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 regarding vehicle scheduling frequency can affect the pandemic dynamics in urban populations. Specifically, we develop a multi-agent simulation framework to model infection dynamics in complex transportation 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 levels of contagiousness, as a function of 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. We also assess the effect of modifying the public transport's running frequency on the virus spread. 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 to model contiguous disease spread in urban environments, including influenza or various COVID-19 variants.

1 Introduction

The objective of the paper is to analyze how decisions made by public transport regulators regarding vehicle frequency in urban populations affect pandemic development dynamics such as influenza or COVID-19. In order to control the spread of a 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

---

[*]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`

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

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 the approach presented by Silva et al. (2020) and model these processes using an agent-based model (ABM) to assess the effects of social distancing interventions in public transportation of a large urban area.

Pandemic widespread can lead to significant changes in requirements for operations of a public transportation system — e.g. see Ge et al. (2022). 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 — which is also a standard approach in the transportation literature. As described by Currie et al. (2020), ABMs are valuable tools for decision makers. For instance, Kerr et al. (2021) have constructed "Covasim"— an agent-based model of COVID-19 dynamics and interventions and state that their model has been used to make informed policy decisions in several countries including the United States, Vietnam, the United Kingdom, and Australia. Another ABM tool, OpenABM-Covid19 (Hinch et al., 2020), assesses COVID-19 non-pharmaceutical interventions including contact tracing and has been designed for demographics in the UK. On the other hand, the work by Speir and Negahban (2020) is an example of a tool that measures the effectiveness of mitigation rules such as social distancing or mask-wearing in cities. Hoertel et al. (2020) introduce a COVID-19 ABM simulator of the urban population of New York City and investigate various pandemic mitigation strategies. While there are multiple all-purpose simulators such as aforementioned 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 to better approximate real-world interactions. 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 (Bezanson et al., 2017) and the `OpenStreetMapX.jl` library. Spatial data come from the OpenStreetMap project along with the "TTC Routes and Schedules" data set from the Toronto Transit Connection (TTC) Open Data portal. All simulation source codes are available on a 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 that take place in society can be a reliable data-driven source of knowledge that supports regulators in defining public transport policies to limit the spread of a virus such as COVID-19.

The remainder of the paper is organized as follows. In Section 2, we present a review of the literature on 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 scale of this

---

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

network is not realistic, but it provides a starting point for the main properties we wish to explore. In Section 4, we introduce a much more general and complex model which more accurately reflects the real world. 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 to 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

The literature on the subject can be broadly classified into three main areas: methods for designing resilient and robust transportation systems, agent-based modelling of pandemic in transportation systems, and optimizing transportation system operations through regulatory decisions. We now provide brief overviews of the most related works in each area.

2.1 Resilience and Robustness of Public Transport Systems

The outbreak of a pandemic can lead to significant changes in the operations of a public transportation system — e.g. see Frias-Martinez et al. (2011); Ge et al. (2022); Müller et al. (2020, 2021). This point is emphasized in Ge et al. (2022), and Yao et al. (2022) quantified the impact through empirical research of vehicle commuters in a collection of Chinese cities. More specifically, Yao et al. (2022) analyzed the data from license plate recognition systems and used clustering methods to understand the behavioral patterns of commuters. They identified commuters' frequent travel patterns before the COVID-19 pandemic and found out that during the pandemic, such patterns changed for over 90% of the vehicles. Interestingly, over half of the studied commuters did not return to their original travel patterns after the pandemic ended.

Serdar et al. (2022) perform a literature review on the resilience of transportation networks. They use agent-based modelling to understand how viruses spread through commuting networks, and conclude that a greater resilience to infection spread can be achieved by combining different transportation modes. Dekker et al. (2022) propose an analytical framework for disruption modelling and argue that tools from complex science – such as Monte Carlo simulation – should be used for rescheduling services when a disruption occurs. Ge et al. (2022) observe that planning and disruption modelling in public transport systems typically follows the multi-layered approach developed by Daduna and Voß (1996). This approach is described in the following four levels: strategic (e.g., line planning), tactical (e.g., timetables), operational planning (e.g., vehicle scheduling) and operations (vehicle dispatching and monitoring). The benefit of this layered approach is that it allows researchers to focus on specific parts of the problem.

In the terminology of Daduna and Voß (1996), we focus on the tactical and operational planning layers. Given a commuting network, we analyze the effect of an urban transportation's general running frequency on infection spread. We analyze a real-world network, however, the model is flexible enough that it can be used for simulating virus spread on a hypothetical network.

2.2 Agent-Based Modelling of a Pandemic in Transportation Systems

Frias-Martinez et al. (2011) construct an agent-based simulator taking into account *social interactions* and *individual mobility patterns* to investigate the 2009 outbreak in Mexico of the influenza

subtype H1N1. This simulator also measures the effect of mobility reduction on the spread of the disease. Aleman et al. (2011) is another work that simulates the behavior of disease during a pandemic. In this model, used by the Ontario Agency for Health Protection and Promotion, the authors consider each individual to be unique and assume that the 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. Hackl and Dubernet (2019) introduce an agent-based transport simulation that as an additional simulation layer includes a virus spread model. The combined model makes it possible to understand the rate of infection in urban areas and the authors utilized it to analyze seasonal influenza outbreaks in Zurich, Switzerland. Laskowski et al. (2011) developed an agent-based model that uses a Geographic Information System (GIS) to study the spread of influenza in an urban area. The model was validated using the 2009 H1N1 influenza outbreak in Kunming, China. In another research ACEMod — an agent-based model has been developed that studies the spatiotemporal dynamics of influenza epidemics in Australia. The model uses *mobility patterns* (employees and students commuting) as well as *human interactions* to offer intervention/mitigation policies — see Cliff et al. (2018).

In a related line of work, Müller et al. (2020, 2021) constructed an agent-based model for pandemic spread in public transportation. They postulate that during the pandemic, public transport should run at its highest frequency and capacity. However, they do not take into consideration that a higher frequency increases the attractiveness of public transportation and encourages more people to use it instead of alternative modes of transportation. They obtain a result that supports that only a complete lockdown can decrease the peak pandemic level.

Gkiotsalitis and Cats (2021) consider a quadratic mixed integer linear programming optimization (QMILP) model for optimal allocation of train schedule including the number of carts and the frequency. Their approach makes it possible to analyze the trade-off between the operating costs of a transportation system, average passenger wait times, and the requirement to provide enough space in the carts to achieve social distancing.

Agent-based modeling (ABM) of a 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 healthcare system, and the economic implications of this virus.

Many researchers have concentrated their efforts on simulating the 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.

To understand how the coronavirus spreads in small towns and cities, Truszkowska et al. (2021) presented an ABM platform which they verified on real data from New Rochelle, New York. Their 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 Cuevas (2020) which helps decision makers develop appropriate strategies.Wallentin et al. (2020) constructed an agent-based model for the city of Salzburg, Austria to support policymakers in taking intervention 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, such as *school/business closures and random testing/quarantine requirements*, Speir and Negahban (2020) designed a customizable

agent-based simulation tool. They then validated their model on New York City, the epicenter of the COVID-19 outbreak in the US. Macal et al. (2020) implemented an agent-based model called CityCOVID to support decision-making about COVID-19 at the level of city, county, and state (currently in the Chicago area). To accomplish this goal, the model analyzes the COVID-19 transmission and understands the behavior of individuals in reaction to the transportation policy regulator interventions. An extension of CityCOVID, focusing on the development of a large-scale synthetic population, is discussed by Kaligotla et al. (2020). In Canada, the Public Health Agency has developed an agent-based simulation to study the impact of non-pharmaceutical interventions on the spread of COVID-19 with the aim of supporting public health decision-makers — see (Ogden et al., 2020). The Covasim agent-based simulator has been applied to inspect the dynamics of the COVID-19 disease and evaluates various intervention policies in Africa, Europe, Oceania, and North America (Kerr et al., 2021).

