically simulate this expanded model and observe the essential processes that take place in it, which we translate into real world conclusions. We have developed the agent-based simulation model using the `Julia` programming language [8] and the `OpenStreetMapX.jl` library. The spatial data comes from the OpenStreetMap project along with "TTC Routes and Schedules" data-set from the Toronto Open Data portal. All simulation source codes are available on GitHub repository[1].

The main contribution of this research study is that it shows that pandemic modelling in the context of public transportation is possible and can bring tangible benefits to policy makers. Numerical simulations on a model that reflects the main processes taking place in society can be a reliable data-driven source of knowledge supporting regulators in defining public transport policies to limit the spread of a virus such as COVID-19.

The remainder of the paper proceeds as follows. In Section 2, we present a literature review regarding COVID-19 in urban areas as well as its economic impacts. In Section 3, we rigorously analyze a small scale network "toy model" using tools from probability theory. The assumptions of

[1] `https://github.com/NykPol/EpidemicInUrbanNetworkToronto`

this model are far from reality, but it incorporates the main properties which we wish to explore. In Section 4, we introduce a far more general and complex model which reflects the real-world more accurately. Next, using discrete-event simulation, we run the simulation on the network of the toy model to verify the consistency between our mathematical and numerical results. Finally, we use simulation experiments to investigate different scenarios of the complex model and find the optimal setup of the public transportation system to minimize infection spread. Finally, we present our conclusions, insights, and possible extensions in Section 5.

2 Literature Background

Agent-based modeling (ABM) of pandemic can be classified into various groups depending on the target disease, population, environment, etc. In this section, we review the use of ABMs for simulating pandemics in urban environments. We focus on influenza transmission, dynamic analysis of COVID-19 outbreaks, the impact of COVID-19 on the health care system, and the economic implications of this virus.

2.1 ABM of Pandemic in Urban Areas

In [9], an agent-based simulator taking into account *social interactions* and *individual mobility patterns* was developed to investigate the 2009 H1N1 outbreak in Mexico. This simulator also measures the effect of mobility reduction on the spread of the disease. Another example which simulates the behaviour of the disease during pandemic outbreaks can be found in [10]. In this model, used by the Ontario Agency for Health Protection and Promotion, authors consider each individual to be unique and assume that transmission and infection rates are non-homogeneous. Furthermore, they consider the use of public transportation in urban cities. The model's output is then passed through software to generate a map of the area where the disease has been spread. In [11], an agent-based transport simulation combined with an epidemic spread model was proposed to understand the rate of infection in urban areas. The authors utilized the model to analyze seasonal influenza outbreaks in Zurich, Switzerland. See also [12] for the development of an agent-based model which employs geographic information system (GIS) to study influenza propagation in an urban area. The model is validated using 2009 H1N1 outbreak in Kunming, China. ACEMod, an agent-based model studies the spatiotemporal dynamics of influenza epidemics in Australia. The model uses *mobility patterns* (worker and student commuting) as well as *human interactions* and offers interventions/mitigation policies; see [13].

2.2 ABM of COVID-19 in Health Case System

Researchers are currently generating numerous agent-based models to investigate the impact of COVID-19 on the health care systems and propose strategies to circumvent the issues arising from the disease. This subsection provides an overview of some of the latest works.

Undoubtedly, testing and test-based interventions play a pivotal role in controlling the disease as people without COVID-19 symptoms can pass on the virus. To this end, an agent-based model, which evaluates the effectiveness of test-based policies, has been proposed by [14]. This tool is comprised of two models. The first model measures the effect of COVID-19 viral load on *secondary infections, false negative and false positive test results.* The second model investigates the propagation of COVID-19 through a group when test-based intervention policies are applied.

The National Institutes of Health (NIH) main campus in Bethesda, Maryland [15] has devised a SIMIO-based simulation model to discover the impact of bus schedule changes and reduced vehicle capacity under a wide range of employees demands. This, as a result, guarantees safe and reliable trips between campus buildings in the post-shutdown phase. Emergency departments (EDs) in the US have faced scarce resources and long waiting times due to COVID-19 pandemic. Researchers in [16] have built a discrete-event simulation model to monitor ED's operations and analyze the the impact of the virus on the ED throughput. In particular, researchers examine if the rates of factors such as patients length of stay and number of unvisited patients are susceptible to staffing changes when they simulate both standard/non-standard (due to COVID-19) patient volumes scenarios. Health systems have cancelled or deferred the non-emergency medical appointments to curb the spread of COVID-19. For instance, [17] proposes a discrete-event simulation providing insight for assigning patients to colonoscopy appointments under reduced capacity while safety measures remain in place. An agent-based model developed in [18] provides guidelines for emergency decision makers to effectively evacuate the emergency department at the Johns Hopkins Hospital during the pandemic. The proposed method first unloads non-COVID-19 patients and then starts evacuating COVID-19 cases. It also assumes that the teams dealing with COVID-19 cases do not interact with other patients and medical groups in other parts of the ED. A stochastic agent-based microsimulation model of the COVID-19 in France has been developed to study the impact of post-lockdown strategies, *including physical distancing, mask-wearing and shielding*, on cumulative incidence and mortality and on occupant beds in ICU; see [19].

2.3 ABM of COVID-19 in Urban Environment

Many researchers have concentrated their efforts on simulating COVID-19 movement across small towns and/or metropolises to enable them to evaluate the performance of interventions through agent-based frameworks. This subsection reviews some of the recent developments in this domain.

In order to understand how coronavirus spreads in small towns and cities, the paper [20] presents an ABM platform verified on real data from New Rochelle, New York. The tool incorporates testing strategies (in-hospital and drive through), different types of treatment (in-home and in-hospital), and intervention approaches (school and business closure/reopening). Another example is the risk assessment of COVID-19 spread in facilities proposed by [21] which helps decision makers develop appropriate strategies. In [22], an agent-based model has been created for the city of Salzburg, Austria, to support policy makers to take interventions measures. The model simulates four scenarios for the after-lockdown phase of COVID-19 and the results are independent of the above-mentioned city.

To measure the efficacy of mitigation rules in metropolises, including *social distancing, mask-wearing, school/business closures, random testing, and quarantines of differing lengths*, a customizable agent-based simulation tool has been designed and validated on New York City, which became the epicenter of the outbreak in the US. The capability of the model is illustrated through evaluating random testing approaches; see [6]. In addition, the paper [7] offers a stochastic agent-based microsimulation model of New York City and study the effect of *quarantine duration, quarantine lifting type, post-quarantine screening, and the use of a hypothetical effective treatment against COVID-19 on the disease's cumulative incidence and mortality, and on ICU-bed occupancy.* In [23], researchers implemented an agent-based model called CityCOVID to support decision making about COVID-19 at the level of city, county, and state (currently the Chicago area). To accomplish this goal, the model analyzes the COVID-19 transmission and understand the behaviour of individuals in reac-

tion to the government interventions. An extension of CityCOVID, focusing on the development of large-scale synthetic population, has been discussed in [24]. In Canada, the Public Health Agency has developed a SEIR-type agent-based simulation [25] to study the impact of non-pharmaceutical interventions on the spread of COVID-19 with the aim of supporting public health decision makers. The compartments used in the model are *isolated cases*, *exposed quarantine* from which spread to *susceptible* individuals is not possible, *asymptotic cases* that might/might not be identified via surveillance. Covasim, an agent-based simulator [26], inspects the dynamics of COVID-19 disease and evaluates various interventions policies in Africa, Europe, Oceania, and North America. Finally, an agent-based tool called OpenABM-Covid19 [5] assesses COVID-19 non-pharmaceutical interventions including contact tracing. The default version of the model is designed for UK demographics but it can be used for other countries through simple re-parameterization.

2.4 Economic Impacts of COVID-19 and ABM

ABMs have been employed to analyze the economic effects of COVID-19 pandemic, even though the number of related reports in the literature is much smaller than those discussing other aspects of the disease. The SEIR agent-based model from [2], analyzes the coronavirus economic impact considering seven scenarios such as *do nothing, lockdown, conditional lockdown, vertical isolation, partial isolation, use of face masks, and use of face masks together with reducing social contacts.* The model has been validated on data from Brazil but is applicable to other contexts as long as the corresponding data is provided. Another example is the ABM of supply chains of 1.6 million firms in Japan which predicts a total production loss of 5.2% in the country's annual Gross domestic product (GDP) if one-month lockdown was implemented in Tokyo; see [27]. Furthermore, the economic, social, and health impacts of COVID-19 in the presence of interventions such as testing, physical distancing, and school/business closure using ABM is the focus of [28]. The study conducted by [29] aims at reducing the health and economic losses resulted from COVID-19 via ABM of the *interrelation between the prevalence of COVID-19 and economic activities.*

3 Theoretical Example

In this section, we investigate a very simple scenario in which agents have only two routes to choose from and all of them start at the same time. Of course, this is not a realistic situation but it captures the essence of the optimization and exhibits similar properties as the simulation of more realistic scenarios.

3.1 Definition of the Toy Model

Suppose that we are given a graph $G = (V, E)$ with vertices $V = \{h, r, w, s_1, s_2\}$, and edges $E = \{hr, rw, hs_1, s_1s_2, s_2w\}$, as well as $n \geqslant 1$ **agents**, which are labelled with a label from set $[n] := \{1, \ldots, n\}$. Initially, all the agents are positioned at their **home** h, and they commute to and from their **work** w each day via two potential routes, $P_1 = (h, r, w)$ and $P_2 = (h, s_1, s_2, w)$ of varying travel times—see Figure 1. Each edge $e \in E$ is associated with the expected **travel time** σ_e, such that $\sigma_i := \sum_{e \in P_i} \sigma_e$ indicates the amount of time it takes for an agent to move along route P_i. Note that we are making the simplifying assumption that the travel times do not depend on the agent. We also assume that there are fixed departure times for leaving both w and h to which all the agents adhere.

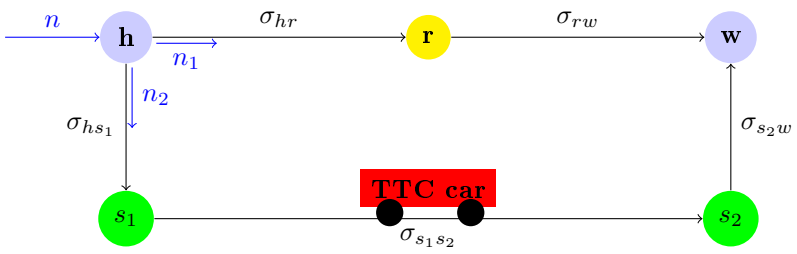

Figure 1: Toy Model

The agents are *mostly* rational in that they each will usually take the path that minimizes their travel time. However, we incorporate randomness into their decision making based on an exponential scaling function of the values σ_1, σ_2. Specifically, set $\sigma_{\min} := \min\{\sigma_1, \sigma_2\}$, and for each $i = 1, 2$, define the **transition probability** $\alpha_i \in [0, 1]$ where

$$\alpha_i := \frac{\exp(-\sigma_i/\sigma_{\min})}{\exp(-\sigma_1/\sigma_{\min}) + \exp(-\sigma_2/\sigma_{\min})}. \tag{1}$$

The results obtained below (that is, Theorem 3.1 and Corollary 3.2) hold for *any* transition probability but the above choice has a few desired properties. First of all, it is a decreasing function of σ_i: the longer travel time, the less likely an agent will select route P_i. Second of all, it only pays attention to the relative difference, that is, if both travel times increase or decrease by the same multiplicative factor, then the transition probability will not change.

During each **commute** $t \geqslant 1$, each agent $a \in [n]$ independently draws one of P_1 and P_2 according to the transition probabilities α_1 and α_2, respectively. If a draws P_i, then they move along P_i or the reversal of P_i, depending on whether they were at h or w after their previous commute.

