Smart and the City: An Exploration of the Impact of Initial Conditions on (Smart) City Planning

Deeksha Singh Vijay Singh

Main academic Supervisor: Professor, Christos, TJORTJIS, International Hellenic University.

A Master Thesis submitted for the Erasmus Mundus Joint Master Degree on Smart Cities and Communities (SMACCs)

August 2022

University of Mons, Heriot Watt University, International Hellenic University, University of the Basque Country

Abstract

Over the last thirty years, literature on smart cities has dominated both academia and industry, however, much of the foundational elements of this field remains ambiguous, subjective, unreproducible and non-quantitatively defined. In light of this, this author first conducted a critical review of existing literature on (smart) city planning to investigate its foundational elements, motivating problem, common themes and barriers. The ensuing knowledge implied that due to a city's inherent chaotic dynamics and presence of tipping points, it is not possible to accurately forecast the possible future states of a city (for specific initial conditions and intervention context) without using non-traditional tools, such as agent-based modeling. This author then implemented a simulation study of four city contexts under stress in order to deconstruct how the crisis trajectories of each was affected by their initial conditions and intervention strategies. Since the motivating problem of the crisis (a disease outbreak) is a symptom of rapid urbanization, much like most cities which need the smart city transformation, the author then ventured to explicate whether each of those four cities would benefit from becoming a "smart" city and, if true, assuming they have the same political and socioeconomic context: would the smart city design for each of those cities be the same? Based on these three modes of analysis, this thesis was able to, via case study, highlight the link between initial conditions, optimal intervention strategies and crisis trajectory, and posit that, if such an impossibly simple model of four improbably similar cities fails to validate the veracity of the one-size-fits-all smart city planning conceit, then, based on the nonlinear dynamics, feedback loops and emergence phenomenon present within real-world cities, such a generalized foundational framework would not be optimal in the real world. Lastly, this thesis also provides a quantitative and holistic analysis framework for smart city planning built upon reducing the dimensionality, and feature set cardinality, of city data.

Keywords: smart city planning, epidemic simulation, complexity theory of cities, agent-based modeling, smart city analysis

Deeksha Singh Vijay Singh

22nd August 2022

Table of Contents

1 Introduction

"Similarly to many material and organic natural systems, cities too are complex self-organizing systems. But what about the uniqueness of cities— of the properties that differentiate them from material and organic entities, how do these are [sic] *related to their complexity and dynamics?"*
- Complexity Theories of Cities Have Come of Age [1]

According to Google Scholar data, around 50 percent of all publications, from 2017 to 2019, related to urban planning, mention smart cities. Despite this, a lot of smart city planning is shrouded in the unknown [2].

Due to the evolutionary nature, multilevel complexity, and socio-technical organizational elements within smart cities, traditional urban planning measures are no longer enough to contain the entirety of the smart city concept, with respect to its theory, function and practice. Thus, there is a call for a new type of smart city planning. However, before such a revolution can take place, it is important that the foundation of the smart city field be stabilized and confirmed by the majority of its practitioners.

However, despite the significant resources and research conducted on trying to define the *what*s and *how*s of smart city design, there is still no convergence upon a definition, metrics or framework that can hold true for the abstracted, general case of transforming a city into a smart city. So, a question arises: is such an abstraction even feasible or logically sound within the real-world context of cities, considering their innate complexity?

This thesis has a threefold aim, to, 1) investigate, via a review of the state of the art: the motivation behind, themes from, and existing barriers within the field of (smart) city planning, 2) deconstruct, via a simulation-based case-study, how exactly crisis trajectory is affected by the initial conditions (of the analysis, city and crisis context) and chosen intervention strategies, and 3) explicate, by virtue of the theoretical and case-study led results, if a one-size-fits-all smart city solution strategy is optimally applicable to the four chosen simulated city contexts. Based on the three avenues of analysis, this author then recontextualizes the implications, born from the designed model, on the smart city field at large.

This thesis begins with a review of the state of the art of (smart) city planning, pointing out the ambiguous foundations within the smart city field and the lack of quantitative and holistic research within it. It also stresses the importance of the field which was motivated by the need to solve the downsides of unsustainable rapid urbanization and the environmental and social problems, which, if left unchecked, would adversely impact the quality of life and wellbeing of the people within those cities.

This thesis then analyzes city planning through three lenses: ecology, complexity and organization. The first two perspectives explain why multilevel complex systems, such as cities, are sensitive to initial conditions, never stabilize long-term and have unpredictable dynamics and future states. The organizational perspective goes on to suggest the importance of urban sustainability as a vision for smart city planning and using evolutionary planning-without-a-plan strategy to deal with the three types of unknowabilities in city planning.

The middle of the thesis contains the design and analysis of a simplified model of a city under stress, concluding that, much like real cities, the model too is sensitive to initial conditions, has feedback loops and subcomponents which co-evolve competitively. An analysis framework is identified for quantitatively and holistically recognizing which hyperlocal parameters most affected the success metric and these results were compared across two cities of different scales.

Lastly, this thesis was able to better justify the instinct that a one-size-fits-all solution would not be recommended when planning for smart cities, as every city (and problem) has a different context and it is more important for a smart city solution to be customized to the individual needs of, and people within, the city.

2 The Smart City

The section examines the existing literature with respect to three key questions 1) what it means to be a smart city, 2) what problems motivated the need for the smart city concept, and 3) how to identify a smart city.

2.1 “Smart” City Definition

The field of smart cities is a burgeoning one with limitless research potential, which, if utilized democratically, has the potentiality to disrupt and elevate the way individuals live and engage with their micro and macro world.

But to do so, one must first contend with the fact that currently there is no generally accepted definition of what exactly constitutes a “smart city” [3]–[6]; This foundational ambiguity is further exacerbated by the perspective, discipline and background of the authors who are invoking the concept of “smart cities” in order to help shed light on a specific problem context [7].

As a primary consequence of this foundational ambiguity, there is often confusion between the concept of “smart cities" with similar - but non-equivalent - concepts such as “intelligent,” “knowledge” or “digital” cities [3], [5], [6]. Which in turn muddies the applicability and reproducibility of the research and resultant models in another context. This intuition is validated by the reality that in the past three decades of research (a timespan which constitutes the majority of smart city-based academic and industry ventures), a large majority of papers dealt with either conceptualizing and qualitatively defining the field (rather than quantifiably or mathematically), or focusing on the underlying technical aspects (rather than its societal implications) [6].

As a secondary consequence of not having a conclusive definition of smart cities is the manner in which this concept is siloed (with respect to application) and, as mentioned in Pereira *et al.*, this may be a reason as to why the majority of smart city research focuses on specific dimensions of the concept rather than treating it holistically [8], [9]. The complexity of a more holistic analysis is often cited as a barrier [3].

Some authors have recognized this issue and attempted to reconcile the different perspectives and sources on the existing body of smart city-based research and applications, offering possible metrics and frameworks for evaluating smart cities. However, despite the considerable work on stabilizing the theoretical framework of the field, there are mutually exclusive views present, even within the prominent literature, on several key metrics and issues.

For example, considering the first and second most influential papers in the field, by co-citation analysis [6]: the 2011 paper by Carigliu, *et al.* and the 2015 paper by Albino *et al.* respectively. The former’s theoretical framework of a smart city - formulated on findings from a project at the Vienna University of Technology, which identified six key dimensions (smart mobility, smart environment, smart people, smart living and smart governance [10]) by which to rank seventy European mid-sized cities - finally concluded that, "a city is smart when investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance" [11].

However, the latter paper by Albino *et al.* posits that a potential reason for lack of consensus for the smart city term may be because any working definition needs to accommodate both "hard" and "soft" domains (the former consist of areas where ICT can "decisively" impact the functioning of the systems (such as buildings, energy grids, etc.) and the latter contain areas which are not as impacted by ICT (education, culture, policy innovation, social inclusion, etc.); after citing the different available definitions for smart cities (including the definition by Caragliu *et al.*), acknowledges that any definition must accommodate the multi-faceted nature of smart cities (i.e., within both hard and soft domains), but concludes that any measurement framework for a smart city must depend on that particular city's vision and objectives, and that existing definitions by self-identified smart cities lack universality, moving on to prove that many smart city ranking systems propagate a non-trivial loss of information due to the inherent complexity of smart cities [3]. The first paper's focus on using the same set of metrics by which to universally measure the smartness of every city clashes against the second paper's claim that even the measurement metrics for smartness should specifically depend on every city's individual context lest important information be lost.

As a quick note, it is important to understand that due to the rapid and continuous growth in "smart" technology, over the span of a few years, many smart-technology-led definitions (i.e., which were not technology-agnostic), much like their specified technologies, fell out of favor [12].

Thus, despite existing devoted research focused on the conceptualization of smart cities, there is still a long way to go before there is any strong convergence on the foundational elements of this field. This is an urgent source of concern, as this would hinder interoperability between independent research within the field and potentially act as a barrier to mainstream application.

2.2 Smart City Motivation

According to the 2018 World Urbanization Prospects report by the United Nations, the world's urban population has increased from 30% in 1950 to 55% in 2018 to an estimated 68% in 2050 [13]. This urban population growth - in part due to the increase in overall population and the upward trend of more people choosing to reside in urban areas - is set to hasten in an age of "rapid urbanization" especially in lower and lower-middle income countries [14]. And, as Ignatieff assures us, this rapid urbanization will result in cities (and megacities) which will be "multilingual, multiracial, and multicultural" [15]. This sentiment is seconded in Echebarria *et al.*'s 2020 review of eighty-four works within the smart city literature, from 1997-2020, wherein the background industry's trend analysis supports the hypothesis that currently the world is undergoing a "third revolution in urban development" wherein cities "stop being passive human settlements and start generating an indigenous force through their creative and innovative potential" [5].

However, as a consequence of this revolution - resultant from accelerated economic and technological growth in urban areas - new challenges have emerged that necessitate the concept of "smarter" cities; Challenges ranging from general environmental issues, such as air pollution, traffic congestion, and poor waste management, to more societal problems, such as dearth of, and unequal access to, resources and increasingly disproportionate social inequality [5], [16], [17].

There is substantial evidence to believe that the current environmental concerns are deeply interlinked with the consequences of human actions and rapid urbanization. As explained through their river delta planning study of the forty largest deltas globally, Sijmons *et al.* were able to identify that all those areas shared "an analogous cluster of spatial puzzles," wherein urbanization, due to its spatial nature, was deeply entwined with the current global environmental problems; this paper offered the opinion that "the web of urbanisation is patent evidence for the fact that humanity can be seen as a formidable force of nature on a planetary level" - this view of humanity's unignorable and nontrivial impact on the environment is echoed by other authors [18], [19]. Sijmons *et al.* go on to offer the stronger claim that those global environmental problems would not be solved unless the urban problems were solved. Thus, any city planning must take this dimension into consideration, and should recognize that it is not feasible to use traditional methods of urbanistic research and design for problems of this magnitude and complexity [20].

The smart city concept was in part motivated by the socio-economic need to positively leverage the benefits of the natural rapid urbanization already underway, while also covering for its downsides via “smart computing technologies”, processes and cyber-physical systems [21]–[23]. The rapid, real-time growth of the aforementioned technologies was a key reason for the industry's confidence in the readiness of smart city solutions [22].

In short, the rapid urbanization and the consequential problems within modern cities, mixed with the growth of novel information and communication technologies (ICT) proved to be the impetus that motivated the smart city concept while also providing a path for a feasible execution [22].

2.3 Smart City Identification

Since a generalized and accepted definition is not currently available, a possible solution could be to agree on common dimensions which, if present in a city, would then identify it as “smart.” However, this too is difficult in execution, as most existing characterization frameworks are either holistic or quantitative, but rarely both.

For example, this potential avenue for recognition of smart cities was studied by Albino *et al.* in their 2015 paper (which evaluates the available post-2008 published peer-reviewed literature) enumerating key dimensions that exist within the realm of smart cities. The paper identified that many researchers in this thematic area agree that “in a dense environment, like that of cities, no system exists in isolation” but noted that despite this assertion, smart cities are seldom evaluated in their totality [3]. Additionally, Albina *et al.* were also quick to emphasize the literature that focuses on the importance of human capital on a smart city‘s goals and execution, writing: “the label “smart city” should refer to the capacity of clever people to generate clever solutions to urban problems” [3]**.** Consequently, based on their literature review analysis, the most commonly identified characteristics of a smart city were: “a city’s networked infrastructure that enables political efficiency and social and cultural development; an emphasis on business-led urban development and creative activities for the promotion of urban growth; social inclusion of various urban residents and social capital in urban development; the natural environment as a strategic component for the future” [3]**.** However, while interdisciplinary and holistic, this qualitative characterization introduces significant ambiguity and subjectivity in the smart city qualification process, with the lack of quantitatively defined key performance indicators making it difficult to compare across cities.

On the other hand, when quantitative analysis is conducted in the smart city field it is seldom holistic and instead only addresses one sub-discipline (typically technology-related) within the issue. For example, several of the studied bibliometric and scientometric research were able to prove that technical engineering was the most explored discipline within the smart city literature [4] and that most co-citation analysis related smart cities to specific technological solutions (such as internet of things, data analytics, etc.) [4], [6], [24].

Thus, there is a noticeable lack of both quantitative and holistic analysis of the smart city concept. There are both practical and theoretical reasons for this research gap.

Considering the practical barriers within the execution of such an analysis: as discussed earlier, smart city researchers tend to talk about the concept from their field of expertise and due to the aforementioned lack of clarity and universality in the foundational elements of the field, it is often not possible to perform quantitative meta-analysis on the results. For example, as reasoned by Echebarria *et al.* in their literature review analysis wherein the authors congregated the existing literature into a descriptive (versus, say an econometric or statistical) meta-analysis cited that the degree of interdisciplinarity and heterogeneity (with respect to their dates of publication, theoretical frameworks and research methods) within the sampled papers did not lend themselves to a more quantitative approach. The authors also go on to assert the need for a holistic analysis of the field and identified holistic analysis as the future the industry was converging towards [5].

As for the theoretical barriers to a more quantitative and holistic analysis: as further explained in the complexity section, a city's properties are the same as that of a natural complex system (i.e., "open, complex and self-organized, and often fractal and chaotic") and, the 2016 paper by Portugali goes as far as to assert that "a city is a simple-complex system" arguing that a city is an artifact (and thus, a simple system) if only its material components (such as buildings, roads, etc.) are considered. However, by virtue of the city's human components (i.e., its urban agents) who have cognitive and planning capabilities and participate in complex interactions, this artifact is made a complex system [25].

Due to this complexity, also echoed by other authors [1], [23], [26]–[30], there is no convenient dependence on the assumption of a separation of scales; separation of scales makes it possible to assume that mean field theory can be accurately applied or that network interdependencies will not lead to fat-tailed behavior [31], [32]. This inability to depend on Euclidean geometry and Gaussian statistics for guidance and ensuing mathematical need to take the road less researched, with fractal geometry, power law statistics and Alexander's living geometry when analyzing cities, is often cited as a barrier to research [31]. As an added consequence, due to this lack of dependence on the separation of scales concept and cities being subject to power law dynamics, any non-holistic analysis for cities might not be accurate long-term.

In order to better understand the ramifications of the above choices, it is important to first understand the base unit: the city.

3 The City from 3 Perspectives

A city is often referred to by its dynamics, complexity and organizational architecture. For example, the 2019 paper by Komninos *et al.* defined cities as "complex systems shaped by bottom-up processes with outcomes that are hard to foresee and plan for." [2]. Thus, in order to better understand the base unit of a smart city, it is vital to first reflect on it as a function of its ecology, complexity and organization.

3.1 Ecological Perspective

This section aims to understand cities from an ecological perspective. The first subsection attempts to categorize what makes a city more like an ecosystem than like a living organism and why that distinction matters. The second subsection then builds upon that qualification by exploring the dynamics-related implications of cities having complexities akin to that of an ecosystem.

3.1.1 A City as a Problem of Organized Complexity

Jane Jacobs' now classic 1961 book *The Death and Life of Great American Cities* cemented the link for many between cities, biology and organized complexity [33].

However, a closer analysis of the surrounding context behind the term ‘organized complexity’ (such as, the four subtypes of this term: artefactual, system, biological and ecological complexities, and the inherent fuzziness of the boundaries between ‘object’ vs ‘system’ or ‘natural’ vs ‘artificial’) is enough to highlight that the city-as-an-organism analogy isn’t convenient or informative enough to be used for understanding, or planning within, cities due to the different *type* of complexity that cities actually display [34].

For example, when considering the nuance between a biological system (as referring to an organism) versus ecology (as it refers to an ecosystem); an organism is finite and defined as being in ‘equilibrium’ if it is stable in its function, i.e., it changes in a manner predictable according to a developmental process. An ecosystem, by contrast, is indefinite in extent, due to it being made up of co-evolving sub-components, and it can never be in a state of equilibrium; even if it is seemingly stable for a short period of time its state is not predictable in the long-term [27]. Thus, a city is not so much an organism, but rather, an ecosystem [35] with a complexity born from its sub-components that function competitively [34].

A city can be identified as an example of system (organized) complexity, which also leans more towards ecological complexity than the alternatives; Generally associated with artificial open systems, examples of system complexity are characterized by changes over time and are typically not pre-designed as a whole from the very start, but rather emerge from the interactions between the different actors within the system. Whereas ecological complexity are *natural* open-ended systems. Using an illustrative example of the subtype: ecosystems are dynamic to a somewhat unpredictable extent and the relationships between the constituent elements may change as they evolve. Formally, the complexity of an ecosystem is multi-faceted, open, adaptive, having nonlinear dynamics and irreversible history [34]. This perspective on the organized complexity subtype is helpful when studying, and intervening for, cities.

If the city is the human habitat, with urbanization as its habit and the physical and functional relationship with the environment as its metabolism analogous, then the city can be viewed as human ecology, which, as Sijmons defines in their 2012 paper, is “a complex ecology that includes language and technology, and that produced and continues to produce its spatial organization as an emergent order” [20].

Thus, the study of city dynamics can be abstracted to an ecological perspective as city dynamics are, when abstracted, a result of the human (AKA urban agents) and non-human (AKA material components) interacting with their environments and each other [25].

3.1.2 A City as Subject to Evolutionary Processes

Evolutionary perspective of urban growth can be applied to smart city planning and evolutionary processes heavily influence smart city planning. This was argued by Komninos *et al.* in their 2019 paper wherein evolutionary processes are "characterized and affected by essential diversifications in the capacity of societies to generate technical innovations that are suitable to their needs" [2]. The impact of evolutionary processes on smart city planning is in part due to the underlying socioeconomic and political context of the city but also due to the behavior of people in the communities within that city.

When cities formulate policies that can best leverage existing funding and research while also attracting more externally-sourced opportunities and investment, they are attempting to optimize drivers of innovation and complexity. Thus, these policy decisions in turn affect the developmental trajectory of the city. However, the "selection environments" or the realistic paths available for leveraging and exploiting innovation at that time period based on public funding, governmental rules, etc. are constrained by the socio-political and economic climate as well as the efficacy of the stakeholders at that time. Or, as explained by Komninos *et al.*: "urban contexts influence the ways in which local governments can create and shape opportunities for innovation" [2].

Moreover, societies are, by their very nature, evolutionary and adaptive; Pelletier characterizes societies as having worldviews and attendant behavior sets which are "variably successful in perpetuating specific configurations of social relations and human/nature interactions within the constraints of particular social and environmental conditions." and goes on to conclude that a civilization's ability to survive long-term is largely dependent on either the stability of those social and environmental conditions or its ability to adapt as those conditions change [36].

As mentioned earlier, a city is made up of human and material components which interact with each other in a complex way. By themselves, the human entities, when considered collectively, form that city's community (or, society). Thus, it can be claimed that considering the lack of long-term equilibrium (or stability) due to the nature of complexity in cities (as explained in the previous section), a city community's ability to survive depends mostly on its ability to adapt to changing circumstances within their individual and global context.

Thus, the evolutionary and adaptive entity (also known as that city's community), in the context of the complexity and multidisciplinary nature of cities [2] gives credence to the finding that city planning is, in turn, shaped by a city's evolutionary processes.