Finally, an agent-based tool called OpenABM-Covid19 (Hinch et al., 2020) 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.

Similarly to the above research, our paper also relies on an ABM model to simulate the urban life of a large metropolis, and so it fits into the growing literature on agent-based modeling of pandemics. However, in our ABM model we focus on pandemic aspects related to public transportation.

2.3   Planning and Optimizing Transportation System Operations

Any design decision in transportation systems results in several externalities including congestion level, pollution, commuter satisfaction, crowdedness and infection likelihood. The goal of a transport system operator is to design routes, timetables, and pricing policies that on the one hand minimize the negative effects and on the other hand keep the transportation system running costs within the budget limit. Pandemic development and urban mobility planning have been proven to be closely intertwined. For instance, in Rahman et al. (2021) it was examined that human mobility was significantly reduced during the pandemic. Shamshiripour et al. (2020) suggested that there may be a shift from the usage of shared mobility options (such as transit services) to transport modes that avoid contact (such as walking, biking, or personal vehicles). This is why researchers have stressed the need for *proactive transportation planning* in a way that promotes an *active and equitable transportation system* to avoid further car reliance in large cities in the post-pandemic times. This problem is addressed in the transportation robustness and smart city literature. For instance, Shi et al. (2020) analyze the data gathered in Shanghai where the goal of controlling COVID-19 outbreaks at the community level was achieved by integrating manual measures with advanced information technology applications.

Huang and Shen (2021) propose an optimization model for the reopening of transit lines and the design of new timetables which take into consideration the risk of infections and social distancing rules. They note that system operators need to either open as many lines as possible to keep the interaction level between the commuters low, or to discourage commuters from using the system by providing a very low frequency of service and/or close some stations. While the first approach is preferable, it might not be feasible with the limited budgets of most municipalities. In another work, Yusuke et al. (2020) constructed a simulation model to discover the impact of bus scheduling changes and reduced vehicle capacity under a wide range of employee demands. As a result, this 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 the COVID-

19 pandemic.

Our work does not directly propose a total redesign of the existing transportation network. Rather, it provides tools that can point to spots where the existing network is missing the ideal infrastructure, e.g., an additional station that could unload some traffic from the overcrowded hubs. We also observe the effects of long waiting times and crowd formulations in a number of the scenarios we study.

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.

3.1 Definition of the Toy Model
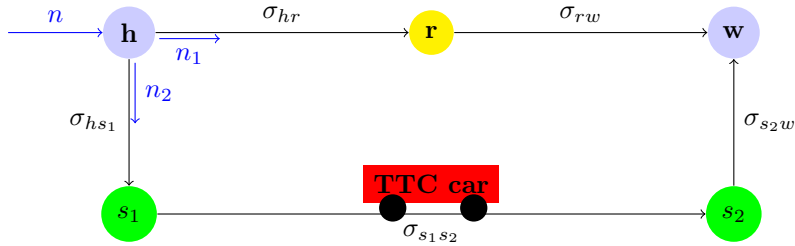
Figure 1: Toy Model

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.

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 the 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

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

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

7

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.

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 $\tau_{c_f} \in \{t_{c_f}, t_{c_f} + 1\}$.*

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 the GitHub repository. For simplicity, for all the scenarios we set $\lambda_1 = 1/100$, $\lambda_2 = 2/100$, $c_s = 0.01$, and $c_f = 0.99$.

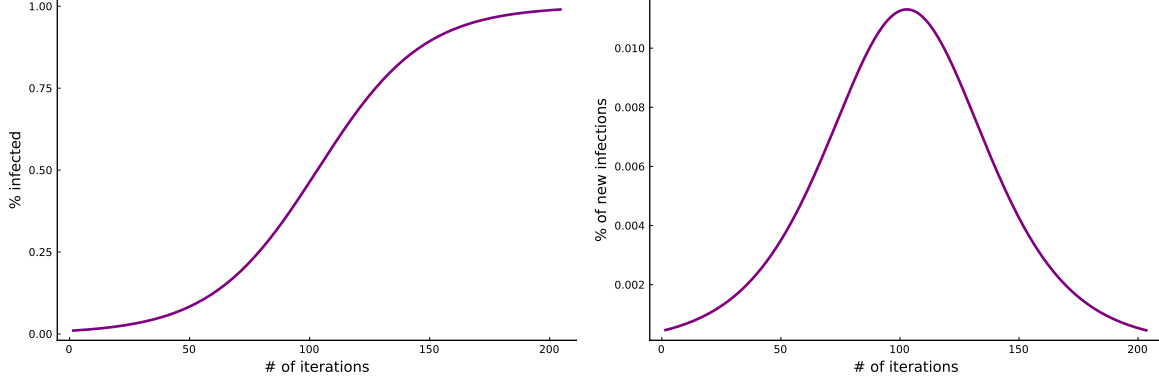Figure 2: Scenario 1: the fraction of the agents infected (left), and the fraction of new infections (right).

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 the agents are infected after $t_f = 204$ iterations.

The fraction of the 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).

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 the 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 the 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 the 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 the agents are infected after $t_f = 217$ iterations. However, as expected, it is worse than Scenario 3—the
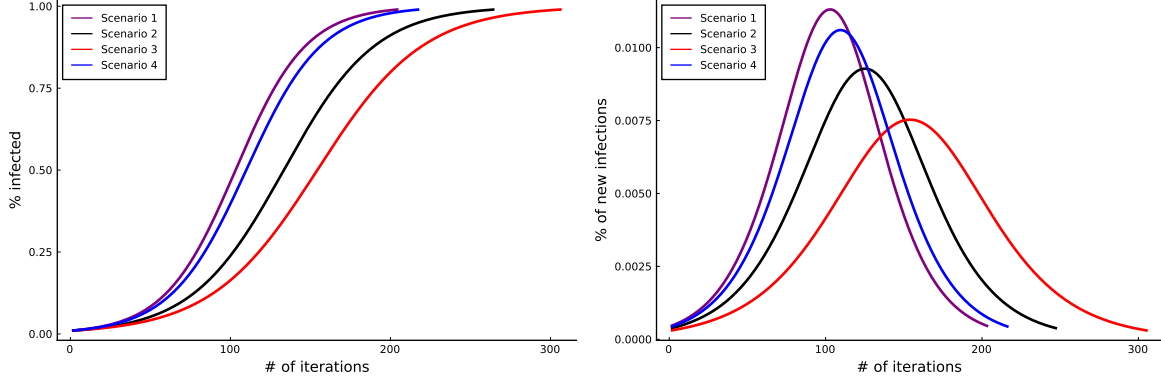
9

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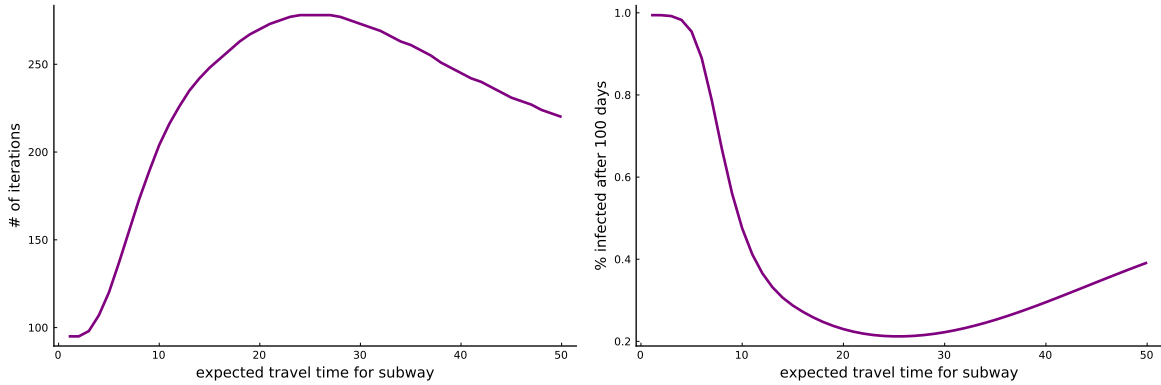

Figure 4: Scenario 5: the number of iterations needed to infect 99% of the agents (left) and the fraction of the agents infected after 100 iterations (right).

