Geographically-Explicit Synthetic Populations for Agent-based Models: A Gallery of Applications

Na Jiang[1 [0000−0001−9161−8954]], Andrew Crooks[1 [0000−0002−5034−6654]], Fuzhen Yin[2 [0009−0003−4981−7345]], and Boyu Wang[1 [0000−0001−9879−2138]]

[1] Department of Geography
University at Buffalo, North Campus, Buffalo, NY 14260, USA
[2] Department of Urban and Regional Planning
University at Buffalo, South Campus, Buffalo, NY 14214, USA
{njiang8, atcrooks, bwang44, fuzhenyi}@buffalo.edu

Abstract. Over the last two decades, there has been a growth in the applications of geographically-explicit agent-based models. One thing such models have in common is the creation of synthetic populations to initialize the artificial worlds in which the agents inhabit. One challenge such models face is that it is often difficult to create reusable geographically-explicit synthetic populations with social networks. In this paper, we introduce a Python based method that generates a reusable geographically-explicit synthetic population dataset along with its social networks. In addition, we present a pipeline for using the population datasets for model initialization. With this pipeline, multiple spatial and temporal scales of geographically-explicit agent-based models are presented focusing on Western New York. Such models not only demonstrate the utility of our synthetic population on commuting patterns but also how social networks can impact the simulation of disease spread and vaccination uptake. By doing so, this pipeline could benefit any modeler wishing to reuse synthetic populations with realistic geographic locations and social networks.

Keywords: Agent-Based Model · Geographically-Explicit Agent-Based Models · Synthetic Population · Python · Mesa.

1 Introduction

Over the last two decades we have seen tremendous growth in the application of agent-based models in a diverse range of fields ranging from archaeology to transportation [8]. In such models, the focus is on exploring how the interactions of individual agents lead to more aggregate patterns emerging. While there are many types of agent-based models, ranging from the predictive to the explanatory [16], one common theme in all models is instantiating an agent population. However, the amount of detail needed to instantiate such models varies. For explanatory models, the populations can be rather abstract and thus not require much detail (e.g., [19]). While for the more predictive models or some case studies, the agents need to be more detailed and grounded on the world around them (e.g., [10, 4]). However, simple models could also have detailed synthetic populations. Thus, the argument we want to put forth here is that agent-based modelers often spend a significant amount of time creating synthetic populations to instantiate their simulations.

So the first aim of this paper is to demonstrate a method that can be used to create geographically-explicit synthetic populations and show through a series of applications how such data can be utlized. These applications are vignettes of where location matters, for example how local interactions lead to the spread of diseases or travel to and from work and the resulting traffic dynamics [8]. To some extent, our argument is that if we are mimicking real geographic areas, the agents should be based on what we know about the area of interest. This argument follows the notion put forward by Mandelbrot [15] if models are generating spatial or physical predictions they must 'look right'.

Initializing populations along with the geographic locations is not the only important task in agent-based modeling, while determining whom the agents interact with is also critical when building such models. Especially, agent-to-agent interactions are one of the hallmarks of agent-based modeling, as it is through such interactions that more aggregate patterns emerge. One way to help foster interactions is through networks [1, 11]. For example, it has been shown that agents' decision-making can be impacted by others who share the same social network [1]. For example, how information diffuses is dependent upon who one trusts (which we will come back to in Section 3.3). Social networks or contact networks (i.e., who do you work with, who is in your family) are also important with respect to how a disease spreads which we will show in Section 3.3. So our second aim of the paper is to demonstrate how social networks can also be incorporated into the generation of synthetic populations.

While there have been efforts to create synthetic populations in fields such as agent-based computational demographics [5] and microsimulation [6], such synthetic populations seldom contain social networks (e.g., [17]) and they are often built for the task at hand (i.e., one-off applications). Others have created synthetic populations with social networks but at the expense of real-world demographic information (e.g., [14, 29]). The reusable synthetic population presented in this paper is intended to free modelers from the task of initializing detailed populations, similar to the notion of how agent-based modeling toolkits allow modelers to focus on what is important, such as developing the model not the graphical user interface (GUI), or charting functions or scheduling agent activities [7].

In what follows due to space limitations we briefly outline our methodology with respect to creating our synthetic population and the general workflow to instantiate our agents (Section 2). We then turn to several applications to demonstrate how the synthetic population can be used to initialize a diverse set of applications at different spatial and temporal scales and also how social networks can be initialized from the beginning. Finally, Section 4 provides a summary of the paper and outlines areas of further work.

2 Methodology

In this section, we first introduce the process of the generation of the synthetic population (Section 2.1) along with the demonstration of the resulting datasets (as shown in Section 2.2). Then, we provide a pipeline on how to utilize the resulting synthetic population datasets to initialize agent-based models from the modeling perspective (see Section 2.3). Both the synthetic population generation and the applications below were created in Python. We chose Python as it is becoming one of the most popular programming languages, which allows us to handle diverse types of geographic data and conduct various spatial analyses. With respect to agent-based modeling, there is also a growing ecosystem of using Python to build such models, especially those utilizing Mesa and Mesa-Geo [13, 27], but we would argue that there are only a handful of geographically-explicit models that have been made freely available for others to learn from. In addition, the applications shown below can further enrich the Python agent-based modeling ecosystem and community.