3.2 Complexity Perspective

Just considering the perspective of Sijmons's 2012 consideration of a city as a representation of human ecology, large scale urbanization processes fall under the purview of complexity theory [20]. However, even a wider context - the evolutionary dynamics inherent within a city, the identification of cities with subtypes of organized complexity and the categorization of a city as an urban ecosystem - all provide an even stronger argument for complexity theory as the next step in the modern study, and planning, of cities.

Complexity Theory of Cities (CTC) is a relatively new field which attempts to better understand urban dynamics, that is: how cities work [1]. In its three decades of study, CTC was able to unify a variety of urban phenomena and properties under a single theoretical umbrella; phenomena and properties, which until then, hadn't been thought to have been dependent on each other (or even influenced by the same base theoretical forces). Moreover, CTC was also able to leverage concepts, defined within complexity science (such as, chaos, emergence and nonlinearities), which have application in urban dynamics and helped clarify that chaos and order in cities are not mutually exclusive [37].

While CTC may have some credible gaps in the current state of the art (the ability to understand what makes a specific city unique being one of them [1]), the field and its practitioners are far from done with recontextualizing urban dynamics (and more recently, urban planning) through the lens of complexity science. It would behoove any smart city practitioner to familiarize themselves with this burgeoning field which has a novel point of view and relevant quantification tools to offer.

The following subsections reframe the city as multilevel systems of systems of systems, a complex adaptive system and an instance of agglomeration effects, and discusses how all those three traits can coexist and add further complexity to a city's dynamics.

3.2.1 A City as Multilevel Systems of Systems of Systems

In Johnson's 2012 paper, they explained that cities are highly entangled, multilevel systems of systems of systems which are hard to understand and predict dynamics of, and have coupled subsystems (wherein changes in one subsystem may affect, or be affected by, changes in another). These traits make the subsystems have "ill-defined boundaries," which adds ambiguity to the ownership process - i.e., not being able to confidently identify whether a property, or consequence, can be uniquely attributed to a specific subsystem [8].

It is important to note that the subsystems that are a part of the city, such as the transportation or energy system, are also innately complex and feed into the overall complexity of the city [38].

As explained earlier in the thesis, and is further validated by Johnson [8], cities are typically planned with an amalgamation of technical, verbal and pictorial analysis; with a projected goal state the city should be in the future, and identifying which interventions can achieve that state. Which, as explained independently by Johnson and Marshall, has mixed success due to the types of 'unknowabilities' birthed by the type of complexity in cities [8], [34]. Thus, many authors advocate the need for network theory, collective intelligence, agent-based computing to deal with such systems (of systems of systems) [8], [39]–[44].

3.2.2 A City as a Complex Adaptive System

A Complex Adaptive System (CAS) is a system with a large number of interconnected subcomponents with nonlinear dynamics in which the behavior of the whole system cannot be predicted from the behavior of its individual subcomponents. A CAS tends to be adaptive and self-organizing, and can learn from past events [45], [46].

A city is a Complex Adaptive System [30], [38], [47], [48], with the people within that city reacting to changing infrastructure, policies and general context, while also interacting with their environment and each other [46].

Furthermore, as a consequence of a city being a CAS, it is argued that the self-organization and emergence tendencies within a city can be leveraged to design strategic interventions in order to bring about certain outcomes and make the city more adaptive [38].

Other consequences of planning for, and studying, cities is that by virtue of its complexity, a city is sensitive to initial conditions [39] and its individual layers (or subsystems) are interconnected in a nonlinear manner that makes the eventual outcomes very difficult to predict [32]. Even in a simple model of a city, the behavior trajectories are chaotic due to the system's inherent unpredictability, resulting in the modeler's inability to forecast (sans a simulation) which specific initial conditions will be the most interesting (i.e., fruitful in giving specific dynamics and/or future states, if those are even possible at all) [27].

So, to recap: a city's future state is highly dependent on the city's initial conditions, but it is impossible to accurately predict the result states possible from a specific iteration of initial conditions and, conversely, it is not possible to backtrack and identify which initial conditions will cause a certain future state, and at tipping points, due to the resulting emergence, the system may even tip into an unpredictable and qualitatively different state [27].

3.2.3 A City as an Instance of Agglomeration Effects

Literature has proof linking the so-called 'agglomeration effects' as a foundational concept for explaining the emergence and persistence of cities globally [49]–[51]. Explained by Ortman *et al.* as the balance between centripetal and centrifugal forces, or in the context of the city: the socio-economic benefits of having a densely populated area versus the associated costs. For cities, agglomeration effects parallel the average socio-economic performance and infrastructure characteristics within cities with changing city size [29].

This result is further supported by other research findings, such as:

- Socio-economic indicators are super-linearly connected (by virtue of agglomeration non-linearities) with population size, enabling larger cities to become centers of innovation, wealth and crime [52]. They have a scaling exponent (on a log-log scale) of $\boldsymbol{\beta} \approx 1.15$ (> 1)
- Material infrastructure metrics are sub-linearly linked to a city's population size, implying a kind of "economies of scale". They have a scaling exponent (on a log-log scale) of $\boldsymbol{\beta} \approx 0.85$ (< 1) [53], [54]

- However, metrics which are dependent on individual human needs (total employment, water or electricity consumption, etc.) have a scaling exponent (on a log-log scale) of $\boldsymbol{\beta} = 1$ [53], [54]
- Local urban dynamics display long-term memory, with cities maintaining their size-enabled advantages (or disadvantages) for decades [52]

The above scaling exponents are useful when trying to ascertain information for infrastructure and socio-economic scores for a target city based on the same metrics' corresponding scores in another base city. However, this comparison is only applicable when comparing cities within the same national boundary [55]. Moreover, it is important to note that these scaling exponents are only applicable in cross-sectional, and not temporal, scale [56], [57].

These scaling exponents are a result of two competing forces: the economies of scale influencing the infrastructure metrics versus the wealth and innovation-creating forces born from social interactions between people in cities which majorly shape the socio-economic metrics [54]. According to the growth models by Bettencourt *et al.,* any city for which economies of scale forces are more influential will eventually stop growing as the population reaches a finite carrying capacity. On the other hand, in cities with growth driven by innovation and wealth creation, there will be an infinite population within a finite time which, when constrained by the existing limited resources, will lead to stagnation and eventual collapse of that city. The only human-led action that can then protect that city is by inspiring accelerating cycles of innovation which will initiate a new period of wealth creation and innovation-driven growth. However, this too is not sustainable long-term as the time to singularity shortens each time a new cycle of innovation is initialized [54].

When placing this disconcerting knowledge within the context of tipping points from the prior sections, it is clear that attempting to identify the exact initial conditions which can facilitate an environment for the urgently needed "good" tipping points (which will initiate a new cycle and stave off collapse) would be no mean feat.

3.3 Organizational Perspective

The following sections shed light on the purpose, unanswered foundational questions and key areas of unknowability within cities (and ergo, also within smart cities).

3.3.1 A City as an Unsustainable Liability

While many authors argue that a smart city's execution strategy must be shaped by the specific goals, vision and mission of that specific city in addition to its culture, context and citizens [2], [3], it might be useful to also highlight a generalized smart city vision in the context of rapid global urbanization. This section argues that (urban) sustainability is one such global metric that every smart city planner must keep in mind during designing, implementing and operating a smart city; a view that aligns with many researchers in the literature [2], [5], [12], [38], [47].

As pointed out by the 2019 United Nations reports, rapid urbanization requires sustainable urbanization in order to fully leverage the positives of this phenomenon and avoid the negatives [13], [14]. Furthermore, as pointed out by Yin *et al.* through their literature review, this umbrella of sustainability must also be able to provide information and decision-making tools for governance, citizens, businesses and the environment [22].

As validated through the earlier section on agglomeration effects, continuing cycles of innovation are necessary for the survival of an innovation-driven city, which is growing at a faster-than-exponential-rate, in order to avoid stagnation and eventual collapse. As West explains in his book *Scale,* that in order "to sustain continuous growth the time between successive innovations has to get shorter and shorter. Thus paradigm-shifting discoveries, adaptations, and innovations must occur at an increasingly accelerated pace. Not only does the general pace of life inevitably quicken, but we must innovate at a faster and faster rate!" [55]. However, even if this unsustainable pace can be somehow maintained, every time there is a new urban problem which requires a disruptive solution (which will then kickstart the new cycle of innovation and stave off the imminent collapse), there will be hysteresis in the changing of public opinion, especially when it's regarding paradigm-shifting ideas or innovations (which often accompany the accelerating cycles of innovation). This hysteresis property makes it difficult to adapt to new problems, or prioritize newer problems with more urgency, quickly because of the accompanying inertia. As identified by the model by Scheffer *et al.,* public opinion changes nonlinearly and has tipping points, this, in turn, is further complicated by the lag between recognition and regulation of a problem, the latter of which is shaped by the political climate. Their model also goes on to explain the significance of the hysteresis points, claiming that "homogenous societies with strong peer control will remain locked into inaction until relatively high problem level. Once active, the reverse switch to inaction is delayed until perceived costs are quite low. Thus there is a tendency to hang on to old problems until they are thoroughly solved" [18]. The resulting spillover costs from this inertia in public opinion might, in turn, pose new problems which could aggravate the already existing environmental and social concerns.

So, the problem and the solution are the very same. For the global community to survive, we need accelerating cycles of innovation, resulting in a faster pace of life (among other consequences), however, this very acceleration facilitates rapid urbanization (due to agglomeration effects [29]) which in turn negatively affects the environment, people and planet. By the very nature of this paradox, this feedback cycle isn't sustainable as it currently stands.

The answer to this need for, and yet fear of, unconstrained growth according to ecological economics, as elucidated by Pelletier, is an organized and restructured attempt at sustainability:

> "The insights of ecological economics, founded on a recognition of the implications of the Laws of Thermodynamics for human organization, point towards a partial recourse to the pathologies of industrial society and clear direction for a more effective form of environmental governance. Although this perspective does not overturn the spectrum of problematic assumptions inherent in the modernist enterprise, it does effectively constrain their most environmentally pernicious potentials by challenging the concept of unconstrained growth. Moreover, it makes explicit the recognition that sustainability is a global, community concern that transcends the capacity of market-mediated instrumental rationality to provide. It also reveals that industrial society can only be assured of long-term viability if we restructure our economic activities with respect to the absolute biophysical limits to sustainability inherent in a finite environment." [36]

Many authors [2], [18], [38], [47], [58] are pursuing this avenue of research. Using sustainability as a barometer for smart city planning is a common theme that the field appears to be converging upon [12].

3.3.2 A City via Uncertain Everyday Decision-making

When talking about smart cities, authors tend to focus on the complexity and inherent uncertainty in the surrounding context [2], [34], [48]. For example, by categorizing a city as a type of open system, Marshall identified three types of ‘unknowabilities’ that arose due to the inherent complexity of such a system [34]. They are, unknowability of…

1. The system as it is
2. Effects of intervention, and
3. Optimal future state

The first unknowability is due to needing to take into consideration the wider (or even global) scope when talking about a local urban ecosystem as a city does not exist in isolation. Moreover, a city is seldom the result of a single planner or design team (thus there is no single source of ownership one can reference, poll or consult when needed), nor is there any consistency or predictability in the patterns of change when studying the interacting components that make up a city which, may in actuality, change their roles at any time.

The second unknowability deals with the inability to accurately predict the long-term effects of any intervention strategies in cities. Unlike biological organisms, which have a reasonably stable lifecycle and it is easier to know the causal effects of (for example) medical strategies when applied to their bodies, cities are not the same.

The last unknowability is due to not knowing the optimal future state that the ideal ‘matured’ city (or ecosystem which, similarly, is made up of many competitively co-evolving species) should have.

Each of the above unknowabilities is one step worse than the one before it, adding ambiguity to the city planning process. For, if it is not possible to know the initial state, the causal effects of a strategy or the optimum goal to steer towards, how can one plan?

This finding is also recognized as a problem by Komininos *et al.* who then argue that the current wealth of a cumulative and interdisciplinary city planning knowledge-base, juxtapositioned with the heterogeneous, uncoordinated and oft-unintegrated digital technologies, in addition to the fragmented and diverse novel producer and user behavior, has moved city planning from the well-trodden roads of traditional planning into a realm of “planning without a plan” wherein cities are grown through evolution (versus detailed plans of the entire city lifecycle outlined from the very beginning). Their proposed “planning without a plan” approach (with all the implied associated uncertainties and chaotic interactions born of multiple organizations, with shifting agendas and priorities, acting in parallel) is led by novel technologies - which are now available for smart city usage - and their influence on the innovation ecosystems, and also leverages opportunities that appear over time, which in time shapes into a final result that could not have been foreseen at the start of the evolutionary process [2].

4 The Model

As has been indicated by the investigation into the existing literature, due to the existing power laws and nonlinear dynamics inherent within cities (which make forecasting impossible at worst and inaccurate at best), non-traditional tools, such as agent-based modeling, are essential for mapping out the sample space of possible future states for a given initialization of initial conditions and intervention strategies.

This section delineates a design for an agent-based model which simulates, via cellular automata, a disease outbreak crisis, within a closed population, in a city. This is done in order to better understand which, if any, of the available initial and/or intervention conditions, within the simulated city, disproportionately affect the crisis trajectory.

The model borrows from existing work [59]–[61] regarding models on, and smart city applications within the field of, epidemic management. The former of which was used to validate the veracity of the designed model by virtue of the base dynamics (the plot for %healthy and %infected over time being S-shaped for a closed population with post-recovery immunity, or having oscillatory dynamics otherwise), and the latter literature was used to validate the feasibility of the selected intervention strategies based on their documented use within real cities.

Furthermore, the model assumes a certain degree of “smartness” in the simulated city: open data and resource transparency that facilitates instantaneous hospital matchmaking and accurate data on infected adults (and their home location), automated contact tracing, and real time notifications to every resident in the simulated city which allows them to follow zoning policies without any lag.

The Overview, Design Concepts and Details (ODD) Protocol [62] was followed below to describe the implemented agent-based model. The first section provides the problem and solution context for the model and its design, which is then compounded upon by a succeeding section on the utilized design concepts, then followed by a section that includes the execution and other miscellaneous information in order to understand and replicate the model. The last two sections highlight the design decisions and assumptions for the model.[1]

4.1 Overview

The following subsections outline the purpose of the model, as well as the simulation world, variables and processes present in it.

4.1.1 Purpose

As seen in the prior sections, a general, widely-accepted smart city definition - i.e., a universally agreed upon one-size-fits-all characterization of what makes a city smart - is currently not available. Thus, this author, instead of expending further resources on this already well-studied avenue of research, attempts to quantitatively understand whether such a “one-size-fits-all definition” is even a desirable quality when talking about smart city implementations.

Since the advent of smart cities was chiefly motivated by a need to solve problems that arose in cities due to rapid urbanization (as seen in the earlier sections), a simulation of one such problem - a disease outbreak - was designed and city response parameters were changed in order to see if the same intervention strategies proved equally fruitful in all the cases with different outbreak-specific parameters (after controlling for the city-specific context, such as population and its dynamics).

This simulated case study was designed in order to 1) analyze a simplified model of a city under stress, 2) identify techniques for quantitatively and reproducibly recognizing which hyperlocal and/or intervention-related parameters, if any, in such models disproportionately affected the success metrics, and 3) understand how smart city design would change in different (simulated) city contexts.

Bluntly speaking, the goal of this thesis is not to equate or draw conclusions about real-life outbreaks from the below simulation, as such an endeavor would veer into oversimplification of a deeply complex phenomenon with little to no real-world impact. Rather, this thesis aims to quantitatively prove the broader claim that the reason no general definition of a smart city exists is because such a definition would be

[1] The model can be found at: https://deekshasinghvs.github.io/disease-outbreak-model/index.html

counterproductive to the conceit of the smart city concept and every smart city implementation must take into consideration the specific city's context, vision, mission and residents.

As explained by Taleb, while a wrong ruler might not be able to accurately output the height of a child, it will nonetheless be able to identify whether that child's height is increasing [63]. By not depending on the model to be 100% accurate and true to real-world conditions (which would be difficult to do, much like a map is not the territory itself, no model can comprehensively represent the dynamics of a real-world system), and instead only behave as a signifier of growth progress, we allow ourselves to move from failing at being predictive to reasonably succeeding at being prescriptive.

If a one-size-fits-all intervention strategy proves successful regardless of the dynamics of the problem in question, then this large-scale simulation study will recommend the need to invest more resources into finding that elusive generalized solution. On the other hand, if such a premise is negated, by case-study, then an argument can be made that instead resources should be redirected to find a bespoke smart city solution for every city based on the needs, dynamics, vision and mission of the city under question.

4.1.2 The Simulation World

Netlogo (version 6.0.2 [64]) was used to simulate the premise with an accompanying Python script which automated the simulation instantiation and execution. The following sections detail the general design within the Netlogo environment.

Netlogo was chosen for the ease and speed of prototyping both spatial and temporal dynamics, its free and open-sourced nature, and ability to integrate with existing automation tools (such as Python) via Application Programming interfaces (APIs). Netlogo's available functionality for simulating independent actors who behave and interact according to predefined micro-level rules in a "physical" space merged with its frontend interface allowing a holistic white-box-esque analysis per discrete-time step (aka a tick) enables researchers to easily identify emergent patterns and macro-level interaction dynamics.

The Netlogo environment provides us with a user-sized grid consisting of - as relevant to our interests - three types of agents: patches, turtles and links. Patches are the digital representation of a physical space, with each patch being a fixed constituent square within the grid, which represents the world being simulated. Turtles are mobile actors who can move on that grid. Links (both undirected and directed) connect two turtles together.

Ticks are a stand-in for discrete-time steps within the simulation world. In addition to viewing the results of the simulation, the frontend interface also allows for user-defined variables which are then instantiated, and used, by the backend.

A view of the simulation world and the entire interface can be seen in Figure 1 and Figure 2 respectively.