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 the 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 the agents and the fraction of the 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 the agents and after 100 iterations only 21.21% of the agents are infected.
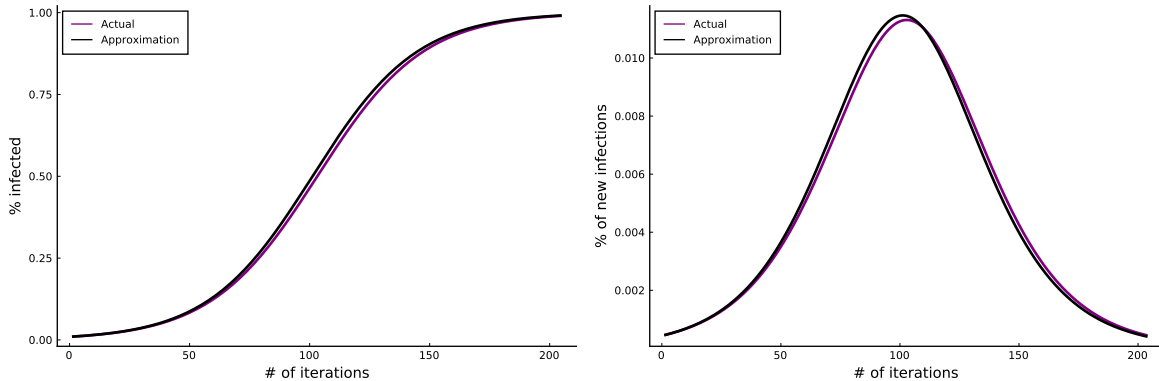
10

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

3.4 Approximation

In Section B (see the appendix), we derive an approximate but closed formula for the number of infected agents after a given number of iterations. We may approximate (3) as follows:

$$\widetilde{x}_{t+1} - \widetilde{x}_t \quad = \quad A(1 - \widetilde{x}_t)\widetilde{x}_t, \qquad \text{where } A := b_1 \lambda_1 \alpha_1^2 + b_2 \lambda_2 \alpha_2^2.$$

On Figure 5 (right), we compare this approximation with the actual values for Scenario 1. We also 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 is derived from Toronto's transportation system. For instance, the term TTC refers to the Toronto Transit Commission. While we focus on one particular city, the simulation can be repeated on an arbitrary city, as the data on OpenStreet map project is publicly available, and our framework is flexible.

All agents have 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

11

Attribute	Type	Class	Attribute	Type	Class
ID	fixed	value	ID	fixed	value
home_location	fixed	value	first_station	fixed	value
work_location	fixed	value	last_station	fixed	value
current_location	dynamic	value	current_station	dynamic	value
direction	dynamic	value	line	fixed	value
infection_status	dynamic	value	route	fixed	set
walking_speed	fixed	value	passengers	dynamic	set
routes	fixed	set	passengers_limit	fixed	value
TTC_car_ID	dynamic	value	arrival_time_interval	fixed	value
passengers_met_in_TTC	dynamic	set	one_trip_max_passengers	dynamic	value
number_of_trips	dynamic	value			

(a) Attributes of the commuter class   (b) Attributes of the TTC car class

Table 2: Objects' attributes

home). When an agent reaches its workplace, it typically works for an average of 8 hours, remaining stationary during this period, and then returns home. Once at home, the agent rests for about 16 hours. To begin the process, 1% of the population is chosen randomly 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 that facilitate describing the simulation model are listed in Table 4 in Appendix C.

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 2a.

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 to 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 into the model (such as the distribution of ages of the agent). Based on the approach of Obuchi et al. (2021), 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 2b.

`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 an 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 kth commuter is thus split into two sums, where the latter aggregates over the expected waiting times of the TTC cars:

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

where all symbols are in accordance with Table 4. Using the A* search algorithm, which is often used in many fields of computer science due to its completeness, and computational efficiency (Zeng and Church, 2009), 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. Liu and Zhou (2016) point out that empirical data show that in reality commuters do not always choose

13

the shortest routes. They propose to model this behavior by using the k-shortest path method to identify a set of routes between starting and destination point and choose among them. In our paper we take a similar approach — we identify three shortest routes between the starting and destination point and randomly select one of them. This provides a more natural distribution of travel paths.

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.

Bulfone et al. (2021) note that 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 an 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 Löhner et al. (2021).

4.1.4   Simulation Dynamics

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

To summarize, the commuter wakes up at home and then selects their route to work among the three fastest alternatives. While traversing their route, they may interact with infected people,
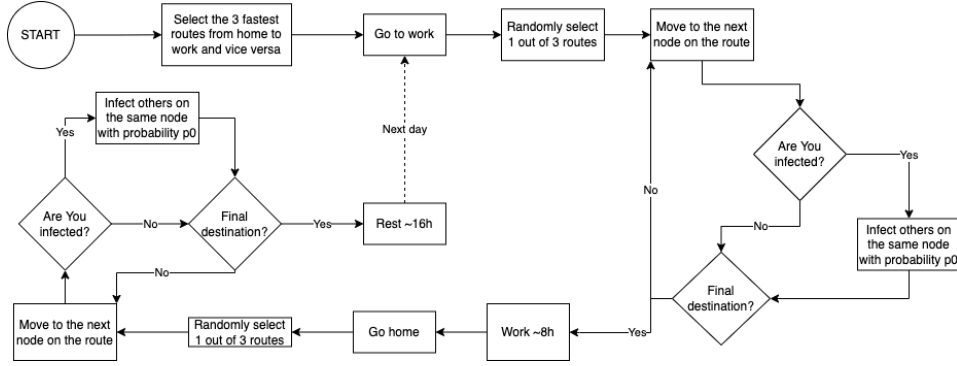
Figure 6: An example of one day simulation dynamics for an agent

which could cause the commuter to become infected herself. After work, they select a route back to their home and then rest for a fixed amount of time. The action diagram for the TTC cars has been presented in Figure 7.
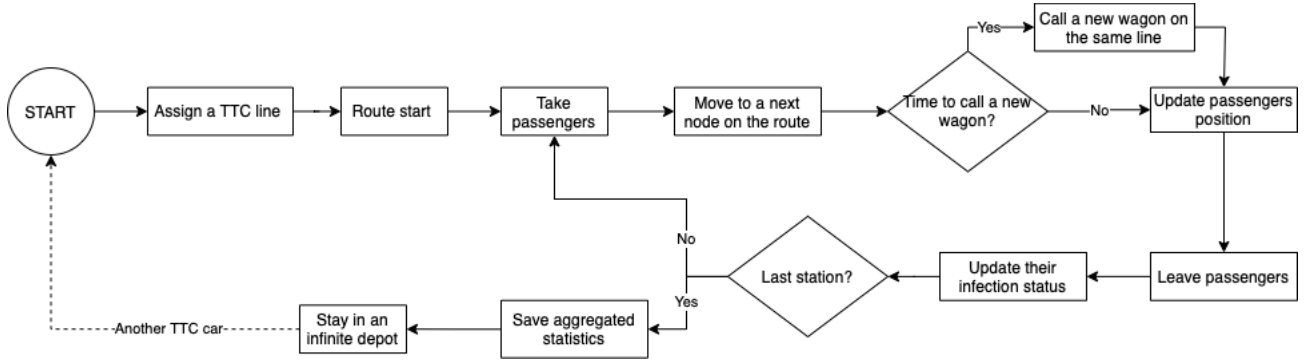

