

A Fully GIS-Integrated Simulation Approach for Analyzing the Spread of Epidemics in Urban Areas

Hedi Haddad¹, Bernard Moulin², Marius Thériault³

¹Department of Computer Science, Dhofar University, Salalah, Sultanate of Oman

²Department of Computer Science and Software Engineering, Laval University, Quebec, Canada

³ÉSAD, Laval University, Canada

Abstract

Human-to-human communicable diseases can be devastating in urban areas where large heterogeneous population groups are living in restricted spaces, causing serious concerns for public health, especially during epidemic outbreaks. Even though Geographic Information Systems (GIS) have been used to study a variety of public health issues in the last decade, their use to study human communicable diseases has been limited to the development of disease clustering, mapping and surveillance systems. These systems don't provide ways to understand and predict the dynamics of diseases spread across an urban region, taking into account the dynamics of human contacts and mobility, which are the main widely recognized mechanisms responsible for diseases' spread. In this paper we address such limits by presenting a GIS-based spatial-temporal simulation approach and software to support public health decision making in the context of communicable diseases in urban areas. The approach fully integrates epidemiological, mobility and GIS-data models at an aggregate population level in order to support spatialized interventions.

1 Introduction

Human-to-human communicable diseases are one of the most major threats to the human society, with their heavy human, societal and economical costs. They can be more devastating in urban areas where large heterogeneous population groups are concentrated in relatively small spaces, making urban centers favorable environments for the spread of various epidemics. This is a source of serious concerns for public health decision makers for several reasons. First, decision makers must act quickly to propose efficient and effective intervention strategies, while coping with the constraints of available countermeasure resources. Consequently, decision makers need often to prioritize who, where and when target with their interventions, which, obviously, requires a good understanding of how these communicable diseases spread taking into account the spatial and temporal dimensions. However, secondly, communicable diseases are complex phenomena, and their dynamics are influenced by a complex set of biological, spatial, temporal, environmental, socio-demographical, behavioral and human mobility factors. Decision makers do not only need to collect data from a wide variety of sources, but they also need to integrate this data in ways that provide a "credible" understanding of the situation. Third, given the inherent complexity and uncertainty of these phenomena, decision makers often perform "What-if" reasoning, by which they elaborate and analyze different candidate intervention strategies and select the best ones. Therefore, there is a need to provide public health decision makers with decision support tools that capture infectious diseases' dynamics and allow them analyzing and prioritizing potential control interventions. Because infectious disease outbreaks are explicit spatiotemporal phenomena, Geographic Information Systems (GIS) can be

an intuitive choice to offer a framework that is well suited for this task. Yet, the use of GIS to analyze infectious diseases has been limited to clustering, mapping and surveillance processes, often using statistical approaches [8]. These systems are certainly useful but they don't provide ways to understand and predict the dynamics of disease spread in an urban region, taking into account the dynamics of human interactions within a spatial environment (commonly referred to as *reaction*) and mobility (commonly referred to as *diffusion*), which are the main widely recognized mechanisms responsible for disease spread. Indeed, this type of dynamic analysis is difficult to realize by means of the GIS functionalities commonly available and often requires the use of other models and tools, in addition to GIS.

To address this challenge, we proposed in [5] a new GIS-based spatial-temporal simulation approach and software to support public health decision making during the first days of an epidemic outbreak in an urban region. In this paper we aim to provide some insights on the methodological foundations of our approach which have not been presented in [5]. The remainder of the paper is organized as follows. Section 2 presents the main theoretical roots that have inspired the proposed approach. Section 3 highlights the approach and the GIS-fully embedded tool. The main limits and potential future extensions are discussed in the conclusion.