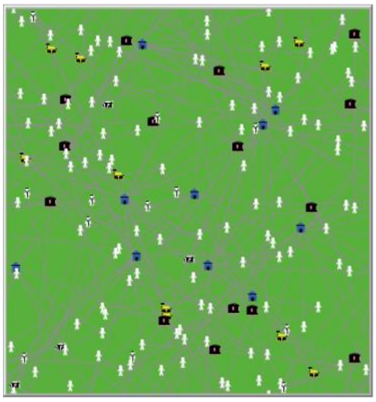

Figure 1: Simulation world for a specific instantiation, at ticks = 0

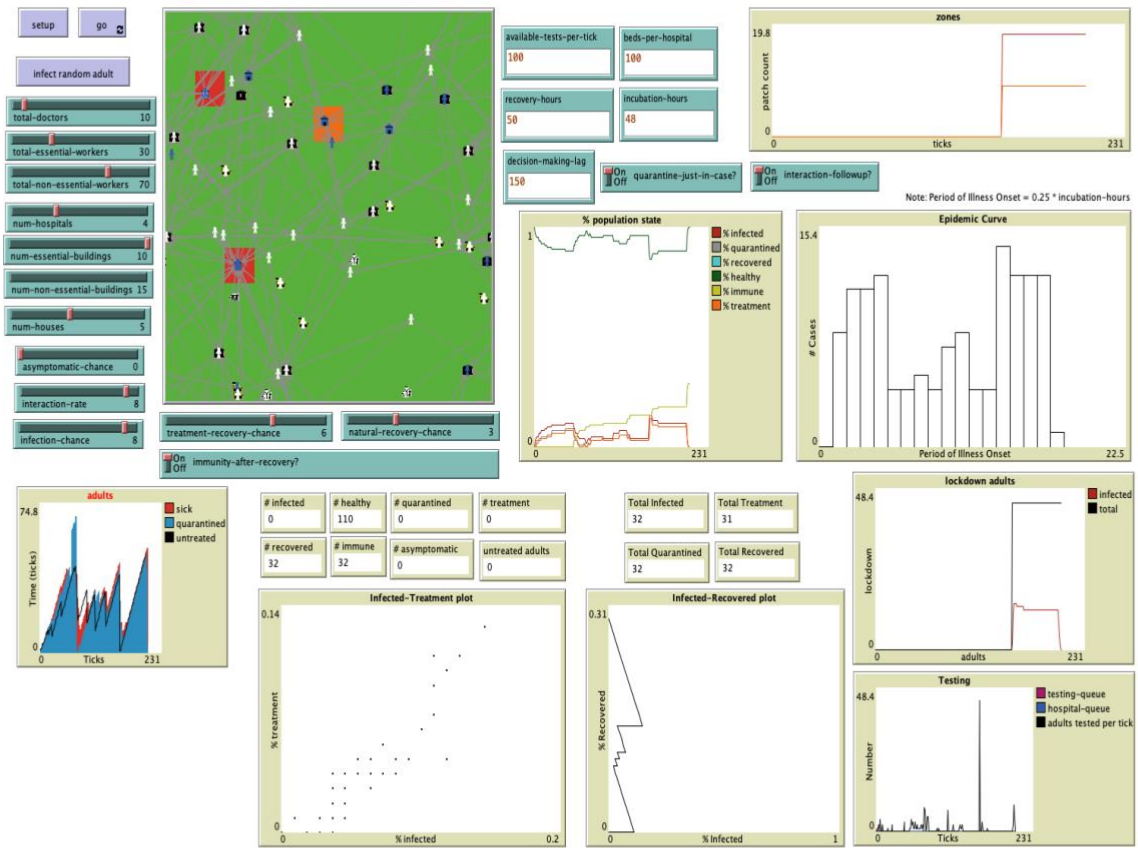

Figure 2: Simulation’s frontend interface for a specific instantiation, at the end of simulation

4.1.3 Entities, state variables and scales

The model's key mobile actors are collectively termed as "adults," i.e., a breed of turtles who each have a specific occupation (essential worker, non-essential worker or doctor) and an assigned location for the building for that occupation, a list consisting of their current state (which can be a logically sound subset of {"immune", "healthy", "infected", "quarantined", "recovered", "treatment"}), location of their house, a Boolean value that denotes if this adult is asymptomatic and three counters for the time (AKA ticks) they have spent sick, untreated and quarantined.

The simulation world in this model wraps horizontally and vertically and consists of 33x33 patches. The patches have two key attributes: *zone* and *infection-cases*. The former denotes the epidemic zone (i.e., red, orange or green) as characterized by the decision-making event, in the event of an outbreak, whereas the latter variable is an aggregated quantity that represents how many adults from (i.e., housed within) that patch have been infected. The latter variable, when sorted and compared, is solely responsible for deciding the city's zones (red zone patches are patches with infection cases which are 75th percentile or above, whereas the orange patches are assigned from the 50th percentile).

For the sake of concision, only links which directly impact the interpretability or influence the final result of the simulation are considered: the directed interaction and treatment links, and the undirected house and occupation links.

Two interaction links (directed each way) are formed between any two adults who interact with each other within the last (user-defined) average incubation hours of the disease. If the interaction follow-up option is toggled on, then for every adult who tests positive for the illness, automated contact tracing can be simulated by having every other adult, with an interaction link with the now-infected adult, automatically alerted and sent for testing for the illness.

Treatment links are directed links from the adult under treatment to the hospital building they are being treated under. The undirected house and occupation links connect every adult to their house and occupation-relevant building respectively. The object diagrams for this simulation are given in Figure 3.

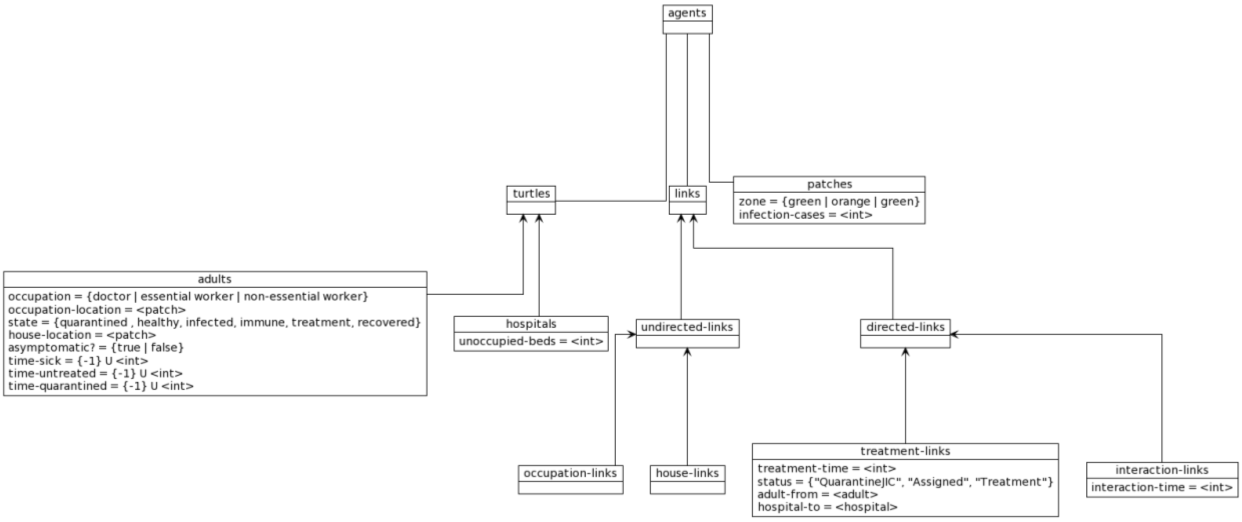

Figure 3: Object Diagram of Model

It is assumed that one tick in the simulation is equal to one hour of real-world time. Every tick increases the number of tests available by the user-defined value (defined within the interface) for *available-tests-per-tick*.

Initial Condition Variables: User-Defined

As can be seen in Figure 2 there are many user-defined variables. The 20 variables are categorized as follows:

1. **Initial Context Initialization**

   These variables denote the hyperlocal city-specific context with respect to the population (both individually and collectively) and infrastructure

   a. ***City-specific variables***

      i. *total-doctors-init*: the total number of doctors in the population. Range of values: 1-100.

      ii. *total-essential-workers-init*: the total number of essential workers in the initial population. Range of values: 1-100.

      iii. *total-non-essential-workers-init*: the total number of non-essential workers in the initial population. Range of values: 1-100.

      iv. *num-hospitals-init*: the total number of hospital buildings in the simulated city, with every doctor (as defined in a.i) randomly assigned to a hospital). Range of values: 1-10.

      v. *num-essential-buildings-init*: the total number of buildings which comprise of essential workplaces in the simulated city, with all essential workers (as defined in a.ii) randomly assigned to a building). By definition, an essential building is any building which scales with $\beta = 0.85$ on a log-log scale with the population. Range of values: 1-10.

vi. *num-non-essential-buildings-init*: the total number of buildings which comprise non-essential workplaces in the simulated city, with all non-essential workers (as defined in a.iii) randomly assigned to a building). By definition, a non-essential building is any building which scales with $\beta = 1$ on a log-log scale with the population. Range of values: 1-10.

vii. *num-houses-init*: the total number of houses available in the city, all adults are randomly assigned a house. Range of values: 1-10.

b. Resource-specific variables

i. *available-tests-per-tick*: the number of tests (for the disease) available per tick. If this number is low, adults who are suspected of infection may have to wait in a first-in-first-out queue in order to get tested. Range of values: set of natural numbers.

ii. *beds-per-hospital*: the total number of beds (initially, at ticks = 0, all beds are unoccupied) per hospital at the start of the simulation. Thus, total number of beds at any moment in the city = number of hospitals x beds-per-hospital. Range of values: set of natural numbers.

c. Individual-specific variables

i. *interaction-rate*: represents the probability (out of 10) that any two adults, who are not from the same household, interact upon meeting (i.e., sharing the same patch location). Note: Interaction rate between adults from the same household is 100%. Range of values: 0-9.

ii. *infection-chance*: represents the probability (out of 10) of an adult getting infected upon interacting with an infected adult. Range of values: 0-9.

iii. *treatment-recovery-chance:* represents the probability (out of 10) that an adult under treatment (from a hospital) successfully recovers after *time-sick* $\geq$ *recovery-hours*. Range of values: 0-9.

iv. *natural-recovery-chance:* represents the probability (out of 10) of an adult successfully recovering (by natural means, without the need for medical intervention) after *time-sick* ≥ *recovery-hours*. Range of values: 0-9.

2. **Problem Response Initialization**

These variables represent the problem context (with respect to the specific dynamics of the outbreak) and the response of the local authorities to the outbreak.

a. Outbreak-specific variables

i. *incubation-hours*: the number of ticks that need to pass between the moment of infection and the onset of symptoms of the disease (necessitating a trip to the hospital for testing), if adult is not asymptomatic. This value also determines the quarantining period and the valid set of interaction-links which direct to adults who would need to be tested as a result of automated contact tracing. Range of values: set of natural numbers.

ii. *recovery-hours:* the average time that needs to pass for recovery from the disease to be a viable state for an infected adult. That is, this is the number of ticks after time of infection that an infected adult can be tested for recovery (depending on the appropriate natural recovery and/or treatment probabilities). Range of values: set of natural numbers.

iii. *immunity-after-recovery?*: is a previously-infected, now-recovered adult immune from the disease and cannot be reinfected with it again? Range of values: {True, False}

iv. *asymptomatic-chance:* represents the probability (out of 10) of an infected adult being asymptomatic (i.e., not having visible symptoms of the illness). Asymptomatic (infected) adults will (unless called upon by contact tracing) not know that they have been infected and thus not be quarantined or otherwise tested. Due to there being no quarantine-related mobility restrictions, they may infect other adults after suitable interaction and infection conditions have been met. Range of values: 0-9.

b. Intervention-specific variables

i. *decision-making-lag:* the ticks (post-first infection in the city) needed for local authorities to make and enact a decision (in this case, zoning). Range of values: set of natural numbers

ii. *interaction-followup?:* Is automated contact tracing available? If true, then all the adults, from an adult who has been tested positive for the disease (connected via the most recent interaction links (i.e., within the last *incubation-hours* period)), will be notified and asked to come in for testing. Range of values: {True, False}

iii. *quarantine-just-in-case?:* If there are not enough tests or unoccupied beds available, should an adult suspected of being infected quarantine just-in-case? If true, then the quarantine exit condition depends on the adult testing as healthy (when the testing queue reaches them). Range of values: {True, False}

4.1.4 Process Overview and Scheduling

As mentioned before, this agent-based model simulates, via cellular automata, a disease outbreak crisis, within a closed population, in a city. This is done in order to better understand which, if any, of the available initial and/or intervention conditions, within the simulated city, disproportionately affects the crisis trajectory.

Thus, the mobile agents (adults) move around the city to work or wander (assuming they are allowed to do so with respect to zoning policies and quarantine requirements), and also interact with each other.

Upon the starting of the outbreak, based on the allowed intervention strategies and (the testing and hospital) resources available, the adults may be required to work from home due to quarantine requirements, quarantine at home or the hospital, go to the hospital for testing and/or treatment, and wait in queues for testing and/or available hospital beds.

The rules for adults are further outlined in the following two sections.

4.2 Design Concepts

The disease outbreak is initiated via the infection of a randomly selected, non-naturally immune, adult within the city's population (via the "infect random adult" button in the interface), and is spread by infected adults interacting with non-infected, non-immune adults (who may or may not be asymptomatic).

Upon infection, and after time sick has eclipsed the incubation time for the disease, provided the adult is not asymptomatic, it is assumed that the adult will fully realize that they are sick as the symptoms of the disease will have become apparent post-incubation period, and the adult will then directly go to the hospital for testing.

Every hospital has a limited number of available beds. Thus, every adult, who wants to get tested, will be directed to the hospital nearest to their current location which has an available bed.

Upon identification of a hospital, the suspected adult is tested (if a test is available), and if they are proven to be infected, they are officially put under treatment and required to quarantine at the hospital, thus taking up a bed until they recover. Otherwise, they are deemed healthy and can return to the default action for adults (to wander or go to work).

If there are no hospitals or tests available, the adult is put in the apropos queue and, if *quarantine-just-in-case?* is true, told to quarantine at home, otherwise, if false, allowed to wander the city like normal.

The hospital and testing queues are dequeued in a first-in-first-out fashion and the start of every “go” procedure begins with dequeuing as many adults in the queues as possible (based on any newly available beds and/or tests).

For every adult who tests positive for the illness, if *interaction-followup?* is true, then automatic contact tracing occurs and every adult the infected adult had interacted with, within the past incubation period, will be required to go to a hospital to get tested.

The interaction between adults from different households will occur according to the *interaction-chance*, but interaction between non-quarantining (at home) adults from within the same household, will occur at a 100% probability.

Zoning policies are decided every user-defined *decision-making-lag* period, and they are decided based on the locations of households with infected adults, with a neighboring radius of 1.

Non-quarantined essential workers and doctors are required to always go to work by default, regardless of their house's zone. Whereas, non-essential workers can only leave their homes if they live in a green or orange zone. Non-quarantined, non-essential workers either wander the city or go to work by default.

Quarantined adults can leave quarantine if they test negative for the illness after they've been quarantined for longer than the disease incubation time (as the test cannot accurately ascertain infection unless time sick > incubation hours). This negative test result can be because the adult was never sick and thus tested healthy with no change, or (if time sick > recovery hours) if the adult recovered either by natural means or by treatment (under a hospital). If *immunity-after-recovery?* is true, then every recovered adult is deemed 100% immune to the disease.

If two adults interact and one of the adults is infected, the other adult may also be infected based on the *infection-chance*. Any infected adult may be asymptomatic according to the *asymptomatic-chance* probability.

4.3 Details

This section outlines the way the world is set up and initialized, whether there is any input data required for the model to work and the sub-procedures and agent rules (with pseudocode) that is necessary for the model logic.

4.3.1 Initialization

After selecting the necessary initial conditions, "setup" should be selected, followed by "infect random adult" which will infect an adult and thus begin the outbreak. Then selecting the "go" procedure will begin the simulation in earnest.

4.3.2 Input Data

Model has no input data, as the simulated environment is assumed to not change.

4.3.3 Submodels

The pseudo-code for non-trivial procedures explained below is given in section 10.5 in the appendix.

The `setup` Procedure

The setup procedure resets the world to the starting position with respect to the initialization of the global variables, patches and turtles.

At the beginning, *outbreak-time* equals -1, as the outbreak has not started yet, and thus, all patches are in the "green" zone with their *infection-cases* attribute set to 0.

The turtles are set up by creating a user-defined number of buildings and adults with their respective occupations. All buildings and adults are randomly placed in the simulated city. For every adult, a house and an occupation-appropriate building is assigned, along with an undirected house and occupation link respectively.

The go Procedure

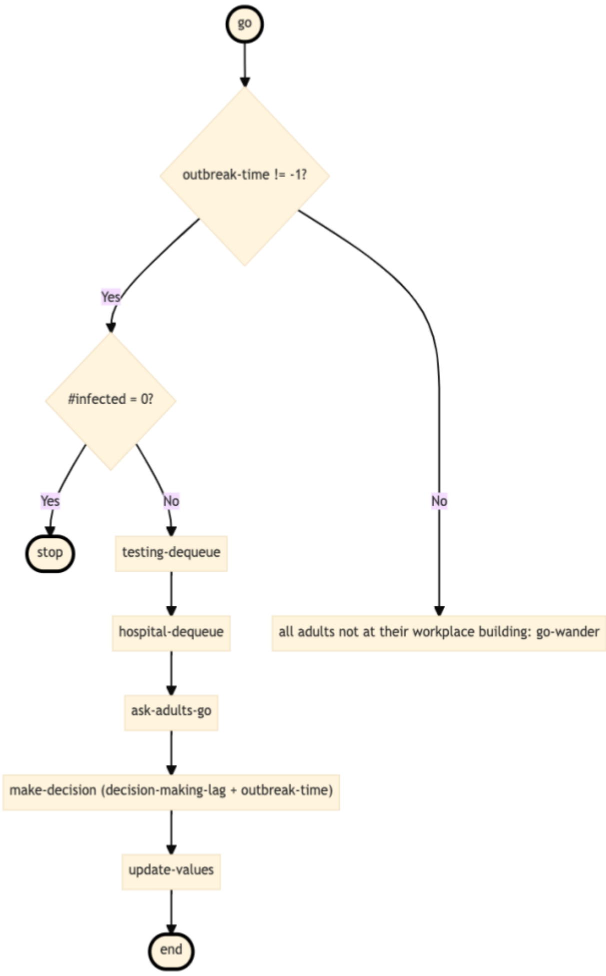

Figure 4: Flowchart of the go procedure

The "go" procedure is repeatedly called (or, in the macro-analysis case studies, for a maximum of 10,000 ticks) after the user clicks on the "go" button in the simulation interface screen for the first time. Clicking on the "go" button signifies that the ticks have officially started and the simulation has begun. By virtue of the programming logic, for every run of the go procedure, the tick counter increments by one.

Figure 4: Flowchart of the go procedureFigure 4 outlines the main flow of logic for the go procedure. Within it, we can see that if the outbreak has not yet started (if *outbreak-time* = -1), that adults who are not currently at their respective assigned workplace buildings (given by every adult's *occupation-location* attribute) are programmatically ordered to randomly wander the simulated city space by calling the *go-wander* procedure.

However, if the outbreak has started, the testing and hospital queues are checked. Since the total number of available tests increase after every tick (by the user-set *available-tests-per-tick* property), the globally defined *testing-queue* - *if* its length is non-zero - can be dequeued (with the implied first-in-first-out strategy) and the *available-tests-per-tick* number of dequeued adults are then tested for the infection. Similarly, the *hospital-queue* global variable is also dequeued, *if* there are any newly available beds from the last tick.

The *ask-adults-go* procedure is a turtle-level function in which every adult's general behavioral rules are outlined. Its behavior is explained in the following section.

Penultimately, the ticks-parameterized *make-decision* procedure is called as a programmatic substitute for the city's local government's decision-making actions and ensuing policies. For every tick past the *outbreak-time,* after a multiple of the user-defined *decision-making-lag* period, zoning decisions are made based on the total *infection-cases* property of each patch. Zoning is based on assigning patches whose *infection-cases* attribute is at least above the 50th or 75th percentile in the orange zone and red zone respectively.

Finally, all the relevant counter attribute values are updated (including incrementing the tick counter by one) in *update-values* and the procedure ends.

Unless constrained externally, the go procedure stops when the number of infected adults (after the outbreak has already begun) equals zero. This property signifies that the outbreak has been handled and the crisis is now over.

The `ask-adults-go` Procedure

Note: the pseudocode for this procedure is included in section 10.5 in the appendix.

This procedure checks if a non-quarantined adult is infected with the illness. If true, then the adult's *time-sick* attribute is compared against the *incubation-hours,* for, if the former is greater, then unless the adult is asymptomatic, the symptoms of the illness are felt by the adult and the adult will go to the hospital (via the "go-to-hospital" procedure). However, if that adult is asymptomatic and if they have been sick for longer than the *recovery-hours,* the adult is checked for recovery (by natural means, as they are not under treatment under a hospital (because otherwise if they had been under treatment, they would have been under quarantine)). Otherwise, if the adult is asymptomatic but has not been sick for long enough to be checked for recovery, or if they have been infected but symptoms have not been apparent yet (due to being < *incubation-hours*), then the adult is programmed to go to work or wander via the "to-work-or-wander" procedure.

All adults not in quarantine may interact with other non-quarantined adults via the "interact" procedure.

If the adult is under quarantine and under treatment, the adult's behavior is controlled by the "ongoing-treatment" procedure which determines if that adult qualifies for getting checked for recovery (and exit quarantine) or if they must continue their quarantine at the hospital uninterrupted.

However, if the quarantined adult (who may or may not be infected) is not under treatment, and if the *time-quarantined* attribute is less than the *incubation-hours* (i.e., the condition that determines if an adult can be accurately tested for the illness): the adult who is quarantining “just in case” is allowed to go home (“go-home” procedure) to quarantine. But if the testing qualification condition is met, then the adult is directed to the “exit-quarantine?*”* procedure which determines whether an adult, post-testing (if a test *is* available) can be assigned as “recovered” or “healthy” and exit quarantine.