2.1 Synthetic Population Generation

Figure 2 shows the processes of the generation of synthetic population, which extends our previous work [12]. Specifically, in this work, we extend the population to the whole of New York State (shown in Figure 1) along with code optimization and refinement. Basic model attributes for geographically-explicit agent-based models require agents to have geographic locations throughout the simulation, for example, home and work locations. Thus, when designing the synthetic population for such models, it's intuitive to use coordinate systems to map the agents to specific locations, which includes the places for agents to live, work or study. To some extent, we would argue that the geographic locations of the synthetic population basically create the possibility for our application to capture the changes of locations that mimic the basic patterns during our daily activities, which will be shown in detail in Section 3.

To create the synthetic population, there are a total of 3 steps as shown in Figure 2. The first 2 steps pertain to the creation of the geographic locations of the agents. In Step 1, we identify the different road types from real road network data (e.g., residential, secondary and primary) with GIS analysis. For example, residential and secondary roads are identified to place home locations, while primary roads are used to place workplace locations. We then created individuals using synthetic reconstruction based on the 2010 US Census data and grouped them into households (as 2020 data is still not available). In Step 2, individuals created in Step 1 are assigned daytime locations based on their age, either work (i.e., work on-site and work remotely) or go to educational sites (i.e., elementary, middle, high schools and daycares). The workplace information was extracted from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) dataset [23]. By using the workplace information, we assigned the workplace locations as

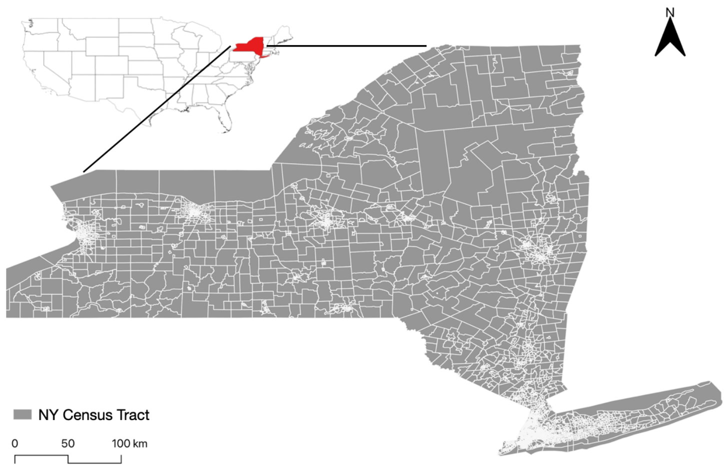

Fig. 1. Census tracts within the study area

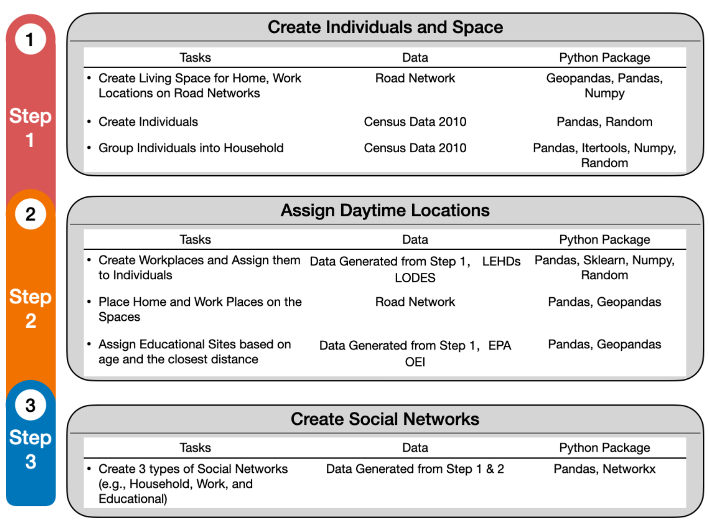

Fig. 2. Steps of Synthetic Population Generation

the daytime locations for adults from the synthetic population. As for the educational sites information (e.g., enrollments and locations), we use data sourced from the US Environmental Protection Agency (EPA) Office of Environmental Information (OEI) [24] to assign younger household members to the closest daycare/schools based on their age (e.g., daycare, elementary, middle, and high school). Further details of this process is provided in our previous work [12].

As discussed above, our synthetic population contains geographic locations by creating spaces on real road networks, while the other main component of the population is their social networks. Within the agent-based model, social networks are used to indicate whom the agent might interact with (see Section 1. In addition, social networks play an important role in information diffusion and disease transmission in our applications. Thus, we created three types of social networks for our synthetic population in Step 3 (shown in Figure 2). Social networks are created based on (1) being in the same household, (2) working in the same place, or (3) attending the same educational site. Household networks are fully connected, while their family members' educational and work networks are generated using a small-world network theory [28] and ideas from Dunbar [9] with respect to group sizes.