2 Main theoretical foundations

Boosted by the pandemic outbreak of H1N1 in 2009, spatially-explicit modeling and analysis of infectious diseases' spread has been subject of substantial research works, most of which focused on the global scale, where it appeared that international air travel can provide good predictions for the worldwide spread of certain communicable diseases [10]. Recently, several works have concluded that urban commuting mobility would be more efficient than global air mobility for limiting human infectious diseases' spread [3], which is raising the need of approaches and tools for studying the phenomenon at a regional level.

The literature of spatially-explicit models [4] is dominated by two main approaches, the *individual-based approach* and the *population-based approach*, tackling the problem from the micro and macro levels, respectively. Both approaches present some limits when it comes to study the phenomenon at an urban scale.

In individual-based approaches, the infection probability depends on individuals' explicit attributes, behavior and social interactions, making them able to fully reproduce the heterogeneity of the phenomenon and, thus, being closer to reality. However, their implementation is complex in terms of time, data processing and computational resources, which commonly can't be afforded by public health agencies, especially in the early days of epidemic outbreaks. Population-based approaches study the phenomenon at an aggregate level, the population, while coarse-grained techniques are used to model individual interactions. The meta-population framework has dominated the literature of population-based approaches in recent years [10], where the world is modeled as a network of connected sub-populations. Sub-populations correspond to geographical regions (countries or urban areas) and connections correspond to movement flows between sub-populations. While this framework is suitable for studying disease spread at a global level, it is based on oversimplified assumptions that make it unsuitable for modeling both the reaction and diffusion processes at an urban scale. Considering the reaction process, common meta-population models assume that individuals inside each population behave identically and have the same probability to contact each other, which has been contradicted by several researches (see [6] for example). Considering the diffusion process, traditional meta-population models assume that individuals move randomly between sub-populations and do not have a memory on their previous locations [10]. However, daily commuting mobility is based on recurrent movements between homes and a limited number of places, which requires memorizing the home locations of individuals in the system. Even though some recent works have extended the meta-population framework to model recurrent commuting mobility and contact of socially-structured populations [1], the mathematical complexity has limited their application to the stage of theoretical explorations, and they have to be combined with numerical simulations in order to study real populations with different types of social groups.

Given the above-mentioned limits of individual-based and population-based approaches, we argued the need of new approaches to model the phenomenon at a meso scale (i.e., aggregate level in urban areas), while being more simple to implement than individual-based models and more close to reality (by its ability to capture the different heterogeneities of the phenomenon), compared to current population-based models. In this context we proposed an integrative approach and a fully GIS-based tool [5]. The approach is inspired by several theories and techniques. Constrained by the paper length, in the following we summarize the three main techniques and theories, respectively, *Public Health Intervention practices*, *Geo-simulation of urban dynamics* and *Time geography*.

2.1 Public Health Intervention practices

The main objective of our approach is to support decision makers in their efforts to propose intervention strategies for controlling disease spread in urban areas. Therefore, it needs to comply with the practices of public health interventions. According to [7], public health practice does not intervene on individuals but targets specific groups called *aggregates* (subgroups of the whole population). The division of populations into aggregates can be done in several ways, depending on the context and purposes of the targeted interventions. Even though grouping population members based on demographic attributes or geographic locations is a common practice, *high-risk aggregates* are the most widely used, corresponding to subgroups of the population with high-risk members in respect to the considered disease [7]. Once target groups are identified, public health policy makers decide the relevant control strategies, including pharmaceutical and /or non-pharmaceutical interventions.

2.2 Geo-simulation of urban dynamics

We believe that the spread of communicable diseases in urban environments should be perceived as an explicit spatial and temporal urban phenomenon. Modeling of urban phenomena is a multidisciplinary topic involving various disciplines such as geography, computer science, geomatics, mathematics and social sciences. In the last decade, *geosimulation* has dominated the literature and became the most widely used approach for modeling urban phenomena. Geosimulation uses a combination of Cellular Automata, GIS and/or Agent-Based simulation techniques to model and reproduce the dynamics of different urban processes, with different interacting actors, at different spatial, temporal and social scales [2]. From this perspective, a geosimulation-based approach is an intuitive choice for our problem.