Figure 7: An example of one day simulation dynamic for a TTC car

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 elements 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, we assume that the commuters from the suburban areas of Toronto are present in the model, but
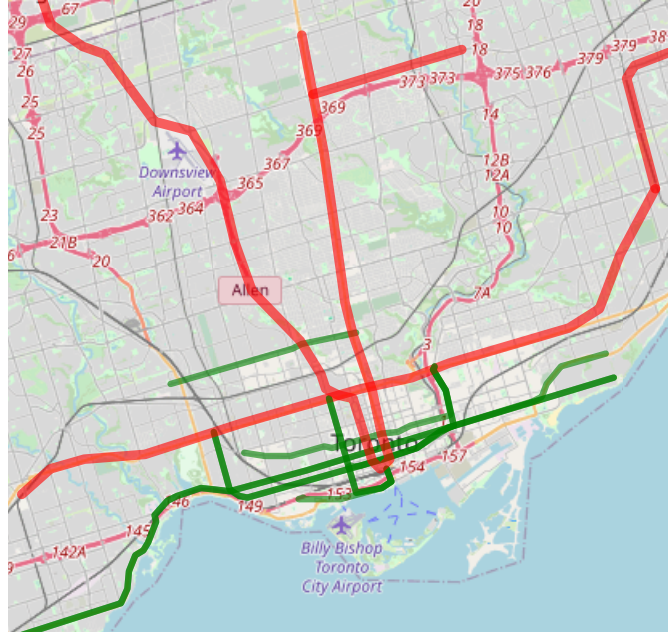
Figure 8: Map of Toronto used in the experiment

we only start tracking their movement once they reach the central part of Toronto. The map was downloaded from the 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, we assume the existence of a direct path between any pair of nodes. 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 the 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 3.

---

[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 3: 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 with the full simulation where the paths taken by the agents are diversified. In terms of the routing decisions, recall that each agent selects one of its three optimal routes uniformly at random in the full simulation. In contrast, in the simplified model each agent chooses between either taking a walk to work or using the TTC, where the travel time depends on the frequency parameter. Thus, if we were to assume that the agents always choose
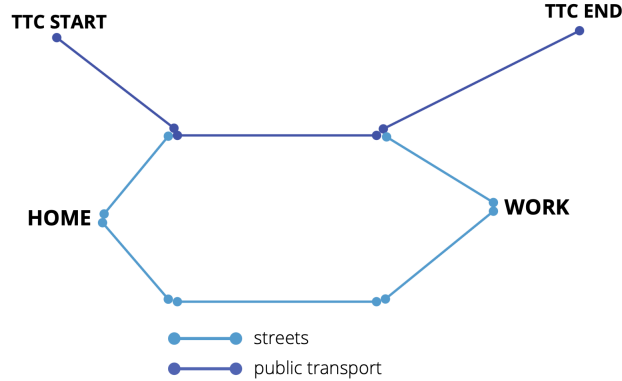
17

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. In our model a short TTC headway such as 2.5 minutes is greatly increasing public transport attractiveness — see Figure 10. This happens because in our model, on one hand, the routes of several commuters are possibly within a walking distance, and on the other hand they optimize their travel times. Hence very short headway is causing a massive redirection of commuter flow from pavements to the transportation system — which causes overcrowding and sharply increases infection probability. This result is consistent with other postulates in literature such as Huang and Shen (2021) postulate decreasing public transport attractiveness in a period of a pandemic. On the other hand, we can see in Figure 10 that very long headway times results in increased infection rates as commuters who live at a great distance from their work and have no other transportation options end up in overcrowded TTC carts. Hence, in our numerical results, 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.

To better understand how the behavior of agents leads to the "sweet spot", it is helpful to consider Figure 11. Observe that when the TTC arrives very frequently, it is extremely attractive for people to take, and so a lot of commuters are infected on their way to the station and inside the TTC. Looking at Figure 11, it could be assumed when TTC is more frequent (the number of TTC cars increases) then the number of passengers increases much more. 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 to get infected easily because of overcrowded public transport. The decrease in the number of TTC users does not compensate for the less frequent TTC.

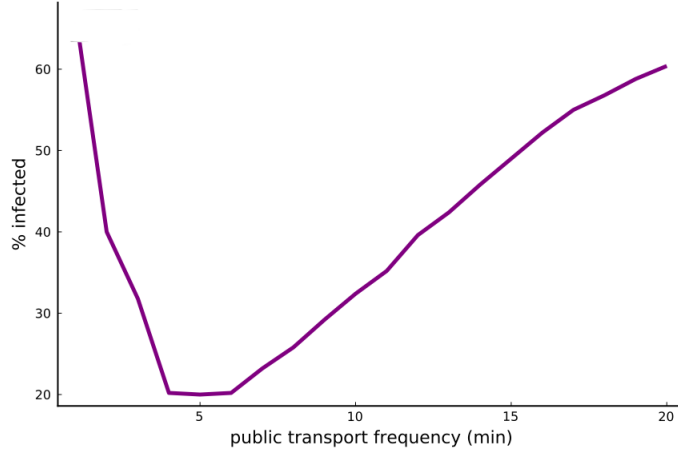We also checked how many commutes (number of trips on the route from home to work and

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

vice versa) it takes to infect 95% of the simplified model's population and compared the results of different TTC frequency setup. 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 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 3 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 the results presented by Coşkun et al. (2021) and Bhadra et al. (2021), as well as the #stayathome strategies promoted all around the globe.

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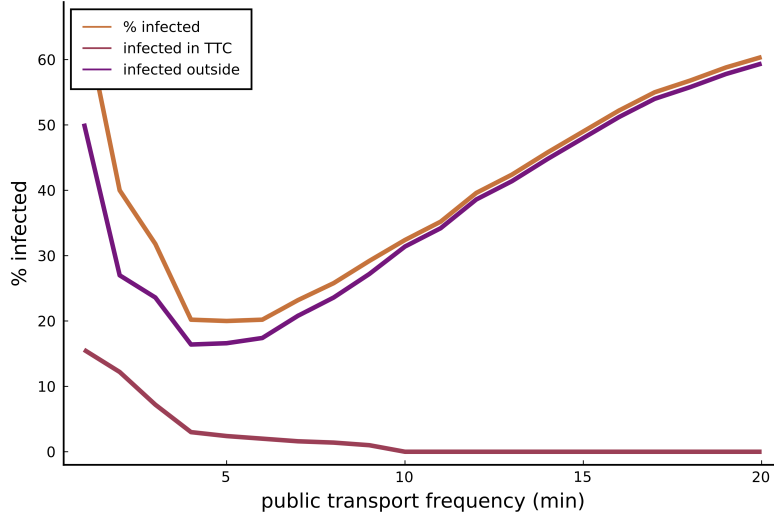
Figure 11: Infected agents in different places of infection at a fixed iteration depending on TTC frequency
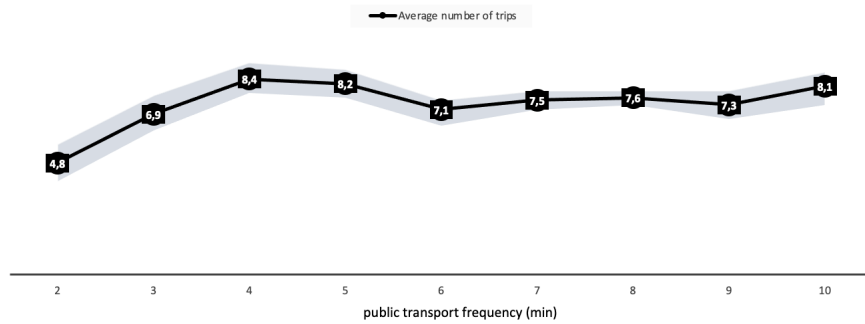

Figure 12: Number of trips on the route from home to work and vice versa depending on TTC frequency

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.