2.2 Synthetic Population Resulting Datasets

Figure 3 shows the results generated by the method described in Section 2.1. Specifically, Figure 3(A) displays all individuals generated by our method; 3(B) uses one census tract from the study area to show examples of households, workplaces and education locations on road networks; and 3(C) shows an example household's members' social networks within one census tract. To display the structure of the resulting datasets, Figure 4(A) shows a sample record extracted from the synthetic population dataset and Figure 4(B) shows the sample of the synthetic social networks extracted from the household network. Table 1 provides the basic information related to each resulting dataset. One point that should be noted is the value of n in each social network dataset is different depending on the size of a specific person's social network, where the minimum size is zero and the maximum size is 14 based on notions from Dunbar [9]. According to 2010 United States Census [22], there were 19,387,102 people and 7,317,755 households in New York State. However, we found the number of individuals in our synthetic population (i.e., 19,074,372) which is not identical to the 2010 US Census data. The reason being is that some tracts have inconsistent counts with respect to the total numbers of males and females or some tracts with group quarters like universities only record the total population and no further demographic information.

2.3 Pipeline of Utilizing Synthetic Population Resulting Datasets in Agent-Based Models

As noted above, one of the aims of our paper is to introduce a pipeline by demonstrating how our synthetic population can be used in a number of applications.

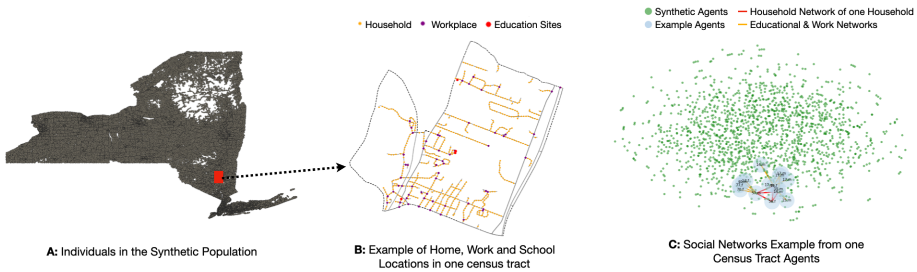

Fig. 3. Synthetic Population Results: (A) All Synthetic Individuals; (B) Household, Workplace and Educational locations within one Census Tract; (C) Agents social networks within one Census Tract.

	id	age	sex	hhold	htype	wp	geometry	long	lat
0	3602900011 0i0	6.0	f	36029000110h335	5	36029000110s0	POINT (-78.82896033753237 42.8319901673968)	-78.828960	42.831990
1	3602900011 0i1	54.0	f	36029000110h66	9	36029010700w572	POINT (-78.82684360311679 42.8441395511133…	-78.826844	42.844140
2	3602900011 0i2	27.0	m	36029000110h185	7	36029000110h185	POINT (-78.85573318586734 42.8396007892089…	-78.855733	42.839601
3	3602900011 0i3	13.0	f	36029000110h335	5	36029000110s0	POINT (-78.82896033753237 42.8319901673968)	-78.828960	42.831990
4	3602900011 0i4	28.0	m	36029000110h186	7	36029000110h186	POINT (-78.83826068581105 42.8553576495999…	-78.838261	42.855358

(A)

	0	1	2	3	4
0	36029000110i0	36029000110i3	36029000110i250	36029000110i1995	36029000110i2356
1	36029000110i3	36029000110i0	36029000110i250	36029000110i1995	36029000110i2356
2	36029000110i44	36029000110i53	36029000110i2204	36029000110i278	36029000110i2151
3	36029000110i53	36029000110i44	36029000110i278	36029000110i2030	36029000110i2204
4	36029000110i58	36029000110i59	36029000110i2740	36029000110i289	36029000110i2192

(B)

Fig. 4. Sample of the Resulting Synthetic Population Dataset: (A) Synthetic Individuals; (B) Social Network.

Table 1. Summary of the generated synthetic datasets (n differs in the size depending on the network, n$\in[0, 14]$).

Dataset	Size	Format	Dimensions ($rows \times columns$)	Description
Synthetic Population	2.83 GB	.csv	$19,074,372 \times 9$	Synthetic individuals and their demographic, location, and work or educational information
Household Network	1.14 GB	.csv	$19,074,372 \times 13$	Individuals and their members live in the same household
Work Network	830 MB	.csv	$8,583,382 \times n$	Individuals and their members in synthetic work networks
School Network	354 MB	.csv	$3,375,721 \times n$	Individuals and their members in synthetic school networks
Daycare Network	94 MB	.csv	$894,496 \times n$	Individuals and their members in synthetic daycare networks

In this pipeline, we consider the resulting population datasets as input parameters to agent-based models. The workflow of inputting these datasets is shown in Figure 5. In general, an *Initialization* function within an agent-based model combines the different input datasets into one dataset that represents all agents in the model. This function first extracts the information (e.g., AgentID, gender, age, work and home locations) to create agents based on the Synthetic Population from Table 1. Following that, the AgentID is used to find the social networks from network datasets (e,g, household, work and educational networks).

As described in Sections 1 and 2.1, geographic locations and social networks play important roles in agent-based modeling. This is also the case here, throughout the simulation, along with the updating of the basic agents' attributes, geographic location changes can also be captured based on the simulation time step. In addition, the model directs the agents to change the interaction group based on their geographic locations. For example, if one agent is employed, then this agent will change its location from home to workplace during the daytime and this agent will interact with the agents that are within the same work social network. Due to the changes in the geographic locations, many types of dynamics can be captured and analyzed by using our pipeline, such as capturing social network changes or formation, tracing location changes and conducting contact tracing of disease outbreaks.