2.3 Time geography

Time geography was initially introduced by Hägerstrand as a framework for explicitly modeling spatio-temporal human activities. In this framework, *places* where activities are carried out are explicitly modeled and represent the spatial bases for activities and human interactions. The sequences of activities constitute the *the space-time paths*, which are individuals' trajectories standing for physical mobility in time and space. Spatial intersections of several space-time paths are called *bundles*, and they identify places where groups of individuals meet and interact, like households and working places.

By adopting the time geography framework, we tried to address two main limits of the meta-population approaches. Firstly, the time geography framework offers an explicit representation of activities, along with their places and times. In the meta-population framework, activities are not explicitly represented, and consequently it is not possible to model the infections that might happen during contacts between people such as in public transportation vehicles. Secondly, most of population-based models use a numerical perspective of the mobility, which is reduced to movement rates between pairs of source and destination localities. These rates are generally expressed as a force of infection and are often compiled from census data records or transport databases, with no details about people's mobility purposes. Several works have also included explicit mobility models, mainly

using gravity and, less frequently, radiation models [11], both approaches being borrowed from a physics perception of mobility [9]. However, the activity-based mobility approach, which is based on time geography, has gained an increased interest in the transportation research community. In contrast to gravity and radiation models, activity-based models assume that all trips are motivated by individual needs, and Activity-based transportation surveys are used to identify these needs for different population groups and how they are typically fulfilled, which allows for building activity-based mobility models, given the demographics of individuals and environmental factors. To the best of our knowledge, aggregated activity-based mobility models have not been used yet with aggregated infectious models; certainly not in the currently existing mathematical meta-population models.

3 The proposed approach

In this section we give an overview of the approach proposed in [5], by emphasizing how its main steps have been inspired by the theories and techniques presented in the previous section. The main idea of the approach is to take advantage of available activity-based transportation surveys to rapidly model the demographic characteristics and activity/mobility patterns of a significant sample of an urban population at an aggregate level, so that the proposed approach fully integrates human epidemiology, human mobility and public intervention models in a GIS system. The proposed simulation tool allows a user to visualize and to assess the spread of a contagious disease in a geographic area simulated in a GIS, taking into account the spatial locations of the residence and usual activities of different population groups. The simulation is initialized with a simulated population sampled from the transportation survey in which the user can easily introduce infected or susceptible people. The tool also allows for the creation and comparison of different "spatialized" pharmaceutical and non-pharmaceutical intervention scenarios in terms of the spatial evolution and distribution of infected people in the studied area. In the following we present the main elements of the approach.

3.1 Socio-demographic groups and their Activity/ Mobility patterns

Inspired by the Time Geography framework, our approach requires the use of activity-based transportation surveys' data which are available in most developed countries to build credible models of aggregated social groups of an urban area as well as models of their activities and daily mobility patterns. Such surveys are periodically conducted by governmental agencies on large representative samples of individuals, while following rigorous statistical procedures. The collected data typically cover three main dimensions, 1) the socio-demographic profiles of the respondents (age, gender, employment, home location, etc.), 2) general information about their households, and 3) their detailed activities and mobility data collected during the day of the survey. Once these individual data are collected, they need to be processed in order to build aggregated models of the socio-demographic groups living in the considered urban area and their activity patterns, taking into consideration the studied communicable disease. In agreement with Public Health practices, the concepts of groups and activities at risk are helpful in this regard. In the context of human-to-human communicable diseases, children and elderly people are the most vulnerable, and we need to distinguish at least the three demographic groups of children, adult and elderly. Risky activities are those that favor different types of contacts between people, like staying at home, studying, working, shopping and taking public transportation. By crossing data about risky demographic groups with risky activities, further socio-demographic groups can be identified along with their activity patterns, like elementary / secondary school students, university students, workers, and retired people.