We assume that some subset of agents $I_0 \subseteq [n]$ is initially **infected**. Suppose that I_t denotes the infected agents after $t \geqslant 0$ rounds. We update I_{t+1} from I_t based on the procedure below. Let us assume that P_i has an **interaction probability** $\beta_i \in [0, 1]$, as well as a **base contagion probability** $\lambda_i \in [0, 1]$ for $i \in \{1, 2\}$.

Suppose $a_1 \in [n]$ is infected at time $t \geqslant 0$, and $a_2 \in [n]$ is not infected after t rounds. Agent a_1 **infects** agent a_2 in round $t + 1$, provided the following events occur in order:

1. In round $t + 1$, agents a_1 and a_2 both select path P_i for some $i \in \{1, 2\}$.

2. Agents a_1 and a_2 **interact**, which given 1., occurs independently with probability β_i by definition.

3. The interaction between agents a_1 and a_2 is **contagious**, which given 1. and 2., occurs independently with probability λ_i.

We say that a_2 is infected during commute $t + 1$, provided there exists at least one agent which infects a_2. In this case, we update I_{t+1} from I_t by adding all those agents within $[n] \backslash I_t$ which became infected in commute $t + 1$. Note that once an agent becomes infected, it remains so for the duration of the process, i.e., $I_t \subseteq I_{t+1}$ for all $t \geqslant 0$.

3.2 Analyzing the Toy Model

Let us assume that $\beta_i = b_i/n$ for non-negative constants b_1, b_2 which do not depend on n. Moreover, we assume that $|I_0| = c_s n$ for some constant $0 < c_s < 1$, where we refer to c_s as the **starting infected proportion**. For convenience, we define the random variable $X_t := |I_t|$ for each $t \geqslant 0$.

Given a target **final infection proportion** c_f such that $0 < c_s \leqslant c_f < 1$, we define τ_{c_f} to be the smallest $t \geqslant 0$ such that $X_t/n \geqslant c_f$. Note that we assume $c_f < 1$ in order to ensure the concentration of our random variables throughout the entire process. We also assume that the parameters $b_1, b_2, \alpha_1, \alpha_2, \lambda_1, \lambda_2, c_s$, and c_f satisfy

$$(1-c_f)(\alpha_1(1-\exp(-b_1\lambda_1\alpha_1 c_s)) + \alpha_2(1-\exp(-b_2\lambda_2\alpha_2 c_s))) > 0. \tag{2}$$

In particular, if all parameters are positive, then this condition is trivially satisfied (recall that $0 < c_s \leqslant c_f < 1$ is already assumed). As we show below (see Corollary 3.2), this will ensure that $\tau_{c_f} < \infty$. In other words, the infection is guaranteed to eventually spread through c_f fraction of the agents. Our goal is thus to predict how long this process will take with a high degree of certainty.

Consider the following recursively defined sequence $(\widetilde{x}_t)_{t=0}^{\infty}$ where

$$\widetilde{x}_{t+1} := \widetilde{x}_t + (1-\widetilde{x}_t)\left(\alpha_1(1-\exp\left(-b_1\lambda_1\alpha_1\widetilde{x}_t\right)) + \alpha_2(1-\exp\left(-b_2\lambda_2\alpha_2\widetilde{x}_t\right))\right), \tag{3}$$

for each $t \geqslant 0$, and $x_0 := c_s$. Clearly, since (2) holds, this sequence is increasing and so, in particular, $c_s \leqslant \widetilde{x}_t \leqslant 1$ for all $t \geqslant 0$. Note that X_t/n is a random variable. However, we will show that one can use $\widetilde{x}_t$ of (3) to approximate X_t/n. This approximation will hold **asymptotically almost surely (a.a.s.)**, that is, it will hold with probability tending to 1 as $n \to \infty$. Having said that, with slightly more work one may compute (or estimate) the failure probability as a function of n (decaying exponentially fast) and state the concentration results for a given value of n. The conclusion would be that with probability at least $1-\epsilon$ for some small constant $\epsilon > 0$, the approximation is accurate even for moderately small values of n.

We are now ready to state the main results in this section. We defer all proofs to Appendix A.

Theorem 3.1. *For each $0 \leqslant t_0 < \tau_{c_f}$, there exists a function $\epsilon_0 = \epsilon_0(n) = o(1)$, such that a.a.s. it holds that*

$$\left|\frac{X_{t_0}}{n} - \widetilde{x}_{t_0}\right| \leqslant \epsilon_0(n).$$

Let us define $t_{c_f} \geqslant 0$ as the smallest $t \geqslant 0$ such that $\widetilde{x}_t \geqslant c_f$. Note that this is a deterministic value, as it only depends on the (deterministic) sequence $(\widetilde{x}_t)_{t=0}^{\infty}$, Moreover, $t_{c_f} < \infty$. To see this, notice that for any $0 \leqslant t < t_{c_f}$, we can apply (2) to ensure

$$\widetilde{x}_{t+1} - \widetilde{x}_t \geqslant (1-c_f)(\alpha_1(1-\exp(-b_1\lambda_1\alpha_1 c_s)) + \alpha_2(1-\exp(-b_2\lambda_2\alpha_2 c_s))) > 0,$$

as $\widetilde{x}_t \geqslant c_s$ for all $t \geqslant 0$. Thus,

$$t_{c_f} \leqslant \frac{c_f - c_s}{(1-c_f)(\alpha_1(1-\exp(-b_1\lambda_1\alpha_1 c_s)) + \alpha_2(1-\exp(-b_2\lambda_2\alpha_2 c_s)))}.$$

The following corollary relates τ_{c_f} and t_{c_f}.

Corollary 3.2. *The following holds a.a.s.:*

(a) If $\widetilde{x}_{t_{c_f}} > c_f$, then $\tau_{c_f} = t_{c_f}$.

(b) If $\widetilde{x}_{t_{c_f}} = c_f$, then $\tau_{c_f} \in \{t_{c_f}, t_{c_f}+1\}$.

b_1	interaction probability b_1/n for route P_1 (walking)
b_2	interaction probability b_2/n for route P_2 (subway)
σ_1	expected travel time for route P_1 (walking)
σ_2	expected travel time for route P_2 (subway)
λ_1	base contagion probability for route P_1 (walking)
λ_2	base contagion probability for route P_2 (subway)
c_s	starting infected proportion
c_f	final infection proportion

Table 1: Parameters of the toy model.

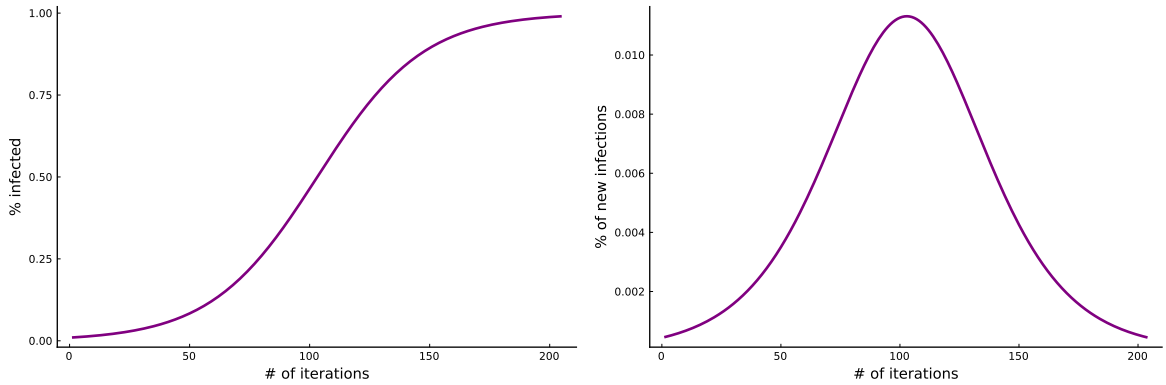

Figure 2: Scenario 1: the fraction of agents infected (left), and the fraction of new infections (right).

3.3 Illustration

Let us recall that there are 8 parameters of the toy model—we list them in Table 1. We investigate a few scenarios below but if one wants to test other sets of parameters, then the `Julia` code can be found on GitHub repository. For simplicity, for all the scenarios we set $\lambda_1 = 1/100$, $\lambda_2 = 2/100$, $c_s = 0.01$, and $c_f = 0.99$.

Scenario 1: $b_1 = 5$, $b_2 = 5$, $\sigma_1 = 15$, $\sigma_2 = 10$.

In this scenario, the subway is 50% faster than walking and so it is used by agents more often, namely, with probability $\alpha_2 \approx 62.25\%$. We claim that 99% of agents are infected after $t_f = 204$ iterations.

The fraction of agents infected after t iterations is presented in Figure 2 (left). We see an "S-shaped" function which can be explained as follows. Initially, the number of infected agents is small and they interact with a small number of non-infected agents. As a result, the new infections are relatively rare. On the other hand, if the number of infected agents is large, then the number of non-infected agents is small. Hence, the number of new infections is also small. The number of new infections is large when there are many infected agents but at the same time there are many non-infected ones that can potentially get infected—see Figure 2 (right).

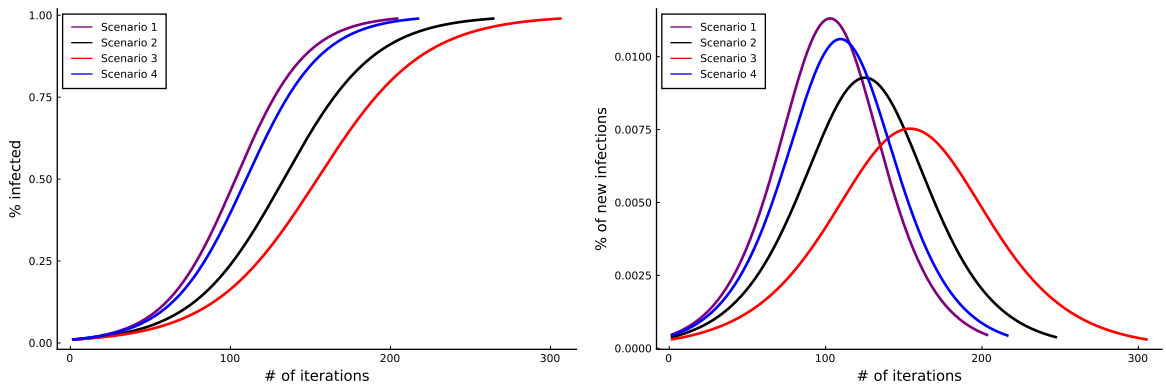

Figure 3: Scenarios 1–4: the fraction of agents infected (left), and the fraction of new infections (right).

Scenario 2: $b_1 = 5$, $b_2 = 5$, $\sigma_1 = 15$, $\sigma_2 = 15$.

In this scenario, the subway slows down and the expected travel time by subway is the same as walking. As a result, agents select each route with the same probability, namely, $\alpha_1 = \alpha_2 = 50\%$. 99% of agents are infected after $t_f = 248$ iterations. The virus spreads slower than in Scenario 1.

Scenario 3: $b_1 = 5$, $b_2 = 3$, $\sigma_1 = 15$, $\sigma_2 = 10$.

In this scenario, the subway's expected travel time is back to the original value (as in Scenario 1) but we assume that agents taking subway interact with a smaller number of other agents: b_2 is reduced from 5 to 3. As expected, it has a positive effect: 99% of agents are infected after $t_f = 306$ iterations. The virus spreads even slower than in Scenario 2.

Scenario 4: $b_1 = 3$, $b_2 = 5$, $\sigma_1 = 15$, $\sigma_2 = 10$.