4.4 Decisions

- Zoning radius is 1 (the 8 nearest patches and the calling patch itself)
- Red zone is declared for the *infection-cases* at least in the 75th percentile among the patches, and orange zone for 50th percentile
- As it is difficult in real life to precisely trace back to which location an infection occurred in, zoning is decided by aggregating the infections at the houses of those infected, rather than the precise location of the transmission
- In the event an adult’s house location is in an official red zone, if the adult is a non-essential worker, then they cannot leave their house and are implied to “work from home.”
- Interaction rate is 100% for adults within the same household
- An adult under treatment is quarantined at the hospital, an adult who has to quarantine just-in-case quarantines in their assigned house
- The simulation is terminated when the number of infected adults equal zero or when ticks equal 10,000, whichever comes to pass first

4.5 Assumptions

- The model assumed a closed population, with no fatalities due to the illness
- Unemployment rate is assumed to be 0% with every adult having an assigned occupation and appropriate workplace building
- Every adult has an assigned home
- Essential workers and doctors are required, regardless of zoning, to go to work
- An adult can go to any hospital for treatment, there is no accounting for health insurance policies which may have some adults preferring some hospitals over others

- Adults have full information about available hospital resources (i.e., unoccupied beds). This is not unrealistic as this can be replicated in the real world by querying through phone calls or web search.
- There is no information transmission lag between governmental policy issuance and public compliance. Thus, every time a decision is made by the local authorities (regarding zoning), it is in effect and complied with immediately
- The disease-causing pathogen does not mutate; however, an adult may be reinfected if the *immunity-after-recovery?* option is toggled off, otherwise the recovered adult is fully immune to the disease
- Natural immunity to the illness is assumed to be null
- All testing is assumed to be 100% accurate. No false positives or negatives.
- Testing results are immediate; there is no lag between an adult taking an available test and knowing the result, both happen within the same tick period.
- A non-asymptomatic infected adult will notice disease symptoms after *time-sick* ≥ *incubation-hour.*
- A non-asymptomatic infected adult upon noticing symptoms of disease will always go for testing at a hospital
- Compliance to zoning and quarantine rules is assumed to be 100%
- A quarantined adult can be accurately tested for the illness after *time-quarantined* ≥ *incubation-hours,* and tested for recovery when *time-quarantined* ≥ *recovery-hours*
- An infected adult can recover (either naturally or via treatment) when *time-sick* ≥ *recovery-hours*

5 Methodology

The analysis was primarily conducted on a macro level, with four case studies identified which varied the initialization parameters outlined in the 4.1.3 section.

The four chosen case studies (as explained further in the succeeding section) were partitioned by their two parent categories, with the two nested case study scenarios under the same parent, based on having the same population. Thus, case 1.1 (2.1) had the same population as case 1.2 (2.2) respectively.

Case analysis strategy was of three types: per case, between cases and all cases. That is, by considering the individual cases (case 1_1, 1_2, 2_1 and 2_2) in isolation, by comparing between cases in different parent categories (cases1 (case 1_1 and 1_2) vs cases2 (case 2_1 and 2_2)) and the entire, unfiltered dataset built from rows from all the four case studies respectively.

Every one of the 256 unique simulation cases within each of the four case study scenarios was run three times in order to account for the randomness in the initial starting spatial position of all the turtle agents. A Python script was created which ran all simulation combinations in parallel in headless mode in Netlogo using the ipyparallel [65], pyNetLogo [66] and NL4Py [67] libraries.

The number of ticks to completion (or 10,000 ticks, whichever came first) was documented for every one of the 3 x 256 = 768 simulations per case study scenario, and the ticks were then aggregated by three metrics: average, minimum and maximum, per unique simulation scenario. A total of 3,072 simulations were run.

Then, for each case analysis strategy: for every aggregated ticks metric, the resultant ticks data was clustered via the K-Means algorithm, with the number of clusters (k) varied from 3 to 10. This resultant dataset, case analysis strategy defined, with each clustered aggregated tick type, further parameterized by the number of clusters, was saved and passed to the next step of the analysis workflow.

Gradient Boosted Trees algorithm was used (using the XGBoost python library [68]) on the above saved datasets in order to determine the feature importance scores (using the permutation-scoring metric on the fitted model), per cluster number, tick aggregation metric and case analysis strategy.

Then, the resultant model from every iteration of the parameterization was used to fit the dataset using only the F most important features (scored by feature importance score), where F was iterated from 3 to 5 (inclusive). The resultant accuracy per F, for each iteration was saved. This was done via the sklearn Python library [69].

Thus, the analysis was parameterized on case analysis strategy, tick aggregation metric, number of clusters and number of important features chosen.

This optionality in parameterization was conducted in order to answer the following meta-analysis questions:

1. Considering the overwhelming amount of data when holistically studying smart cities, is there a way to pre-process and "bin" values without losing a significant amount of information?
2. Which aggregation method should be used when analyzing the success metric (i.e., ticks to termination) on a macro-scale?
3. How many features (i.e., user-defined variables) of the simulated city in crisis can most accurately estimate the crisis trajectory (i.e., the ticks to completion)?
4. How does the final result change when moving from intra-city level to inter-city level macro analysis?

The justifications for analysis methodology are explained in section 5.2.

5.1 Macro Analysis

The macro analysis was conducted through four simulation case studies. Table 1 outlines the constant and varying elements of each case study, the fixed and variable initialization values for each attribute in every case study scenario is listed in section 10.1 in the Appendix.

For base population A, two key case study groups can be identified: 1) Best case: Resources, i.e., when the city has enough resources available at every tick to deal with the outbreak, and 2) Worst case: Resources, i.e., when the city has a dearth of resources to deal with the outbreak and there are delays due to testing and/or hospital queues.

In another city with population B, which is calculated to be 20% more than A, the same best- and worst-case scenarios for resources are simulated, however, all but two of the properties (*treatment-recovery-chance* and *natural-recovery-chance*) in the two new cases are a scaled-up version of the respective properties in the prior seen cases. The scaling occurs in linear fashion (i.e., increasing linearly by 20%) for the city-specific variables, but occurs in a log-log fashion for every fixed-valued property with a (super or sub-linear) scaling factor as decided by the urban scaling laws as explained in section 3.2.3. Section 10.2 in the appendix contains the scaling factors and final values used for the relevant properties.

Table 1: Table outlining the four simulation cases

Category	**Subcategory**	**Macro-Analysis Cases**			
		1. At Closed Population A		**2. At Closed Population B (> A)**	
		1.1. Best Case: Resources	**1.2. Worst Case: Resources**	**2.1. Best Case: Resources**	**2.2. Worst Case: Resources**
Initial Context Initialization	*City*	Fixed	Fixed	Fixed[2]	Fixed[1]
	Resource	Fixed - **High**	Fixed - **Low**	Fixed - **High**[1]	Fixed - **Low**[1]
	Individual	Fixed	Fixed	Fixed[1]	Fixed[1]
Problem Response Initialization	*Outbreak*	Variable	Variable	Variable	Variable
	Intervention	Variable	Variable	Variable	Variable

Note that it is assumed that parent cases 1 and 2 occur in different cities within the same country. This assumption is important to note as the scaling laws that are the foundation of this analysis are 1) not valid temporally upon population growth within the same city (only cross-sectionally across different cities with differing populations), and 2) only comparable for cities within the same country.

Furthermore, it is defined that an essential building is any building for which the scaling exponent, $\beta < 1$ (such as car dealer, petrol stations, etc.) and a non-essential building is one where $\beta = 1$ (such as post offices, pharmacies, etc.) [53], [54].

5.2 Decisions

This section details the rationale for the decisions made in the analysis process.

Why Ticks to Completion as the Success Metric?

[2] Not the same values as in cases 1, but scaled based on the scaling exponent and the new, increased population (see section 10.2 in the appendix for more details)

Since the final “ticks” value depends on when in the simulation the number of infected adults becomes null (or 10,000 ticks is reached first), it is a useful success metric for the crisis. Additionally, since this is a quantitative value that can be compared, in an apples-to-apples manner, across different city contexts and it has a real-world equivalent (1 tick = 1 hour), it was chosen.

Why Clustering?

In order to reduce the number of unique ticks to completion values that the analysis process needed to contend with, K-means clustering was chosen to group the different rows according to their ticks to completion. This way, it was also possible to broadly and visually analyze the different clusters according to their mean.

Number of clusters, k, was parameterized from 3 to 10 because for any $k > 10$, the resultant clustered means were not dissimilar enough to justify their inclusion and increased cost of complexity.

Why Gradient Boosted Trees?

As gradient boosted trees [70] provide high predictive accuracy, work in classification scenarios and with heterogeneous data types, it was identified as a useful tool for this scenario. Since this model’s analysis is not meant to be online, the associated computational expense was not a barrier to use. Moreover, since the datasets used were not following the traditional training-testing-cross-validation framework, and instead needed to be holistically and uniquely analyzed only once (and the model wouldn’t be an input for future classification) in order to score the feature importance, the associated risk of overfitting was also not a barrier to adoption.

Why Permutation-based Feature Importance Scoring?

Permutation-based feature scoring was selected as most other impurity-based scoring metrics are biased in favor of high cardinality features over features which were binary or otherwise had a small set of possible values [71].

For example, when the gains metric was used for feature scoring, although the accuracy of the selected features was high, features with a low number of possible values (such as *quarantine-just-in-case?*) were negatively impacted despite their actual importance in the model. This was not the case with permutation-based feature scoring.

While the selected scoring metric could have problems with highly-correlated features, upon calculating a correlation matrix [72], this was deemed to not be a cause of concern for this analysis.

6 Results

The case datasets were compared as follows:

1. **By case:** each case scenario was studied in isolation. That is four case datasets: case1_1, case1_2, case2_1, case2_2
2. **Between cases:** each child case scenario was aggregated with their sibling (under the same parent) case scenario. Resulting in two case datasets: cases1 (case1_1 and case1_2), cases2 (case2_1 and case2_2)
3. **All Cases:** all four cases were considered in one dataset

6.1 Do Initial Conditions Matter?

As seen in the complexity section, real cities are sensitive to initial conditions, so the question arises, is this trait also true when analyzing the model's results?

6.1.1 Analysis Initial Conditions

The analysis data was collected along several dimensions which were also varied in order to identify which specific initialization along each dimension would afford the results the best (aggregated) accuracy.

Number of Clusters Selection

After grouping the resultant dataset by case, aggregation metric and number of features chosen (in that order), the row with the highest accuracy and the lowest number of clusters was chosen from each group. All 105 groups of rows calculated that number of clusters, k = 3, as the initialization was optimal, with an average accuracy of 86.2%.

Aggregation Metric Selection

When grouping the resultant (k = 3 filtered) dataset by case and aggregation metric, as can be seen from Table 2, max-ticks (which is the maximum of the ticks to completion values from all three iterations of the same unique case scenario) is able to provide a minimum average accuracy, across all feature subset sizes, of 88.3% for all cases and feature subset sizes.

As a side note, the improvement between the accuracy values of max-ticks versus the other two aggregation metrics is even more apparent when considering the minimum average accuracy per case comparison in the original dataset, where k is iterated from 3 to 10. The minimum average accuracy improves from 64.7% to 75.3% when the max-ticks metric is chosen.

This makes logical sense as this metric can be seen as an adequate (within our limited scope) representation of the "worst case" scenario for a specific unique simulation case, answering: what is the longest period a city in crisis can take to deal with the outbreak?

Table 2: Selection of aggregation metric which provides the best accuracy score when grouped by case and aggregation metric (at number of clusters = 3)

Subcase	Avg-ticks	Max-ticks	Min-ticks
Case	**Average(Accuracy)**		
All Cases	77.1	88.3	86.2
Cases1	83.6	89.0	86.5
Cases2	78.6	89.5	86.8
Case1_1	82.3	90.0	89.8
Case1_2	81.1	92.0	87.8
Case2_1	83.1	92.7	88.4
Case2_2	80.1	91.8	86.2
Minimum(Avg(Accuracy)	**77.1**	**88.3**	**86.2**

Important Features Subset Selection

As can be seen in Figure 5 which contains the plot between the cardinality of the most important features set which were included in the test set (with the rest excluded) versus the accuracy of their model-fitted predicted values, at a) number of clusters, k=3 and b) k not constrained to any specific value, with both plots having max-ticks as the aggregation metric.

It is easy to discern from the top plot that, for the most part, at number of clusters = 3, only considering the three most important features (as decided by the permutation-based feature importance algorithm) is enough to identify the cluster label (and thus, by proxy, estimate the ticks to completion range for that K-Means clustered row) for a simulation scenario with an average accuracy of 92.7% and, sans the all-cases scenario, give a case comparison accuracy of more than 89%.

It needs to be noted that the accuracy value improvement for every comparison scenario barring the all-cases scenario, where the size of the feature set > 3 is within a ~5% range

Figure 5: Number of features tested (legend) vs avg. accuracy, grouped by case study comparison scenario at aggregation metric = max-ticks at… a) (top) k=3, b)(bottom) k ∈ {3-10}.

Any bar plots not shown in the figure have an accuracy < 80%, and were thus ignored for visual clarity

of the chosen feature set size's accuracy counterpart. Thus, when balancing resources, computational storage and time and analysis expediency, it makes logical, scientific and fiscal sense to choose a smaller feature set based on the above proposed methodology.

6.1.2 City and Problem Context Initial Conditions

As seen in Figure 6 and Figure 7, on changing the problem context's initial conditions (outbreak and intervention-specific scenarios resp.), the maximum ticks to termination changes.

There is significant improvement in the outbreak-specific variables, wherein, simply toggling on the *immunity-after-recovery?* variable may result in ticks growing from 165 ticks to 8495 ticks.

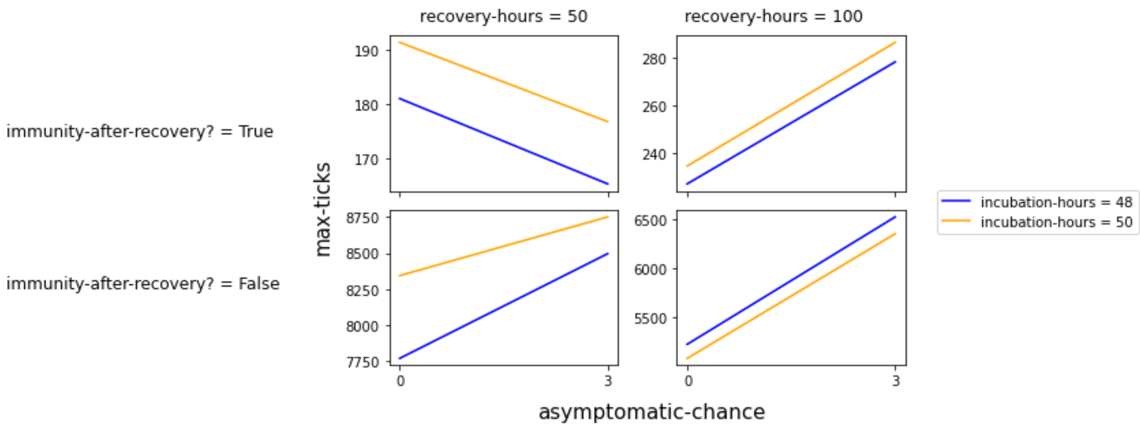

Figure 6: Max-ticks vs Outbreak-specific variables for all cases

Figure 7 also hints that the *quarantine-just-in-case?* intervention strategy is most effective when *interaction-followup?* is also enabled, with the combination being 4.15 times more effective with respect to maximum ticks (at *decision-making-lag* = 100). The next best combination for optimizing ticks to termination is ensuring only *interaction-followup?* = True.

Interestingly, if automated contact tracing is not possible, there is no general rule of thumb for selecting a *decision-making-lag* (DML) value. As seen in the bottom plot, reducing the *decision-making-lag* does not always result in a shorter simulation completion time (ticks(DML = 150) < ticks(DML = 100)), which may seem counterintuitive but manually validates the finding that *quarantine-just-in-case?* is more important (ranking as the second most important intervention-specific variable in the all cases analysis) than *decision-making-lag*. This finding also validates the thesis's underlying hypothesis that a perfect or optimal solution on paper might not be the most resource-friendly or effective strategy. Based on this simulation, it could be argued that more resources should be directed towards implementing policies to quarantine just-in-case (especially in the event of lack of testing resources) versus updating the zoning policy.

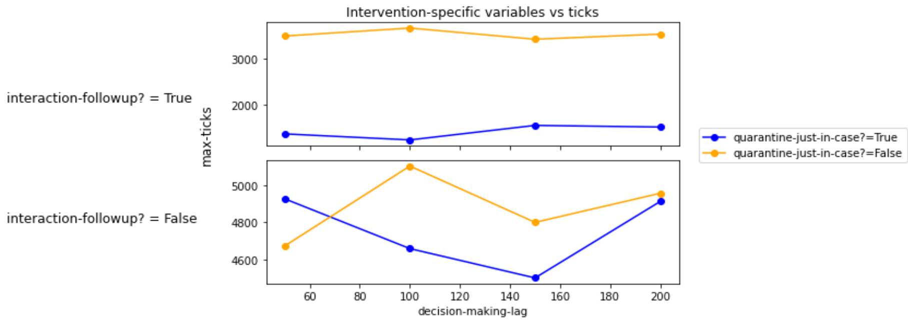

Figure 7: Maximum ticks vs intervention-specific variables for the all cases dataset

6.2 Intervention Strategies in Cities

As can be seen from Figure 8 and Figure 9, which contain the scores and names respectively of the three most important features, per case comparison type (at k=3, aggregation metric = max-ticks).

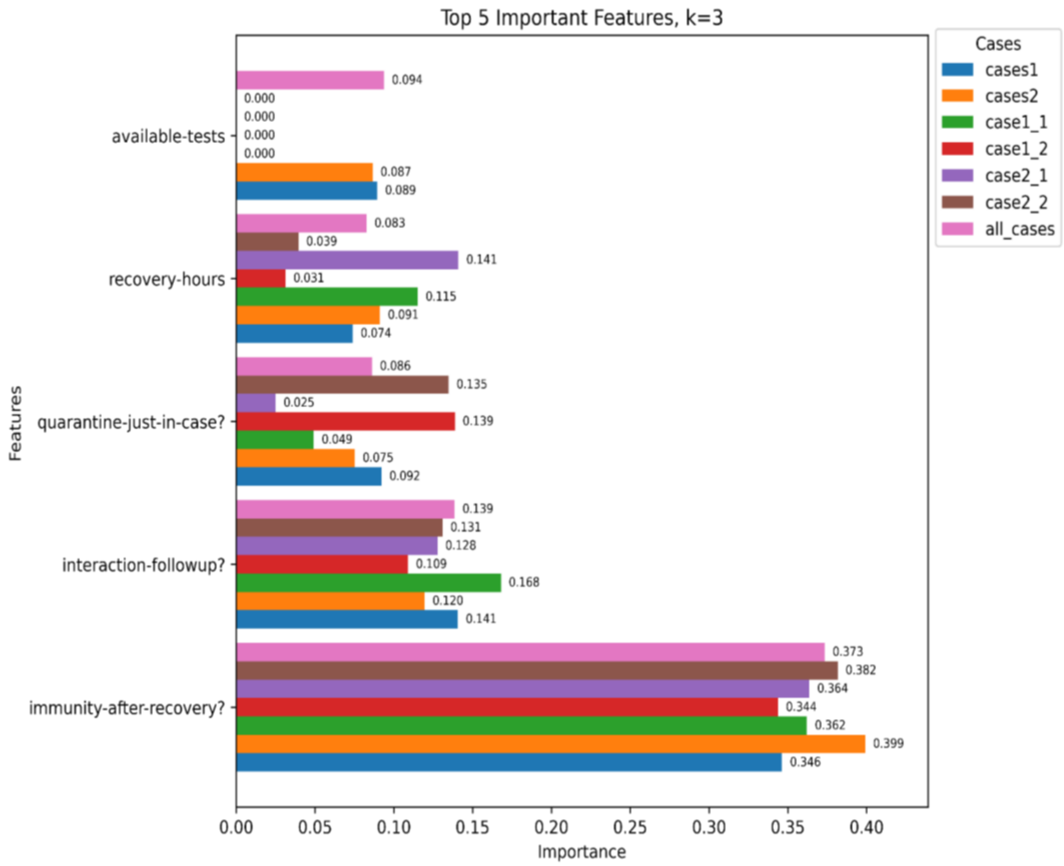

Figure 8: Importance scores

For all case comparison scenarios, *immunity-after-recovery?* property is universally the most important. As seen from the number of features per case vs accuracy table in section

10.4 in the appendix, by only including this sole feature in the clustered dataset the trained XGBoost model is able to, with more than 80% accuracy, predict the cluster, and thus, the ticks to termination for that initialization.