Without loss of generality, we applied our approach using the 2006 Quebec's Origin-Destination Transportation Survey, and we identified 9 socio-demographic groups with their activity-mobility patterns, corresponding to infants (less than 5 years old who stay at home or attend a nursery), young students (people between 6 and 15 years old who attend elementary, primary or secondary schools), adult students (between 16 and 64 years old

who carry out (pre) university studies), adult workers, adults who are neither workers nor students, and retired people (more than 65 years old).

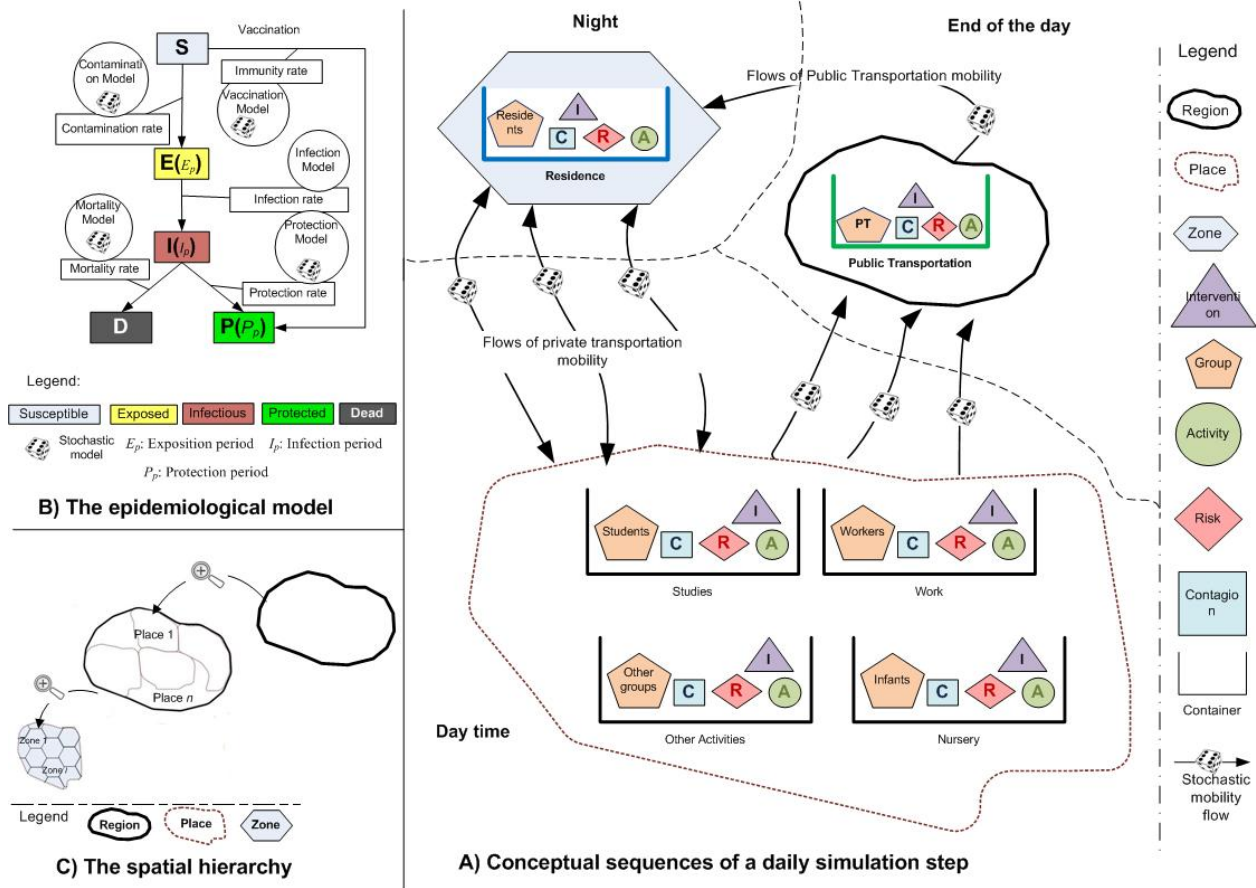


Figure 1: A) The conceptual logic of simulating diseases spread using different types of containers and groups' mobility patterns, B) The stages of the used epidemiological model, C) The used spatial hierarchy

