

Improving Agent Based Models and Validation through Data Fusion

Marek Laskowski, Bryan C.P. Demianyk, Marcia R. Friesen, Robert D. McLeod and Shamir N. Mukhi¹

*Internet Innovation Centre, University of Manitoba
¹The Canadian Network for Public Health Intelligence (CNPHI)
1015 Arlington Street, Winnipeg, MB*

Abstract

This work is contextualized in research in modeling and simulation of infection spread within a community or population, with the objective to provide a public health and policy tool in assessing the dynamics of infection spread and the qualitative impacts of public health interventions. This work uses the integration of real data sources into an Agent Based Model (ABM) to simulate respiratory infection spread within a small municipality. Novelty is derived in that the data sources are not necessarily obvious within ABM infection spread models. The ABM is a spatial-temporal model inclusive of behavioral and interaction patterns between individual agents on a real topography. The agent behaviours (movements and interactions) are fed by census / demographic data, integrated with real data from a telecommunication service provider (cellular records) and person-person contact data obtained via a custom 3G Smartphone application that logs Bluetooth connectivity between devices. Each source provides data of varying type and granularity, thereby enhancing the robustness of the model. The work demonstrates opportunities in data mining and fusion that can be used by policy and decision makers. The data become real-world inputs into individual SIR disease spread models and variants, thereby building credible and non-intrusive models to qualitatively simulate and assess public health interventions at the population level.

Keywords: *Agent Based Modeling; Personal Contact Patterns.*

Introduction

Complex networks underlie the transmission dynamics of many epidemiological models of disease spread, in particular agent based models (ABM). Network-based epidemiological models use a percolation-like principle to simulate disease spread through the population [1]. Agent based models are being increasingly employed due to their potential to capture complex emergent behaviours during the course of an simulated epidemic, where these behaviours arise from the non-linearities of human-human contacts. ABMs may employ an explicit or implicit social contact network defined by structured agent interactions. In the explicit case, a disease model (e.g., susceptible - exposed - infected - recovered, SEIR type) can be implemented directly on the network, although in the case of ABM, these resemble simulation models rather than the steady state analysis of network based models mentioned in [1].

In all cases, though, the fidelity of an agent-based model relies in part on the credibility of the social contact network data that feeds it. Potential data sources include census and

demographic data (coarse) and finer-grained data made available by various means of polling personal electronics such as cell phones. In related work, it was demonstrated that data to model a social contact network can be collected through web services or wireless sensory devices or “motes” worn by individuals in the target population and subsequently used in an infectious disease spread model [2]. Such an approach has been previously undertaken to gather data, for example in an organization (workplace or school). The resulting estimated social contact network was used to model an influenza-like illness (ILI) within the setting [3], based on a standard SEIR type model. In this time-stepped model, infection spreads between two vertices (individuals) along the weighted edges of the network which represent the amount of social contact between the two individuals. Estimating social contact networks in larger populations, (metropolitan scale or larger), is an area of research still in its relative infancy.

In cases where precise contact network data is unavailable, an alternative is to mine data as done by EpiSimS [4] which uses United States Department of Transportation information to estimate the schedules of the agents in several metropolitan areas. This presumes that the choices of places for agents to interact is constrained by the transportation network (model), which itself is a complex network. Schedules for the agents are synthesized from census and USDOT data. A simulation is then run during which a synthetic contact network is constructed from the interactions of the agents and their locations. The resulting dynamic bipartite graph [4] is used to simulate disease spread in the manner stated earlier, except on a much larger scale. Both EpiSimS and another well-validated infectious disease simulator, BioWar [5], initially perform validation on model components separately. This is an important component of plausibly reasoned argument, supporting the statement that the model as a whole functions as specified.

The objective of the present work is to investigate methods to begin validating ABMs in varying stages of development by comparing extracted contact networks to known theoretical social contact network models. Ideally, networks which embed some notion of space or time will be essential drivers of disease spread in the real world. Thus, extracted networks may need to be weighted, for example, to associate weight with the time period during which two agents were in contact. The first such model is of a rural community in the province of Manitoba, Canada. The emphasis in this work is in integrating data from emerging sources that can be used within discrete time and space disease spread ABMs. The contagions of interest are influenza like illnesses (ILI) or other respiratory infections that are primarily contracted through direct or proximal contact.

Methods

In the first part of the study, we discuss a small scale ABM of two adjacent communities in the Rural Municipality of Stanley, Manitoba with a combined population of approximately 16,500 residents: Winkler, Manitoba at 10,000 residents and Morden, Manitoba at 6500 residents. This is a spatial temporal model with demographic data coming from Statistics Canada [6]. From this perspective, agents are provided with schedules, and a model of disease spread is run. Figure 1 illustrates the topography of the region of interest.