This time we test the scenario in which agents meet less frequently when they walk: b_1 is reduced from 5 to 3. Not surprisingly, it also has a positive effect in comparison to Scenario 1: 99% of agents are infected after $t_f = 217$ iterations. However, as expected, it is worse than Scenario 3—the base contagion probability for the subway is twice the corresponding probability for walking and so reducing interactions on the subway has a larger impact.

A comparison of all scenarios can be found on Figure 3. The fraction of agents infected after t iterations is presented on the left and new infections are presented on the right.

Scenario 5: $b_1 = 5$, $b_2 = 5$, $\sigma_1 = 15$, $1 \leqslant \sigma_2 \leqslant 50$.

Finally, we investigate the process for various values of σ_2, namely, we consider $\sigma_2 \in [50]$. When $\sigma_2 = 50$ (the subway is more than 3 times slower than walking), most agents walk, often interacting with each other and infecting themselves quickly. Similarly, when $\sigma_2 = 1$ (the subway is 15 times faster than walking), most agents take subway infecting one another even faster (recall that the base contagion probability for the subway is twice that of walking). In Figure 4, we present the number of iterations needed to infect 99% of agents and the fraction of agents infected after 100 iterations, both as a function of σ_2. The "sweet spot" turns out to be when $\sigma_2 = 25$: it takes 278 iterations to infect 99% of agents and after 100 iterations only 21.21% of agents are infected.

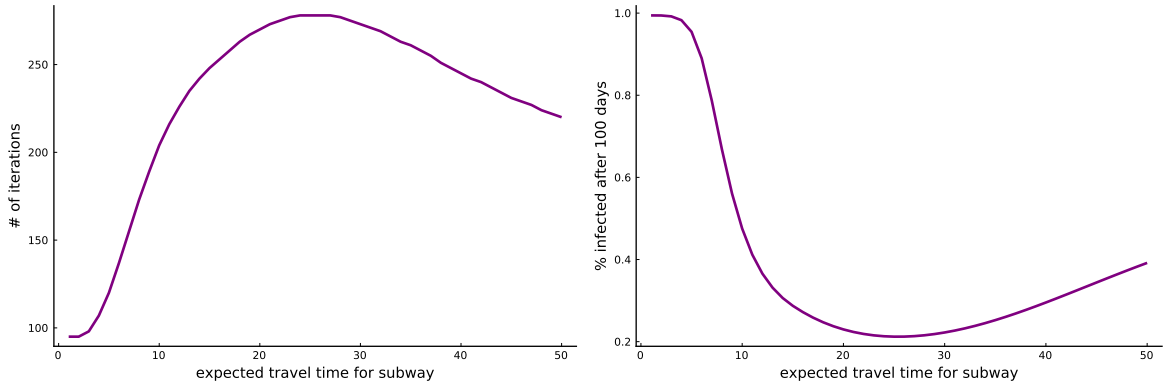

Figure 4: Scenario 5: the number of iterations needed to infect 99% of agents (left) and the fraction of agents infected after 100 iterations (right).

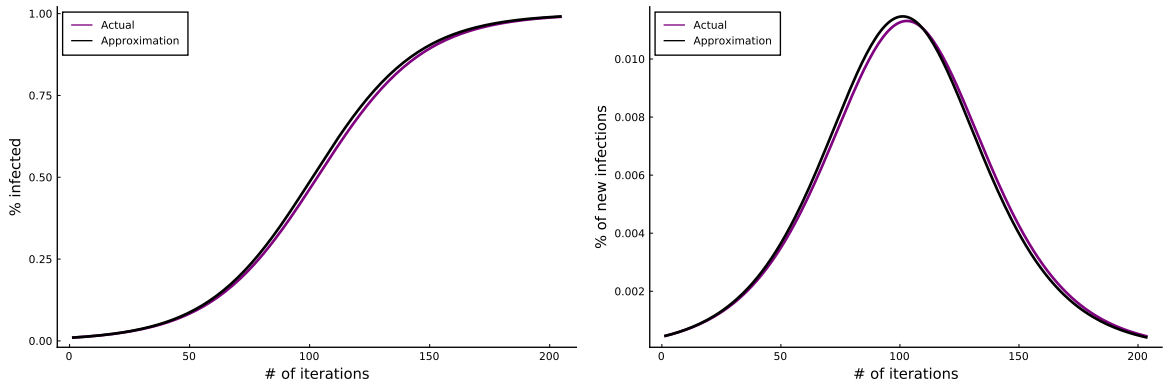

Figure 5: The fraction of agents infected (left), and the fraction of new infections (right): actual vs. approximated values.

3.4 Approximation

In this section, we derive an approximate but closed formula for the number of infected agents after a given number of iterations. First, let us note that $\exp(-x) = 1 - x + O(x^2)$, and so we may approximate (3) as follows:

$$\begin{aligned}
\widetilde{x}_{t+1} - \widetilde{x}_t &= (1-\widetilde{x}_t)\left(\alpha_1(1-\exp\left(-b_1\lambda_1\alpha_1\widetilde{x}_t\right)) + \alpha_2(1-\exp\left(-b_2\lambda_2\alpha_2\widetilde{x}_t\right))\right) \\
&\approx (1-\widetilde{x}_t)\widetilde{x}_t\left(b_1\lambda_1\alpha_1^2 + b_2\lambda_2\alpha_2^2\right) \\
&= A(1-\widetilde{x}_t)\widetilde{x}_t, \qquad \text{where } A := b_1\lambda_1\alpha_1^2 + b_2\lambda_2\alpha_2^2.
\end{aligned}$$

On Figure 5 (right), we compare this approximation with the actual values for Scenario 1. Since the base contagion probabilities are typically very small, the exponent is also small and so the approximation $\exp(-x) \approx 1 - x$ is relatively good.

Next, we approximate the difference equation by the differential equation:

$$x'(t) = A(1-x(t))x(t)$$

with the initial condition $x(0) = c_s$. We get that

$$\widetilde{x}_t \approx x(t) := \frac{1}{1 + \exp(-At)(1/c_s - 1)}.$$

As expected, this approximation is working well as depicted on Figure 5 (left). By solving $x(t) = c_f$, it follows that

$$t_f \approx \frac{1}{A} \ln\left(\frac{c_f}{1 - c_f} \cdot \frac{1 - c_s}{c_s}\right).$$

4 Numerical Experiments

4.1 Model Details

To conduct our experiments, we use a multi-agent discrete event simulation model, which we implemented in `Julia`. We selected this programming language because of its performance, and its built-in simulation and distributed computing capabilities. The general logic of the model is that agents are randomly placed in downtown Toronto. For this reason, the nomenclature in this paper derives from the city's transportation system. Namely, the term TTC refers to the Toronto Transit Commission. The simulation can be repeated for any other city since the data available on OpenStreet map project is easily available and the framework is flexible.

Each agent has their own work and home locations which they periodically travel between. Specifically, at the beginning of the process, each agent selects six static routes, three for each direction of their commute. Their routes involve the streets the agents may walk on, as well as the public transport available. Before each trip, each agent independently and uniformly at random chooses between one of their three routes in the appropriate direction (i.e., home to work, or work to home). When an agent arrives at their work location, they work an average of 8 hours (thus remaining put), before returning home, at which point they again rest for approximately 16 hours. To begin the process, 1% of the population is randomly chosen to be infected. We refer to each of these agents as a "patient zero." The agents then begin moving, and infect each other with some probability p_0 outdoors and a much higher probability indoors. The simulation stops when 95% of the agents have become infected.

Symbols listed in Table 2 facilitate describing the simulation model in the following sections.

K	population of agents
k	commuting agent, $k \in K$
$\|K\|$	the total number of agents (cardinality of K)
V	set of intersections and public transportation stations (vertices)
v	single intersection or public transportation station, $v \in V$ (vertex)
E	set of road/sidewalk and public transportation routes sections (directed edges)
e	single road/sidewalk, $e \in E$ (directed edge); an edge e can be also represented as an ordered pair of vertices that is we can write $e = (v_i, v_j)$ where $i \neq j$
V_P	set of intersections (vertices) accessible to pedestrians, $V_P \subseteq V$
V_T	set of stations and stops (vertices) used by public transportation system, $V_T \subseteq V$; we assume that $V_P \cup V_T = V$ and $V_P \cap V_T = \varnothing$, that is, (V_P, V_T) is a partition of V
$G(V, E)$	urban network (directed, weighted, strongly connected graph) with weights representing the time needed to traverse an edge;
W	set of public transportation system (or: TTC) cars
w	public transportation (or: TTC) car, $w \in W$
$\|W\|$	the total number of public transportation cars (cardinality of W)
$w(v)$	public transportation car available at the node $v \in V$ (it is assumed that a node can be served only by a single transportation car, the exchange stations are assuming that adjacent platforms connected by a sidewalk)
s_k	route taken by an agent k, $s_k = (e_k^{(1)}, \ldots, e_k^{(n_k)})$
t_e	weight of an edge $e \in E$, measured in time needed to traverse e (in seconds)
$\widehat{t_k}$	estimated time to traverse the route of the agent k
$\bar{t_k}$	actual time to traverse the route of the agent k
$\hat{d}_k^{(w)}$	estimated transportation wait time of the agent k for the car w
$\phi^{(w)}$	number of infected agents in a car $w \in W$
$\psi^{(v)}$	number of infected pedestrians at the vertex $v \in V_P$
p_0	base probability of infection
p	effective probability of infection

Table 2: List of symbols used in the paper

4.1.1 Types of Agents

There are two types of **agents** in the model: **people (commuters)** and **public transport (TTC) cars**. Each agent type is characterized by a set of **attributes**, each of which is categorized as fixed or dynamic. A **fixed attribute** is assigned to an agent at the very beginning of the simulation and remains unchanged. In contrast, a **dynamic attribute** changes throughout the course of the simulation. We also indicate in the supporting tables whether each attribute is a value or a set of values.

Each person (or: agent, commuter) has the set of characteristics as listed in Table 3.

Attribute	Type	Class
ID	fixed	value
home_location	fixed	value
work_location	fixed	value
current_location	dynamic	value
direction	dynamic	value
infection_status	dynamic	value
walking_speed	fixed	value
routes	fixed	set
TTC_car_ID	dynamic	value
passengers_met_in_TTC	dynamic	set
number_of_trips	dynamic	value

Table 3: Attributes of the commuter class

ID is a unique identifier of a person. home_location and work_location are two randomly selected vertices from a city graph. current_location is a vertex pointing to the current position of a person. direction indicates whether a person is going from home to work or vice versa. infection_status helps understand if a person has already been infected, and if so, whether it occurred on the street or inside a public transportation vehicle. walking_speed describes how many meters per second a person can go on foot. For simplicity, this parameter is the same for all people but it can be made more realistic if more data is injected to the model (such as distribution of ages of the agent). Based on [30] the value for this parameter was set at 1.25 m/s, which is exactly the walking speed observed in Japan during the COVID-19 pandemic in 2020. routes are lists of graph edges an agent passes through before reaching their target location. As mentioned earlier, each person has three favourite routes in each direction, which we discuss in detail in Subsection 4.1.2. TTC_car_ID is assigned to a person once they are on-board a specific public transport vehicle. passengers_met_in_TTC is the set of individuals an agent met on their ongoing ride in the same TTC vehicle, i.e., it resets to the empty set once a TTC ride ends, and potentially starts filling up again during the next one. If at the end of a commuter's TTC trip there are infected people within the set, then the new infection_status is calculated and updated. By doing this in this way we ensure that an agent cannot be infected during the trip, which means that they cannot infect others during the same commute. All TTC infections are resolved when agents leave a TTC car. number_of_trips counts the number of one-direction trips a commuter makes during a simulation.

Each TTC car has the attributes presented in Table 4.