3.2 The aggregated spatial model

Once the groups of the urban area are identified, with their activity and mobility models, they must be spatialized. For this purpose, we use an aggregated spatial model that is based on the concept of *container*. We define a container as an abstract concept that represents the location in which social groups of people can carry out typical activities that can be characterized using a transportation survey. Every container has a *spatial extension*, which corresponds to its physical location in the studied geographic urban area. We use a hierarchical decomposition of the geographic urban area, as illustrated in Figure 1-C. At the most detailed level we use a spatial tessellation of space as a set of *Zones* which usually correspond to local administrative boundaries of interest to public health decision makers (such as census tracts). *Zones* can be aggregated into *Places*, which can be aggregated into a *Region* representing the whole urban area of interest. We therefore model containers at different spatial scales, which offers more flexibility to deal with any missing transportation data or any design choice. The types of containers are chosen with respect to the activity data available in (or computable from) transportation surveys. We draw a distinction between *anchor activities* (staying at home, working, studying, and taking public transportation) and *other activities* (like shopping, socializing, etc.). Every anchor activity is associated with a type container such as *Work container* for working, *Primary school container* for primary

studies and *Residence container* for activities carried out at home. Since transportation surveys' data generally provides detailed information about the residence location (usually the census tract) but not about other activities, we associate residence containers with zones and anchor activity containers with places. However, if data is available, all the anchor activity containers can be associated with zones. The *Container of other activities* aggregates the locations of all the other activities that are different from the anchor activities. For simplification purposes and supposing that these activities affect the disease transmission less considerably, we associate the container of other activities with the regional level of the spatial hierarchy, corresponding to the whole urban area of study. Again, if possible and needed, this container could be associated with places or zones. Finally, we aggregate in a *Public transit container* all the activities that are related to individuals' public transit, since people spend some time with each other in public vehicles where disease transmission is facilitated. For simplification purposes we only use one Public transit container at the regional level, since we can easily compute aggregated contact attributes from transportation surveys. But if more detailed data is available, we could distinguish several public transit containers such as Bus container, Metro container and Regional Train container.

3.3 The spatialized contagion model

Containers are used to model the places where activities take place and consequently where different socio-demographic groups interact with each other, offering occasions for disease transmission from infectious to susceptible individuals. Every socio-demographic group is subdivided into five compartments that correspond to typical disease infection stages (Susceptible, Exposed, Infected, Protected and Dead) (Figure 1-B). Every type of container has a local contagion model, which specifies how to compute the probability of infection after a contact between susceptible and infectious individuals, taking into account individual and environmental risk factors, including activity and spatial environment risk factors. Every type of container has different parameters for the local contagion model.

3.4 The spatialized intervention model

Considering the intervention scenarios, we distinguish *Population-Oriented* and *Activity/Mobility-Oriented* scenarios. Population-Oriented interventions target population groups and correspond to *Vaccination* and *Quarantine* scenarios. Both scenario types are considered to be regional and characterized by occurrence days and targeted population groups. In contrast, Activity/Mobility-Oriented scenarios target the population activities, and they are currently supported by the *Public Settings Closure*, where the decision maker can specify the closure of certain activity locations (such as schools or even public transit) on certain days. However, all intervention scenarios could be used to target groups at different scales such as geographic zones (residence containers) and places (anchor activities' containers).