The second most frequent important feature (for all cases other than case2_1 and case2_2, for which it is the third most important feature) is *interaction-followup?* which determines whether automated contact tracing is taking place when an adult tests positive for the infection. Inclusion of this feature in the testing dataset improves the accuracy of the prediction to more than 85% for every case.

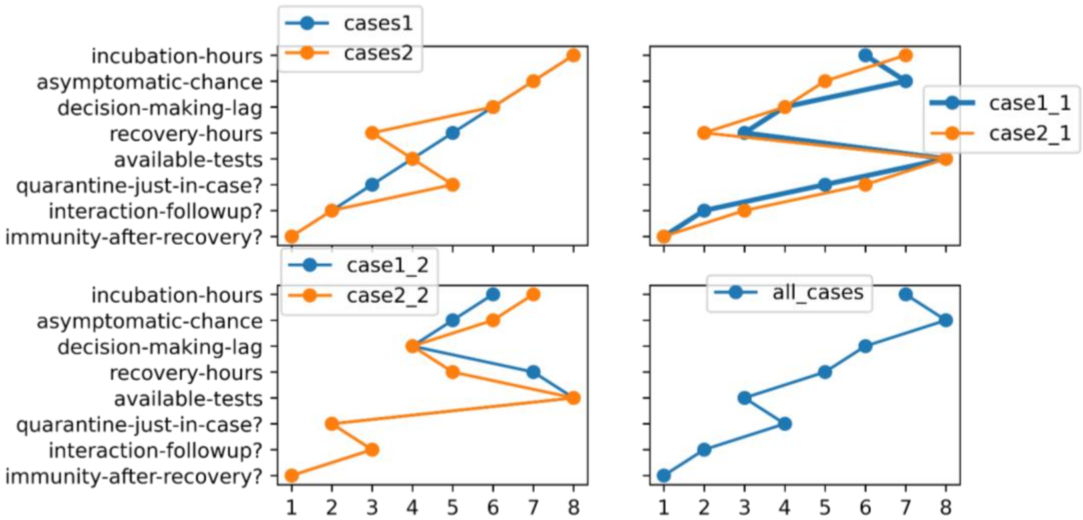

Figure 9: Ranked important features per case scenario

For the low resource cases (case1_2 and case2_2), it can be hypothesized that *quarantine-just-in-case?* should rank higher than *interaction-followup?* based on the rationale that compared to needing automated contact tracing, such cities would more importantly prefer that adults suspected of infection be required to quarantine just-in-case and not wander the city while potentially spreading the illness even further into the population. A suspected adult is defined as someone who has gone to a hospital for testing either because of automated contact tracing or due to feeling symptoms of the disease (which would then imply that their *time-sick* > *incubation-hours*). Even if *interaction-followup?* is unavailable, for the latter case of suspicion, a policy of mandatory quarantining for suspected adults should prove beneficial towards controlling the crisis (and reducing the ticks to completion).

However, it can be then argued that in the case with population B (>A by 20%) and corresponding higher *interaction-chance*, the *relative* importance score of *interaction-followup?* would be higher compared to that from the corresponding case in population A, as the contact tracing functionality of *interaction-followup?* depends on the probability of interaction and *incubation-hours* of the disease and a higher probability of interaction would statistically beget more infected adults within the same incubation period, and if a specific percentage of the population is asymptomatic and/or incubation period is high, then an infected adult would not be made aware that they are to quarantine and would continue wandering the city and spreading the illness among their interaction circle. Thus, within two populations - after controlling for the illness - automated contact tracing functionality should be relatively more important and urgent to enable when the population (and ergo, *interaction-chance*) increases.

This intuition is proven true by the model and feature importance rankings, with case1_2 and 2_2 having *quarantine-just-in-case?* and *interaction-followup?* as the second and third most important features respectively, and case2_2's *interaction-followup?* feature importance score be 16.4% of the total (almost on par with its score for *quarantine-just-in-case?)*, which is more than that for case1_2's (14.5%, a comparatively lower score compared to its 18.7% score for *quarantine-just-in-case?*).

On that note, while *available-tests-per-tick* ranks high in the holistic cases (i.e., cases1, cases2, and all cases), it is not *the* most important feature. This implies that while resources to manage the crisis are *a* key element of any crisis management strategy, they are not the most pressing crisis management technique, with automated contact tracing (*interaction-following?*) and understanding of the disease dynamics (*immunity-after-recovery?*) being more important.

Interestingly enough, *decision-making-lag*, which theoretically should have been an important metric that affected the final ticks to termination, did not rank in the top 5 most important features for any of the holistic cases but did rank fourth in the individual cases. However, even for the individual cases, the inclusion of this feature in the important features columns in the testing set did not result in any significant improvement (< 0.8%) in the final model prediction accuracy.

Note: Importance scores and rankings for all features by case comparison type are included in section 10.3 in the appendix.

6.3 Influence of Hyperlocal Parameters on Crisis Impact

In order to better understand the dynamics of each case on a micro-scale, the best and worst initializations in each case were chosen (based on, respectively, the lowest and highest ticks to completion values, wherein ticks from each of the three iterations per unique case were averaged), with the constraint that the initializations for all best cases and that for all worst cases be the same. The two chosen initializations can be seen in the interface screenshots.

6.3.1 Best Case Comparison

On a micro-scale, it can be hypothesized that on average, infected adults within cities, which have limited testing resources, will have *time-sick* ≈ *time-untreated*, as the lack of testing resources will ensure that not many people recover via treatment, and instead will have to depend on their immune system for natural recovery (i.e. *natural-recovery-chance*).

Furthermore, if *quarantine-just-in-case?* is true then the average time under quarantine for all adults should be more than the average *time-sick* for infected adults (this value disparity should also worsen when *interaction-followup?* is true). However, for cities which don't have to worry about testing resources, for ticks > *incubation-hours*: on average, it can be rationalized as most of the recovery will happen via medical treatment, that *time-sick* > *time-untreated* and *time-quarantined* ≤ *incubation-hours*.

This instinct is proven true when comparing the best-case scenario "adults" plot as well as the "Total Treatment" and "Total Recovered" values for Case1_1 (2_1) with that of Case1_2 (2_2) in Figure 10 (Figure 12) and Figure 11 (Figure 13).

Furthermore, the epidemic curve (in which the x-axis time-step is parameterized by the *incubation-hours*) also subtly changes. In Case1_1 (or Case2_1), the epidemic curve implies a propagated source pattern of spread, which occurs when a pathogen is spread from one susceptible person to another, and the peaks, separated by one incubation period, get successively higher and higher until the infection is controlled or the number of susceptible persons decrease [73]. Note: in all the Epidemic Curve plots, 1 time step = (*incubation-hours* / 4). However, in Case1_2 (or Case2_2) the epidemic curve's pattern of spread is more visually similar to that of a point source with secondary transmission (or index case with limited spread). This is when one person is able to infect other susceptible people, starting the outbreak, but control measures are able to reduce the number of secondary transmissions. This helps validate the efficacy of the selected intervention strategies. This curve, combined with the fact that the *quarantine-just-in-case?* control option was toggled on and that the majority of the infected adults recovered by natural means in both cases (see “Total Recovered” and “Total Treatment”), proves that in cases with low testing resources, it is possible to control the outbreak by economically feasible control measures. Counterintuitively, the total number of infected adults in the latter low resource cases is comparable to (or even better than) their high resources counterparts.

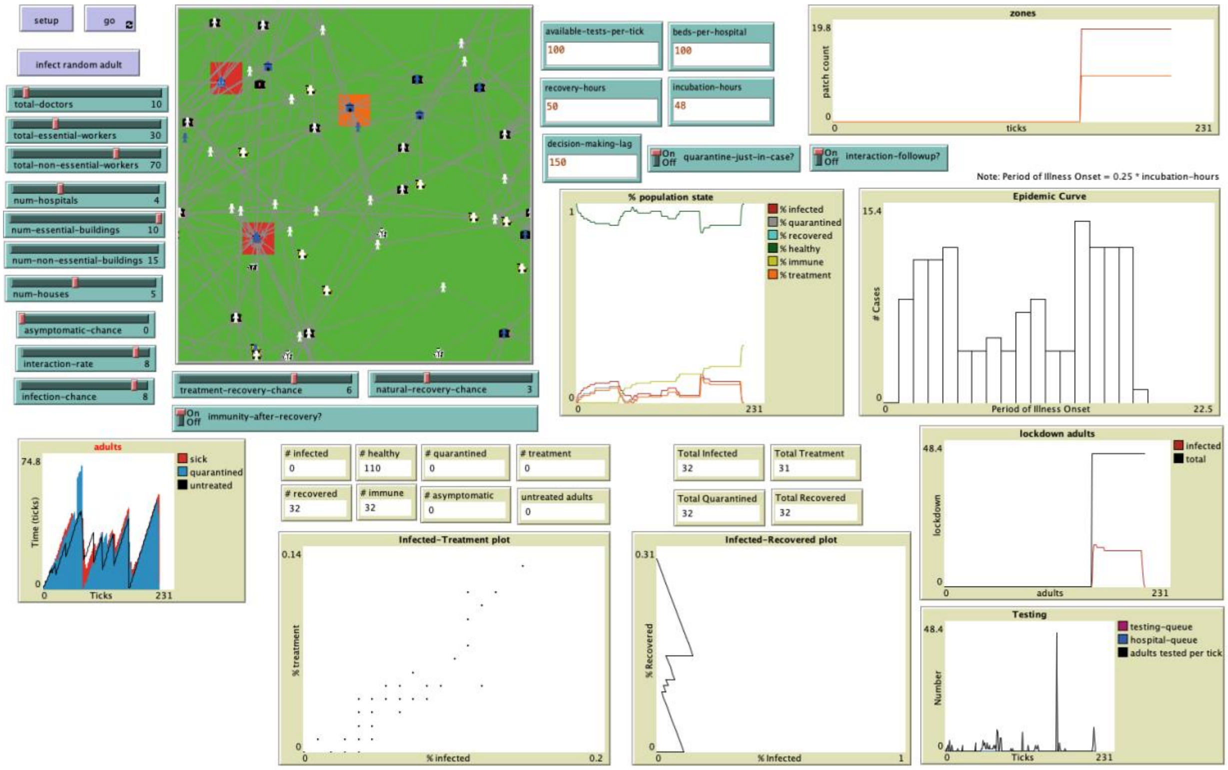

Figure 10: Best case scenario for Case 1_1 group

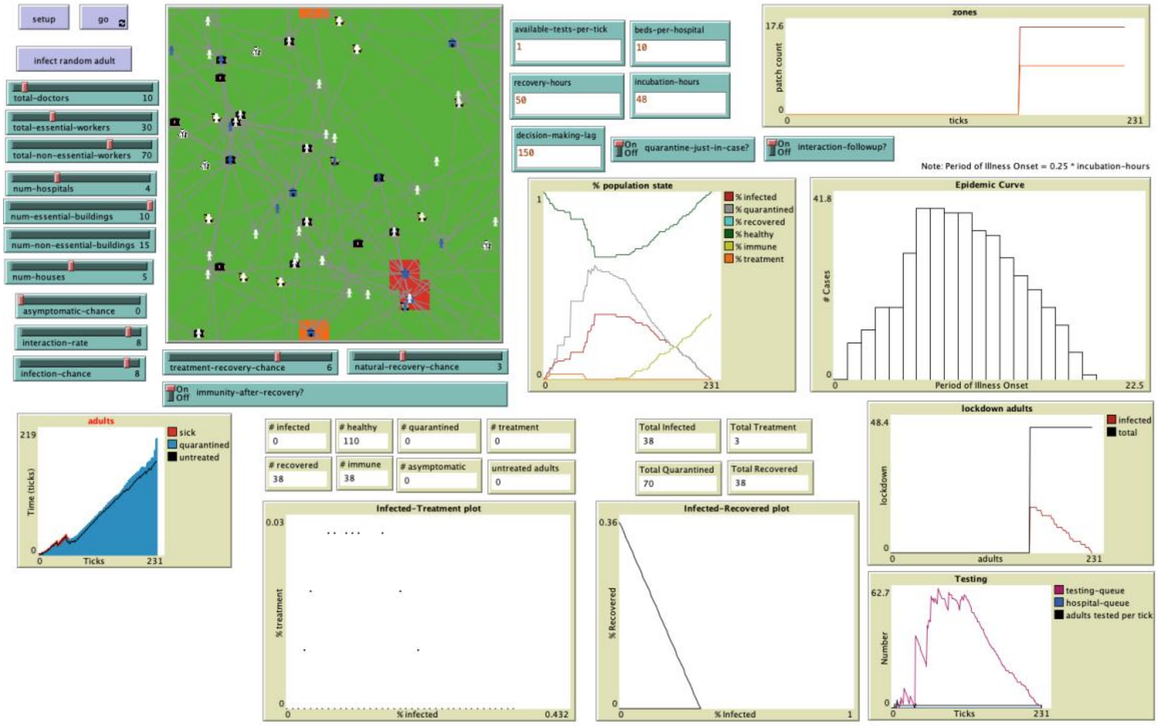

Figure 11: Best case scenario for Case 1_2 group

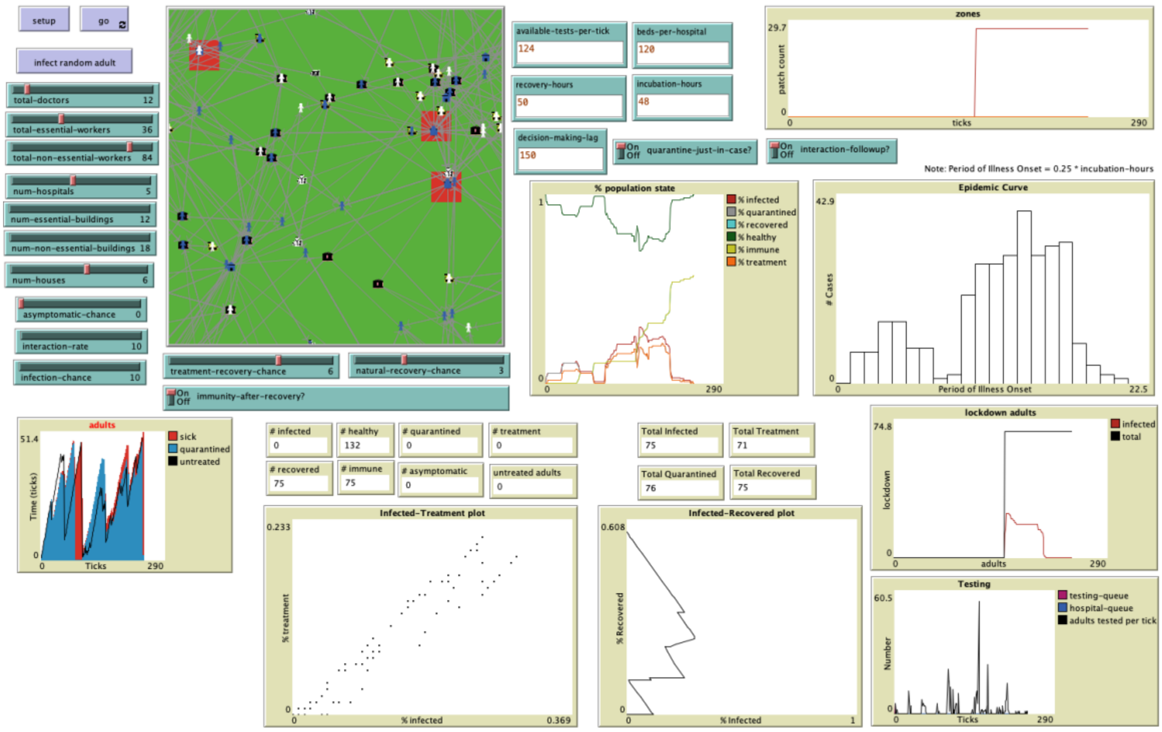

Figure 12: Best case scenario for Case 2_1 group

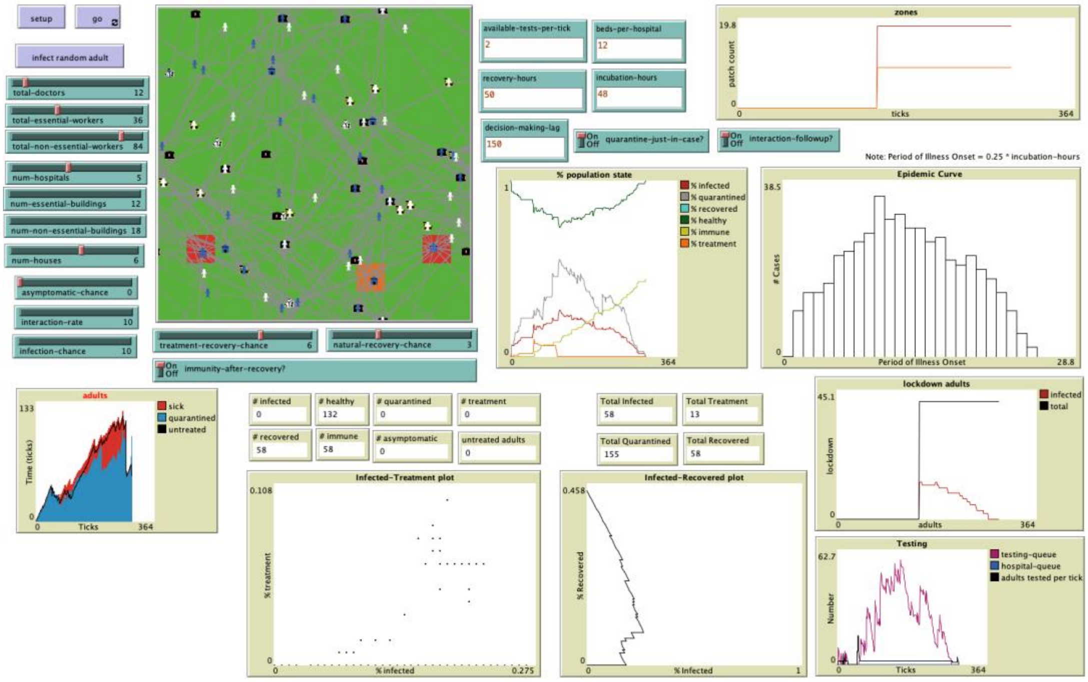

Figure 13: Best case scenario for Case 2_2 group

6.3.2 Worst Case Comparison

In all the worst-case scenarios, the outbreak never ends. However, by capping the results at 30,000 ticks, it is possible to draw some conclusions about the overall dynamics of the crisis.

Firstly, it can be hypothesized that cities with higher interaction scores and infection rates (such as both cases under population B) will have a higher total recovery to total infection ratio than their population A counterparts. Moreover, it can be rationalized that cities with high testing resources (such as case1_1 and case2_1) will have an overall higher total recovery value as their residents are not solely dependent on natural recovery (and quarantining just-in-case is disabled in all worst cases). This rationalization is validated by the simulation, with cases 2_1 > 1_1 > 2_2 > 1_2 with resp. ratios of 0.995, 0.9946, 0.9926, 0.9912. Section 10.6 in the appendix can be referred to for the initialization information and interface results.