Figure 1: The topography of the rural region of interest

The towns of Morden and Winkler are roughly seven miles apart in southwest Manitoba. One of the reasons for selecting this area is that it is representative of many North American rural municipalities. Figure 1 also illustrates the location of three cellular service towers with MTS Allstream as the service provider. The ABM is discussed in terms of model validation using data that is mined from anonymized cell phone use records. In addition to cell phone usage, the model is also improved using a Smartphone application that provides greater fidelity of proximity contacts using Bluetooth enabled devices as proxies for people.

There are two primary obstacles to fusing data to a model. The first is the collection of the data, with assurances that the data collected is meaningful and accurate, and mining or interpreting the data for parameters or characteristics useful to the model. The second difficulty is integrating the data into the model itself, running simulations and attempting to qualify (and ideally, quantify) the outputs. In many instances, the results of the simulations may be self fulfilling, as in, overcrowding in isolated and impoverished communities leads to increased infection spread. The interventions that one could model may provide guidance for policies that may then be considered. For example, an intervention associated with reducing infection spread may be a recommendation to stay home while ill; in overcrowded residential communities a more effective intervention may be quarantine or a modified quarantine policy whereby an infected person may be advised to seek temporary housing in a facility set up specifically for that purpose. While somewhat self-evident, modeling with real data may help to elucidate these types of options or interventions.

The Model and ABM Simulator

The model described here is a milestone in the process of designing and implementing an ABM simulation framework geared towards high fidelity modeling of human institutions of varying scales. The broad design goals of this framework, called Simstitution, are based on the collective experience of the authors gained while developing Agent Based Models of human institutions. Originally, models of hospital emergency departments [7] and cities [8] were implemented upon “one-shot” simulators, that is, a simulator strongly coupled to the specific modeling application [9]. A one-shot simulator is comparatively easy to implement, and gives the modeler fine control over the simulator processes, enabling them to fulfill their requirements. Typically, in order to minimize development effort, the designer will make assumptions which ease the implementation of the model at hand, without consideration for how these assumptions will constrain or complicate re-purposing the simulator to implement a different model. From a software engineering perspective, part of the reason that one-shot models are so easy to produce is that little or no effort goes into making the software reusable or

extendible. The large number of one-shot simulators observed in the literature [9] is problematic because by their nature they are difficult to re-use. The reusability of the simulator in turn affects the reliability of the simulator; the more researchers that (re)use a particular simulator the more chances that bugs will be identified and fixed. Furthermore, when a number of models produce reasonable results using a common simulator, confidence in the credibility of the simulator is increased. Publishing results from a series of models built upon a common simulator framework, combined with verification of model components (or sub-models), is a common path for building confidence in simulator frameworks for epidemiological modeling [4][5].

Simstitution Design Goals

Although there are several frameworks [10]-[15] which can be used to develop agent based models, these are dwarfed by the number of one-shot or otherwise domain-specific simulators, suggesting that no framework has yet hit upon a “sweet-spot” between flexibility, extendibility, and specific support classes for human-centric domains [9]. Human-centrism includes the notion that Agents are spatially oriented and situated since humans are physical entities that occupy and traverse space, rather than existing in some abstract information domain. Simulator support for a range of human time steps on the order of seconds to hours or days is also desirable.

Other design features include adherence to software engineering principles to improve re-use and maintainability of the framework, as well as extendibility especially where machine learning can be leveraged for automated generation of agent policy [16][17].

For rapid model construction, a next generation ABM framework should facilitate the incorporation of real-time data such as from database leading to increasingly data-driven simulation. A tool for visualization and interacting with the model in a graphical manner (GUI) also facilitates model development, validation, and debugging. Visualization is also key for communicating results with subject matter experts and stakeholders [18]. Such a visualization tool can also be extended to serve as a tool for model construction or editing model parameters imported from real data.

The accessibility of agent behavior development to persons with a non-programming background can be improved by first providing a scripting layer on top of the compiled code, and then perhaps adding a visual or block (e.g. OpenBlocks [19]) programming (drag and drop) on top of that. Over time a library of useful scripted behaviors can be built up.

The increasing availability of parallel or distributed computing systems also suggests that contemporary or future agent based simulator frameworks have support for distributed, parallel, or cluster computing. The increasing availability of cluster-based compute resources (a consequence of Moore’s Law), sensitivity to real-time computational constraints, and medical data privacy issues augur well for cluster-based computing. As a result, the Simstitution design emphasizes scalability with respect to multiple processors and discrete memory spaces over efficiency in executing one particular type of model.

Currently, there is an emergence of general-purpose computing on graphics processing units (GPGPU) as excellent accelerators for data parallel applications with regular data access patterns. This leads to opportunities for accelerating agent based simulation as well. However, optimization is still challenging, as the data access patterns are still somewhat irregular for

most ABMs. Currently, GPUs are very well suited to ABMs that resemble cellular automata, percolation, game-of-life, or particle swarm models. Without doubt, higher level ABM (social autonomous interacting agents) simulations will also benefit from the compute resources of GPUs as the technology evolves (optimizing compilers, etc.).