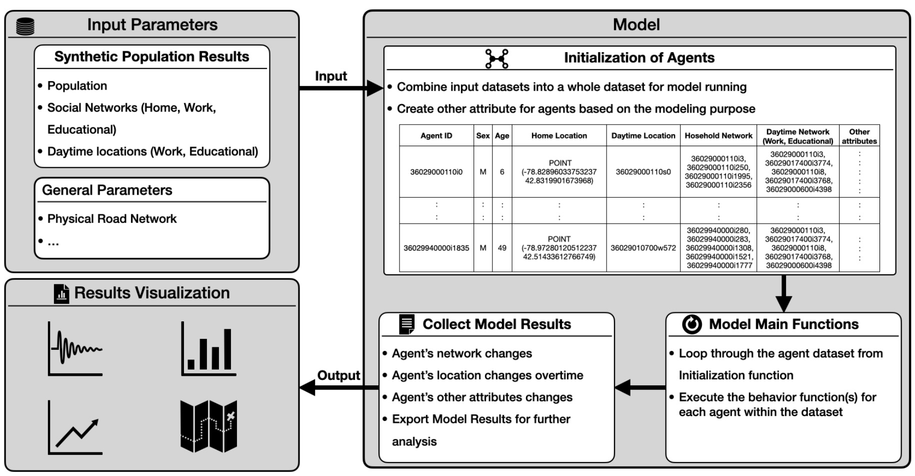

Fig. 5. Pipeline of Utilizing Synthetic Population Resulting Datasets in Agent-Based Models

3 A Gallery of Applications

As discussed in Section 1, three different geographically-explicit agent-based models with various spatial and temporal scales have been built by using the synthetic population generated by our method (see Section 2.1). This section will briefly introduce the design and structure of each model along with a sample of simulation results to demonstrate the flexibility and utility of our synthesized population. Section 3.1 introduces a traffic dynamic model that simulates the morning commute in the Buffalo urban area and Section 3.2 shows a disease spread model in Eire County. Models from Section 3.3 also use Eire County as the study area, but it simulates the people's opinion changes on the COVID-19 vaccination. All models are built using Python to build geographically-explicit agent-based models. In addition, we choose to share all models at: `https://github.com/njiang8/abm-3-application`, which will potentially provide the research foundation for those who share similar research interests.

3.1 Commuting model, Erie County

Commuting is a feature of our daily lives, which refers to going to and from work. Through daily commuting, we can wittiness rush hour traffic and other forms of congestion. As the synthetic population has home and work locations it is relativity straightforward to use this information to initialize a commuting model. Here we just show a simple example that simulates the daily commute of residents living in one census tract from Erie County of New York State and captures their location changes through the simulation (i.e. commuting, at home, at work or at an educational site).

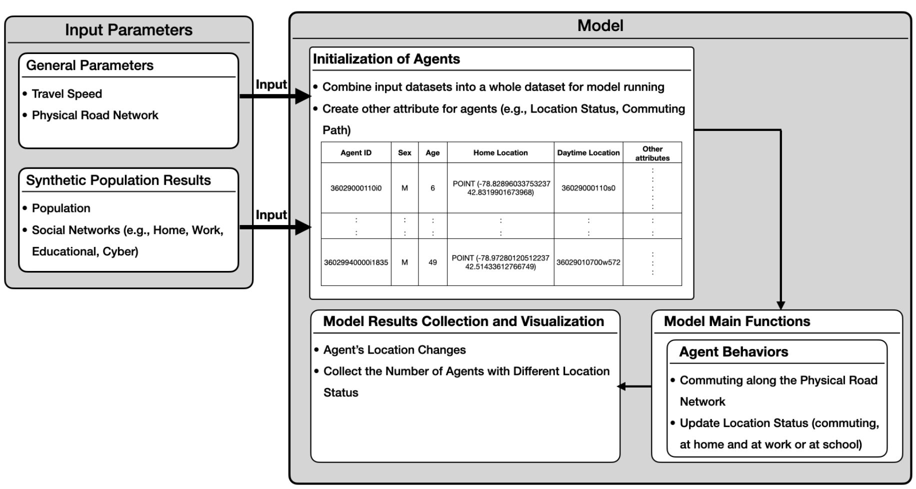

Fig. 6. Erie County Commuting Model

Figure 6 shows the structure of the model, the process is that the model takes the synthetic population dataset as the input to calculate the commute path using an $A*$ algorithm based on home location and daytime location (e.g., work). The model moves the agents on the physical road network based on the travel speed and the commute path. One step in this model represents 5 minutes, in which the model updates the agent location on the physical road network and location statuses. At the starting point of the model, all agents' location status are at home, which we assume the model starts at 5:55 am. When it is the turn for a specific agent to commute, this agent's location status will change to "*commuting*". To determine the time for the agent to start commuting, we randomly assign a time using a normal distribution between 6 am to 9 am for agents to start their morning commute. When the agents reach their daytime locations (e.g., work), their location status will change to "*at daytime location*". In addition, we assume all agents spend eight hours at their daytime location, thus, the afternoon commute will start eight hours after the agents reach their daytime locations. Similarly, when agents reach home, their location status will change back to "*at home*". Figure 7 shows the daily location status changes captured by the model, which can be considered as one of the general patterns of life we see within cities (e.g., home and work commute).