Attribute	Type	Class
`ID`	fixed	value
`first_station`	fixed	value
`last_station`	fixed	value
`current_station`	dynamic	value
`line`	fixed	value
`route`	fixed	set
`passengers`	dynamic	set
`passengers_limit`	fixed	value
`arrival_time_interval`	fixed	value
`one_trip_max_passengers`	dynamic	value

Table 4: Attributes of the TTC car class

`ID` is a unique identifier of a TTC car. `first_station` and `last_station` are the first and the last station of a specific public transport line. `current_station` is the current position of a vehicle. `line` is a unique identifier of each direction of a public transport line, i.e., the attribute `line` in the simulation is a combination of a line and its direction. `route` of a subway or streetcar vehicle is presented as a sequence of stations a TTC car passes through (the same for all vehicles with the same `line` attribute). As was already mentioned, in our implementation we artificially add the directions for every line (e.g. Line 1(A->B) Line 1(B->A)) for implementation purposes, so each direction has its unique route. `passengers` is a set of people who are currently inside the TTC car. `passengers_limit` is the maximum number of passengers which can fit in one vehicle at any point in time. `arrival_time_interval` is the average number of minutes that pass between two TTC cars on the line. `one_trip_max_passengers` is the maximum number of passengers who were inside a TTC car at one point in time. This parameter is helpful to measure whether a line is overloaded. It is worth noting that a TTC car disappears after reaching the last station. Nevertheless, it calls a new one on the same line, which keeps the efficiency of the public transport system at the same level because the arrival time interval does not change. It has its real-world interpretation, i.e., that every car is disinfected at the last station.

4.1.2 Routing

A route of agent k can be represented in terms of n_k consecutively adjacent edges:

$$s_k = (e_k^{(1)}, \ldots, e_k^{(n_k)}) \tag{4}$$

Notably, we assume that the commuting agents do *not* know precisely how long a route which includes public transportation will take. They instead make assumptions about the waiting time of the TTC cars and take this extra time into account when determining their routes. The estimated time needed to traverse the route of the k^{th} commuter is thus split into two sums, where the latter aggregates over the expected waiting times of the TTC cars:

$$\widehat{t_k} = \sum_{e \in s_k} t_e + \sum_{\substack{(v_1,v_2) \in s_k: \\ (v_1,v_2) \in V_P \times V_T}} \hat{d}_k^{(w(v_2))} \tag{5}$$

where all symbols are in accordance with Table 2. Using the A* search algorithm, which is often used in many fields of computer science due to its completeness, optimality, and optimal efficiency [31],

we find the top three shortest routes in each direction for agent k with respect to the edge weights of (5). Thus, each agent has six different paths representing routes. Note that the factual travel time of a route will in most cases differ from its estimated traversal time.

Each time an agent goes from home to work or vice versa, they draw one of their three relevant routes uniformly at random. We introduce this randomness, as we work under the assumption that the commuters may not always wish to take the fastest route available, due to a variety of reasons.

4.1.3 Infection Scheme

In our model, the commuters can get infected with a virus similar to COVID-19. Specifically, at any point in the process, there are a number of infected commuters, each of which has a chance to infect an uninfected agent when an interaction occurs. An interaction between two commuters can occur in two different ways: during the wait at an intersection or by sharing public transportation. We assume that once an agent becomes infected, they remain so for the duration of the process. Moreover, the infections occur independently, i.e., the joint probability of non-infection over a number of encounters is the product of the events individual probabilities. The dynamics of the infection process is a consequence of the implementation choices, as documented in the open-source code available at the project's GitHub repository. In summary this is the result of the chosen Discrete Event Simulation framework.

As found in [32], the odds of indoor transmission of COVID-19 was almost 19 times higher compared to outdoors. Based on this publication, in our study, we assume a base probability of infection p_0, which is applied to encounters at intersections, while for the TTC, it is 19 times of that value. Note that either of these values can be adjusted easily in the `Julia` implementation of the framework. As a result, the probability Q_{k_1,k_2} that an infected agent k_1 infects an uninfected agent k_2 during a particular encounter can be summarized as follows:

$$Q_{k_1,k_2} = \begin{cases} p_0 & \text{if } k_1 \text{ and } k_2 \text{ meet at an intersection,} \\ 19 \cdot p_0 & \text{if } k_1 \text{ and } k_2 \text{ meet on a TTC car.} \end{cases} \tag{6}$$

At each stage of the trip of agent k, the probability depends on the number of infected individuals that were encountered, and the location of the encounter. Let us assume that k is uninfected when it meets $\psi^{(v)}$ infected agents at intersection v, or $\phi^{(w)}$ infected agents while on a TTC car w. In this case, if Q_{total} denotes the overall probability that k is infected during the current stage of their trip, then

$$Q_{\text{total}} = \begin{cases} 1-(1-p_0)^{\psi^{(v)}} & \text{if } k \text{ meets } \psi^{(v)} \text{ infected agents at } v \in V_P \\ 1-(1-19\cdot p_0)^{\psi^{(w)}} & \text{if } k \text{ meets } \psi^{(w)} \text{ infected agents while on } w \in W. \end{cases} \tag{7}$$

Also it is important to remember that all TTC infections are resolved when agents leave a TTC car, which means that an agent can be infected during a trip, but cannot spread the infection during the same trip (see also [33]).

4.1.4 Simulation Dynamic

We explain the simulation dynamics using a separate action diagram for each agent type. These diagrams allow us to explain the procedure of actions taken by the various agents throughout the course of one exemplary day. We first present the diagram for the commuters (Figure 6):

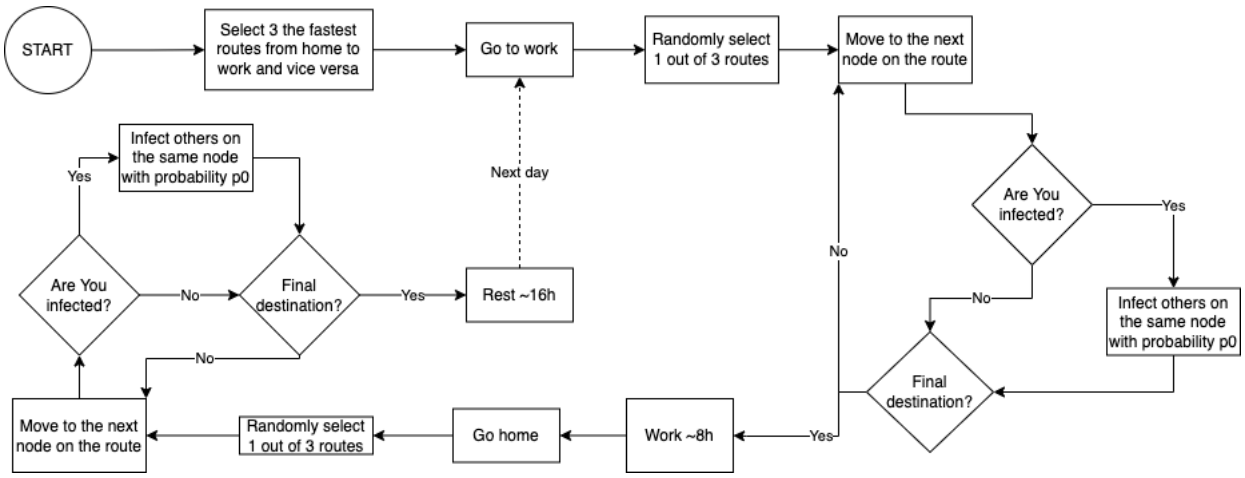

Figure 6: An example of one day simulation dynamic for an agent

To summarize, the commuter wakes up at home, and then selects their route to work. While traversing their route, they may interact with people who are infected, which could cause themselves to become infected. After work, they select a route back to their home, and then rest for a fixed amount of time. Below is the action diagram for the TTC cars (Figure 7):

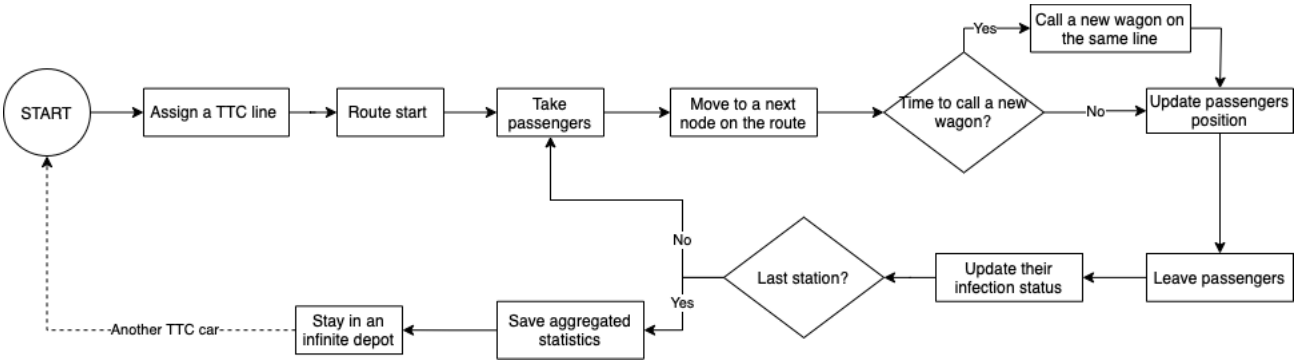

Figure 7: An example of one day simulation dynamic for a TTC car

A TTC car is fixed to a single route upon its creation. While traversing its route, it stops at each vertex to drop off and pick up passengers (commuters). When a passenger leaves the subway or streetcar vehicle, it updates their infection status based on the number of infected co-passengers they interacted with. When the TTC car reaches the last station, it drops off all its passengers, computes the relevant statistics gathered on the way (e.g., maximum number of passengers) and then stays in a depot for the remainder of the process. Additionally, using its arrival time interval attribute, it calls a new car on the same route at the same starting location. The reason that this is done in this way is that it is convenient for implementation purposes.

4.2 Simulation Setup

We evaluated the simulation model on a simplified representation of downtown Toronto and its transportation system, i.e., TTC (Toronto Transit Commission). Figure 8 presents the exact ele-

ments of the city utilized in the experiment. The simulation map includes the streets of the city, as well as the TTC lines of streetcars (marked in green in Figure 8) and subway (marked in red). In the simulation, we track the agents only as they reach the downtown area of the city. Therefore,

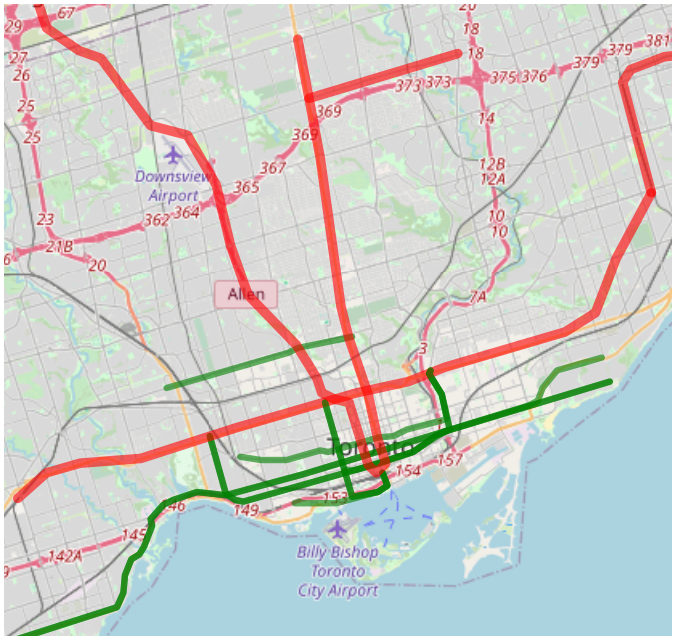

Figure 8: Map of Toronto used in the experiment

we assume that the commuters from the suburban areas of Toronto are present in the model, but we only start tracking their movement once they reach the central part of Toronto. The map was downloaded from OpenStreetMap project[2]. From the perspective of the simulation, the most crucial elements of the map are the streets, sidewalks and intersections. The types of roads that were taken into account are 1) primary, 2) secondary, 3) tertiary, 4) unclassified, and 5) residential.