Naturally limiting the degree of accessibility of the environment limits what Agents can perceive and interact with in the environment (including other Agents). Localizing agent perception, not only fits in well with the Agent paradigm, it also limits to what extent information needs to be shared between processes in a distributed model, which should facilitate using spatial decomposition as a guide for distributing computational load.

These disparate goals require balance in feature choice and design.

Simstitution design details

Simulated entities within Simstitution fall into either of two major categories; Agents (SimAgent), which are the autonomous entities that make decisions and interact with the environment; and instances of the SimRegion class, which represent spatially partitioned subdivisions of the environment. Note from Figure 2 that the SimObject is abstract, and exists because SimAgent and SimRegion have much of their interfaces in common.

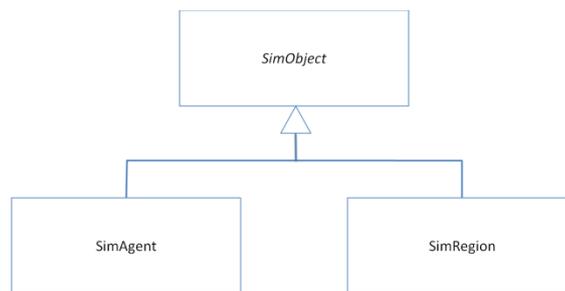


Figure 2: Class diagram for core Simstitution class hierarchy

One of the core design tenets of Simstitution is that the spatial division is closely intertwined with the division of computational work across processors and discrete memory boundaries. Therefore, SimRegion is unit of spatial decomposition as well as a convenient unit of computation. In the latter role, it can be considered as a container for agents that need to have their next state computed. Figure 3 illustrates the details of this relationship. A particular instance of SimRegion can be the parent container of SimAgents or SimRegions but not both types at the same time. This restriction will in practice result in tree hierarchies of SimRegions, with SimAgents contained in the leaf SimRegions, and the “top region” at the root of the tree. The SimRegion spatial decomposition granularity becomes increasingly fine away from the root and towards the “leaf regions” of the tree.

Time advances in the simulation when the simulator advances the time of the top region (root of the tree) by some discrete time step. The top region will then advance the time of its children by the same time step in a recursive fashion such that the tree is traversed in a depth first manner until all the SimAgents in the leaf regions have been simulated for that time step. The simulator will restart this process again, until a certain number of time steps have elapsed.

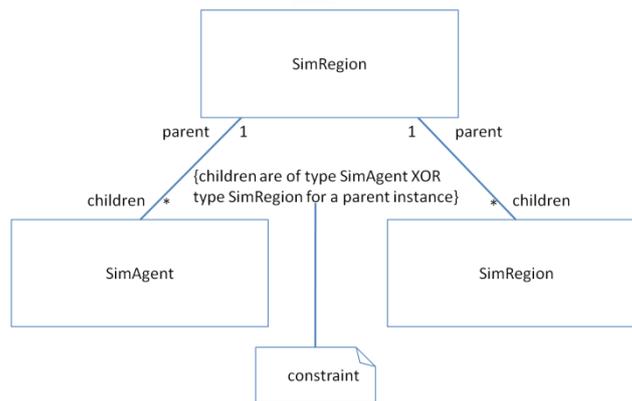


Figure 3: Relationships between core class instances, forming a tree

IndividualPolicy is a modular unit that affects the behavior of the subscribed SimAgent, which may also require the IndividualPolicy to store encapsulated SimAgent state data specific to that IndividualPolicy. Examples are a schedule policy which causes the SimAgent to observe a particular day/night work/home schedule, or in the case of a hospital being modeled, a doctor policy which causes the SimAgent to treat patients within a hospital. Within a SimRegion, each possible concrete derived IndividualPolicy class has a corresponding GroupPolicy for that SimRegion. The GroupPolicy acts as a factory for the corresponding IndividualPolicy and, if required, facilitates coordination between one or more derived IndividualPolicy classes (ex. healthcare worker policy in a hospital that coordinates interaction between nurse and doctor IndividualPolicies). Implicit here is the assumption that the properties of the local environment constrain the behavior of agents (ex. airport security lineup, swimming pool, hospital, bank, etc.). The associations between SimRegion, SimAgent, GroupPolicy, and IndividualPolicy are shown in Figure 4.

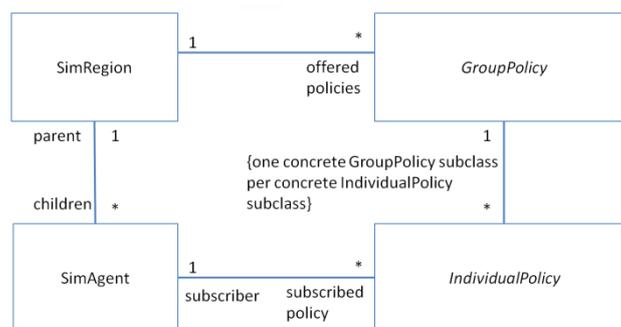


Figure 4: Relationships involving modular agent policies

Communication or interaction between SimAgents exclusively uses messages passed between SimAgents. Messages received by a SimAgent are relayed to its IndividualPolicies which can lead to an internal change of state, or an action to be taken which could lead to additional messages being sent to other IndividualPolicies on the same subscribed SimAgent, or messages sent to other SimAgents. Message passing fits well with the agent paradigm, since the alternative implies a direct mapping between external events and internal agent state which violates the principle of agent autonomy [20].