Moreover, those results are confirmed looking at 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
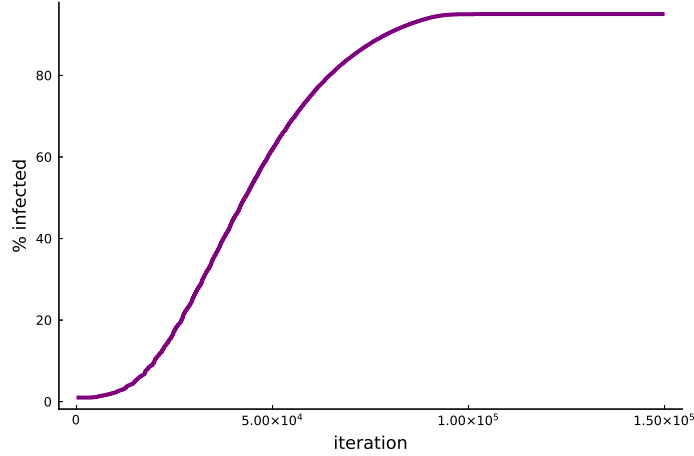
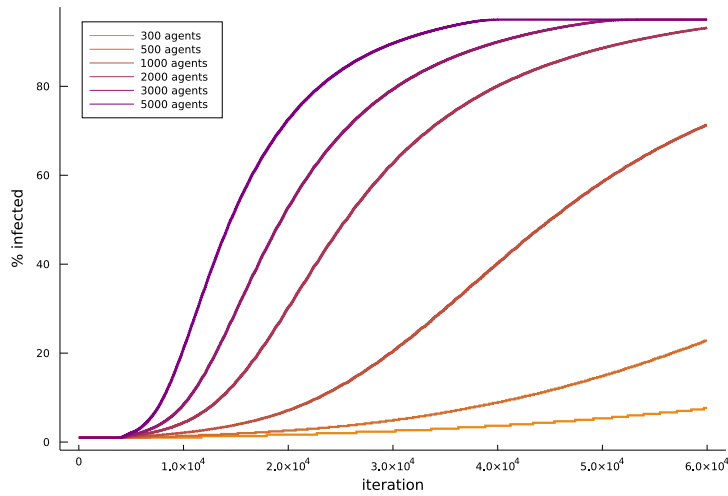Figure 13: Exemplary epidemic curve in the simulation model

Figure 14: Population density vs pace of infection

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, and 2) sparse agents concentration outside, i.e., less people interactions at intersections, the 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) the TTC frequency (OX axis), and 2) the number of infected agents that participated in the interaction (bar colors). The numbers on the blocks refer to the number of encounters of a 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 the number of interactions in small groups. It is the
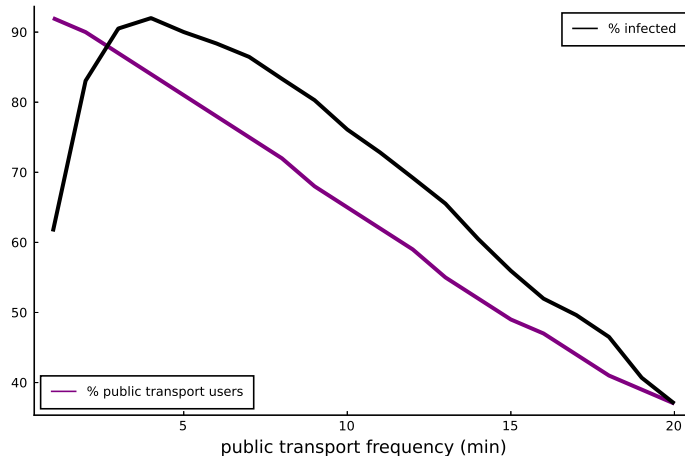
21

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

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 public transportation, the 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, and 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 at 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 others, which leads to an 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 traveling relatively small distances, so they are more willing to walk. Thus, they are more evenly distributed on the map, leading to fewer potentially "dangerous" meetings.

Another crucial decision taken by the transportation system operator is the maximum car capacity. Namely, it is possible to introduce a restriction that only some percentage of the total subway car area can be occupied by passengers, i.e., a passenger limit would be introduced. The goal of such restriction would be to avoid crowds inside the TTC cars. In Figure 20 passenger limit's im-
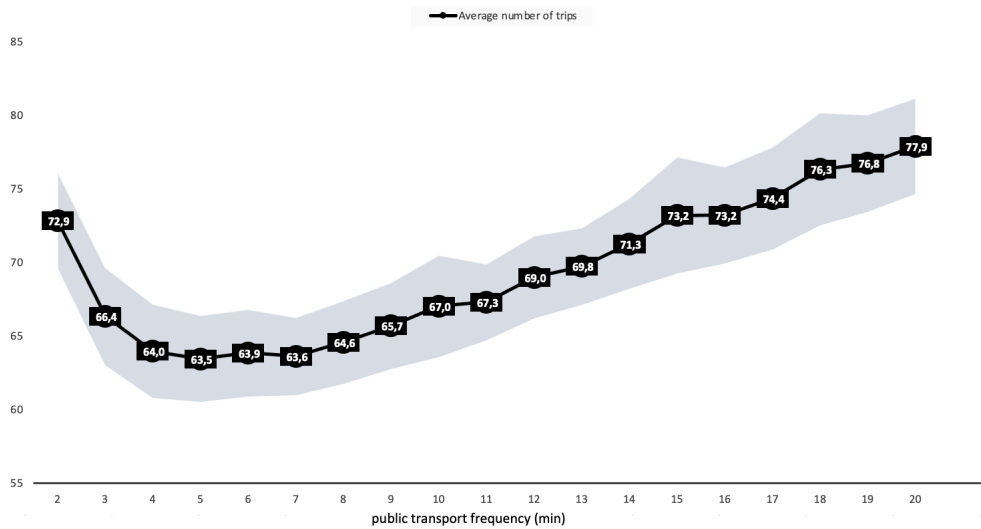
22

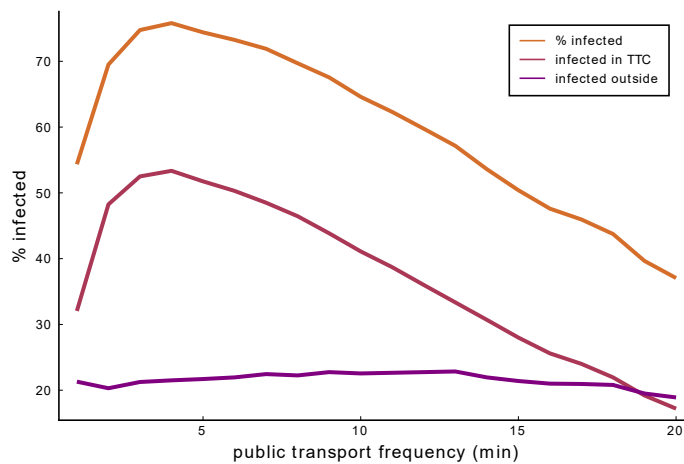Figure 16: Number of trips on the route from home to work and vice versa depending on TTC frequency