After preparing and filtering the map, we represented it as a strongly connected (directed) graph. That is, there is a vertex (directed) path between any two of its vertices. (For more details see, for example, [34] or any other textbook on graph theory.) This avoids a situation where agents are randomly placed in a location where they can never leave.

We later modified the map by adding public transportation nodes and vertices to the resulting graph. The two networks were connected via those walking vertices that were located nearby TTC stations. The real-world data regarding public transport schedule and routes in Toronto was downloaded from City of Toronto's Open Data Portal[3]. The final representation of the whole system is a graph, where vertices are stops and intersections, edges are routes between stops and intersections, and weights are the number of seconds between two different graph vertices. It is worth noting that each TTC line and its direction has a unique set of edges and vertices. Next, the agents were placed on the graph and the simulation was run. The simulation hyper-parameters can be found in Table 5.

[2]https://www.openstreetmap.org/
[3]https://open.toronto.ca/dataset/ttc-routes-and-schedules/

Parameter	Value
`number_of_agents`	2 000
`number_of_patients_zero`	20
p_0	0.001
`walking_speed`	1.25m/s
`passengers_limit`	10 people
`arrival_time_interval`	from 1 to 25 minutes by 1 minute
`number_of_simulations`	100
$\hat{d}_k^{(w)}$	half of a TTC interval (frequency)

Table 5: Simulation hyperparameters

The simulation was written in `Julia` language whose distributed computing capabilities enabled an integration with AWS S3 and performing all necessary computations in the cloud. The cluster consisted of 320 computing cores. The results are presented in the next section.

4.3 Experiment Results

The results section is divided into three parts. We begin with the results of a simplified model that is based on the mathematical toy model presented in Section 3. We use this simplified model as an intermediate step in understanding the complexity of the full model. The second part of the section deals with the validation of the full simulation model. Finally, we explain the results of the experiments, and discuss the primary outcome of our study.

4.3.1 Simplified Model Results

We first make a number of key simplifications and alterations in the simulation model. The goal of these changes is to ensure that the simplified model resembles behaviour of the mathematical toy model from Section 3.

1. A simplified graph, as presented in Figure 9.
2. All agents have the same home and work location and leave their homes/works simultaneously.
3. Assigning probabilities of choosing a route (entirely walking the whole way or utilizing one TTC edge) is based on an exponential scaling function as in the case of the mathematical toy model.
4. Small population of agents due to the small graph.

The smaller graph and homogeneous home and work locations result in a model very similar to the mathematical toy model, as the agents are forced to travel in large groups and reach consecutive nodes simultaneously. This contrasts the full simulation where the paths taken by the agents are diversified. In terms of the routing decisions, recall that each agent selects of one its three optimal routes uniformly at random in the full simulation. In contrast, in the simplified model the agents each have the same two routes, one of which involves the TTC, and thus has a travel time which depends on the frequency parameter. Thus, if we were to assume that the agents always choose

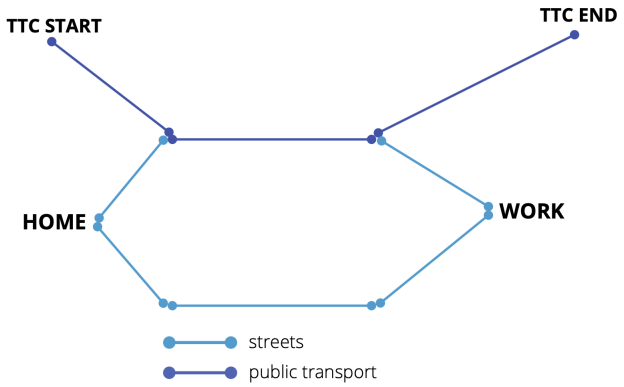

Figure 9: Simplified graph used in toy simulation model

their optimal route, then increasing the frequency sufficiently high would force all the commuters to forgo taking the TTC. Instead, we use the exponential weighting as in the toy model to determine route selection probabilities.

Let us explain how the frequency of running public transportation impacts the pace of the epidemic in the simplified model. Firstly, we observe that due to the routing scheme of the model, the more frequent the trains are, the more people there are that want to ride them. Conversely, by increasing the interval between public transportation trains, the TTC becomes less attractive, and so more commuters will choose to walk, thus crowding at intersections. As we can see in Figure 10, the curve presenting the percentage of infected agents after a fixed time frame is U-shaped with a visible "sweet-spot" (the optimal lowest number of infected commuters) at 4 minutes.

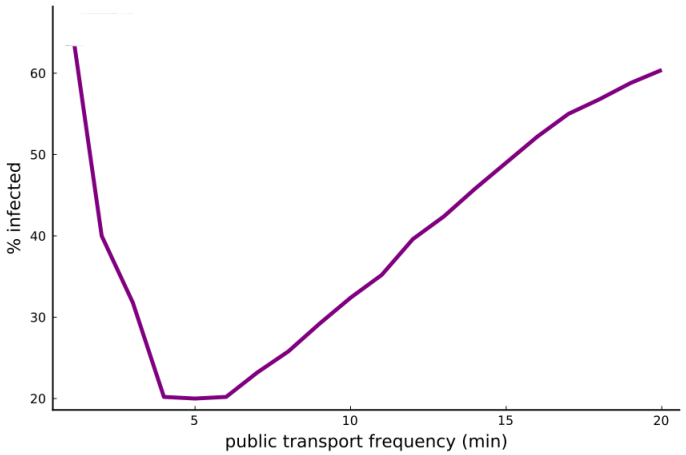

Figure 10: Infected agents at a fixed iteration depending on TTC frequency

In order to understand this "sweet spot", it is helpful to consider Figure 11. Observe that when the TTC arrives very frequently, it is extremely attractive for the people to take, and so a lot of commuters are infected on their way to the station and inside the TTC. When the frequency increases to 4 minutes, the people are appropriately spread amongst the routes to minimize the average probability of infection, and so the described "sweet spot" occurs. Note that by increasing

the frequency past the "sweet spot", public transport runs rarely so those who choose the TTC get infected easily because of overcrowded public transport. The decrease in the number of TTC users does not compensate for the less frequent TTC.

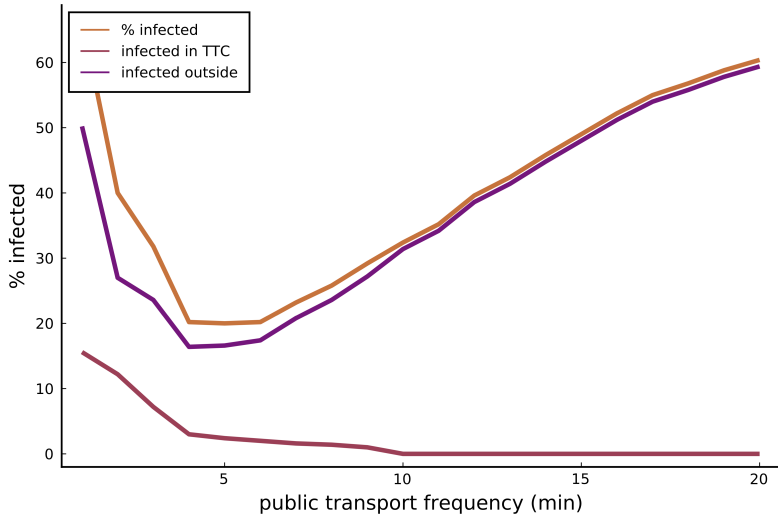

Figure 11: Infected agents in different places of infection at a fixed iteration depending on TTC frequency

We also checked how many commutes it takes to infect 95% of the simplified model's population. As observed in Figure 12, this number is lower for very frequent public transport and higher when the TTC is rare. This is the case because public transport is extremely popular when the trains almost immediately arrive at the station. Since the probability of infection in public transport is significantly higher than on the streets, many people use the TTC and quickly get infected. On the other hand, when public transport arrives infrequently, many people choose the walking path instead of the TTC, and so it takes longer (in terms of number of commutes) to infect each other. Nevertheless, it takes the most commutes to infect 95% of the population when people are spread between public transport and streets in a way that neither of them is overloaded. Thus, the same conclusion as previously can be drawn from the Figure 12 as the maximum number of commutes is achieved when public transport frequency is 4 minutes.

4.3.2 Full Model Validation

After the experiment was conducted, we verified the elementary properties of the model. We have examined the model's take on epidemic curves and increased in population while keeping the area they move around intact (hence increasing population density or propensity to leave their homes).

First of all, the model's infection curve is a classic s-shaped curve as in Figure 13, for every tested parametrization. The figure presents the outcome in the scenario where all hyper-parameters are set as in Table 5 and public transportation runs every 20 minutes.

The other model property we have found is that the more agents are on the map, the faster the epidemic develops (see Figure 14). While for 300 agents, the epidemic has not had enough time to develop till the end of the presented period, all three phases of an epidemic were completed for the population of 5000 agents. This outcome agrees with [35], [36], and the #stayathome strategies promoted all around the globe.

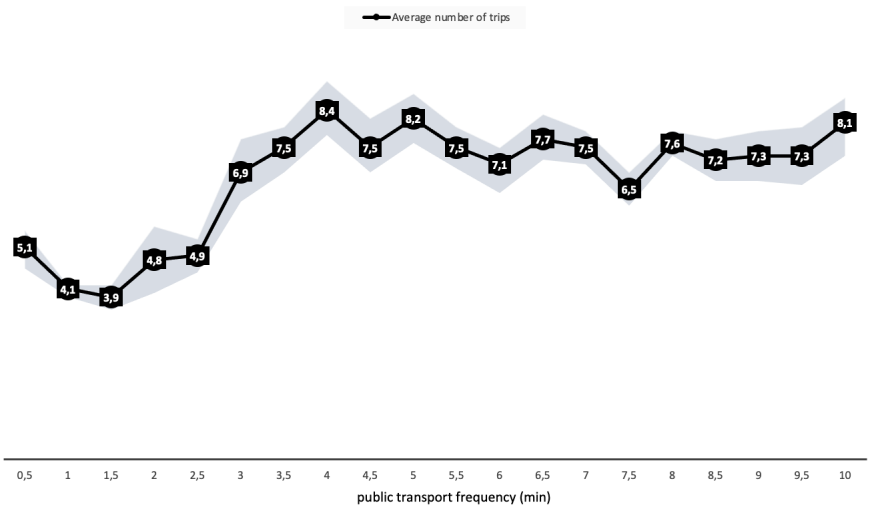

Figure 12: Number of trips on the route from home to work and vice versa depending on TTC frequency

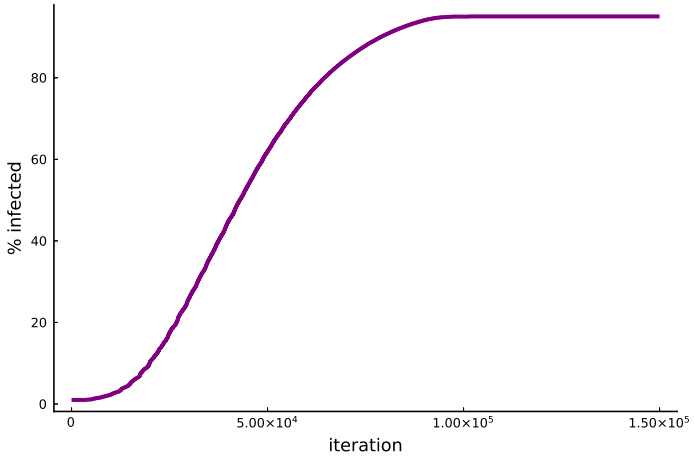

Figure 13: Exemplary epidemic curve in the simulation model

4.3.3 Full Model Results

In the full simulation model we also found that the frequency of running public transportation has an impact on the pace of the epidemic, however, the further conclusions differ significantly.