Details of Small Town Model - Morden

The current work incorporates the framework features mentioned in the previous section, and includes visualization capabilities to observe emergent model behavior during execution. The model is fairly basic so the SimRegion tree only consists of two layers; the root or top SimRegion (Morden) and the leaf SimRegions which represent the home, school, and work locations that agents occupy. The leaf SimRegions are arranged in a grid with empty spaces between structures to allow for SimAgent travel. Agents are assigned work, school, and home locations based on demographic data [6].

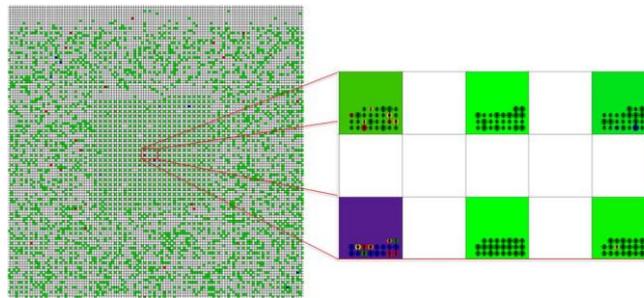


Figure 5: Screenshot of running simulation. Morden (left), close-up of 6 classrooms (right).

Figure 5 shows a screenshot of the Morden simulation at a particular time step. On the left side the entire city is shown. On the right is a detailed view of six classrooms in the center of town in which individual SimAgent details can be seen. Details include the gender and age of the SimAgent, as well as disease status. Disease status is the most interesting, and is indicated by the color of the SimAgent icon. The icon changes color, with green indicating a susceptible state. Once the agent is infected it turns yellow, orange, and red depending on how long they have spent in the infected state. Finally, recovered SimAgents turn blue. The leaf SimRegions are depicted as colored squares where the color of the square shows the aggregated disease state of the SimAgents within that region. SimRegions with no SimAgents contained inside are white. Those with one or more SimAgents display a blended color tile based on the aggregated disease state of the SimAgents inside.

Four concrete IndividualPolicy subclasses were used to generate the SimAgent behavior in the Morden model. The SchedulePolicy determines whether a particular agent wants to be at its assigned work, school, or home, depending on the demographic profile of the particular SimAgent, and the current time which advances in increments of one hour. The SchedulePolicy sends messages containing the desired destination to the SimAgent’s MovementPolicy which handles the actual movement. The InfluenzaPolicy maintains the particular SimAgent’s disease state, and if in the Infected state, sends “infection” messages to other SimAgents in the same SimRegion, which is how disease spreads between SimAgents. Finally, the BluetoothTrackingPolicy emulates the Bluetooth Smartphone contact app, and is the source of the synthetic contact data. Currently the corresponding GroupPolicies were used to facilitate aggregation of data in a spatially explicit manner to achieve the tiling effect in Figure 5.

Framework Roadmap – Next Steps

The next step or milestone will be to extend the framework by developing modules to simulate finer granularity time and space, namely facilities for Agents finding paths and steer through complex environments at time steps on the order of seconds. One such prototypical institutional environment would be improved hospital models [7].

Following that, we intend to scale up the number of agents, leveraging parallelism where possible to determine whether spatial partitioning will facilitate execution speedup and if so, under what conditions.

In a concurrent development process (possible due to the modularity of the design) tools are also being created to facilitate the integration of increasingly detailed data such as street maps and demographics of places such as Morden.

In order to promote the ideals of software re-use, once the core Simstitution simulator has reached a reasonable level of functional maturity, the code will be made available to other researchers under a general public license.

Results

Augmenting Data Sources

In addition to demographic data, the two sources of augmenting data here are associated with coarse grained data from anonymized cellular records and a finer grained Smartphone application programmed to log close-proximity Bluetooth devices. Data from cellular records typically provide service providers with input for network planning, investments, and management of evolving needs. This type of data also has considerable application to public health interests, although at this time it is difficult to derive its direct benefit in contrast to more explicit inputs such as those associated with census and demographic data, due to both technology and policy issues.

Cellular Data

Data from four consecutive weekdays in November 2010 was extracted from the data provided by the cellular service provider. The data includes the cell tower GPS and antenna sector (if applicable) that the mobile device is associated with, the AAA record (every time the phone accesses the network excluding voice and SMS), and time stamp of the access. Even at four days, this represented just over 14 GB of data. Once processed for the connections with the towers of interest (Figure 1), this amounted to just under 500,000 records. Although statistical in nature, the data can be further processed to estimate flux of persons between the two neighboring towns. Within an infection spread model, this type of information helps in estimating patterns of movement that contribute to infection spread. Once stored in a database, queries allowed for extracting anonymized device activities. Figure 6 illustrates the breakdown of mobile devices accessing the towers in Morden and/or Winkler. For an individual, a duty cycle can be estimated, illustrating the percentage of time a person is likely to be in one region or another. The timestamp can also be used to infer primary community of residence. Users counts here indicate that approximately 2650 users remained in Morden, approximately 485 users remained in Winkler, while 2285 users spent time in both Morden as well as Winkler over the four day data collection period.