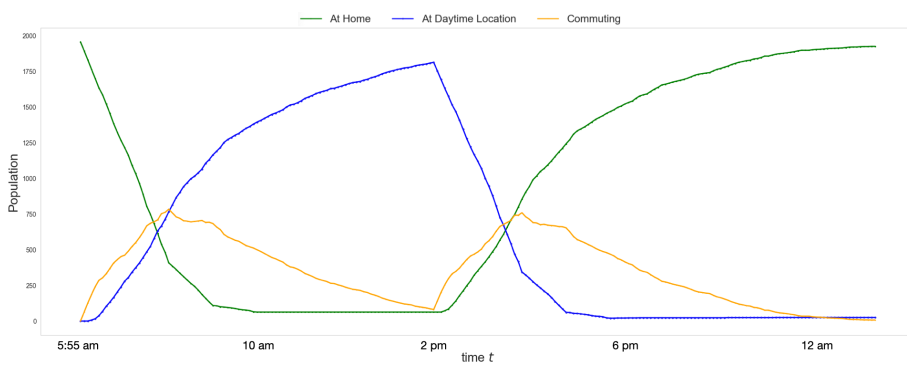

Fig. 7. Location Status

3.2 Large Scale Disease Spread Model

Over the last few years, COVID-19 has drawn attention around the world. To better understand the spread of different threads of COVID-19 and prevent future diseases that would be similar to COVID-19, we utilize the synthetic population and its social networks to build an agent-based model to simulate a COVID-like hypothetical disease spread through different social networks (i.e., home, work and education). With these social networks, the model can be utilized to conduct contact tracing. In addition, different types of viruses can be introduced to the model, which allows us to capture different disease dynamics within one model.

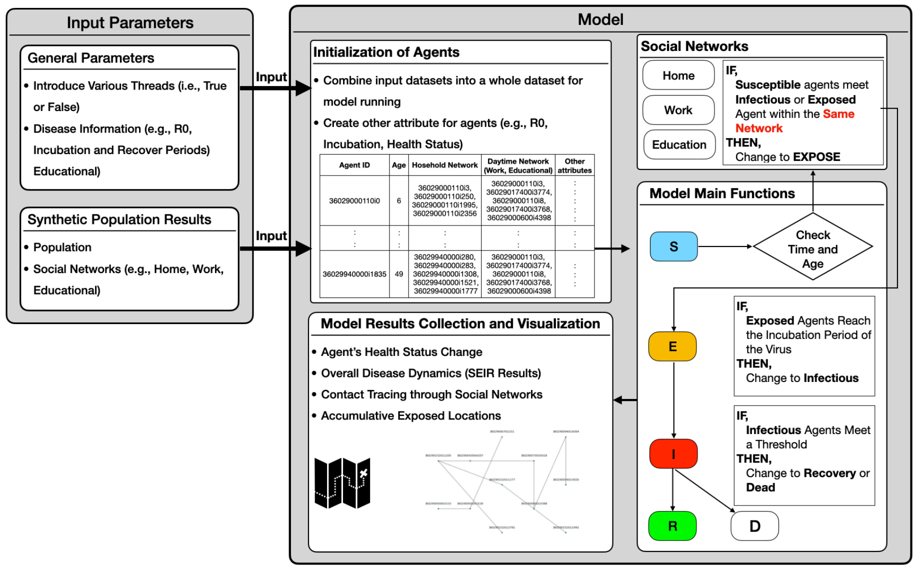

Fig. 8. Large Scale Disease Spread Model Structure.

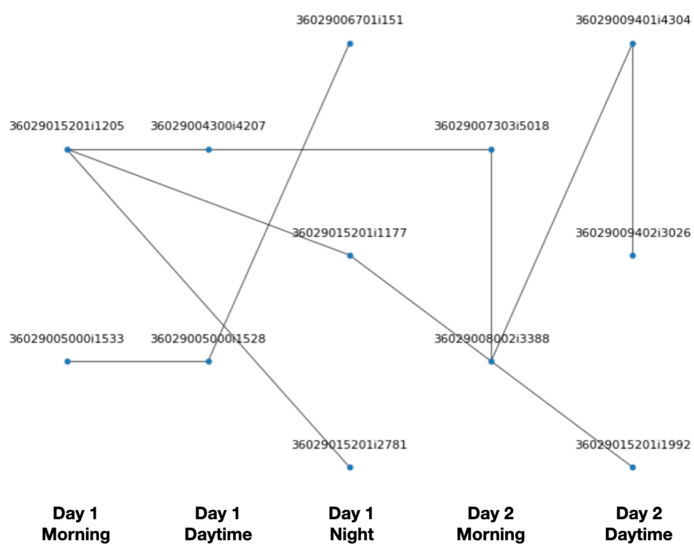

Fig. 9. Disease Model with Contact Tracing via Social Networks.