Figure 17: Places of infection

pact on the number of infections turns out not to be equal between the potential TTC frequencies. After ca. one work week of the simulation, the number of infections is strongly dependent both on the passenger limit, and TTC frequency. If the subway runs frequently enough, e.g., every 2 minutes, none of the restriction variants are effective — the subway is never that crowded anyway. At the same time, there are so many TTC cars and passengers that they get infected while being in smaller groups, just more frequently. Under moderate TTC frequency, e.g., every 5 minutes, only a very strict limit on the number of passengers can impact the infection rate in a meaningful way. With such a running frequency, the TTC is a popular choice among travelers but cannot serve all passengers that need a ride. This is why a mild restriction would not be particularly helpful here as the number of passengers in the car and the station would be at its maximum and there will be a lot of fully loaded trains (and not too many pedestrians on the streets). Last but not least, with a TTC running frequency as low as every 10-15 minutes, the subway becomes less popular, and even
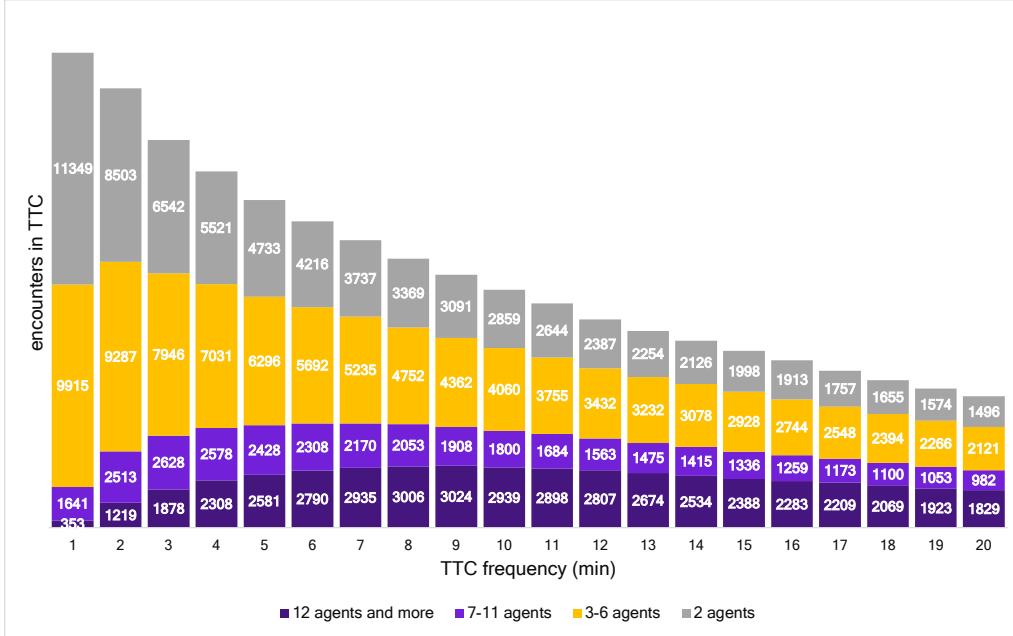
Figure 18: Number and structure of encounters in terms of crowd size (each bar colour represents the number of agents involved in a single encounter)

a mild passenger limit helps to lower the number of infections — during the rush hour, citizens will spend more time waiting for a TTC car in a more spacious and less risky place in terms of infection probability.

Let us consider an initiative by a transportation system regulator (e.g. a municipality) where an information campaign is held and unnecessary physical contact in closed public spaces is discouraged. In the model, such a measure can be included via the maximum number of interactions the agents make in the subway. The more interactions the agents are allowed to make during their trip, the larger the percentage of infected is after ca. one week of travel. In practice, the result of such a campaign would be strictly dependent on the people's reaction to the regulators' warning which can vary for different countries, municipalities, cultures, and, in general, the trust the citizens put in the authorities. These factors can be reflected by the number of interactions made, i.e., the lowest interaction numbers would occur in places of strong trust in the government, while the highest — in the places where the government's warnings are not heard by the inhabitants. Figure 21 shows that the more restrained the passengers are with each other and the less crowded parts of the TTC car they choose, the lower the percentage of infected people. When travelers are keen to interact with each other or just be in each other's direct proximity, the infections go up significantly.

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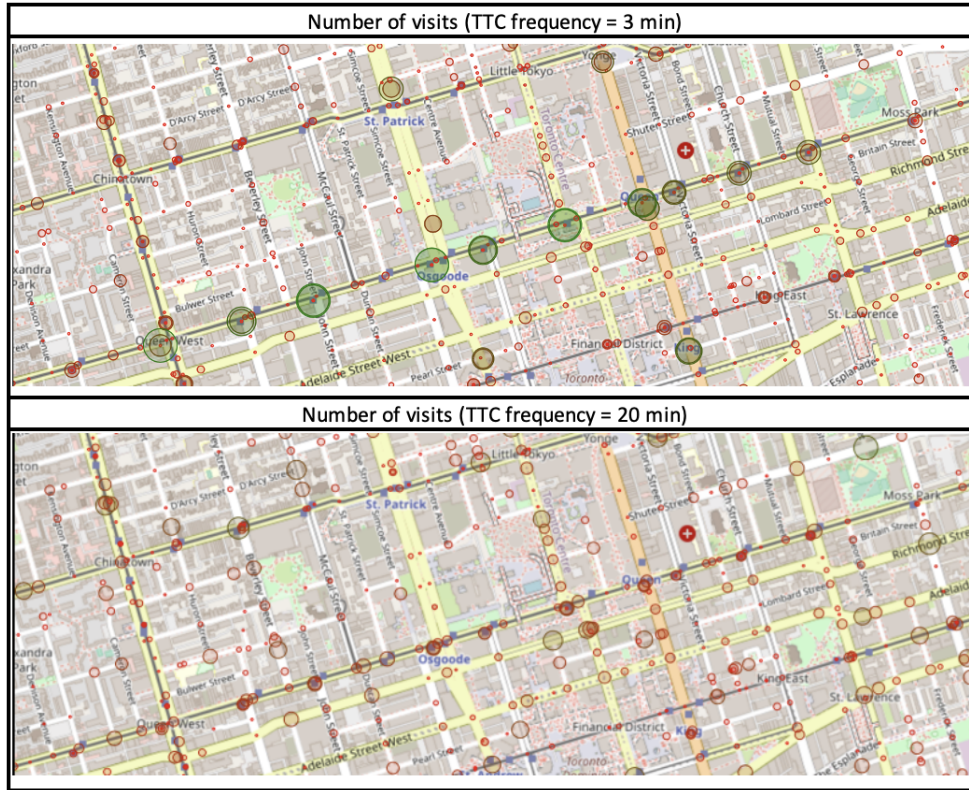
Figure 19: Individual nodes visits depending on TTC frequency on a small fragment of the map
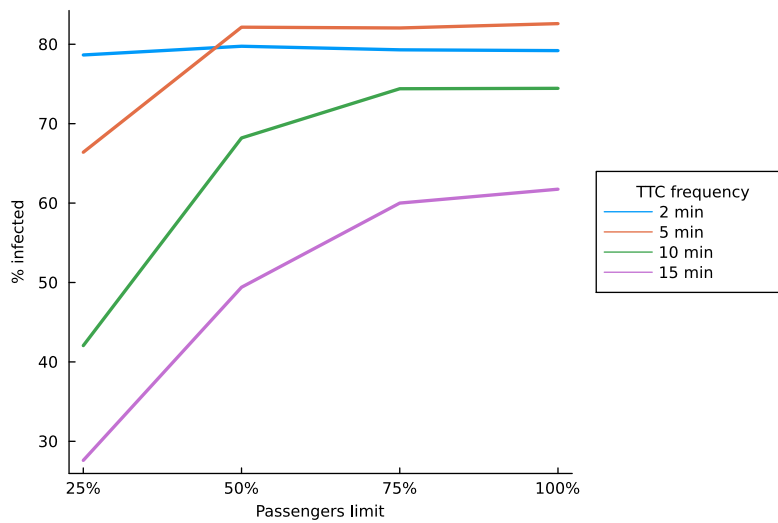

Figure 20: Impact of government restrictions on TTC passenger limits on the number of infections under different TTC frequencies (x-axis indicates the percentage of permissible passengers)

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
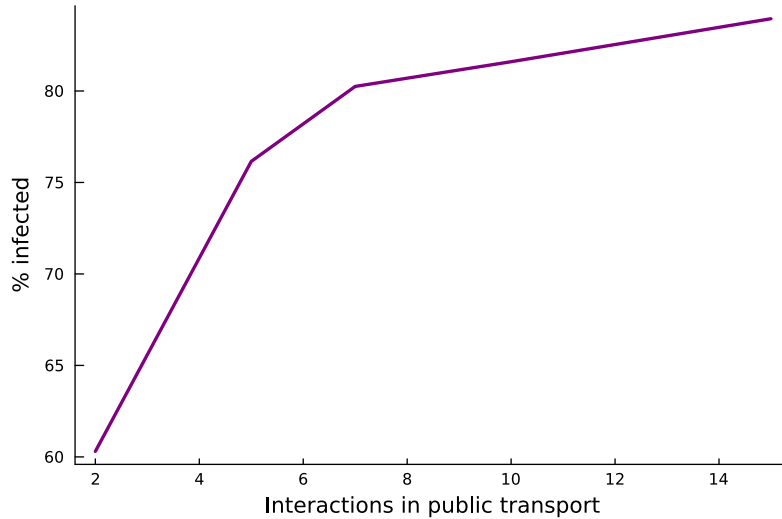
Figure 21: The impact of the number of interactions in TTC (e.g., due to the government's information campaign) on the percentage of infected citizens

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.