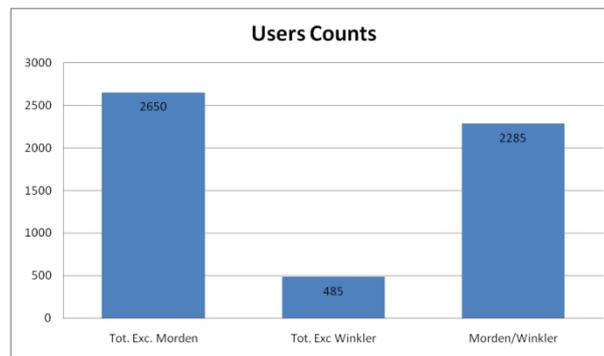


Figure 6: Morden and/or Winkler Mobile User Aggregates

This data can be refined further based upon those with access records in both Morden and Winkler. Figure 7 illustrates the breakdown of users who access cell towers in both communities over the duration of a single connection of their cellular device to the network. The actual device accesses between the two communities break down as approximately 65/35, reflecting durations more accurately.



Figure 7: Breakdown of users with records in both communities

Bluetooth Smartphone Data

The second source of data was a Smartphone application designed to poll its local environment on regular intervals for close-proximity Bluetooth enabled devices. The application is representative of automated and non-intrusive proximity data collection methods where it is tacitly assumed that consumer electronics serve as proxies for their users. This assumption has limitations, including the disproportionate distribution of cellular devices within a given population to certain demographic subsets; yet, arguably these techniques have increasing credibility as more and more people carry electronic devices. To date, a pilot test has been undertaken with four Smartphones collecting data on close-proximity Bluetooth-enabled devices for just over a three month period. During this time approximately 500,000 records were collected. Platforms to date include Blackberry Storm and HTC Hero devices. Data includes the MAC and any assigned meta-identity of both the probe device and the polled (probed) device, the timestamp, and a location if the probe device is GPS-enabled.

Figure 8 illustrates samples of the data collected and residing on the database. Some records provide more information than others and as such, several records are perhaps more interesting than others. The second highlighted row indicates a device called “General Motors”, scanned while the Agent 2 probe was on a local highway. Many other devices are much more easily identified and more easily associated with actual persons. Culling of Bluetooth devices that are not obviously a person is possible but not undertaken here at this time.

id	date	agent	agentmac	device	devicemac	longitude	latitude	altitude	observations
5863	2010-10-25 12:38:31	Agent3	F4093BDC7C	BlackBerry 9520	307C3F6AA7	(N/A)	(N/A)	(N/A)	(N/A)
34643	2010-10-18 09:24:32	Agent3	F4093BDC7C	Antenna Bayer-KouJack's MacBook	6033482A90	(N/A)	(N/A)	(N/A)	(N/A)
14933	2010-10-12 14:26:19	Agent1	F4093C43B9	Emanuel Abrikov's MacBook Pro	C8C2CB4E21	(N/A)	(N/A)	(N/A)	(N/A)
25501	2010-10-13 13:09:16	Agent2	F4093CF258	Sam	001CA41852A9	(N/A)	(N/A)	(N/A)	(N/A)
31712	2010-10-15 13:29:36	Agent1	F4093C43B9	Kevin Barros-Naval's MacBook Air	90840DF48E44	(N/A)	(N/A)	(N/A)	(N/A)
28567	2010-10-14 08:04:23	Agent2	F4093CF258	HF99	0013E60F0ED8	(N/A)	(N/A)	(N/A)	(N/A)
42901	2010-10-20 09:39:46	Agent3	F4093BDC7C	Scott Webb's MacBook Pro	58805F823A98	(N/A)	(N/A)	(N/A)	(N/A)
10489	2010-10-12 09:16:50	Agent3	F4093BDC7C	[AJ]ITS-LAPTOP	00027A579141	(N/A)	(N/A)	(N/A)	(N/A)
3636	2010-10-08 18:03:14	Agent2	F4093C43B9	Pireot v5.11C	00121CECC098	(N/A)	(N/A)	(N/A)	(N/A)
1103	2010-10-07 15:25:55	Agent2	F4093C43B9	General Motors	001E2C2C2C21	(N/A)	(N/A)	(N/A)	(N/A)
56109	2010-10-21 12:39:46	Agent1	F4093C43B9	3gn	C8C2CB4E21	(N/A)	(N/A)	(N/A)	(N/A)
7556	2010-10-09 16:27:09	Agent3	F4093BDC7C	CS10a	002E9FC0D6F	(N/A)	(N/A)	(N/A)	(N/A)
5904	2010-10-08 18:59:06	Agent3	F4093BDC7C	0000	0017D5747C6	(N/A)	(N/A)	(N/A)	(N/A)
20003	2010-10-13 09:24:42	Agent3	F4093BDC7C	hwster	001F6262883	(N/A)	(N/A)	(N/A)	(N/A)
2923	2010-09-18 02:36:38	Agent2	F4093CF258	www.medford-c3	002A3E35094	(N/A)	(N/A)	(N/A)	(N/A)
9358	2010-10-10 15:18:52	Agent3	F4093BDC7C	SF4-M610	001D2E8E5A8	(N/A)	(N/A)	(N/A)	(N/A)
30855	2010-10-14 16:27:25	Agent1	F4093C43B9	Roznie Araya MacBook Pro	00260B0C018	(N/A)	(N/A)	(N/A)	(N/A)
32927	2010-10-16 13:50:13	Agent1	F4093C43B9	M350	001D2E8A1752	(N/A)	(N/A)	(N/A)	(N/A)
473	2010-10-06 16:17:46	Agent1	F4093C43B9	M350A	001D2E8C1EB	(N/A)	(N/A)	(N/A)	(N/A)