Figure 14 contains the % infection - % treatment plots for each case wherein the color of each point on the plots is parameterized by the tick it occurred on, during the 30,000 ticks run.

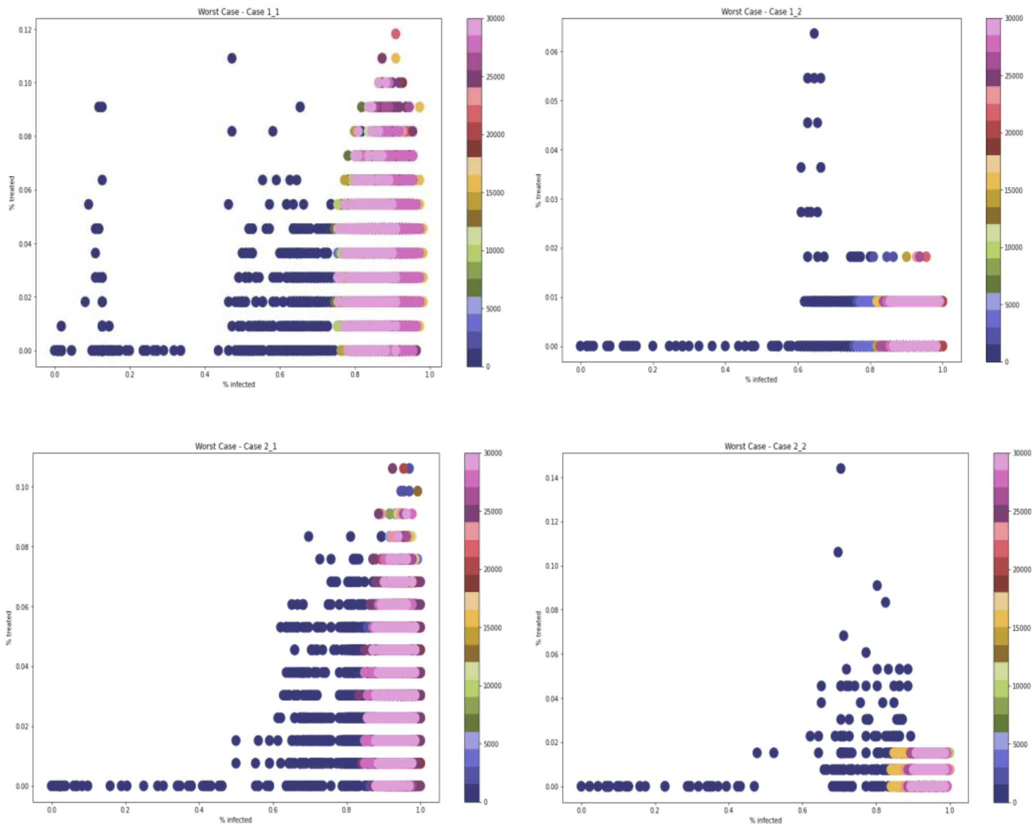

Figure 14: Worst Case Infection-Treatment Plots parameterized by time

It is visually evident that compared to their sibling counterparts, cases 1_1 and 2_1 are spread more on the vertical axis, which makes logical sense as the abundant testing resources in both cases ensures that infected adults can be treated without having to wait in the testing (or hospital) queues.

Moreover, it is interesting to note that in all four cases, there occurs some kind of convergence, which can be visually identified by the plot location that the points, colored within the red and pink spectrum, are more abundant around. For cases 1_1 and 1_2, this convergence happens when % infected is around the [0.8, 1] range, and for cases 2_1 and 2_2 around the [0.85, 1] range. Thus, it can be hypothesized from these results that a city's hyperlocal parameters significantly affect the dynamics of the crisis trajectory.

The peak of the percentage of treated adults in all four cases occurs early in cases 1_2 and 2_2, with the latter converging points occurring within the lower right quarter of the plots. This could be because during the early days of the outbreak, there are existing reserves of testing resources which would account for the initially high treatment percentages. This would imply that a city's existing reserves of testing resources may help alter the crisis initial trajectory significantly, giving the city the opportunity to enforce other intervention strategies.

7 Discussion

As seen from the results, much like real-world cities, there exists a link between the simulated city and crisis's initial conditions, the intervention strategies used, and the impact and duration of the crisis trajectory.

Firstly, a city and crisis's initial conditions affect the optimal intervention strategy, as seen in section 6.2. This result is then compounded, by the second result that initial conditions influence the duration and impact of the crisis trajectory, as seen in sections 6.1.2 and 6.3 respectively. Finally, this feedback trifecta is completed by the results from Figure 7 (in section 6.1.2) and the worst-case and best-case comparisons in section 6.3, which show that intervention strategies (after controlling for initial conditions) influence the duration and impact of the crisis trajectory respectively. The worst-case and best-case comparison in particular have the same initial conditions applied to the same four city contexts, but a difference in the intervention strategy makes the crisis duration go from a small finite duration (best-case) to a potentially never-ending disease outbreak (in the worst-case).

Since the dynamics of the four simulated cities are, for the current scope and purpose, similar to that of a real city and all four of the simulated cities' motivating problem (a disease outbreak) is a symptom of rapid urbanization [74], much like most cities which require the smart city transformation, this section first explores if improving the healthcare system of each of the four cities and transforming it into "smart" healthcare would improve the crisis trajectory. Then this section posits the implications of the previous result on a smart city design for each of the four cities, and what that would mean for a one-size-fits-all generalized smart city planning solution. Finally, this section concludes with possible threats to validity.

7.1 Implications

There are several direct consequences from the model, however, this section deals with the three key implications, born of the results, which are the most applicable to the field of smart cities and smart city planning. The first section explains how the simulated cities can be integrated with smart technologies and frameworks, which is then followed by a section which explores the planning of a smart healthcare system for each of the four simulated cities. The final section answers the motivating question of this thesis: should there be a one-size-fits-all smart city solution?

7.1.1 Inherent Source of Smartness in Simulated Cities

The four simulated cities have some inherent sources of smartness, by virtue of automated contact tracing, real time notifications (which is why there is no lag in zoning policy compliance) and holistic information transparency, resource sharing and open data.

Firstly, as seen in the above results, automated contact tracing (and implicitly, real-time notifications to residents) ranks as the most important intervention-specific feature in the most general scenario: the all cases dataset. These two features can be implemented as subprocedures through an urban operating system (UrbanOS). An UrbanOS enables dynamic governmental decision-making as well as acts as a single-source-of-truth for residents to interact with their city. For example, Ignatieff, whose UrbanOS concept is built on the premise of urban ethics and an implicit understanding that a city's operating system is owned by everyone and its resilience highly depends on the success of the city's leadership [15], defines the key institutional elements of an UrbanOS as enabling: "pathways to documentation for all new arrivals; equality before the law and fair policing; interethnic coalition building in governance; intercommunal fairness in the distribution of contracts and patronage; open real estate ladders; and open job ladders." On the other hand, authors [2], [75], [76] have also defined UrbanOS in terms of consolidating all the city's ICT, by - to paraphrase Komninos *et al.* - integrating all the network infrastructure, sensors and other hardware devices, software - and equally important - *people* across the different domains and urban systems. By integrating citizens as a stakeholder (via participatory governance or other means) back into smart city planning, the execution of smart cities moves more towards the motivating ideal.

Secondly, despite there not being a clear consensus on smart city definitions or general frameworks, there *are* literature narratives which the field generally agrees upon as best practices, with open data as a key area of unanimous assent.

Data and APIs which are open, and secure, ensure that residents are able to access and build upon their own information in a democratic fashion. The above model was only possible due to the underlying assumption of resource sharing, open data and hospital information transparency. If this assumption was not made, then the crisis trajectory would be harder to predict, control and optimize. The UrbanOS equivalent in the model (i.e., matchmaking adults with hospitals with available beds and automated contact tracing) is only possible with the implicit understanding that all data is accessible, open and secure. By ensuring open-sourced and freely available datasets and APIs, innovation is also improved as residents can then build products leveraging these inputs to better their circumstances.

As gleaned in a 2015 paper, Yin *et al*. categorizes available smart city architecture as 1) data-centric, and 2) multidisciplinary smart city. The former category recognizes that data is the foundation for all future realizations of smart cities, and that almost all available literature on smart city architecture identifies data sensing and data transmission as the smart city's fundamental starting point. It is important to understand that the *reason* data is king when talking about smart cities is because the key challenge within smart cities is understanding the interactions between the city and its people [77]; and programmatically, a city can be understood as its sensed data. Which is also why almost all proposed smart city infrastructures include the use of sensors [22].

Moreover, the latter multidisciplinary category, in the paper by Yin *et al.*, highlights the smart city as a systematic concept, wherein, ICT solutions not only help solve multidisciplinary problems, but it is also vital to use multidisciplinary knowledge to understand and solve problems in the urban domain [22]. Relatedly, many papers further validate the model's assumption by rationalizing that since ICT is used to facilitate the operation, maintenance and strategy of smart cities, and, despite strong corporate interests within this field, that it is vital that these ICT systems (and related data) remain open-access [3], [12], [75].

Tangentially, capitalist interest should be minimized as much as possible - on both the smart city branding and the overall design. The former because corporations have a conflict of interest when defining smart cities, due to wanting to play to their company's strengths and thus are incentivized to position themselves as a thought leader in the field for better Search-Engine Optimization and brand loyalty. On the other hand, corporate influences, due to their proprietary processes and technologies, make it difficult to ensure that the datasets and APIs are open-sourced and freely accessible, thus it is preferable to not heavily depend on corporations for smart city design. This view on the corporate smart city model was also identified as a major theme in a scientometric review of more than 7,800 smart city papers which were published over a span of three decades [12].

7.1.2 Is a Smart Healthcare Transformation Possible?

According to the previous section, the four simulated cities have the same level of inherent smartness. By further assuming that all four cities also have the same socio-political and economic climate as well as the same goals, resident demographics and problems in the non-healthcare sectors, the question arises: would enabling "smart healthcare" in these cities help improve their crisis trajectories (with respect to crisis impact and duration), based on the known results on their individual intervention strategy importance?

Firstly, as can be seen by the feature importance ranking results, for city scenarios with low resources (case1_2 and case 2_2), control measures (such as, quarantining just-in-case) rank higher than research and innovation measures (such as studying the disease to learn more about its average recovery time, which would then determine the minimum quarantining period an adult under treatment would have to wait before taking a test). Secondly, as seen in the 6.3.2 section, for such cities, the initial days of the crisis are especially vital as at that time, there are existing reserves of testing resources which would allow for treating, by percentage, more adults than would be possible later in the crisis trajectory. By treating more adults earlier rather than later, the flattened infection curve may allow the city the opportunity to enforce other intervention strategies. Thirdly, it is important to note that control measures without social security as a backup, especially earlier on in the outbreak lifecycle (when the public may not even recognize it as a problem – as seen by the research on inertia in public opinion in section 3.1.1) may worsen public trust and quality of life, the ensuing spillover costs may include unrest, noncompliance with policies and even crime.

These three observations imply that a smart healthcare design for cities with low testing resources would have to prioritize early detection and treatment, smart community measures (with a focus on engendering community trust and awareness) and strengthening the social security measures so that strict policies (such as mandatory quarantining just in case) can be enforced without negatively affecting an adult's quality of life, mental and social wellbeing, or livelihood. Furthermore, control measures would imply the need for resource transparency, open data and investing in ICT for developing applications that offer real-time notifications about automated contact tracing, zoning policies and infection alerts by location. In such cities, it is inevitable that, due to the need for stricter control measures, there will be a negative effect on businesses, management and processes within the city. Thus, a smart healthcare framework that works on reducing the total time in crisis by early detection and treatment of the illness, as well as strict control measures, would be recommended.

In contrast, for cities which don't have to worry about limited resources, a healthcare system which prioritizes data sensing and collection, as well as research on the disease would yield a better outcome; with *recovery-hours, interaction-followup?* and *decision-making-lag* generally ranking as relatively more important than the quarantine just-in-case strategy. This implies that such cities would prioritize an ICT architecture which makes use of real-time data from sensors and hospitals. In such cities, it would be important that policies (zoning and otherwise) reflect the latest research on the disease, as this way it would be possible to reduce the scale of impact on the everyday processes, management and businesses within the cities. In the worst-case scenario in such cities, the cyclically linked % infected and % treated plots would vary often and by a significant degree. Thus, a healthcare system which first focuses on analyzing the disease and finetuning policies to best reduce the impact on everyday adults, while flattening the infection curve would be optimal.

Thus, even when abstracted, it can be seen that even for such simple simulations of cities, the suggested smart healthcare strategy changes.

7.1.3 Should There Be A One-Size-Fits-All Smart City Solution?

As we have seen above, even for the simplified model, the initial conditions (with respect to the analysis process, problem and city contexts) influence the crisis trajectory in an unpredictable way. Even after fixing the analysis's initial conditions: not at all initial condition variables (AKA features) are created equal, with a select number of features being more important than others in influencing the crisis's completion time. *But,* even these so-called “analyzable” important features lists change, based on the scope of the analysis (any two selected case comparison scenarios have feature lists with non-overlapping rankings). Furthermore, as seen in best- and worst-case study, even with the problem’s initial conditions fixed, a city’s hyperlocal parameters can disproportionately affect not only the crisis duration but also the crisis’s impact in an unpredictable manner. To further add to the analysis anarchy, the effective intervention strategies may be different based on the unique city and problem context, and may not be intuitive.

Due to the complex nature of the city and its agents, the inherent feedback loops and emergence born from the actions of the actors interacting with their environment, it is nigh impossible to predict a general solution to a complex dynamical systems equation without considering the specifics involved, for the city and problem under question - something which was explained by the literature review and validated by the designed model. If a simple model with understandable behavioral rules is so difficult to analyze, predict and generalize, despite having an omniscient bird’s eye view with the ability to see the effects of every action (pulled from a finite set of possible actions) on the future city (with ticks), then what about the scaled-up actuality that is a real-life city with its three areas of unknowability?

This instinct is validated by the previous section on designing the priorities of a smart healthcare system for the four simulated cities. The previous section shows that even though the four cities have more features in common than not, due to their few differing initial conditions, their healthcare plans are different. If such minimal diversity in so few dimensions gives rise to so much variance within the plan for smart healthcare, which is a subcomponent of smart city [60], [78], then what about the breadth of diversity found within real cities along multiple dimensions? It wouldn’t be incorrect to then hypothesize that smart city planning would not benefit from a generalized, one-size-fits-all approach, based on the facts that (smart) cities have subcomponents which co-evolve competitively and interact nonlinearly (as seen in section 3.1.1) and diversity within cities will only increase in the future [15].

So, where do we go from here?

7.1.4 Smart City Planning, Reimagined

Firstly, taking a cue from existing literature, while many authors agree that smart cities should adapt their implementation plan to their city’s vision and goals, and that the motivation, design, and execution would look different in different smart cities [3], it is important to note that the priorities and the core understanding of *how* a smart city works also varies with the specific city in question [2]. Often strategies or interventions that work in a specific city might not work in another city's context due to the inherent multilevel nature of cities, as they are systems of systems of systems and therefore have hard-to-predict behavior. However, a possible method to test an intervention's probable success in a new context is via simulation and agent-based modeling tools [8]. This thesis recommends the use of non-traditional tools which don't only depend on gaussian statistics as power laws within cities muddy the veracity, and results of, the analysis process.

Komninos *et al.* reiterate these sentiments in their 2019 paper, defining smart city planning as "a process that highlights the uniqueness of each city trajectory, is based on rapidly changing digital technologies, and is ready to value opportunities offered over time rather than copycat planning, locked-in optimal models and one-size-fits-all solutions." and confirms that "transferring growth models from one region to another is questionable as there is no "optimal" development model, and new successful trajectories and developmental paths emerge spontaneously and unexpectedly in space" (Komninos et al., 2019). Thus, it is important to lean on non-traditional tools of analysis (such as agent-based modeling and fractals) when dealing with such systems.

Secondly, it is also recommended to leverage a hyperlocal and hybrid approach for analysis, which leverages both bottom-up and top-down strategies at different levels. To begin with, using a bottom-up strategy, Johnson's 2012 paper leverages the dynamic nature of the methodology for building formal vocabulary proposed by Gould *et al.* which argues that such a process should be guided by 'The Principle of Usefulness' [79] wherein only terms which are useful for a particular city and its strategic policies should be included, with the adding of new terms allowed as the city and its goals evolve. Such a formal vocabulary could, theoretically, grow forever [8]. Once important hyperlocal features have been identified by this principle, a top-down strategy can be used to further qualify key dimensions for later online analysis - much like the analysis methodology followed in this thesis, wherein data was collected via simulation along multiple and varying dimensions, and then important features were ranked and selected by a machine learning model.

Thirdly, while the smart city planning should be customized to the specific goals of the city and its residents, it’s important to recognize that, based on the trends within the literature, all city planning must be directed towards facilitating sustainability.

7.2 Threats to Validity

Despite the knowledge that models such as these are sensitive to initial conditions (for both analysis methodology and possible end results), due to resource constraints, it was necessary to limit the scope of the analysis in a few dimensions, such as setting the number of maximum ticks to 10,000, repeating each unique simulation only three times, only dealing with a closed population and having a finite set of initial conditions. Thus, it could be argued that, due to the feedback loops, emergence and inherent nonlinear dynamics within the model, the results may drastically change were the scope to expand.

However, considering that this simple model was not meant to be a one-to-one accurate simulation of outbreak dynamics, and was instead meant to be a barometer for proving that if such simple model - with so much of its complexity smoothened, could demonstrate sensitivity to initial conditions, unpredictable crisis trajectories, and show that a one-size-fits-all intervention strategy would not apply within even the model’s (comparatively) simplistic constraints - then how could these properties not be true for a city, which is much more complex system, with scaled-up drivers of growth and innovation, and heterogenous in terms of types of agents as well as decision-making priorities? In this sense, the simplicity of the model acts as a tool for putting into perspective the magnitude of the difference.

Additionally, as seen from the literature review section, it is nigh impossible to predict which initial conditions should be chosen and/or would prove interesting in terms of dynamics and end states. Thus, while theoretically, studying tipping points could prove more insightful than the initializations chosen above, such tipping points are hard to forecast without simulating every single combination possible and analyzing phase plots and creating bifurcation diagrams. Thus, creating a chicken-or-an-egg paradox.

The designed model does not include the impact of external sources and drivers of innovation, such as outside funding, international opportunities and programs for enrichment and development, etc. However, while this exclusion makes the model context more isolated and sandboxed, and could greatly reduce the ticks to completion data, it is important to note that such drivers of innovation (and complexity) can never be taken for granted and depends on an individual city's ability to apply and positively leverage such opportunities. But, this exclusion is even more important when comparing across different cities which have different socioeconomic and political climates (and thus, different access levels to such opportunities) and could also influence the results of a cross-sectional analysis.

8 Conclusion

The thesis began by exploring the foundational elements of the smart city concept according to the existing literature. Following this attempt at overviewing the state of the art within the field, this report took a step back to fully examine the city from three interlinked perspectives, the: ecological, complexity science and organizational. This effort to shed light on the backend engines of city science was done in order to then identify hints for novel methodologies to rethink, re-contextualize and reframe the smart city concept. The succeeding section then explained the design and rationale for an agent-based-model which simulated a disease outbreak in a closed population parameterized by user-defined problem and city context initial conditions, with sections on an analysis methodology and results following it. The penultimate part then discussed the potential implications and consequences of this model's results.

This thesis contains four key contributions, which are: 1) a theoretical review of the state of the art in smart city planning, 2) an agent-based model which highlighted that initial conditions affect optimal intervention strategies and the crisis trajectory (and that the three are in fact linked), 3) a case-study led hypothesis, validated by a qualitative literature review, which posits that a one-size-fits-all generalized framework is not optimal when planning for smart cities and provides some recommendations for a new style of planning, and lastly, 4) a key meta-result, this thesis also explains and uses an analysis methodology for holistic and quantitative identification of important features in (smart) city planning models.