As in the previous versions of the model, we observed that due to the routing scheme, the less frequent the trains are, the least people want to ride them. As seen in Figure 15, the percentage of agents that used public transportation at least once in their home-to-work or work-to-home routes declines as the gap between subsequent trains increases. The more rare the trains are, the more attractive walking becomes in comparison. While the TTC users number drops, the curve presenting the percentage of infected agents in a fixed time frame is inverted-U-shaped with a visible peak at 4 minutes and more than 90% of infected population. This is a curve that is widely different than the ones obtained in the previous sections of the study.

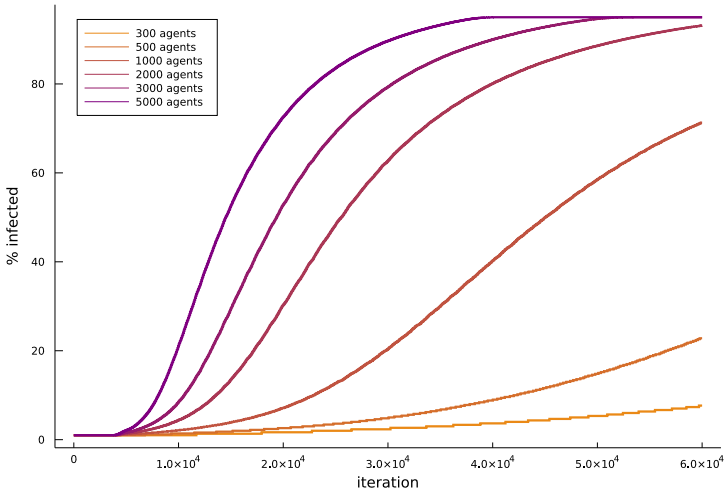

Figure 14: Population density vs pace of infection

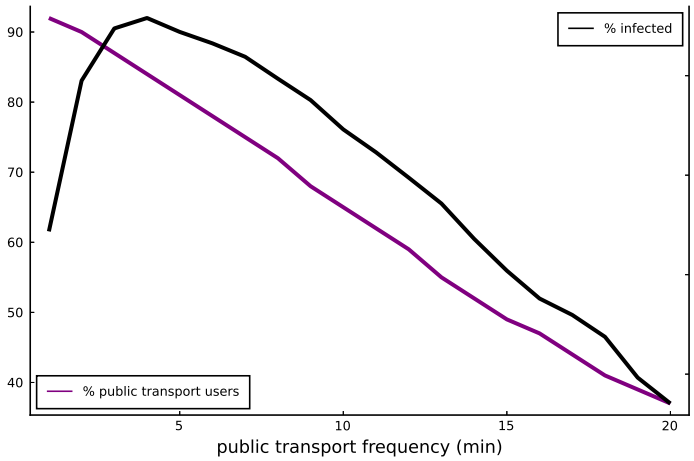

Figure 15: Percentage of TTC users and infected agents at a fixed iteration depending on TTC frequency

Moreover, those results are confirmed looking at the Figure 16, which shows how many commutes it takes to infect 95% of the population. The point when pace of infection is the fastest is around 4-5 minutes, which is consistent with the previously described results. The number of commutes divided by 2 could be interpreted as number of working days, because it represents a trip: home->work->home. Thus, at the most "dangerous" point it takes around 31 working days (63.5/2) to infect 95% of the population, while at "the safest" point, when public transport arrives to the station every minute, it takes around 43 working days (86.5/2), which is approximately 40% longer comparing to the worst scenario.

In Figure 17 you can see the source of the infection peak. The chart shows the location of infection for every infected agent and the total percentage of infected agents at every TTC frequency. It turns out that while the infections on the city streets are relatively flat due to 1) low infection probability, 2) sparse agents concentration outside, i.e., less people interactions at intersections, the

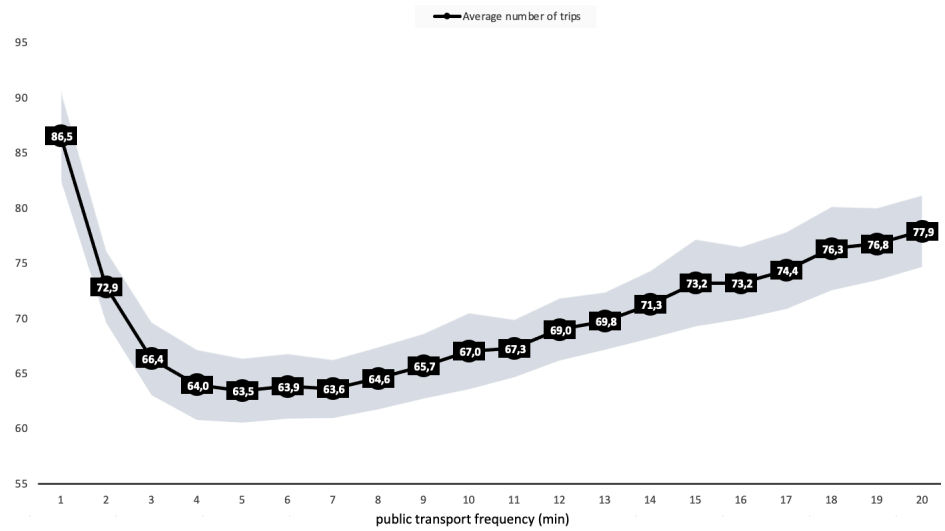

Figure 16: Number of trips on the route from home to work and vice versa depending on TTC frequency

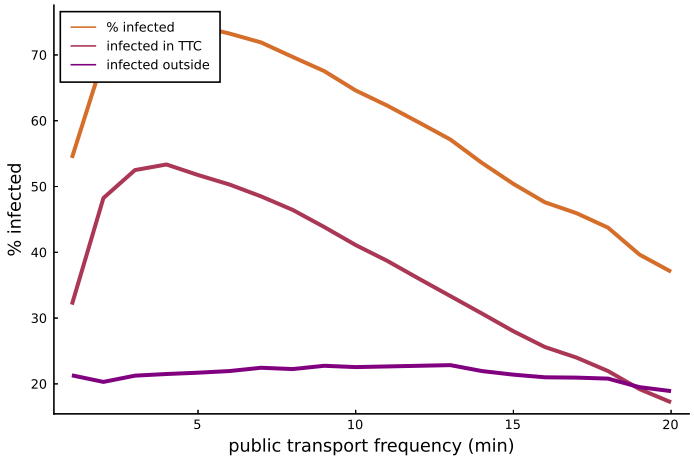

Figure 17: Places of infection

peak in total infections is perfectly aligned with a peak of infections in TTC.

The main reason behind the result is captured in Figure 18. The barplot presents the number of encounters in TTC by 1) TTC frequency (OX axis), 2) number of infected agents that participated in the interaction (bar colours). The numbers on the blocks refer to the number of encounters of given crowd size and with a given TTC frequency.

Firstly, the number of meetings decreases as trains become less frequent and fewer agents use public transportation. However, mind how the structure of the encounters changes—the less frequent the trains are, the lower the percentage and number of interactions in small groups. It is the case for all crowd sizes smaller than 7. Simultaneously, crowd sizes of 7-11 agents and 12 and more are non-monotonous. Their share is relatively small in the left part of the graph due to the frequently running trains—even though many agents come to stations to use the public transportation, the

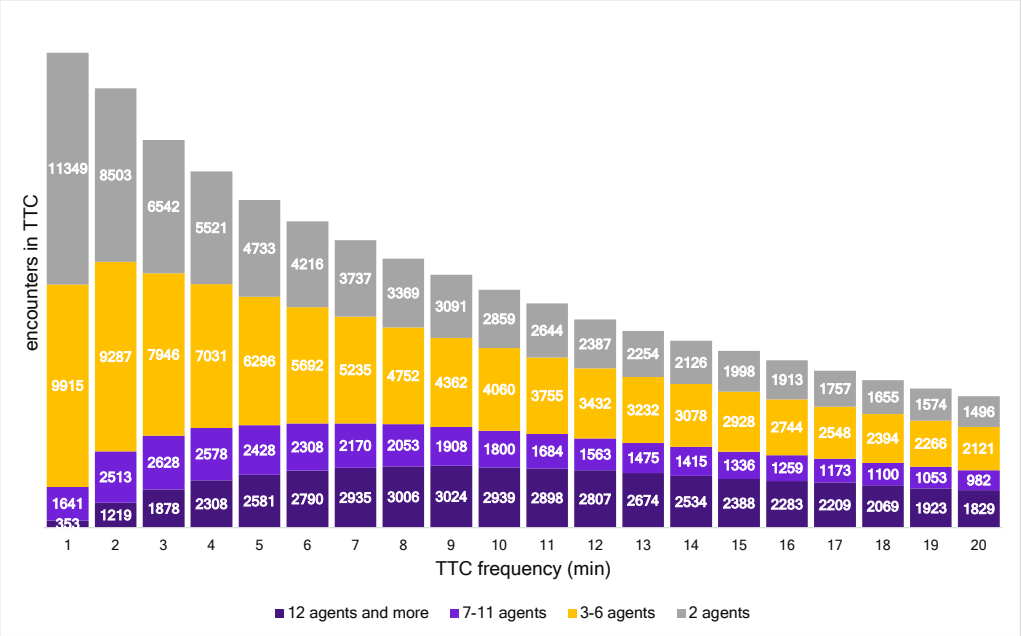

Figure 18: Number and structure of encounters in terms of crowd size (each bar colour representing number of agents taking part in one encounter)

train comes quickly enough to prevent them from forming a large crowd. As the gap between subsequent trains rises, TTC loses its capability to unload all passengers in time. Therefore, people crowd at stations and in trains leading to high infection rates. In the case of the infrequent trains, the number of people willing to ride TTC is so low that the high share of high-crowd encounters does not result in a high infection pace. These forces lead to an infection peak in a spot, where 1) most agents are willing to take TTC due to its relatively high frequency, 2) TTC is not effective enough in separating passengers between trains.

The reason why this mechanism occurs is the time mismatch between train and agents arrivals to stations. While some agents just missed the train, the other ones are coming and gradually form a considerable crowd if the train does not come soon enough. In contrast, if they walked in hordes and reached the station simultaneously as they do in the toy version of the model, the crowd size would always be the percentage of the population that choose to take the TTC route. Mind also that the map size in the toy model is unrealistically tiny, which is why agents can form large crowds in the street. Hence, the model would suggest a severe danger of walking and is not realistic.

Those findings could also be confirmed by looking at the number of visits of individual nodes on the map. In Figure 19 the size of a circle around the point represents the number of visits. When public transport is very frequent (TTC frequency = 3 minutes), TTC nodes are much more crowded than the other, which leads to the increase in the number of infections as many people meet each other in a closed space like a TTC car. On the other hand, when public transport is sporadic (TTC frequency = 20 min), there is no specific outbreak area. In this case, TTC is not attractive for people travelling relatively small distances, so they are more willing to walk. Thus, they are more evenly distributed on the map, leading to fewer potential "dangerous" meetings.