Figure 8: A sample of data collected

The Bluetooth contact data is conjectured to be a type of data that can be described by empirical laws. The distribution used follows the Pareto Law. Pareto's law is given in terms of the cumulative distribution function (CDF), i.e. in this case the number of contacts (N_c) with duration larger than or equal to a duration is an inverse power of the duration as expressed below:

$$P[N_c > D] \sim D^{-p}$$

From the Pareto distribution, a power law exponent was calculated and varied from 1.4 to 1.75 for the four probe devices used (R^2 values were consistently above 0.95). A power law exponent less than 2 implies that there is no first moment or mean associated with the distribution. As the data obtained from the probe devices is finite, a mean can be calculated, though. An interesting but not surprising parameter that can be extracted from the Pareto principle is the 80/20 rule. From the data collected, the 80/20 rule was applied to indicate the number of contacts that comprised 80% of the total contact duration. From this, it was estimated that 80% of a person's time is spent with a number of personal contacts that varied between 7 and 20, for the four probe devices. This was extracted from the number and duration of contacts with approximately 5,000 unique Bluetooth devices probed. This is consistent with intuition that although the total number of daily contacts may be large, the majority of one's time is spent with only a small number of people.

Evolving the ABM

This section discusses how models, in this case the ABM can be improved and validated to some degree through inclusion of as many data sources as practical. The first and most obvious would be using as accurate demographic data as possible. The ABM developed here is based on data obtained through the federal census by Statistics Canada. In addition, models of schools have been refined to provide for reasonable class sizes, data which are estimated here but would benefit from using real data of this type. With this model, a disease spread simulation was run and provided a baseline for modeling the spread of a respiratory

infection or ILL. Figure 9 illustrates the spread of a disease among a urban community, represented by Morden, in isolation.

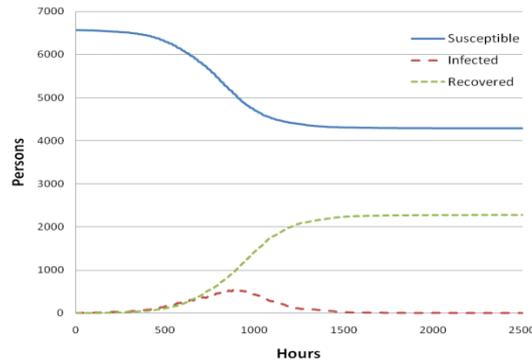


Figure 9: SIR Disease Spread Simulation

In the first effort to improve the basic ABM, it was instrumented in terms of agent contacts and durations which should reflect the patterns in data extracted from the Bluetooth probe devices. The objective was to see how well the model reflected real person-person networks. For the baseline simulations of the single town ABM typical contact patterns for all agents were instrumented. The results of this analysis are summarized as follows:

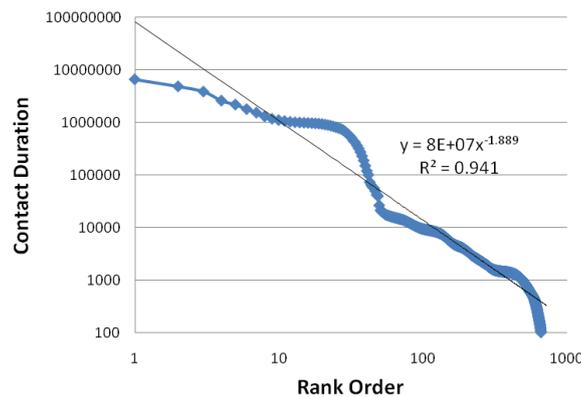


Figure 10: Rank ordering of all agents (aggregated)