Figure 8 shows the agent-based disease model's components and how the model takes the synthetic population and its social networks as the input for initialization. The model takes our synthetic population dataset including its social networks as input parameters along with other parameters related to a hypothetical disease (e.g., R0, incubation and recovery period) for initialization. Then, a Susceptible-Exposed-Infectious Removed (SEIR) model is integrated into the agents to represent their health status. In this model, one time step represents eight hours, where one day is divided into 3 time periods (i.e., at home (either sleeping or getting up), being at work (or going to school, and being back at home). In each time step, the agents interact (i.e., spread the disease or get infected) through their networks. For example, while at work, agents will only interact with agents in their work network. While agents are at home, they only interact with their household network members. When the disease is spreading within the model, the model has been programmed to trace the sources of the disease. For example, the model can track the agents who caused other agents to be infected through interactions among their social networks. In this model, two agents are randomly selected as infected at the start of the simulation. Figure 9 shows the contact tracing captured by the model, while Figure 10 (A) shows the disease dynamic of two diseases, while Figure 10 (B) displays the exposed and infectious dynamics for both diseases.

3.3 Vaccination Model for Erie County

Vaccination is one of the most efficient approaches to mitigate the spread of disease, especially, when preventing some newly discovered diseases like COVID-19 [18]. In this model, we simulate the first dose of the COVID-19 vaccination progress within Erie County based on how people's vaccination opinions change which is impacted by their social networks. Besides the traditional social networks (e.g., household, education and work), we also added a cyber social network for agents to represent online communication interactions (i.e., social media). To build the cyber social networks, we identified social media users for both teens and adults based on their demographic attributes such as age, gender, and location [2, 25]. Then, we generated a scale-free network among social media users to match the characteristics of the cyber networks found by other studies [21, 3].

Figure 11 shows the structure of this model, where the model takes the synthetic population and the network described above as the input to initialize the simulation. In addition, parameters related to vaccination progress, such as initial opinions and susceptibility to vaccination are also added to the model. In this model, one time step represents a week. In each time step, agents interact with their neighbors from different social networks and update their opinions toward vaccines and calculate the probability of getting vaccinated. Then the agents decide whether to get vaccinated or not. The model captures the overall vaccine rate for all agents and agents under various age groups. With respect to the opinion changes toward vaccination, different social networks impact the agents' opinions with different weights. Through the experiments, assigning a

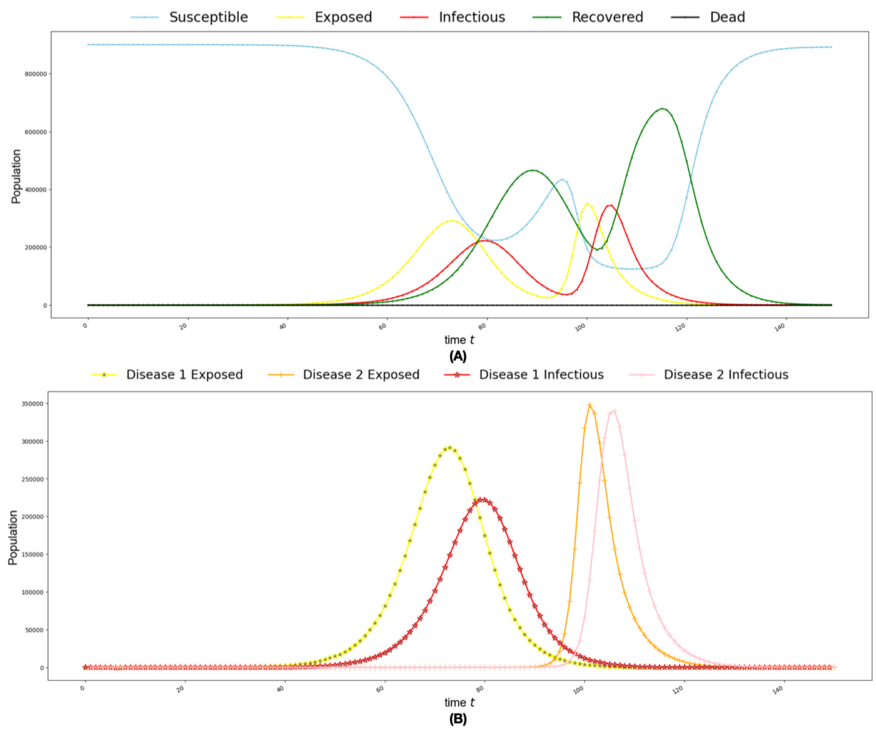

Fig. 10. Disease Dynamics for Two Diseases.

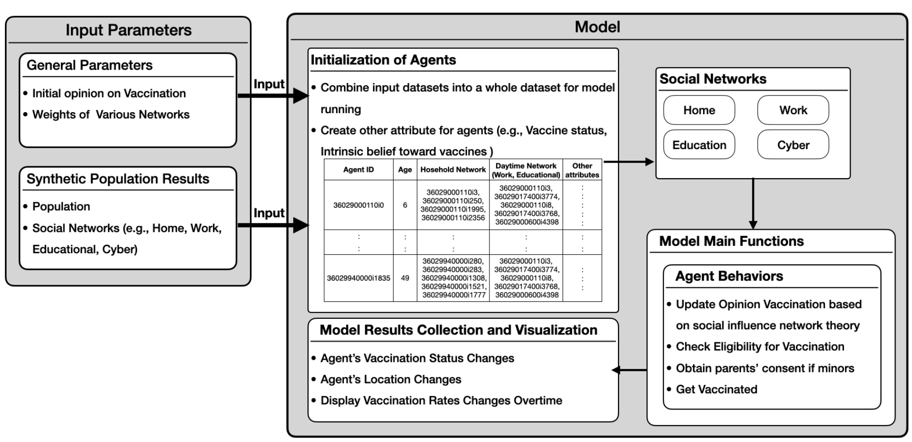

Fig. 11. Vaccination Opinion Dynamic Model.