Figure 19: Individual nodes visits depending on TTC frequency on a small fragment of the map

5 Conclusions

This paper shows that public transportation schedules can be optimized in terms of epidemic safety. However, one has to be cautious of the assumptions they are making. Namely, we have presented two vastly different sets of results obtained under different assumptions and conditions. The more intricate model resulted in an infection peak instead of a sweet spot which shows that the optimization in the area cannot be optimized strictly under only one criterion. This conclusion agrees with the reality—people would not be infecting each other if they did not meet anywhere, especially in closed spaces such as a train station and a train car. Nevertheless, such a solution would be highly impractical, which is why it is crucial for a regulator to choose a train frequency that is 1) practical for the citizens, and 2) is not too expensive to maintain, i.e., choose a frequency on the right side of the infection peak.

The presented model has been calibrated with sample data from Toronto Public Transportation. Note that for simulations we have selected only a subset of the TTC system to illustrate the emergent phenomena. However, the proposed approach can be extended onto a massive transportation model of the entire city. We believe that our results can help decision maker to understand trade-offs when deciding between various frequencies for public transport in times of an ongoing epidemic.

Acknowledgments

The research was funded by the NSERC Alliance project with Security Compass entitled "COVID-19: Agent-based framework for modelling pandemics in urban environment", and complemented with SOSCIP COVID-19 Response Program.

References

[1] Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The lancet*, 395(10223):497–506, 2020.

[2] Petrônio CL Silva, Paulo VC Batista, Hélder S Lima, Marcos A Alves, Frederico G Guimarães, and Rodrigo CP Silva. Covid-abs: An agent-based model of covid-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos, Solitons & Fractals*, 139:110088, 2020.

[3] Christine SM Currie, John W Fowler, Kathy Kotiadis, Thomas Monks, Bhakti Stephan Onggo, Duncan A Robertson, and Antuela A Tako. How simulation modelling can help reduce the impact of covid-19. *Journal of Simulation*, 14(2):83–97, 2020.

[4] Cliff C Kerr, Robyn M Stuart, Dina Mistry, Romesh G Abeysuriya, Katherine Rosenfeld, Gregory R Hart, Rafael C Núñez, Jamie A Cohen, Prashanth Selvaraj, Brittany Hagedorn, et al. Covasim: an agent-based model of covid-19 dynamics and interventions. *PLOS Computational Biology*, 17(7):e1009149, 2021.

[5] Robert Hinch, William JM Probert, Anel Nurtay, Michelle Kendall, Chris Wymatt, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, et al. Openabm-covid19-an agent-based model for non-pharmaceutical interventions against covid-19 including contact tracing. *medRxiv*, 2020.

[6] Speir Connor and Negahban Ashkan. Analyzing covid-19 control strategies in metropolitan areas: A customizable agent-based simulation tool. In *2020 Winter Simulation Conference (WSC)*, 2020.

[7] Nicolas Hoertel, Martin Blachier, Carlos Blanco, Mark Olfson, Marc Massetti, Frederic Limosin, and Henri Leleu. Facing the covid-19 epidemic in nyc: a stochastic agent-based model of various intervention strategies. *MedRxiv*, 2020.

[8] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017.

[9] E. Frias-Martinez, G. Williamson, and V. Frias-Martinez. An agent-based model of epidemic spread using human mobility and social network information. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 57–64, 2011.

[10] Dionne M Aleman, Theodorus G Wibisono, and Brian Schwartz. A nonhomogeneous agent-based simulation approach to modeling the spread of disease in a pandemic outbreak. *Interfaces*, 41(3):301–315, 2011.

[11] Jürgen Hackl and Thibaut Dubernet. Epidemic spreading in urban areas using agent-based transportation models. *Future internet*, 11(4):92, 2019.

[12] Marek Laskowski, Bryan CP Demianyk, Julia Witt, Shamir N Mukhi, Marcia R Friesen, and Robert D McLeod. Agent-based modeling of the spread of influenza-like illness in an emergency department: a simulation study. *IEEE Transactions on Information Technology in Biomedicine*, 15(6):877–889, 2011.

[13] Oliver M. Cliff, Nathan Harding, Mahendra Piraveenan, E. Yagmur Erten, Manoj Gambhir, and Mikhail Prokopenko. Investigating spatiotemporal dynamics and synchrony of influenza epidemics in australia: An agent-based modelling approach. *Simulation Modelling Practice and Theory*, 87:412–431, 2018.

[14] Regnier Eva, M. Sanchez Susan, and J. Sanchez Paul. Testing-based interventions for covid pandemic policies. In *2020 Winter Simulation Conference (WSC)*, 2020.

[15] Legard Yusuke, Ivey Nathan, Rodriguez Antonio, and Wolski Joseph. Utilizing simulation to evaluate shuttle bus performance under passenger counts impacted by covid-19. In *2020 Winter Simulation Conference (WSC)*, 2020.

[16] Tang Shi, Rojas-Cordova Alba, McDonald Samuel, Furmaga Jakub, Piel Carl, Courtney Mark, and Diercks Deborah. Data-driven staffing decision-making at a large emergency department in response to covid-19. In *2020 Winter Simulation Conference (WSC)*, 2020.

[17] VanDeusen Adam, Liao Che-Yi, Venkat Advaidh, Cohn Amy, Kurlander Jacob, and Saini Sameer. Evaluating patient triage strategies for non-emergency outpatient procedures under reduced capacity due to the covid-19 pandemic. In *2020 Winter Simulation Conference (WSC)*, 2020.

[18] Haghpanah Fardad, Ghobadi Kimia, and W. Schafer Benjamin. How to evacuate an emergency department during pandemics: A covid-19 agent-based model. In *2020 Winter Simulation Conference (WSC)*, 2020.

[19] Nicolas Hoertel, Martin Blachier, Carlos Blanco, Mark Olfson, Marc Massetti, Marina Sánchez Rico, Frédéric Limosin, and Henri Leleu. A stochastic agent-based model of the sars-cov-2 epidemic in france. *Nature Medicine*, 26(9):1417–1421, 2020.

[20] Agnieszka Truszkowska, Brandon Behring, Jalil Hasanyan, Lorenzo Zino, Sachit Butail, Emanuele Caroppo, Zhong-Ping Jiang, Alessandro Rizzo, and Maurizio Porfiri. High-resolution agent-based modeling of covid-19 spreading in a small town. *Advanced Theory and Simulations*, page 2000277, 2021.

[21] Erik Cuevas. An agent-based model to evaluate the covid-19 transmission risks in facilities. *Computers in Biology and Medicine*, 121:103827, 2020.

[22] Gudrun Wallentin, Dana Kaziyeva, and Eva Reibersdorfer-Adelsberger. Covid-19 intervention scenarios for a long-term disease management. *International Journal of Health Policy and Management*, 9(12):508–516, 2020.

[23] M. Macal Charles, Ozik Jonathan, T. Collier Nicholson, Kaligotla Chaitanya, M. MacDonell Margaret, Wang Cheng, J. LePoire David, Chang Youngsoo, and J. Martinez-Moyano Ignacio. Citycovid: A computer simulation of covid-19 spread in a large-urban area. In *2020 Winter Simulation Conference (WSC)*, 2020.

[24] Kaligotla Chaitanya, Stevens Abby, Mucenic Bogdan, Ozik Jonathan, T. Collier Nicholson, Whan Choe Kyoung, P. Rimer Sara, Hotton Anna, and M. Macal Charles. Development of large-scale synthetic population to simulate covid-19 transmission and response. In *2020 Winter Simulation Conference (WSC)*, 2020.

[25] Nick H Ogden, Aamir Fazil, Julien Arino, Philippe Berthiaume, David N Fisman, Amy L Greer, Antoinette Ludwig, Victoria Ng, Ashleigh R Tuite, Patricia Turgeon, et al. Artificial intelligence in public health: Modelling scenarios of the epidemic of covid-19 in canada. *Canada Communicable Disease Report*, 46(8):198, 2020.

[26] Cliff C Kerr, Robyn M Stuart, Dina Mistry, Romesh G Abeysuriya, Gregory Hart, Katherine Rosenfeld, Prashanth Selvaraj, Rafael C Nunez, Brittany Hagedorn, Lauren George, et al. Covasim: an agent-based model of covid-19 dynamics and interventions. *medRxiv*, 2020.

[27] Hiroyasu Inoue and Yasuyuki Todo. The propagation of economic impacts through supply chains: The case of a mega-city lockdown to prevent the spread of covid-19. *PloS one*, 15(9):e0239251, 2020.

[28] Frank Dignum, Virginia Dignum, Paul Davidsson, Amineh Ghorbani, Mijke van der Hurk, Maarten Jensen, Christian Kammler, Fabian Lorig, Luis Gustavo Ludescher, Alexander Melchior, et al. Analysing the combined health, social and economic impacts of the corovanvirus pandemic using agent-based social simulation. *Minds and Machines*, 30(2):177–194, 2020.

[29] Takeshi Kano, Kotaro Yasui, Taishi Mikami, Munehiro Asally, and Akio Ishiguro. An agent-based model of the interrelation between the covid-19 outbreak and economic activities. *Proceedings of the Royal Society A*, 477(2245):20200604, 2021.

[30] Shuichi P Obuchi, Hisashi Kawai, Manami Ejiri, Kumiko Ito, and Kenji Murakawa. Change in outdoor walking behavior during the coronavirus disease pandemic in japan: A longitudinal study. *Gait & posture*, 88:42–46, 2021.

[31] Wei Zeng and Richard L Church. Finding shortest paths on real road networks: the case for a. *International journal of geographical information science*, 23(4):531–543, 2009.

[32] Tommaso Celeste Bulfone, Mohsen Malekinejad, George W Rutherford, and Nooshin Razani. Outdoor transmission of sars-cov-2 and other respiratory viruses: a systematic review. *The Journal of infectious diseases*, 223(4):550–561, 2021.

[33] Rainald Löhner, Harbir Antil, Ashok Srinivasan, Sergio Idelsohn, and Eugenio Oñate. High-fidelity simulation of pathogen propagation, transmission and mitigation in the built environment. *Archives of Computational Methods in Engineering*, 28(6):4237–4262, 2021.

[34] Bogumił Kamiński, Paweł Prałat, and François Théberge. *Mining Complex Networks*. Chapman and Hall/CRC, 2021.

[35] Hamit Coşkun, Nazmiye Yıldırım, and Samettin Gündüz. The spread of covid-19 virus through population density and wind in turkey cities. *Science of the Total Environment*, 751:141663, 2021.

[36] Arunava Bhadra, Arindam Mukherjee, and Kabita Sarkar. Impact of population density on covid-19 infected and mortality rate in india. *Modeling Earth Systems and Environment*, 7(1):623–629, 2021.

[37] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random graphs*, volume 45. John Wiley & Sons, 2011.

A Additions to Section 3

In order to prove Theorem 3.1, we first prove the following lemma.

Lemma A.1. *There exists a function $\epsilon_1 = \epsilon_1(n) = o(1)$, such that for any $0 \leqslant t < \tau_{c_f}$, a.a.s. it holds that*

$$\left| X_{t+1} - X_t - (n - X_t) \left(\sum_{i=1}^{2} \alpha_i \left(1 - \exp\left(-\frac{b_i \lambda_i \alpha_i X_t}{n} \right) \right) \right) \right| \leqslant \epsilon_1.$$

Note that ϵ_1 does not depend on t.