8.1 Future Work

The current model does not fully study the impact of the outbreak on the non-essential businesses, this economic analysis could also help shed light on the consequences of zoning.

Future research could range from focusing on expanding the dimensions in the model, such as changing the population from closed to semi-closed to open, to implementing new outbreak control measures, such as mask mandates.

If the population is not closed, in order to control fatalities and add realism to the model, a vaccine distribution strategy should also be implemented.

Research should also be conducted on the impact of citizen decision-making on the results, for example, by using a stochastic distribution for psychological tendencies to go to the hospital or voluntary mask wearing, and analyzing how the end results then change. In terms of control measures, it would be interesting to understand the impact on the results post-integration of a participatory governance scheme for automated decision-making in the urban operating system.

While it is known that such models and analysis methods are not an exact science, and some complexity is lost, it is important to consider a city in a holistic way that respects its complexity. For, while such an analysis might not be the most accurate or comprehensive answer, it is *a* starting point of analysis where previously there had been none. By also leveraging external opportunities for growth and development, and including sustainability as a planning goal, evolutionary planning strategies and citizen participation into the process, analysis like this will move from the sandbox to the real world.

"What is at stake here is more than just keeping the show on the road. When we achieve the fragile common good called community, when the real economy of a city begins to approximate, however imperfectly, the moral economy, we achieve an important victory in relationship to globalization itself. We cease to feel that we are the prisoner of forces stronger than ourselves. We cease to feel like pawns in someone else's game. To have a moral community in a city is to recover sovereignty and mastery. It is to have the sense that we can work together to shape our common life to humane, not inhuman ends."

- The Moral Operating System of a Global City [15]

9 References

[1] J. Portugali, *Complexity theories of cities have come of age: an overview with implications to urban planning and design*. Heidelberg New York: Springer, 2012.

[2] N. Komninos, C. Kakderi, A. Panori, and P. Tsarchopoulos, “Smart city planning from an evolutionary perspective,” *J. Urban Technol.*, vol. 26, no. 2, pp. 3–20, 2019.

[3] V. Albino, U. Berardi, and R. M. Dangelico, “Smart Cities: Definitions, Dimensions, Performance, and Initiatives,” *J. Urban Technol.*, vol. 22, no. 1, pp. 3–21, Jan. 2015, doi: 10.1080/10630732.2014.942092.

[4] L. Zhao, Z. Tang, and X. Zou, “Mapping the Knowledge Domain of Smart-City Research: A Bibliometric and Scientometric Analysis,” *Sustainability*, vol. 11, no. 23, Art. no. 23, Jan. 2019, doi: 10.3390/su11236648.

[5] C. Echebarria, J. M. Barrutia, and I. Aguado-Moralejo, “The Smart City journey: a systematic review and future research agenda,” *Innov. Eur. J. Soc. Sci. Res.*, vol. 34, no. 2, pp. 159–201, Apr. 2021, doi: 10.1080/13511610.2020.1785277.

[6] A. Sharifi, Z. Allam, B. Feizizadeh, and H. Ghamari, “Three Decades of Research on Smart Cities: Mapping Knowledge Structure and Trends,” *Sustainability*, vol. 13, no. 13, Art. no. 13, Jan. 2021, doi: 10.3390/su13137140.

[7] A. Camero and E. Alba, “Smart City and information technology: A review,” *Cities*, vol. 93, pp. 84–94, Oct. 2019, doi: 10.1016/j.cities.2019.04.014.

[8] J. Johnson, “Cities: Systems of Systems of Systems,” in *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning and Design*, J. Portugali, H. Meyer, E. Stolk, and E. Tan, Eds. Berlin, Heidelberg: Springer, 2012, pp. 153–172. doi: 10.1007/978-3-642-24544-2_9.

[9] G. V. Pereira, P. Parycek, E. Falco, and R. Kleinhans, “Smart governance in the context of smart cities: A literature review,” *Inf. Polity*, vol. 23, no. 2, pp. 143–162, 2018, doi: 10.3233/IP-170067.

[10] Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanovic, N., and Meijers, E., “Smart cities Final Report ranking of European Medium-sized Cities Final Report 2007,” Centre of Regional Science, Vienna UT, 2007. Accessed: Feb. 28, 2022. [Online]. Available: http://www.smart-cities.eu/ http://www.smart-cities.eu/download/smart_cities_final_report.pdf

[11] A. Caragliu, C. Del Bo, and P. Nijkamp, “Smart Cities in Europe,” *J. Urban Technol.*, vol. 18, no. 2, pp. 65–82, Apr. 2011, doi: 10.1080/10630732.2011.601117.

[12] C. Zheng, J. Yuan, L. Zhu, Y. Zhang, and Q. Shao, “From digital to sustainable: A scientometric review of smart city literature between 1990 and 2019,” *J. Clean. Prod.*, vol. 258, p. 120689, Jun. 2020, doi: 10.1016/j.jclepro.2020.120689.

[13] United Nations, “World Urbanization Prospects: The 2018 Revision - Key Facts,” United Nations, 2019. Accessed: May 31, 2022. [Online]. Available: https://population.un.org/wup/Publications/Files/WUP2018-KeyFacts.pdf

[14] United Nations, Department of Economic and Social Affairs, and Population Division, *World urbanization prospects: the 2018 revision*. 2019.

[15] Michael Ignatieff, “The Moral Operating System of a Global City: Los Angeles,” *Carnegie Council for Ethics in International Affairs*, Jan. 31, 2014. https://www.carnegiecouncil.org/media/series/100/the-moral-operating-system-of-a-global-city-los-angeles (accessed May 31, 2022).

[16] T. Nam and T. A. Pardo, “Conceptualizing smart city with dimensions of technology, people, and institutions,” in *Proceedings of the 12th annual international digital*

government research conference: digital government innovation in challenging times, 2011, pp. 282–291.

[17] J. H. Lee, R. Phaal, and S.-H. Lee, “An integrated service-device-technology roadmap for smart city development,” *Technol. Forecast. Soc. Change*, vol. 80, no. 2, pp. 286–306, 2013.

[18] M. Scheffer, F. Westley, and W. Brock, “Slow Response of Societies to New Problems: Causes and Costs,” *Ecosystems*, vol. 6, no. 5, pp. 493–502, 2003.

[19] M. Equihua *et al.*, “Ecosystem antifragility: beyond integrity and resilience,” *PeerJ*, vol. 8, p. e8533, Feb. 2020, doi: 10.7717/peerj.8533.

[20] D. Sijmons, “Simple Rules: Emerging Order? A Designer’s Curiosity About Complexity Theories,” in *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning and Design*, J. Portugali, H. Meyer, E. Stolk, and E. Tan, Eds. Berlin, Heidelberg: Springer, 2012, pp. 281–309. doi: 10.1007/978-3-642-24544-2_16.

[21] D. Washburn and U. Sindhu, “Helping CIOs Understand ‘Smart City’ Initiatives,” *Smart City*, p. 17, 2010.

[22] C. Yin, Z. Xiong, H. Chen, J. Wang, D. Cooper, and B. David, “A literature survey on smart cities,” *Sci. China Inf. Sci.*, vol. 58, no. 10, pp. 1–18, Oct. 2015, doi: 10.1007/s11432-015-5397-4.

[23] D. Tokody and G. Schuster, “Driving forces behind Smart city implementations-The next smart revolution,” *J. Emerg. Res. Solut. ICT*, vol. 1, no. 2, pp. 1–16, 2016.

[24] A. Kousis and C. Tjortjis, “Data mining algorithms for smart cities: A bibliometric analysis,” *Algorithms*, vol. 14, no. 8, 2021, doi: 10.3390/a14080242.

[25] J. Portugali, “What Makes Cities Complex?,” in *Complexity, Cognition, Urban Planning and Design*, Cham, 2016, pp. 3–19. doi: 10.1007/978-3-319-32653-5_1.

[26] B. Jiang and T. Jia, “Zipf’s law for all the natural cities in the United States: a geospatial perspective,” *Int. J. Geogr. Inf. Sci.*, vol. 25, no. 8, pp. 1269–1281, Aug. 2011, doi: 10.1080/13658816.2010.510801.

[27] M. Batty and S. Marshall, “The Origins of Complexity Theory in Cities and Planning,” in *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning and Design*, J. Portugali, H. Meyer, E. Stolk, and E. Tan, Eds. Berlin, Heidelberg: Springer, 2012, pp. 21–45. doi: 10.1007/978-3-642-24544-2_3.

[28] U. Ekman, “Smart City Planning: Complexity,” *International Journal of E-Planning Research (IJEPR)*, Jul. 01, 2018. www.igi-global.com/article/smart-city-planning/204622 (accessed Feb. 07, 2021).

[29] S. G. Ortman, J. Lobo, and M. E. Smith, “Cities: Complexity, theory and history,” *PLOS ONE*, vol. 15, no. 12, p. e0243621, Dec. 2020, doi: 10.1371/journal.pone.0243621.

[30] O. Poocharoen, P. Thiengburanathum, and K. C. Lee, “The smart city as a complex adaptive system: the ebbs and flows of humans and materials,” in *Smart Cities in Asia*, Edward Elgar Publishing, 2020.

[31] B. Jiang, “A City Is a Complex Network,” *ArXiv150908452 Phys.*, Feb. 2016, Accessed: Jul. 06, 2021. [Online]. Available: http://arxiv.org/abs/1509.08452

[32] A. F. Siegenfeld and Y. Bar-Yam, “An Introduction to Complex Systems Science and Its Applications,” *Complex. Hoboken*, vol. 2020, 2020, doi: http://dx.doi.org.ezproxy1.hw.ac.uk/10.1155/2020/6105872.

[33] J. Jacobs, *The death and life of great American cities*. New York, NY, 1961.

[34] S. Marshall, “Planning, Design and the Complexity of Cities,” in *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning and Design*, J. Portugali, H. Meyer, E. Stolk, and E. Tan, Eds. Berlin, Heidelberg: Springer, 2012, pp. 191–205. doi: 10.1007/978-3-642-24544-2_11.

[35] S. Marshall, *Cities Design and Evolution*, 1st ed. Routledge, 2009. Accessed: Jun. 06, 2022. [Online]. Available: https://www.routledge.com/Cities-Design-and-Evolution/Marshall/p/book/9780415423298

[36] N. Pelletier, "Of laws and limits: An ecological economic perspective on redressing the failure of contemporary global environmental governance," *Glob. Environ. Change*, vol. 20, no. 2, pp. 220–228, May 2010, doi: 10.1016/j.gloenvcha.2009.12.006.

[37] J. Portugali, "Complexity Theories of Cities: Achievements, Criticism and Potentials," in *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning and Design*, J. Portugali, H. Meyer, E. Stolk, and E. Tan, Eds. Berlin, Heidelberg: Springer, 2012, pp. 47–62. doi: 10.1007/978-3-642-24544-2_4.

[38] Rob Roggema, "The Future of Sustainable Urbanism: Society-Based, Complexity-Led, and Landscape-Driven," *Sustainability*, 2017, doi: doi:10.3390/su9081442.

[39] M. Batty, *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. Cambridge, MA, USA: MIT Press, 2005.

[40] I. Blecic, A. Cecchini, and G. A. Trunfio, "A general-purpose geosimulation infrastructure for spatial decision support," in *Transactions on Computational Science VI*, Springer, 2009, pp. 200–218.

[41] I. Blecic, A. Cecchini, and G. A. Trunfio, "Cellular automata simulation of urban dynamics through GPGPU," *J. Supercomput.*, vol. 65, no. 2, pp. 614–629, 2013.

[42] I. Blecic, A. Cecchini, and G. A. Trunfio, "How much past to see the future: a computational study in calibrating urban cellular automata," *Int. J. Geogr. Inf. Sci.*, vol. 29, no. 3, pp. 349–374, 2015.

[43] S. Kalkowski, "Collective Intelligence and its Potential in Smart Cities," *Read. Artif. Intell. Appl.*, p. 21, 2014.

[44] J. Ravetz, *Deeper City: Collective Intelligence and the Pathways from Smart to Wise*. Routledge, 2020.

[45] John H. Miller and Scott E. Page, *Complex Adaptive Systems*. Princeton University Press, 2007. Accessed: Aug. 15, 2022. [Online]. Available: https://press.princeton.edu/books/paperback/9780691127026/complex-adaptive-systems

[46] L. Zomer, E. Moustaid, and S. Meijer, "A meta-model for including social behavior and data into smart city management simulations," in *2015 Winter Simulation Conference (WSC)*, Dec. 2015, pp. 1705–1716. doi: 10.1109/WSC.2015.7408289.

[47] J. Lobo *et al.*, "Urban Science: Integrated Theory from the First Cities to Sustainable Metropolises," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3526940, Jan. 2020. doi: 10.2139/ssrn.3526940.

[48] M. Svítek, P. Skobelev, and S. Kozhevnikov, "Smart City 5.0 as an Urban Ecosystem of Smart Services," in *Service Oriented, Holonic and Multi-agent Manufacturing Systems for Industry of the Future*, Cham, 2020, pp. 426–438. doi: 10.1007/978-3-030-27477-1_33.

[49] J. V. Henderson, *Urban development*. New York, NY: Oxford University Press, 1991.

[50] M. Fujita, P. Krugman, and A. J. Venables, *The spatial economy*. London, England: MIT Press, 2001.

[51] J. K. Brueckner, *Lectures on urban economics*. London, England: MIT Press, 2011.

[52] L. M. A. Bettencourt, J. Lobo, D. Strumsky, and G. B. West, "Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities," *PLOS ONE*, vol. 5, no. 11, p. e13541, Nov. 2010, doi: 10.1371/journal.pone.0013541.

[53] C. Kühnert, D. Helbing, and G. B. West, "Scaling laws in urban supply networks," *Phys. Stat. Mech. Its Appl.*, vol. 363, no. 1, pp. 96–103, Apr. 2006, doi: 10.1016/j.physa.2006.01.058.

[54] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West, “Growth, innovation, scaling, and the pace of life in cities,” *Proc. Natl. Acad. Sci.*, vol. 104, no. 17, pp. 7301–7306, Apr. 2007, doi: 10.1073/pnas.0610172104.

[55] G. West, *Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies*, First Edition. New York: Penguin Press, 2017.

[56] G. Xu, Z. Zhou, L. Jiao, T. Dong, and R. Li, “Cross-sectional Urban Scaling Fails in Predicting Temporal Growth of Cities.” arXiv, Oct. 17, 2019. Accessed: Jun. 22, 2022. [Online]. Available: http://arxiv.org/abs/1910.06732

[57] L. Bettencourt, V. Yang, J. Lobo, C. Kempes, D. Rybski, and M. Hamilton, “The interpretation of urban scaling analysis in time,” *J. R. Soc. Interface*, 2020, doi: 10.1098/rsif.2019.0846.

[58] V. D. Florio, “Social translucence as a transition-inducing mechanism in complex socio-ecological systems,” *Ecol. Complex.*, vol. 44, p. 100883, Dec. 2020, doi: 10.1016/j.ecocom.2020.100883.

[59] C. Yang and U. Wilensky, “NetLogo epiDEM Basic model.” Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 2011. Accessed: Aug. 20, 2022. [Online]. Available: http://ccl.northwestern.edu/netlogo/models/epiDEMBasic

[60] D. Costa and J. P. Peixoto, “COVID-19 Pandemic: A Review of Smart Cities Initiatives to Face New Outbreaks,” *IET Smart Cities*, Jul. 2020, doi: 10.1049/iet-smc.2020.0044.

[61] M. Greer, R. Saha, A. Gogliettino, C. Yu, and K. Zollo-Venecek, “Emergence of oscillations in a simple epidemic model with demographic data,” *R. Soc. Open Sci.*, vol. 7, no. 1, p. 191187, doi: 10.1098/rsos.191187.

[62] V. Grimm, U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback, “The ODD protocol: A review and first update,” *Ecol. Model.*, vol. 221, no. 23, pp. 2760–2768, 2010, doi: https://doi.org/10.1016/j.ecolmodel.2010.08.019.

[63] N. N. Taleb, “A Map and Simple Heuristic to Detect Fragility, Antifragility, and Model Error,” *SSRN Electron. J.*, 2011, doi: 10.2139/ssrn.1864633.

[64] Wilensky, U, “NetLogo.” Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999. [Online]. Available: http://ccl.northwestern.edu/netlogo/.

[65] IPython, “ipyparallel.” GitHub repository, 2015. [Online]. Available: https://github.com/ipython/ipyparallel

[66] M. Jaxa-Rozen and J. H. Kwakkel, “PyNetLogo: Linking NetLogo with Python,” *J. Artif. Soc. Soc. Simul.*, vol. 21, no. 2, p. 4, 2018.

[67] C. Gunaratne and I. Garibay, “NL4Py: Agent-based modeling in Python with parallelizable NetLogo workspaces,” *SoftwareX*, vol. 16, Dec. 2021, doi: 10.1016/j.softx.2021.100801.

[68] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, New York, NY, USA, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[69] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[70] J. H. Friedman, “Greedy function approximation: A gradient boosting machine.,” *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.

[71] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[72] “Permutation Importance with Multicollinear or Correlated Features,” *scikit-learn*. https://scikit-

learn/stable/auto_examples/inspection/plot_permutation_importance_multicollinear.html (accessed Aug. 20, 2022).

[73] The Association of Faculties of Medicine of Canada, “Chapter 7 Applications of Research Methods | AFMC Primer on Population Health,” in *AFMC Primer on Population Health*, Accessed: Aug. 11, 2022. [Online]. Available: https://phprimer.afmc.ca/en/part-ii/chapter-7/#_Toc253401506

[74] C.-J. Neiderud, “How urbanization affects the epidemiology of emerging infectious diseases.,” *Infect. Ecol. Epidemiol.*, vol. 5, p. 27060, 2015, doi: 10.3402/iee.v5.27060.

[75] S. Zygiaris, “Smart City Reference Model: Assisting Planners to Conceptualize the Building of Smart City Innovation Ecosystems,” *J. Knowl. Econ.*, vol. 4, no. 2, pp. 217–231, Jun. 2013, doi: 10.1007/s13132-012-0089-4.

[76] A. Luque-Ayala and S. Marvin, *Urban Operating Systems: Producing the Computational City*. The MIT Press, 2020. doi: 10.7551/mitpress/10869.001.0001.

[77] A. J. Jara, D. Genoud, and Y. Bocchi, “Big data in smart cities: from poisson to human dynamics,” in *2014 28th International Conference on Advanced Information Networking and Applications Workshops*, 2014, pp. 785–790.

[78] J.-H. Kim and J.-Y. Kim, “How Should the Structure of Smart Cities Change to Predict and Overcome a Pandemic?,” *Sustainability*, vol. 14, no. 5, Art. no. 5, Jan. 2022, doi: 10.3390/su14052981.

[79] P. Gould, etc., J. H. Johnson, and G. H. Chapman, *The structure of television*. London, England: Pion, 1984.

[80] D. Helbing *et al.*, “Power Laws in Urban Supply Networks, Social Systems, and Dense Pedestrian Crowds,” in *Complexity Perspectives in Innovation and Social Change*, Dordrecht: Springer Netherlands, 2009, pp. 433–450. doi: 10.1007/978-1-4020-9663-1_17.