3.5 Geo-simulation of the spread of the communicable disease

We use a geo-simulation approach to simulate the dynamics of disease spread over an urban region (Figure 1-A). At the beginning of the simulation, socio-demographic groups are assigned to their residence containers. At the beginning of every simulation step (corresponding to one day), the simulation engine checks if an intervention scenario must be applied at the current step (vaccination, etc.). When appropriate, such a scenario is loaded and its constraints are applied. Then, the simulation engine moves the different groups from their residential containers to their activity then to public transit containers where they spend some time and interact with other moving groups en route towards their residential containers. The simulation engine applies the local contagion model in order to simulate the disease transmission inside each activity container and inside the public transit container. At the end of the day, the simulation engine moves all the active groups back to their residence containers where they spend the night. It also applies a local contagion model in order to simulate the disease transmission inside

residential containers and then carries out all the computations required for the initialization of the next simulation day (computation of the numbers of susceptible, exposed, infectious, protected and dead people of the day at each residence container). It is worth mentioning that daily activity and mobility flows are constrained by several activity parameters such as daily presence and quarantine rates. In order to model the constraints of interactions between activity patterns of household members, we also define interaction relationships between some activities. For example, the mobility of children who attend nursery is affected by the mobility of their parents, because infected parents who are in quarantine will not bring their children to the nursery.

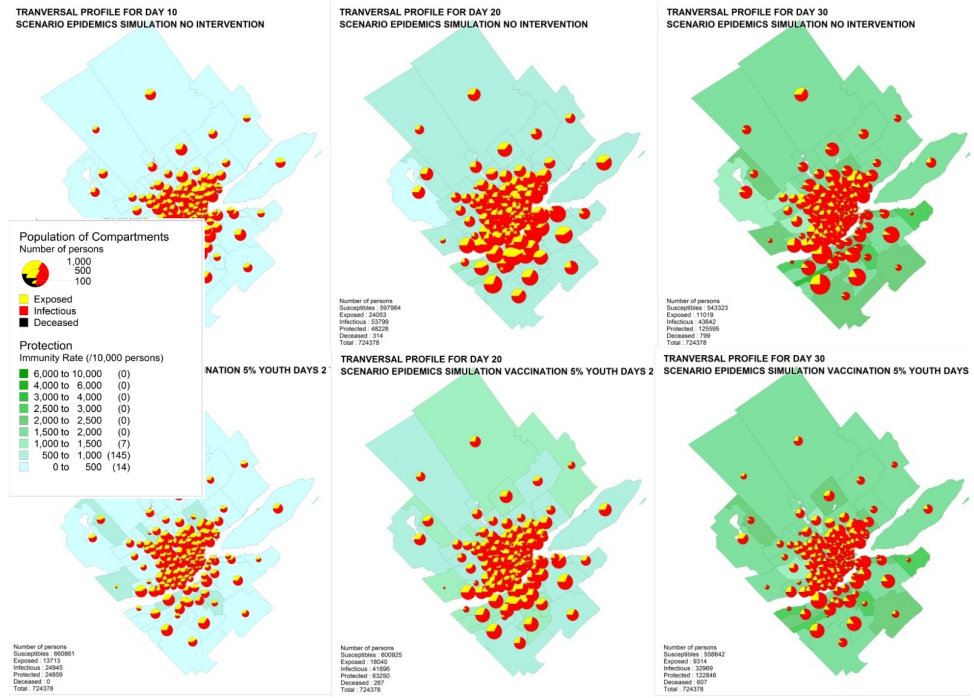


Figure 2: A snapshot of the graphical animations generated by the tool comparing spatial infections' distributions without (top) and with (bottom) interventions