From a transportation policy point of view, it can be argued that an extreme increase in public transportation frequency can slow the spread of the virus by leading to more encounters between small groups of passengers. Making public transportation less frequent can be treated similarly as it discourages passengers from unnecessary travel and encourages short-distance passengers to walk instead. However, this strategy has negative implications. It may promote non-ecological car travel and exclude non-car-owners from urban life. In addition, it could be ineffective in the case of a more contagious strain of the virus, as tends to result in a larger crowd formation.

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 to a massive transportation model of the entire city. We believe that our results can help decision makers to understand trade-offs when deciding between various frequencies for public transport in times of an ongoing epidemic.

The presented research can be extended in many ways. Firstly, the ABM approach makes it possible to apply the same model for real-world data on commuter movement patterns (e.g. from GPS devices). We are considering only two means of transportation (walking vs. TTC), while other options such as driving a car or a bicycle could be included. In the proposed model commuters optimize only their travel times — this could be extended to a multi-criteria decision model where travel time could be weighted against their travel costs or environmental awareness. Similarly, the transportation system regulator (e.g. a municipality) needs to take several factors such as budget availability, transportation system capacity, or carbon footprint.

Statements and Declarations — Competing Interests

The authors have no conflicts of interest to declare.

Acknowledgments

References

Aleman, D. M., Wibisono, T. G., & Schwartz, B. (2011). A nonhomogeneous agent-based simulation approach to modeling the spread of disease in a pandemic outbreak. *Interfaces*, *41*(3), 301–315, `https://doi.org/10.1287/inte.1100.0550`.

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98, `https://doi.org/10.1137/141000671`.

Bhadra, A., Mukherjee, A., & Sarkar, K. (2021). Impact of population density on COVID-19 infected and mortality rate in India. *Modeling Earth Systems and Environment*, *7*(1), 623–629, `https://doi.org/10.1007/s40808-020-00984-7`.

Bulfone, T. C., Malekinejad, M., Rutherford, G. W., & Razani, N. (2021). Outdoor transmission of SARS-CoV-2 and other respiratory viruses: a systematic review. *The Journal of Infectious Diseases*, *223*(4), 550–561, `https://doi.org/10.1093/infdis/jiaa742`.

Cliff, O. M., Harding, N., Piraveenan, M., Erten, E. Y., Gambhir, M., & Prokopenko, M. (2018). Investigating spatiotemporal dynamics and synchrony of influenza epidemics in Australia: an agent-based modelling approach. *Simulation Modelling Practice and Theory*, *87*, 412–431, `https://doi.org/10.1016/j.simpat.2018.07.005`.

Coşkun, H., Yıldırım, N., & Gündüz, S. (2021). The spread of COVID-19 virus through population density and wind in Turkey cities. *Science of the Total Environment*, *751*, 141663, `https://doi.org/10.1016/j.scitotenv.2020.141663`.

Cuevas, E. (2020). An agent-based model to evaluate the COVID-19 transmission risks in facilities. *Computers in Biology and Medicine*, *121*, 103827, `https://doi.org/10.1016/j.compbiomed.2020.103827`.

Currie, C. S., Fowler, J. W., Kotiadis, K., Monks, T., Onggo, B. S., Robertson, D. A., & Tako, A. A. (2020). How simulation modelling can help reduce the impact of COVID-19. *Journal of Simulation*, *14*(2), 83–97, `https://doi.org/10.1080/17477778.2020.1751570`.

Daduna, J. & Voß, S. (1996). Efficient technologies for passenger information systems in public mass transit. In *Proceedings of the first INFORMS conference on information systems and technology. INFORMS, Washington* (pp. 386–391).

Dekker, M. M., et al. (2022). A next step in disruption management: Combining operations research and complexity science. *Public Transport*, *14*(1), 5–26, `https://doi.org/10.1007/s12469-021-00261-5`.

Frias-Martinez, E., Williamson, G., & Frias-Martinez, V. (2011). 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* (pp. 57–64).: IEEE.

Ge, L., Voß, S., & Xie, L. (2022). Robustness and disturbances in public transport. *Public Transport*, (pp. 1–71)., `https://doi.org/10.1007/s12469-022-00301-8`.

Gkiotsalitis, K. & Cats, O. (2021). Optimal frequency setting of metro services in the age of COVID-19 distancing measures. *Transportmetrica A: Transportation Science*, `https://doi.org/10.1080/23249935.2021.1896593`.

Hackl, J. & Dubernet, T. (2019). Epidemic spreading in urban areas using agent-based transportation models. *Future internet*, *11*(4), 92, `https://doi.org/10.3390/fi11040092`.

Hinch, R., et al. (2020). OpenABM-Covid19 — an agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. *medRxiv*, `https://doi.org/10.1371/journal.pone.0245787`.

Hoertel, N., Blachier, M., Blanco, C., Olfson, M., Massetti, M., Limosin, F., & Leleu, H. (2020). Facing the COVID-19 epidemic in NYC: a stochastic agent-based model of various intervention strategies. *MedRxiv*, `https://doi.org/10.1101/2020.04.23.20076885`.

Huang, Y. & Shen, Z. M. (2021). Optimizing timetable and network reopen plans for public transportation networks during a COVID19-like pandemic. *arXiv preprint arXiv:2109.03940*, `https://doi.org/10.48550/arXiv.2109.03940`.

Janson, S., Luczak, T., & Rucinski, A. (2011). *Random graphs*, volume 45. John Wiley & Sons.

Kaligotla, C., Stevens, A., Ozik, J., Collier, N., Macal, C., Martinez-Moyano, I., Mucenic, B., Hotton, A., & Choe, K. (2020). Development of large-scale synthetic population to simulate COVID-19 transmission and response. In K. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, & R. Thiesing (Eds.), *Proceedings of the 2020 Winter Simulation Conference*.

Kerr, C. C., Stuart, R. M., Mistry, D., Abeysuriya, R. G., Rosenfeld, K., Hart, G. R., Núñez, R. C., Cohen, J. A., Selvaraj, P., Hagedorn, B., George, L., Jastrzębski, M., Izzo, A. S., Fowler, G., Palmer, A., Delport, D., Scott, N., Kelly, S. L., Bennette, C. S., Wagner, B. G., Chang, S. T., Oron, A. P., Wenger, E. A., Panovska-Griffiths, J., Famulare, M., & Klein, D. J. (2021). Covasim: an agent-based model of COVID-19 dynamics and interventions. *PLOS Computational Biology*, *17*(7), e1009149, `https://doi.org/10.1371/journal.pcbi.1009149`.

Laskowski, M., Demianyk, B. C., Witt, J., Mukhi, S. N., Friesen, M. R., & McLeod, R. D. (2011). 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, `https://doi.org/10.1109/titb.2011.2163414`.

Liu, J. & Zhou, X. (2016). Capacitated transit service network design with boundedly rational agents. *Transportation Research Part B: Methodological*, *93*, 225–250.

Löhner, R., Antil, H., Srinivasan, A., Idelsohn, S., & Oñate, E. (2021). High-fidelity simulation of pathogen propagation, transmission and mitigation in the built environment. *Archives of Computational Methods in Engineering*, *28*(6), 4237–4262, `https://doi.org/10.1007/s11831-021-09606-6`.