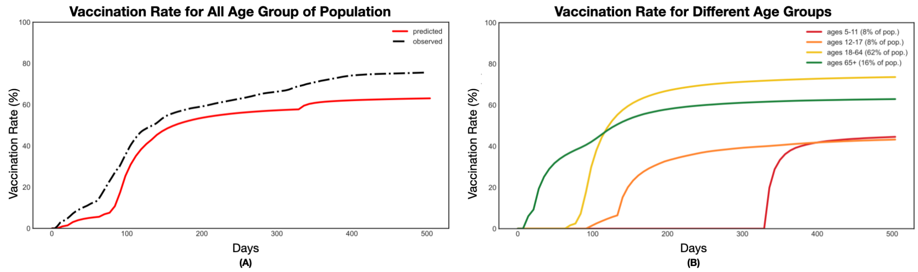

Fig. 12. Simulation Vaccination Rate v.s. Real Vaccination Records: (A) All Population; (B) Different Age Groups of Population

higher weight to traditional social networks (i.e., household and work or educational) than cyber social networks allows the model to generate results that are more aligned with the real-world situation. Specifically, we found when giving equal weight to the household, work and educational social networks and a lower weight on the cyber social network (e.g., one-third of their household and work or educational social networks weights) produce results that are aligned to the actual vaccination rates of Erie County. Figure 12 (A) shows the simulation vaccination rate compared with the real vaccination rate of Erie County, where the simulation results are 8.8 % lower than the real vaccination rate. Due to the synthetic population having age attributes, this model can capture the vaccination rate of different age groups (shown Figure 12 (B)).

4 Summary and Outlook

In this paper, we have introduced a method that creates a reusable synthetic population dataset with social networks, which builds upon our previous work [12] but extended to the entire state of New York. In addition, we have demonstrated a pipeline of utilizing the population dataset and social networks can be used to initialize agents for a diverse range of applications at different spatial and temporal scales. For example, the simulations of commuting (Section 3.1), the spread of different diseases (Section 3.2) and the vaccination process (Section 3.3). At the same time, these models also contribute to the growing ecosystem of Mesa-Geo models [27].

The synthetic population generation method and the applications presented in this paper are based on New York state. However, one of our current works is using the same method to create a synthetic population dataset in the state of Missouri for other agent-based modeling applications (e.g., [26]). Furthermore, we would like to extend this method to the whole of the US. The idea of creating a reusable synthetic population could also be used to populate digital twins or more aptly digital shadows [20] as we do not think that the agents would actually drive the system with a two way flow of information like in digital twins but to populate the digital models such as 3D cityscapes with realistic agents.

Looking ahead, one area of research is to add to the basic synthetic population, for cyberspace types of networks like we have shown in Section 3.3 as this would allow for further explorations into models that span the cyber-physical realm such as opinion dynamics and voting styles of models. It should also be noted that while we have established household, working and educational networks, we have omitted purely social connections. Our rationale for this is that such networks could emerge from the modeling application (e.g., as in [14]). We chose not to do this here, as the focus was on generating a baseline synthetic population that can be applied in different applications. Even with these areas of further work, we hope this paper has demonstrated the utility of geographically explicit synthetic populations and how the same population can be used in diverse applications.

5 Acknowledgments

The authors of this paper would like to thank the Department of Geography, the College of Arts and Sciences and the Center for GeoHazards Studies at the University at Buffalo for providing support for this work.

References

1. Alam, S.J., Geller, A.: Networks in agent-based social simulation. In: Heppenstall, A., Crooks, A.T., See, L.M., Batty, M. (eds.) Agent-Based Models of Geographical Systems, pp. 199–218. Springer, New York, NY (2012)
2. Auxier, B., Anderson, M.: Social media use in 2021. Pew Research Center **1**, 1–4 (2021)
3. Bailey, M., Farrell, P., Kuchler, T., Stroebel, J.: Social connectedness in urban areas. Journal of Urban Economics **118**, 103264 (2020)
4. Benenson, I., Omer, I., Hatna, E.: Entity-based modelling of urban residential dynamics: The case of Yaffo, Tel Aviv. Environment and Planning B **29**(4), 491–512 (2002)
5. Billari, F.C., Prskawetz, A. (eds.): Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour. Springer, New York, NY (2003)
6. Birkin, M., Wu, B.: A review of microsimulation and hybrid agent-based approaches. In: Heppenstall, A., Crooks, A.T., See, L.M., Batty, M. (eds.) Agent-Based Models of Geographical Systems, pp. 51–68. Springer, New York, NY (2012)
7. Crooks, A.T., Castle, C.: The integration of agent-based modelling and geographical information for geospatial simulation. In: Heppenstall, A., Crooks, A.T., See, L.M., Batty, M. (eds.) Agent-Based Models of Geographical Systems, pp. 219–252. Springer, New York, NY (2012)
8. Crooks, A., Malleson, N., Manley, E., Heppenstall, A.: Agent-based modelling and geographical information systems: a practical primer. Sage, London, UK (2018)
9. Dunbar, R.I.M.: The social brain hypothesis. Evolutionary Anthropology: Issues, News, and Reviews **6**(5), 178–190 (1998). https://doi.org/https://doi.org/10.1002/(SICI)1520-6505(1998)6:5¡178::AID-EVAN5¿3.0.CO;2-8