3.6 Proof of concept

As a proof of concept of our approach, we implemented the *P2PCoDiGeosim* tool (Person to Person Communicable Disease Geosimulation) within the MapInfo GIS using the MapBasic programming language. It is fully integrated within the GIS system and coupled with a Graphic Display Module that allows for the graphic visualization of the simulation results. In our software, the population activity and mobility data are compiled from the 2006 OD survey of Quebec metropolitan area. The compiled and enriched data corresponds to a total adjusted and weighted population of 724378 individuals, constituting a credible representation of the whole population of Quebec area at that time. Residential zones are associated with census tracts, leading to the subdivision of the simulated area into about 180 census tracts (residential containers). The tool innovates in offering a complete integration of mobility and infection models in a GIS-based software for decision support in the public health domain. The full integration of the simulator in a GIS allows a public health decision maker to simply set various intervention scenarios and to view simulation results in different display formats authorized by the GIS (graphs, charts, etc.), which facilitates the comparison of the outcomes of different scenarios. The system also offers the option of visualizing the spatial spread of the disease in an animated way (see an example illustrated in Figure 2).

4 Conclusion

Our work attempted to propose a new approach and associated GIS-based simulation tool suitable for modeling the spread of communicable diseases in urban areas, while being simpler to implement than individual-based models and closer to reality than population-based models. Indeed, the approach operates at an aggregate level which is plausible since it is based on a significant sample of the real population whose characteristics are computed from transportation surveys' data, in contrast to common meta-population models. The approach is simple and can be promptly put into use, since it does not require lengthy processing of population and mobility data. Transportation surveys are widely available, at least in North America and Europe and the approach and tool can be adapted to local variations of these data sets.

Considering future work, applying the approach on various disease spread cases is certainly one of our main priorities. Moreover, we restricted some intervention scenarios at the regional level for simplification purposes. Such a restriction needs to be lifted to provide decision makers with more flexibility by allowing the specification of intervention scenarios at other spatial scales.

References

- [1] A. Apolloni, C. Poletto, J. Ramasco, P. Jensen, and V. Colizza. Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theoretical Biology and Medical Modelling*, 11(3), 2014.
- [2] I. Benenson and P. Torrens. *Geosimulation: Automata-Based Modeling of Urban Phenomena*,. Wiley, 2006.
- [3] S. Charaudeau, K. Pakdaman, and P. Boelle. Commuter mobility and the spread of infectious diseases: Application to influenza in france. *PLoS ONE*, 5(1), 2014.
- [4] D. Chen, B. Moulin, and J. E. Wu. *Analysing and Modeling Spatial and Temporal Dynamics of Infectious Disease*. Wiley, 2015.
- [5] H. Haddad, B. Moulin, M. Thriault, and D. Navarro-Velazquez. Integrated epidemiologic simulation for person to person contagion through urban mobility within gis. In *HealthGIS 2012, The First ACM SIGSPATIAL International Workshop on the Use of GIS in Public Health*, pages 63–71. ACM, November 2012.
- [6] I.-Z. Kiss, D.-M. Green, and R.-R. Kao. The effect of network mixing patterns on epidemic dynamics and the efficacy of disease contact tracing. *Journal of the Royal Society, Interface*, 5(24):791–799, 2008.
- [7] D.-J. Porche. Population-based public health practice. In *Public and community health nursing practice: a population-based approach*, pages 2–16. SAGE Publications, Inc, 2003.
- [8] C. Smith, S. Le Comber, H. Fry, M. Bull, S. Leach, and A. Hayward. Spatial methods for infectious disease outbreak investigations: systematic literature review. *EuroSurveill*, 20(39), 2015.
- [9] J. Toole, Y.-A. de Montjoye, M. Gonzalez, and A. Pentland. Modeling and understanding intrinsic characteristics of human mobility. In B. Goncalves and N. Perra, editors, *Social Phenomena, Computational Social Sciences*, pages 15–35. Springer International Publishing, Switzerland, 2015.
- [10] L. Wang and X. Li. Spatial epidemiology of networked metapopulation: An overview. *Chinese Science Bulletin*, 59(28):3511–3522, 2014.
- [11] A. Wesolowski, W. Prudhomme, N. Eagle, A.-J. Tatem, and C.-O. Buckee. Evaluating spatial interaction models for regional mobility in sub-saharan africa. *PLoS Computational Biology*, 11(7), 2015.