10 Appendix

10.1 Combinations for Simulated Initializations

- **Case 1.1: Best Case - Resources**
- Fixed: city and individual specific initialization
    - total-doctors = 10
    - total-essential-workers = 30
    - total-non-essential-workers = 70
    - num-hospitals = 4
    - num-essential-buildings = 10
    - num-non-essential-buildings = 15
    - num-houses = 5
    - interaction-rate = 8
    - infection-chance = 8
    - treatment-recovery-chance = 6
    - natural-recovery-chance = 3
- Fixed - High: city resource specific variables
    - available-tests-per-tick = 100
    - beds-per-hospital = 100
- Variable: problem-specific variables
    - incubation-hours-options = {48, 50}
    - recovery-hours-options = {50, 100}
    - immunity-after-recovery? = {true, false}
    - asymptomatic-chance-options = {0, 3}
- Variable: intervention-specific variables
    - decision-making-lag-options = {50, 100, 150, 200}
    - interaction-followup? = {true, false}
    - quarantine-just-in-case? = {true, false}

- **Case 1.2: Worst Case - Resources**
- Fixed: city and individual specific initialization
    - total-doctors = 10
    - total-essential-workers = 30
    - total-non-essential-workers = 70
    - num-hospitals = 4
    - num-essential-buildings = 10
    - num-non-essential-buildings = 15
    - num-houses = 5
    - interaction-rate = 8
    - infection-chance = 8
    - treatment-recovery-chance = 6
    - natural-recovery-chance = 3
- Fixed - Low: city resource specific variables
    - available-tests-per-tick = 1
    - beds-per-hospital = 10
- Variable: problem-specific variables

- incubation-hours-options = {48, 50}
- recovery-hours-options = {50, 100}
- immunity-after-recovery? = {true, false}
- asymptomatic-chance-options = {0, 3}

- Variable: intervention-specific variables
  - decision-making-lag-options = {50, 100, 150, 200}
  - interaction-followup? = {true, false}
  - quarantine-just-in-case? = {true, false}

- **Case 2.2: Best Case - Resources**
- Fixed: city and individual specific initialization
  a. total-doctors = 12
  b. total-essential-workers = 36
  c. total-non-essential-workers = 84
  d. num-hospitals = 6
  e. num-essential-buildings = 12
  f. num-non-essential-buildings = 18
  g. num-houses = 5
  h. interaction-rate = 10
  i. infection-chance = 10
  j. treatment-recovery-chance = 6
  k. natural-recovery-chance = 3
- Fixed - High: city resource specific variables
  a. available-tests-per-tick = 124
  b. beds-per-hospital = 120
- Variable: problem-specific variables
  a. incubation-hours-options = {48, 50}
  b. recovery-hours-options = {50, 100}
  c. immunity-after-recovery? = {true, false}
  d. asymptomatic-chance-options = {0, 3}
- Variable: intervention-specific variables
  a. decision-making-lag-options = {50, 100, 150, 200}
  b. interaction-followup? = {true, false}
  c. quarantine-just-in-case? = {true, false}

- **Case 2.2: Worst Case - Resources**
- Fixed: city and individual specific initialization
  a. total-doctors = 12
  b. total-essential-workers = 36
  c. total-non-essential-workers = 84
  d. num-hospitals = 6
  e. num-essential-buildings = 12
  f. num-non-essential-buildings = 18
  g. num-houses = 5
  h. interaction-rate = 10
  i. infection-chance = 10
  j. treatment-recovery-chance = 6
  k. natural-recovery-chance = 3
- Fixed - Low: city resource specific variables
  a. available-tests-per-tick = 2
  b. beds-per-hospital = 12

- Variable: problem-specific variables
    a. incubation-hours-options = {48, 50}
    b. recovery-hours-options = {50, 100}
    c. immunity-after-recovery? = {true, false}
    d. asymptomatic-chance-options = {0, 3}
- Variable: intervention-specific variables
    a. decision-making-lag-options = {50, 100, 150, 200}
    b. interaction-followup? = {true, false}
    c. quarantine-just-in-case? = {true, false}

10.2 Macro-Analysis Scaling

Category	Scaling Exponent	Notes	Macro-Analysis Cases: Best Case: Resources — 1.1. At Population A	Macro-Analysis Cases: Best Case: Resources — 2.1. At Population B	Macro-Analysis Cases: Worst Case: Resources — 1.2. At Population A	Macro-Analysis Cases: Worst Case: Resources — 2.2. At Population B
total population	1.2	Linear scale	110	132	110	132
num-doctors	1.2	Linear scale. In reality, this value should be 1.15 on a log-log scale [53] but since doctors don't affect the end ticks dynamics in a medical-function sense, this is assumed linear	10	12	10	12
num-essential-workers	1.2	Linear scale	30	36	30	36
num-non-essential-workers	1.2	Linear scale	70	84	70	84
num-houses	$\beta = 1$	Log-log scale [54]	5	6	5	6
num-essential-buildings	$\beta = 0.85$	Log-log scale [53]	10	12	10	12
num-non-essential-buildings	$\beta = 1$	Log-log scale [53]	15	18	15	18
num-hospitals	$\beta = 1$	Log-log scale [53]	4	5	4	5
interaction-rate	$\beta = 1.15$	Log-log scale [80]	8	10	8	10
infection-chance	$\beta = 1.15$	Log-log scale [54]	8	10	8	10
beds-per-hospital	$\beta = 1$	Log-log scale [53]	100	120	10	12
available-tests-per-tick	$\beta = 1.15$	Log-log scale. Assumed so because GDP and innovation scale superlinearly on a log-log scale [54], which would theoretically also scale this value superlinearly	100	124	1	2

10.3 Feature Importance

10.3.1 Importance Scores per Case Comparison

Feature\Case Type	cases1	cases2	case1_1	case1_2	case2_1	case2_2	all_cases
immunity-after-recovery?	0.346	0.399	0.362	0.344	0.364	0.382	0.373
interaction-followup?	0.141	0.120	0.168	0.109	0.128	0.131	0.139
quarantine-just-in-case?	0.092	0.075	0.049	0.139	0.025	0.135	0.086
recovery-hours	0.074	0.091	0.115	0.031	0.141	0.039	0.083
decision-making-lag	0.061	0.052	0.071	0.049	0.045	0.054	0.059
available-tests	0.089	0.087	0.000	0.000	0.000	0.000	0.094
asymptomatic-chance	0.043	0.043	0.039	0.036	0.027	0.029	0.040
incubation-hours	0.039	0.032	0.045	0.034	0.024	0.028	0.041
total-doctors	0.000	0.000	0.000	0.000	0.000	0.000	0.017

10.3.2 Rankings per Case Comparison

Feature\Case Type	cases1	cases2	case1_1	case1_2	case2_1	case2_2	all_cases
immunity-after-recovery?	1	1	1	1	1	1	1
interaction-followup?	2	2	2	3	3	3	2
quarantine-just-in-case?	3	5	5	2	6	2	4
available-tests	4	4	8	8	8	8	3
recovery-hours	5	3	3	7	2	5	5
decision-making-lag	6	6	4	4	4	4	6
asymptomatic-chance	7	7	7	5	5	6	8

incubation-hours	8	8	6	6	7	7	7
total-doctors	9	9	9	9	9	9	9

10.4 Number of Features per Case vs Accuracy (max-ticks, k=3)

# Features	Avg. Accuracy						
	all cases	**cases1**	**cases2**	**case1_1**	**case1_2**	**case2_1**	**case2_2**
1	83.9	82.4	85.5	80.9	84.0	84.8	85.2
2	85.3	86.3	85.5	87.5	87.5	87.1	85.5
3	85.8	89.8	90.6	93.0	96.1	97.3	96.1
4	91.2	91.8	90.6	93.8	96.1	97.3	96.1
5	95.1	94.5	95.1	94.9	96.5	97.3	96.1

10.5 Pseudo-Code - ask-adults-go

```
procedure ask-adults-go:
for every adult, a:
{
    if !("quarantined" in a.state):
    {
        if ("infected" in a.state):
        {
            #  a is infected and not assigned to a hospital and not
under treatment with a hospital (Which is a given as if they were
infected and assigned to a hospital => the adult would be in
quarantine)

            if (a.time-sick > incubation-hours)
            {
                if !(a.asymptomatic?):
                {
                # infected, symptoms are showing as a has been sick
for longer than incubation-hours => go to hospital

                let res (go-to-hospital self)

                }
                else:
                {
                    # infected, symptoms should show but adult is
asymptomatic, so they don't notice they are infected

                    if((a.time-sick > recovery-hours) and (is-
recovered? self))
                    {
                        assign-recovered self
                    }

                    to-work-or-wander self

                }
          }
          else:
          {
            # infected but symptoms are not apparent yet
```

```
            to-work-or-wander self
          }
        }
        else:
        {
          to-work-or-wander self
        }

        interact self
    }
    else:
    {

        # adult is in quarantine

        if ("treatment" in a.state)):
        {
            # quarantined and under treatment (both under a hospital)
=> assumption: member? "infected" state = true (a is/was infected)
            ongoing-treatment self
        }
        {
            # quarantined adult (who may or may not be infected) is
not under treatment nor assigned to a hospital

            if (a.time-quarantined < incubation-hours):
            {
                # remain in quarantine since testing cannot be done
until time-quarantined > incubation-hours
                go-home self
            }
            else:
            {
                # check if adult can exit quarantine
                if(exit-quarantine? self):
                {
                    unassign-quarantine self
                    starting-positions self

                }
            }
```

```
        }
    }
      assign-color self
}
```

10.6 Worst-Case Scenario Plots

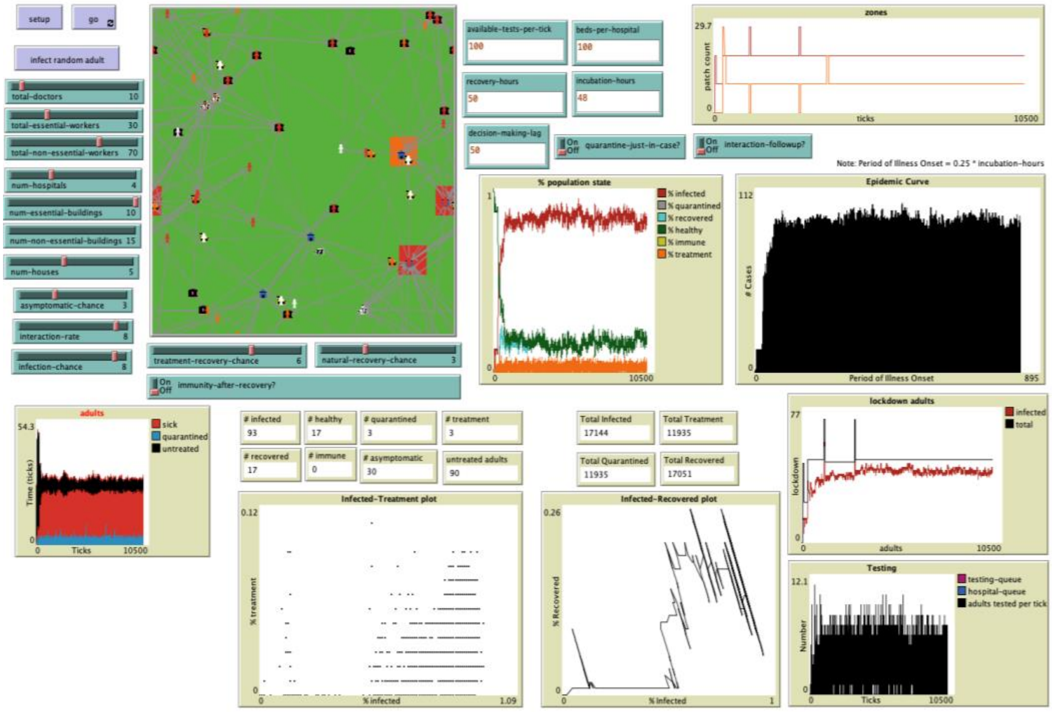

Figure 15: Worst case scenario for Case 1_1 group

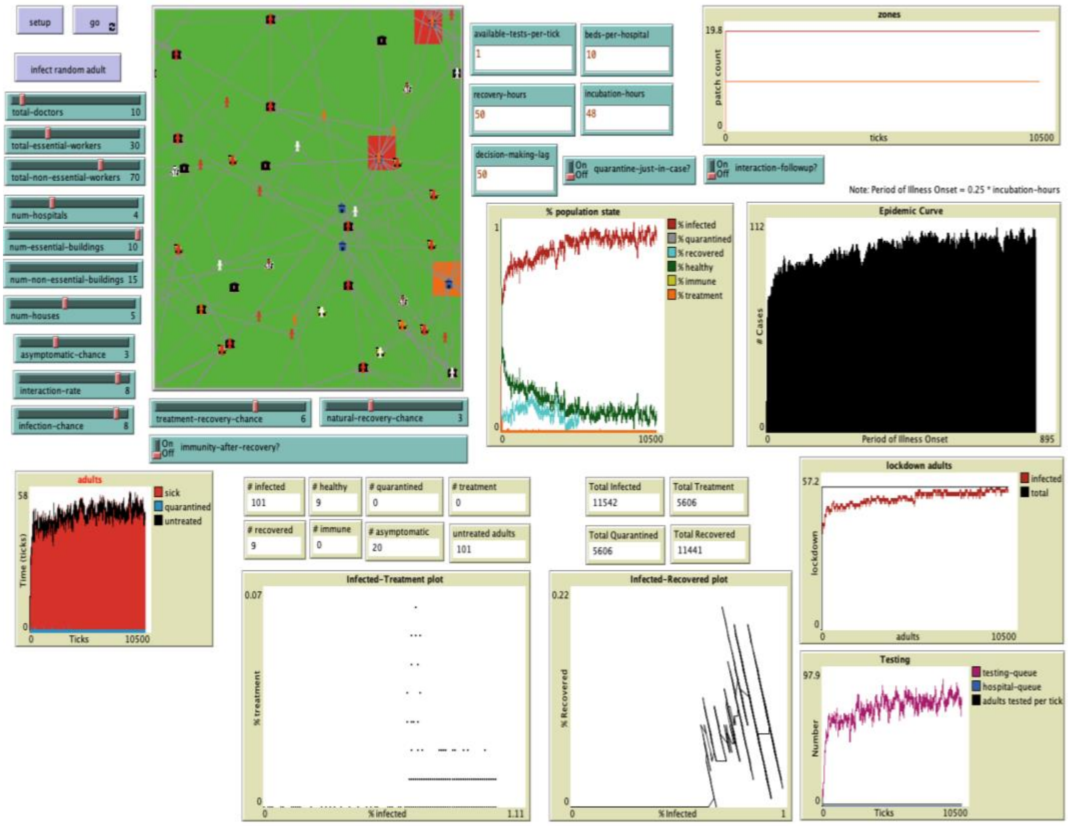

Figure 16: Worst case scenario for Case 1_2 group

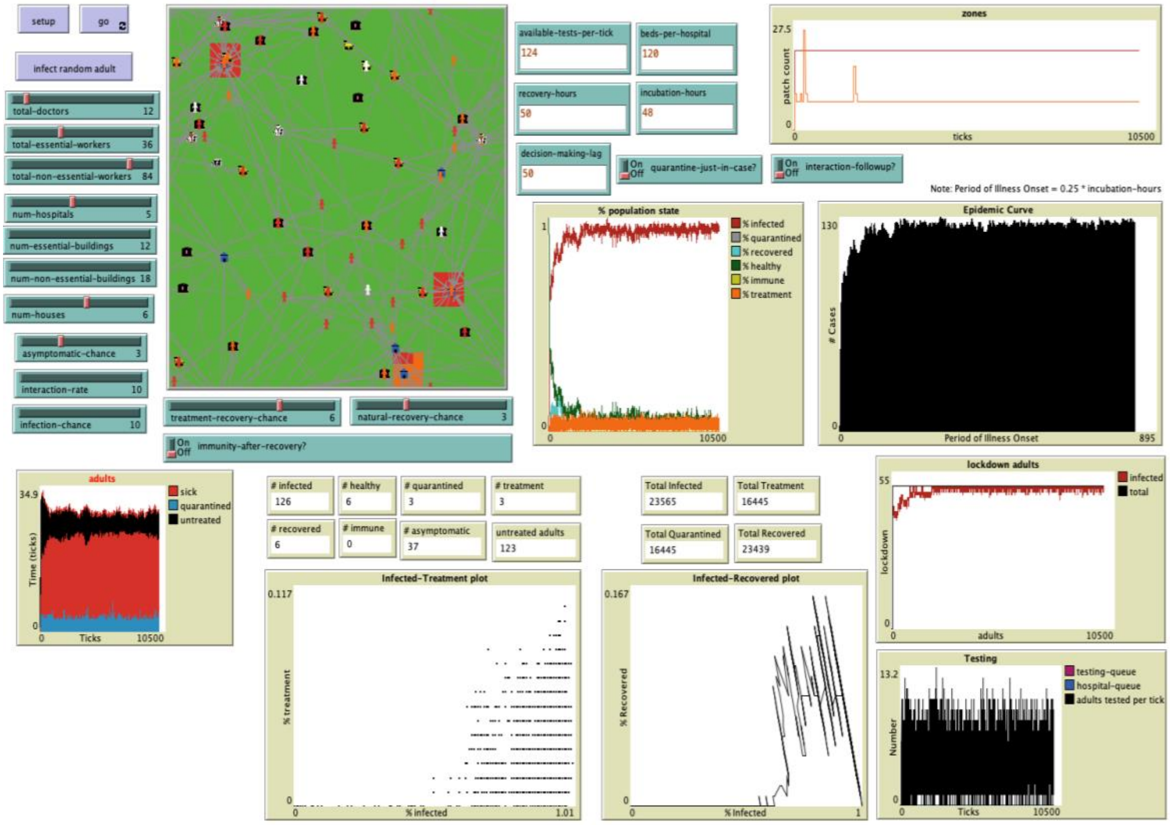

Figure 17: Worst case scenario for Case 2_1 group

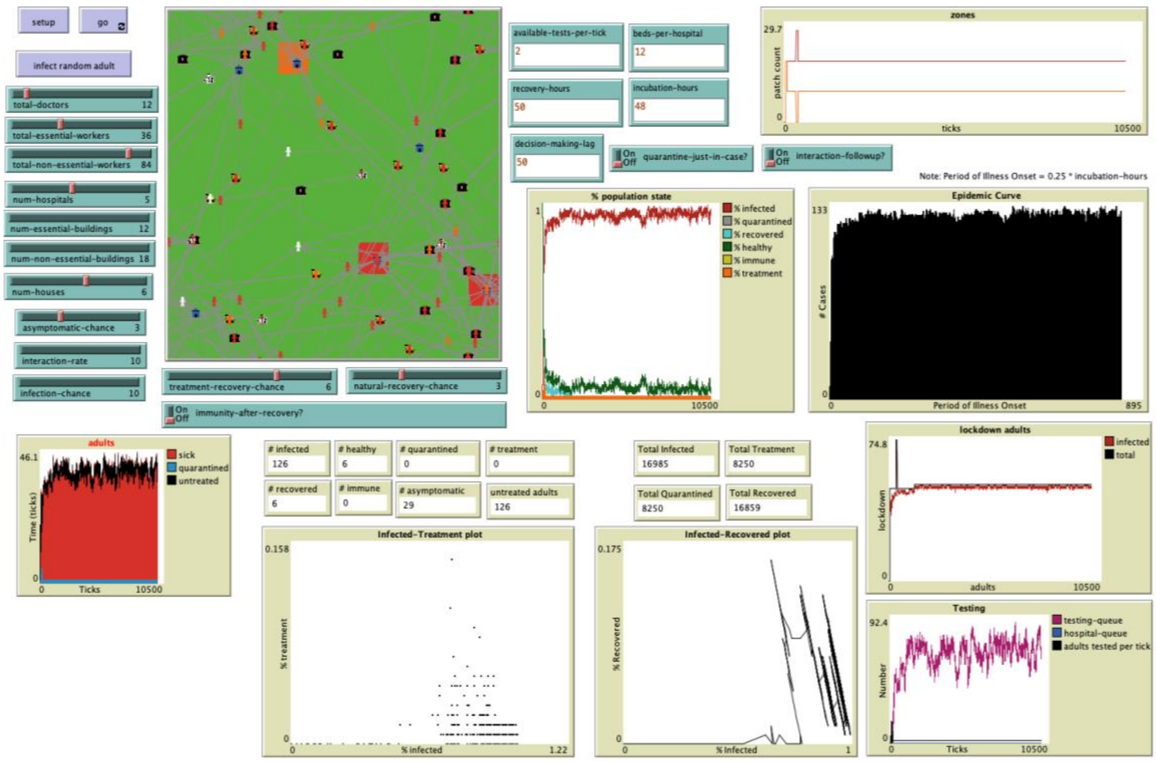

Figure 18: Worst case scenario for Case 2_2 group