Figure 10 illustrates the rank ordering aggregated over all agents. The rank order exponent (Zipf’s law) is approximately 1.9. This yields a estimated power law exponent of approximately 1.53. The implication is that an agent’s contact pattern would follow a power law distribution (heavy tail) without finite moments. This result is expected from both the Bluetooth proximity pilot as well as well intuitive perceptions of real face-to-face contact patterns. This instrumentation of the ABM helps validate it as approximating real world contact patterns. From these ABM simulations and the aggregated rank orderings, an 80/20 rule can also be estimated. In this case, 80% of the contact durations are spent with approximately 4% of a person’s contacts (25/670). This again is consistent with data extracted from the Bluetooth data collection pilot. Figure 11 illustrates the rank ordering of contact parameterized by demographic. Intuitively these profiles appear reasonable. School age children spend considerable time with three groups, household members, school classmates, and friends. The knee in the curve of school age children is between 20 and 32. For samples of age groups the exponents associated with Zipf’s law are presented in Table I. Perhaps it is also intuitive that a 2 year old and a 70 year old have similar contact patterns, presumably though

the 2 year old eats more dirt. Also the distribution of the adults perhaps reflects the famous quote by American philosopher and naturalist Henry David Thoreau who said, “The mass of men lead lives of quiet desperation”. This type of parameter extraction is also consistent with actual survey results reported in [21].

The consequence of the rank ordering implies that the coefficient associated with the corresponding Pareto distribution would be between 0 and 1. The lack of a finite mean in the corresponding contact PDF approximation would imply that a few long duration contacts are a significant vector of infection spread. In these cases the (heavy) tail wags the dog.

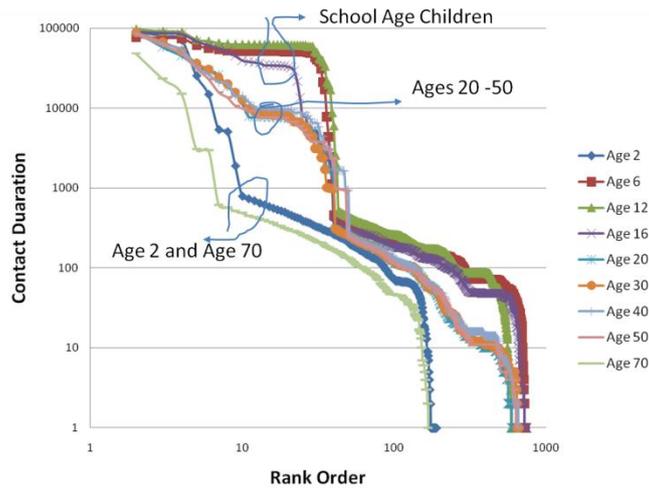


Figure 11: Rank ordering of agents of different demographics

TABLE I. ZIPF EXPONENTS FOR VARIOUS DEMOGRAPHICS

Age	Zipf Exponent	R ²
2	-1.86	0.76
6	-1.51	0.80
12	-1.85	0.78
16	-1.66	0.80
20	-2.0	0.94
30	-1.87	0.96
40	-1.95	0.95
50	-1.95	0.97
70	-1.50	0.85

Discussion

Other means of validating the data from a simulation like this ABM includes its relation to other types of published data. For example, in [21] contact patterns are analyzed as derived from a large population survey that indicated that for their preliminary modeling “5- to 19-year-olds are expected to suffer the highest incidence during the initial epidemic phase of an emerging infection transmitted through social contacts measured here when the population is completely susceptible”. These expectations are consistent with the contact patterns generated by our ABM.

In the second instance of enhancing the ABM, it was recognized that Morden does not exist in isolation and as such, flux of persons into and out of the area is required. This is not unlike large scale efforts where simulations are based upon data extracted from airline travel, for example. In this case the data - albeit voluminous - is reasonably extractable. It is more difficult to obtain inter-community travel in rural settings. In this environment, there are few if any directly available data sets but rather opportunities for inferencing from more disparate sources.

Although an ABM running a bounded topography may be applicable to geographically isolated communities, in semi-rural settings there is considerable interaction with surrounding towns that need be accounted for. From Figure 6, an indication of interactions between Morden and Winkler can potentially be inferred from cellular tower access. The data suggests that of the cell phone carrying persons (approximately 4000) with primary residence in Morden, approximately 34% are seen to have records in both Winkler and Morden, with that person spending on average 65% of their time in Morden and 35% in Winkler. Similarly of the approximately 1400 phone carrying persons with primary residence in Winkler, approximately 65% are seen to have records in both Winkler and Morden, with that person spending on average 65% of their time in Winkler and 35% in Morden.

These very coarse estimates nonetheless allow one to begin modeling multiple communities and their interactions. One can burrow deeper into the data and determine periods of time a representative individual would spend in each community. Further simulations will include representative agent movement trajectories extracted from the cell records integrated into the simulator. Figure 12 illustrates a typical duty cycle associated with randomly selected users and their access to cellular towers in Morden and Winkler. The first two user data duty cycle plots reinforces routine activity theory as users are primarily seen in Morden during the night with intertown tower records primarily during the day. The third user's behavior is considerably more erratic. In either case these types of trajectories are required in improving interacting ABMs.