10. Eubank, S., Guclu, H., Anil Kumar, V., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N.: Modelling disease outbreaks in realistic urban social networks. Nature **429**(6988), 180–184 (2004)
11. Hamill, L., Gilbert, N.: Social circles: A simple structure for agent-based social network models. Journal of Artificial Societies and Social Simulation **12**(2), 3 (2009)
12. Jiang, N., Crooks, A.T., Kavak, H., Burger, A., Kennedy, W.G.: A method to create a synthetic population with social networks for geographically-explicit agent-based models. Computational Urban Science **2**(1), 7 (2022)
13. Kazil, J., Masad, D., Crooks, A.: Utilizing python for agent-based modeling: The Mesa framework. In: Thomson, R., Bisgin, H., Dancy, C., Hyder, A., Hussain, M. (eds.) Social, Cultural, and Behavioral Modeling. pp. 308–317. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-61255-9_30
14. Kim, J.S., Jin, H., Kavak, H., Rouly, O.C., Crooks, A., Pfoser, D., Wenk, C., Züfle, A.: Location-based social network data generation based on patterns of life. In: MDM. pp. 158–167. IEEE (2020)
15. Mandelbrot, B.B.: The Fractal Geometry of Nature. W.H. freeman, New York (1983)
16. Manson, S., An, L., Clarke, K.C., Heppenstall, A., Koch, J., Krzyzanowski, B., Morgan, F., O'Sullivan, D., Runck, B.C., Shook, E., et al.: Methodological issues of spatial agent-based models. Journal of Artificial Societies and Social Simulation **23**(1) (2020)
17. Nowok, B., Raab, G.M., Dibben, C.: synthpop: Bespoke creation of synthetic data in r. Journal of Statistical Software **74**(11), 1–26 (2016). https://doi.org/10.18637/jss.v074.i11, `https://www.jstatsoft.org/index.php/jss/article/view/v074i11`
18. Polack, F.P., Thomas, S.J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J.L., Pérez Marc, G., Moreira, E.D., Zerbini, C., et al.: Safety and efficacy of the bnt162b2 mrna covid-19 vaccine. New England journal of medicine **383**(27), 2603–2615 (2020)
19. Schelling, T.C.: Dynamic models of segregation. Journal of mathematical sociology **1**(2), 143–186 (1971)
20. Sepasgozar, S.M.: Differentiating digital twin from digital shadow: Elucidating a paradigm shift to expedite a smart, sustainable built environment. Buildings **11**(4), 151 (2021)
21. Ugander, J., Karrer, B., Backstrom, L., Marlow, C.: The anatomy of the facebook social graph. arXiv preprint arXiv:1111.4503 (2011)
22. US Census Bureau: Decennial Census (2010, 2000) (2010), retrieved September 9, 2019, from `https://www.census.gov/data/developers/data-sets/decennial-census.html`.
23. US Census Bureau: US Census Bureau Center for Economic Studies Publications and Reports Page (2019), `https://lehd.ces.census.gov/announcements.html`. Accessed Sep. 09, 2019
24. USEPA Office of Environmental Information: Educational Institutions, US, 2015, ORNL, SEGS (2015), `https://tinyurl.com/2n9h8hmp`. Accessed July. 25, 2023
25. Vogels, E.A., Gelles-Watnick, R., Massarat, N.: Teens, social media and technology 2022. Pew Research Center (2022)
26. Wang, B., Crooks, A.: Agent-Based Modeling of Consumer Choice by Utilizing Crowdsourced Data and Deep Learning. In: Beecham, R., Long, J.A., Smith, D.,

Zhao, Q., Wise, S. (eds.) 12th International Conference on Geographic Information Science (GIScience 2023). vol. 277, pp. 81:1–81:6. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany (2023)
27. Wang, B., Hess, V., Crooks, A.: Mesa-Geo: A GIS extension for the Mesa agent-based modeling framework in python. In: Proceedings of the 5th ACM SIGSPATIAL International Workshop on GeoSpatial Simulation, (GeoSim '22). p. 1–10. Association for Computing Machinery (2022). https://doi.org/10.1145/3557989.3566157
28. Watts, D., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684), 440 (1998)
29. Züfle, A., Wenk, C., Pfoser, D., Crooks, A., Kim, J.S., Kavak, H., Manzoor, U., Jin, H.: Urban life: a model of people and places. Computational and Mathematical Organization Theory **29**(1), 20–51 (2023)