Macal, C., Ozik, J., Collier, N., Kaligotla, C., MacDonell, M., Wang, C., LePoire, D., Chang, Y., & Martinez-Moyano, I. (2020). CityCOVID: A computer simulation of COVID-19 spread in a large-urban area. In K. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, & R. Thiesing (Eds.), *Proceedings of the 2020 Winter Simulation Conference*.

Müller, S. A., et al. (2021). Predicting the effects of COVID-19 related interventions in urban settings by combining activity-based modelling, agent-based simulation, and mobile phone data. *PloS one*, *16*(10), e0259037, `https://doi.org/10.1371/journal.pone.0259037`.

Müller, S. A., Balmer, M., Neumann, A., & Nagel, K. (2020). Mobility traces and spreading of COVID-19. *MedRxiv*, (pp. 2020–03)., `https://doi.org/10.1101/2020.03.27.20045302`.

Obuchi, S. P., Kawai, H., Ejiri, M., Ito, K., & Murakawa, K. (2021). Change in outdoor walking behavior during the coronavirus disease pandemic in Japan: A longitudinal study. *Gait & posture*, *88*, 42–46, `https://doi.org/10.1016/j.gaitpost.2021.05.005`.

Ogden, N. H., et al. (2020). Artificial intelligence in public health: Modelling scenarios of the epidemic of COVID-19 in Canada. *Canada Communicable Disease Report*, *46*(8), 198, `https://doi.org/10.14745/ccdr.v46i06a08`.

Rahman, S. M., Ratrout, N., Assi, K., Al-Sghan, I., Gazder, U., Reza, I., & Reshi, O. (2021). Transformation of urban mobility during COVID-19 pandemic–lessons for transportation planning. *Journal of Transport & Health*, *23*, 101257, `https://doi.org/10.1016/j.jth.2021.101257`.

Serdar, M. Z., Koç, M., & Al-Ghamdi, S. G. (2022). Urban transportation networks resilience: indicators, disturbances, and assessment methods. *Sustainable Cities and Society*, *76*, 103452, `https://doi.org/10.1016/j.scs.2021.103452`.

Shamshiripour, A., Rahimi, E., Shabanpour, R., & Mohammadian, A. K. (2020). How is COVID-19 reshaping activity-travel behavior? Evidence from a comprehensive survey in Chicago. *Transportation Research Interdisciplinary Perspectives*, *7*, 100216, `https://doi.org/10.1016/j.trip.2020.100216`.

Shi, X., Sun, W., Voß, S., & Jin, J. (2020). Smart city: A perspective of emergency and resilience at a community level in Shanghai. In *International Conference on Computational Logistics* (pp. 522–536).: Springer.

Silva, P. C., Batista, P. V., Lima, H. S., Alves, M. A., Guimarães, F. G., & Silva, R. C. (2020). 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, `https://doi.org/10.1016/j.chaos.2020.110088`.

Speir, C. & Negahban, A. (2020). Analyzing COVID-19 control strategies in metropolitan areas: A customizable agent-based simulation tool. In K. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, & R. Thiesing (Eds.), *Proceedings of the 2020 Winter Simulation Conference*.

Truszkowska, A., et al. (2021). High-resolution agent-based modeling of COVID-19 spreading in a small town. *Advanced Theory and Simulations*, (pp. 2000277)., `https://doi.org/10.1002/adts.202000277`.

Wallentin, G., Kaziyeva, D., & Reibersdorfer-Adelsberger, E. (2020). COVID-19 intervention scenarios for a long-term disease management. *International Journal of Health Policy and Management*, *9*(12), 508–516, `https://doi.org/10.34172/IJHPM.2020.130`.

Yao, W., Yu, J., Yang, Y., Chen, N., Jin, S., Hu, Y., & Bai, C. (2022). Understanding travel behavior adjustment under COVID-19. *Communications in Transportation Research*, (pp. 100068)., `https://doi.org/10.1016/j.commtr.2022.100068`.

Yusuke, L., Nathan, I., Antonio, R., & Joseph, W. (2020). Utilizing simulation to evaluate shuttle bus performance under passenger counts impacted by COVID-19. In K. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, & R. Thiesing (Eds.), *Proceedings of the 2020 Winter Simulation Conference*.

Zeng, W. & Church, R. L. (2009). Finding shortest paths on real road networks: the case for A*. *International journal of geographical information science*, *23*(4), 531–543, `https://doi.org/10.1080/13658810801949850`.

A   Proofs of theorems from Section 3

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 Janson et al. (2011)). 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 the agents that choose the path P_j on the commute s. Moreover, for $s \geqslant 0$, define Y_s^j to be the number of the 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,

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

To simplify (10), observe that conditional on I_t, the random variables $|I_t \cap D_{t+1}^j|$ and $|D_{t+1}^j \cap \neg I_t|$ are distributed as $\text{Bin}(|I_t|, \alpha_j)$ and $\text{Bin}(|\neg I_t|, \alpha_j)$, 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. $\ell := |I_t \cap D_{t+1}^j| = (1 + o(1))\alpha_j X_t$ and $k := |\neg I_t \cap D_{t+1}^j| = (1 + o(1))\alpha_j(n - X_t)$. Thus, a.a.s.

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

On the other hand, Y_{t+1}^j conditional on I_t and D_{t+1}^j is distributed as a binomial $\text{Bin}(k, p)$ with parameters $k = |D_{t+1}^j \cap \neg I_t|$ and $p = 1 - (1 - \beta_j \lambda_j)^{|I_t \cap D_{t+1}^j|} = 1 - (1 - \beta_j \lambda_j)^\ell$. 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.

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

(Formally, one should stochastically lower and upper bound Y_{t+1}^j by $\text{Bin}(k_-, p_-)$ and $\text{Bin}(k_+, p_+)$ 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.,

$$|Y_{t+1}^j - \mu_j| \leqslant \epsilon \mu_j \tag{13}$$

for each $j = 1, 2$. On the other hand, $X_{t+1} - X_t = \sum_{j=1}^2 Y_{t+1}^j$, 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. $\qquad\square$

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),

$$\frac{X_{t+1}}{n} \leqslant \frac{X_t}{n} + \left(1 - \frac{X_t}{n} \right) \left(\sum_{i=1}^{2} \alpha_i \left(1 - \exp\left(-A_i X_t/n \right) \right) \right) + \delta$$

$$\leqslant \widetilde{x}_t + \delta + (1 - \widetilde{x}_t + \delta) g(X_t/n) + \delta$$

$$\leqslant \widetilde{x}_t + (1 - \widetilde{x}_t) g(X_t/n) + \delta + \delta$$

$$\leqslant \widetilde{x}_t + (1 - \widetilde{x}_t) g(\widetilde{x}_t) + 2A\delta + 2\delta + \delta$$

$$\leqslant \widetilde{x}_t + (1 - \widetilde{x}_t) g(\widetilde{x}_t) + 5A\delta$$

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

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

32

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, $X_0/n = \tilde{x}_0$ 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.

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

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

Proof of Corollary 3.2. Let us first fix $c := \tilde{x}_{t_{c_f}+3}$, 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.,

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

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.

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

On the other hand, $\tilde{x}_{t_{c_f}-1} < c_f$. Thus, for all n sufficiently large, $\epsilon_1(n) < c_f - \tilde{x}_{t_{c_f}-1}$. 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 $\tilde{x}_{t_{c_f}} > c_f$ (part (a)). Using (18) for $k = t_{c_f}$, we get that a.a.s.,

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

Since $\tilde{x}_{t_{c_f}} > c_f$, for all n sufficiently large, $\epsilon_1(n) < \tilde{x}_{t_{c_f}} - c_f$. 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 $X_{t_{c_f}}$. $\quad\square$

B   Approximation for the number of infected agents

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:

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

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

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

C List of symbols used in the paper

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 4: List of symbols used in the paper