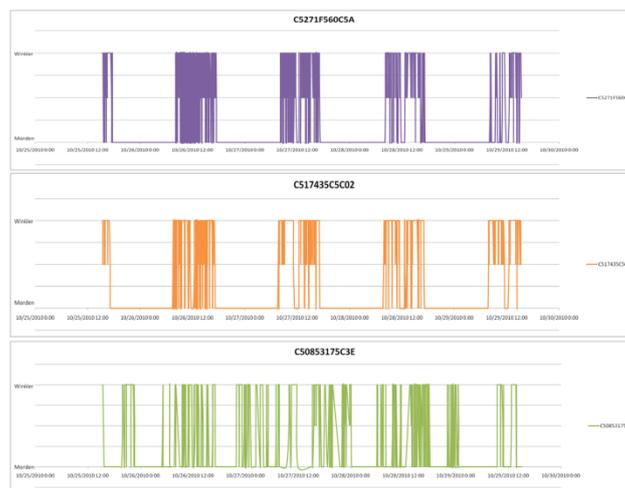


Figure 12: Temporal sequence diagram of a user spending accessing towers in Morden and Winkler

Model evolution is depicted in Figure 13 where external sources are integrated as they become available. At present, these are done in a manual fashion but are amenable to automation

References

- [1] Newman M, Spread of epidemic disease on networks. *Physical Review*. 2002 Jul;66(1):016128.
- [2] Demianyk BCP, Sandison D, Libbey B, McLeod RD, Eskicioglu R, Guderian R, Friesen MR, Ferens K, Mukhi, SN. Technologies for generating personal social network contact graphs. 2010. *IEEE HealthCom 2010*; 2002; Lyon, France.
- [3] Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH. A high-resolution human contact network for infectious disease transmission. *PNAS*, in press 2010.
- [4] Stroud P, Del Valle S, Sydoriak S, Riese J, Mniszewski S. Spatial dynamics of pandemic influenza in a massive artificial society. *J Artificial Societies and Social Simulation*. 2007; 10(4)9, <http://jasss.soc.surrey.ac.uk/10/4/9.html>.
- [5] Carley K, Altman N, Casman E, Fridsma D, Yahja A, Chen L, Kaminsky B, Nave D. BioWar: Scalable agent-based model of bioattacks. *IEEE Trans on Systems, Man, and Cybernetics*. 2006;36:252-265.
- [6] Statistics Canada [Internet]. <http://www.statcan.gc.ca/>
- [7] Laskowski M, McLeod RD, Friesen MR, Podaima BW, Alfa, AS. Models of emergency departments for reducing patient waiting times. *PLoS ONE*. 2009;4(7):e6127.
- [8] Borkowski M, Podaima BW, McLeod RD. Epidemic modeling with discrete space scheduled walkers: Possible extensions to HIV/AIDS. *BMC Public Health*. 2009; 9(Suppl 1): S14, doi:10.1186/1471-2458-9-S1-S14.
- [9] Uhrmacher A, Weyns D, editors. *Multi-Agent systems: Simulation and applications*. New York: CRC Press; 2009. .
- [10] Multi-agent Simulation Environment [Internet]. www.simsesam.de
- [11] Modeling & Simulation – Subproject James II [Internet]. <http://www.mosi.informatik.uni-rostock.de/mosi/projects/cosa/james-ii/>
- [12] Luke S, Cioffi-Revilla C, Panait L, Sullivan K, Balan G. MASON: A Multi-Agent Simulation Environment. *Simulation: Trans of the Society for Modeling and Simulation International*. 2005;82(7):517-527.
- [13] SPADES: System for parallel agent discrete event simulation [Internet]. <http://spades-sim.sourceforge.net/>
- [14] XJ Technologies simulation software and services [Internet]. <http://www.xjtek.com/>
- [15] SWARM development group wiki [Internet]. <http://www.swarm.org/>
- [16] Miller J. Active nonlinear tests (ANTs) of complex simulation models. *Manage Sci*. 1998;44(6):820-30.
- [17] Laskowski M. An agent based decision support framework for healthcare policy, augmented with stateful genetic programming. 2010. Ph.D. Thesis, U of Manitoba.
- [18] Bonabeau E. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Science*. 2002;99(Suppl 3):7280-7287, <http://www.pnas.org/content/99/suppl.3/7280.full#xref-ref-3-1>
- [19] MIT open blocks download page [Internet]. <http://education.mit.edu/openblocks>
- [20] Parunak HVD. Go to the ant: Engineering principles from natural multi-agent systems. *Annals of Operations Research*. 1997;75(0):69-101.
- [21] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Med*. 2008;5(3):e74. doi:10.1371/journal.pmed.0050074