Proof of Lemma A.1. We shall apply a standard form of Chernoff bound (see, for example, Corollary 2.3 in [37]). Let $X \in \mathrm{Bin}(k, p)$ be a random variable with the binomial distribution with parameters k and p, and suppose $0 < \epsilon \leqslant 3/2$ and $\mu := \mathbb{E}X = kp$. In this case,

$$\mathbb{P}[|X - \mu| \geqslant \epsilon \cdot \mu] \leqslant 2 \exp\left(-\frac{\epsilon^2 \cdot \mu}{3} \right). \tag{8}$$

Given $s \geqslant 0$ and $j \in \{1, 2\}$, define D_s^j as the set of agents which choose path P_j in commute s. Moreover, for $s \geqslant 0$, define Y_s^j to be the number of agents which become infected during commute s while taking P_j. Clearly, $Y_s^j = |D_s^j \cap I_s \cap \neg I_{s-1}|$ for $s \geqslant 1$. Let us now condition on I_t and D_{t+1}^j for $0 \leqslant t < \tau_{c_f}$ and fix agents $a, b \in [n]$ where $a \neq b$. Observe then that if $a \in [n] \backslash I_t$, $b \in I_t$, and both a and b select P_j then

$$\mathbb{P}[\text{Agent } a \text{ is infected in commute } t+1 \text{ by agent } b \,|\, I_t, D_{t+1}^j] = \beta_j \lambda_j.$$

To see this, observe that if both agents select path P_j, then agent b infects a provided a and b interact, and their interaction is contagious. This occurs with probability precisely $\beta_j \lambda_j$. As the agents of $I_t \cap D_{t+1}^j$ infect a independently of one another, we have that

$$\mathbb{P}[\text{Agent } a \text{ is infected in commute } t+1 \,|\, I_t, D_{t+1}^j] = 1 - (1 - \beta_j \lambda_j)^{|I_t \cap D_{t+1}^j|}. \tag{9}$$

Now, $\beta_j = b_j / n$, so in particular, $\beta_j \lambda_j = o(1)$. Thus, if $a \in [n] \backslash I_t$ and $a \in D_{t+1}^j$ then

$$\mathbb{P}[\text{Agent } a \text{ is infected in commute } t+1 \,|\, I_t, D_{t+1}^j] \sim 1 - \exp\left(-\frac{b_j \lambda_j |I_t \cap D_{t+1}^j|}{n} \right).$$

Thus,

$$\mathbb{E}[Y^j_{t+1} \,|\, I_t, D^j_{t+1}] \sim |D^j_{t+1} \cap \neg I_t| \left(1 - \exp\left(-\frac{b_j \lambda_j |I_t \cap D^j_{t+1}|}{n}\right)\right) \tag{10}$$

To simplify (10), observe that conditional on I_t, the random variables $|I_t \cap D^j_{t+1}|$ and $|D^j_{t+1} \cap \neg I_t|$ are distributed as $\mathrm{Bin}(|I_t|, \alpha_j)$ and $\mathrm{Bin}(|\neg I_t|, \alpha_j)$, respectively. Moreover, $c_s n \leqslant |I_t| \leqslant c_f n$, as $0 \leqslant t \leqslant \tau_{c_f}$, so we may apply (8) with say $\epsilon(n) := 1/n^{1/3} = o(1)$ to conclude that a.a.s. $\ell := |I_t \cap D^j_{t+1}| = (1 + o(1))\alpha_j X_t$ and $k := |\neg I_t \cap D^j_{t+1}| = (1 + o(1))\alpha_j (n - X_t)$. Thus, a.a.s.

$$\mathbb{E}[Y^j_{t+1} \,|\, I_t, D^j_{t+1}] \sim \alpha_j (n - X_t) \left(1 - \exp\left(-\frac{b_j \lambda_j \alpha_j X_t}{n}\right)\right) = \Theta(n). \tag{11}$$

On the other hand, Y^j_{t+1} conditional on I_t and D^j_{t+1} is distributed as a binomial $\mathrm{Bin}(k, p)$ with parameters $k = |D^j_{t+1} \cap \neg I_t|$ and $p = 1 - (1 - \beta_j \lambda_j)^{|I_t \cap D^j_{t+1}|} = 1 - (1 - \beta_j \lambda_j)^{\ell}$. Moreover, a.a.s.

$$\mu_j := \alpha_j (n - X_t) \left(1 - \exp\left(-\frac{b_j \lambda_j \alpha_j X_t}{n}\right)\right) \sim kp, \tag{12}$$

where the approximation for p follows from (9). As a result, by taking $\epsilon(n) = 1/n^{1/3}$ and applying (8), we get that a.a.s.

$$\mathbb{P}[|Y^j_{t+1} - \mu_j| \geqslant \epsilon \mu_j \,|\, I_t, D^j_{t+1}] \leqslant \exp(-\Omega(n^{1/3})).$$

(Formally, one should stochastically lower and upper bound Y^j_{t+1} by $\mathrm{Bin}(k_-, p_-)$ and $\mathrm{Bin}(k_+, p_+)$ with some deterministic functions $k_\pm$ and $p_\pm$, such that $p_+/p_- \to 1$ and $k_+/k_- \to 1$ as $n \to \infty$.) Thus, a.a.s.,

$$|Y^j_{t+1} - \mu_j| \leqslant \epsilon \mu_j \tag{13}$$

for each $j = 1, 2$. On the other hand, $X_{t+1} - X_t = \sum_{j=1}^2 Y^j_{t+1}$, and

$$\mu_1 + \mu_2 \sim (n - X_t) \left(\sum_{i=1}^2 \alpha_i \left(1 - \exp\left(-\frac{b_i \lambda_i \alpha_i X_t}{n}\right)\right)\right),$$

so the proof is complete. □

Proof of Theorem 3.1. Let us set $A_i := b_i \lambda_i \alpha_i \geqslant 0$ for $i = 1, 2$, and $A := \max\{1, A_1, A_2\}$ for convenience. Given $0 \leqslant t_0 < \tau_{c_f}$, our goal is to show that there exists a function $\epsilon_0 = \epsilon_0(n) = o(1)$ such that a.a.s.

$$\left|\frac{X_{t_0}}{n} - \widetilde{x}_{t_0}\right| \leqslant \epsilon_0(n).$$

In order to prove this, we first prove the following implication. Let us take $0 \leqslant t < \tau_{c_f}$, and assume that $0 \leqslant \delta = \delta(n) \leqslant 1$ satisfies

$$\left|\frac{X_t}{n} - \widetilde{x}_t\right| \leqslant \delta, \tag{14}$$

and

$$\left|X_{t+1} - X_t - (n - X_t) \left(\sum_{i=1}^2 \alpha_i \left(1 - \exp\left(-\frac{b_i \lambda_i \alpha_i X_t}{n}\right)\right)\right)\right| \leqslant \delta. \tag{15}$$

Under these assumptions, we claim that $|X_{t+1}/n - \widetilde{x}_{t+1}| \leqslant 5A\delta$. Observe first that by (14), $X_t/n \leqslant \widetilde{x}_t + \delta$, so

$$\exp(-A_i X_t/n) \geqslant \exp(-A_i(\widetilde{x}_t + \delta)).$$

Now, using the inequality $1 - x \leqslant \exp(-x)$, we get that

$$\exp(-(z_1 + z_2)) \geqslant \exp(-z_1)(1 - z_2) \geqslant \exp(-z_1) - z_2$$

for $z_1, z_2 \geqslant 0$. It follows that

$$\exp(-A_i(\widetilde{x}_t + \delta)) \geqslant \exp(-A_i\widetilde{x}_t) - A\delta. \tag{16}$$

Define $g(z) := \sum_{i=1}^{2} \alpha_i(1 - \exp(-A_i z))$ for $z \in \mathbb{R}$. Observe then that $g(z)$ is an increasing function of z and so by (14) and (16),

$$g(X_t/n) \leqslant g(\widetilde{x}_t + \delta) \leqslant g(\widetilde{x}_t) + 2A\delta. \tag{17}$$

Now, applying (15),

$$\begin{aligned}
\frac{X_{t+1}}{n} &\leqslant \frac{X_t}{n} + \left(1 - \frac{X_t}{n}\right)\left(\sum_{i=1}^{2} \alpha_i \left(1 - \exp\left(-A_i X_t/n\right)\right)\right) + \delta \\
&\leqslant \widetilde{x}_t + \delta + (1 - \widetilde{x}_t + \delta)g(X_t/n) + \delta \\
&\leqslant \widetilde{x}_t + (1 - \widetilde{x}_t)g(X_t/n) + \delta + \delta \\
&\leqslant \widetilde{x}_t + (1 - \widetilde{x}_t)g(\widetilde{x}_t) + 2A\delta + 2\delta + \delta \\
&\leqslant \widetilde{x}_t + (1 - \widetilde{x}_t)g(\widetilde{x}_t) + 5A\delta
\end{aligned}$$

where the latter inequalities follow since $g(z) \leqslant 1$ for $z \geqslant 0$, and by (14) and (17). Yet, $\widetilde{x}_{t+1} := \widetilde{x}_t + (1 - \widetilde{x}_t)g(\widetilde{x}_t)$, so

$$\frac{X_{t+1}}{n} \leqslant \widetilde{x}_{t+1} + 5A\delta.$$

An analogous argument shows that $X_{t+1}/n \geqslant \widetilde{x}_{t+1} - 5A\delta$. Thus, assuming (14) and (15) for $\delta \leqslant 1$, it follows that

$$\left|\frac{X_{t+1}}{n} - \widetilde{x}_{t+1}\right| \leqslant 5A\delta.$$

In order to complete the argument, take $\delta(n) = \epsilon_3(n)$, where $\epsilon_3 = o(1)$ is from Lemma A.1. Clearly, we may assume that $5^{t_0}A\delta \leqslant 1$ for n sufficiently large. Now, $X_0/n = \widetilde{x}_0$ by assumption. Moreover, since t_0 is constant, we may apply Lemma A.1 to each $0 \leqslant t \leqslant t_0$ to ensure that (15) holds a.a.s. for all $0 \leqslant t \leqslant t_0$ *simultaneously*. Thus, proceeding by induction, we get that a.a.s.

$$\left|\frac{X_{t_0}}{n} - \widetilde{x}_{t_0}\right| \leqslant 5^{t_0}A\delta.$$

Since t_0 is a constant, $5^{t_0}A\delta = o(1)$, so the argument is complete. □

Proof of Corollary 3.2. Let us first fix $c := \widetilde{x}_{t_{c_f}+3}$, which clearly satisfies $c > c_f$. By applying Theorem 3.1, it is easy to show that $\tau_c > t_{c_f} + 1$. Thus, we may apply Theorem 3.1 at the values from $\{t_{c_f} - 1, t_{c_f}, t_{c_f} + 1\}$. Specifically, there exists $\epsilon_1 = o(1)$ such that a.a.s.,

$$\left|\frac{X_k}{n} - \widetilde{x}_k\right| \leqslant \epsilon_1(n) \tag{18}$$

for each $k \in \{t_{c_f} - 1, t_{c_f}, t_{c_f} + 1\}$.

Now, for $k = t_{c_f} - 1$, (18) implies that a.a.s.

$$\frac{X_{t_{c_f}-1}}{n} \leqslant \widetilde{x}_{t_{c_f}-1} + \epsilon_1.$$

On the other hand, $\widetilde{x}_{t_{c_f}-1} < c_f$. Thus, for all n sufficiently large, $\epsilon_1(n) < c_f - \widetilde{x}_{t_{c_f}-1}$. It follows that a.a.s.

$$\frac{X_{t_{c_f}-1}}{n} < c_f,$$

and so $\tau_{c_f} \geqslant t_{c_f}$.

Suppose now that $\widetilde{x}_{t_{c_f}} > c_f$ (part (a)). Using (18) for $k = t_{c_f}$, we get that a.a.s.,

$$\frac{X_{t_{c_f}}}{n} \geqslant \widetilde{x}_{t_{c_f}} - \epsilon_1.$$

Since $\widetilde{x}_{t_{c_f}} > c_f$, for all n sufficiently large, $\epsilon_1(n) < \widetilde{x}_{t_{c_f}} - c_f$. It follows that a.a.s.

$$\frac{X_{t_{c_f}}}{n} > c_f,$$

and so $\tau_f \leqslant t_{c_f}$. The same argument can be applied for part (b) but with $X_{t_{c_f+1}}$ instead of $X_{t_{c_f}}$. $\square$