

# A Flexible Agent-Based Framework for Infectious Disease Modeling

Florian Miksch<sup>1,2,\*</sup>, Christoph Urach<sup>3</sup>, Patrick Einzinger<sup>2</sup>, and Günther Zauner<sup>2</sup>

<sup>1</sup>University of the Philippines Cebu, Department of Computer Science, Cebu City, Philippines  
fmiksch.up@gmail.com

<sup>2</sup>dwh Simulation Services, Vienna, Austria  
{florian.miksch,patrick.einzinger,guenther.zauner}@dwh.at

<sup>3</sup>Vienna University of Technology,  
Institute for Analysis and Scientific Computing, Vienna, Austria  
christoph.urach@tuwien.ac.at

**Abstract.** Agent-based modeling is a method to model a system by autonomous entities. The proposed framework models single persons with personal behavior, different health states and ability to spread the disease. Upon simulation, the epidemic emerges automatically. This approach is clear and easily understandable but requires extensive knowledge of the epidemic's background. Such real-world model structures produce realistic epidemics, allowing detailed examination of the transmission process or testing and analyzing the outcome of interventions like vaccinations. Due to changed epidemic propagation, effects like herd immunity or serotype shift arise automatically. Beyond that, a modular structure splits the model into parts, which can be developed and validated separately. This approach makes development more efficient, increases credibility of the results and allows reusability and exchangeability of existing modules. Thus, knowledge and models can be easily and efficiently transferred, for example to compute scenarios for different countries and similar diseases.

**Keywords:** epidemic simulation, modular concept, disease propagation.

## 1 Introduction

Simulation of epidemics has a long history in mathematics but also in medical fields. However, calculations vary a lot in addressed problems and accuracy. In the past years, new methods have emerged and became possible due to increasing computational power. The intention of this work is to integrate old and new methodologies in a newly developed framework to provide a flexible, standardized and easy-to-handle approach for modeling a wide class of infectious diseases. This framework consists of a model that relies on agent-based modeling, which is a promising young technique because it aims to simulate dynamic systems based on fundamental rules [1–3]. Its modular concept assures high flexibility. The approach should be able to produce accurate results by modeling epidemics in a realistic way, which helps to deal with uncertain dynamics and effects. It can be used for studies on a specific disease and is

---

\* Corresponding author.

suitable for current research on general issues concerning analysis of epidemics. The approach also supports handling of disputed effects like herd immunity [4, 5] and serotype shift [6–8], which make it hard to predict the outcomes of vaccination strategies. In an agent-based model they automatically occur upon the fundamental rules and are not part of the model structure.

## **2 Underlying Concepts**

### **2.1 Epidemic Modeling**

Classical approaches for epidemic modeling often use ordinary differential equation respectively system dynamics, where aggregated variables represent the population of interest that is being split by health and other attributes [9–13]. These models are easy to handle from a mathematical point of view but they are on a high abstraction level which makes it hard to go into details or calculate parameters from observed data. In contrast to that, another concept is simulating simplified individual people with their behavior and observing how an epidemic arises [3, 14, 15]. Such an approach is often referred to as agent-based modeling and creates an epidemic based on knowledge on individuals.

### **2.2 Agent-Based Modeling**

Agent-based modeling is a method that emerged in the 1990s. It tries to model the world as it is observed, based on individual entities, which are called agents. These agents exist within an in-silico environment; they have attributes and behavior, can change themselves and also affect each other [1, 16–18]. For epidemics, this approach means to model single persons in their environment, give them the ability to be healthy or infected and assure that relevant contacts, which allow transferring the disease from infectious to susceptible persons, happen. Generally, it is important to incorporate all information relevant for the spread of an infectious disease such as personal attributes, social behavior concerning contacts and aspects about the disease. Then, one can simulate the model and observe the propagation of the disease. The outcome of single individuals in the model is usually not of interest which requires further computations. Results are commonly given by statistical analyses on the simulated population.

It should be clear that agent-based modeling is not a well-defined method. It is rather a general concept that provides freedom for the modeler but requires extended research for specific models addressing different problems [18].

## **3 The Model Framework**

### **3.1 Aim**

We intend building a framework for epidemic simulation that should be flexible enough to be applied on different populations, different situations and on different diseases with different transmission paths. To support this aim, we focus on a modular

concept, introducing three different modules for population, contacts and disease, which can be developed and validated separately. Additionally, a protocol module tracks the agents and provides the results.

The following chapters show an approach how to set up such a framework and how to read results and to interpret the dynamics of the system. The aim is to show the ideas, structures, challenges and potentialities, but it is not a programming instruction.

### 3.2 The Population Module

The agents, which represent people, are the most fundamental part of the model. In a first step, the population module has to create these people and equip them with basic attributes like age and gender and other relevant information about personal background or previous medical history. They might have behavior like getting older, eventually die, and women might give birth to children. Considering immigration and emigration assures an accurate population structure over long simulation periods.

The module has to create the desired number of people and assign them these attributes according to their joint distribution. Additionally, it has to provide interfaces in a way that other modules can add more attributes and rules. Commonly, each interface is implemented as a reference to an object that handles all attributes and behavior assigned by a module.

### 3.3 The Contact Module

An infectious disease spreads by transmission from one person to another. It can be transmitted in various ways like for smear infections, sexually transmitted or airborne diseases. Pathogens can also be transmitted by food, water, or by animals like mosquitos. The duty of the contact module is to model all kind of events that can lead to transmission of the actual disease.

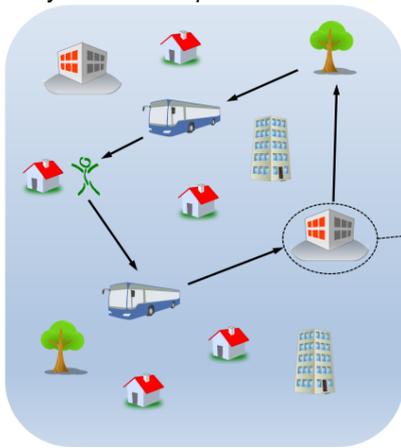
We are going to present a place-based approach to build a social contact system, which differs from partly used contact networks [14, 19, 20]. This is done in 3 steps and is mostly suitable for diseases that spread directly through airborne and smear transmissions. Figure 1 shows an exemplary visualization of such a structure.

(1) At first, all places where infectious contacts might happen need to be created. Studies often explicitly consider households, workplaces, schools, transportation and leisure places [21]. Based on national data, the modeler needs to identify how many places of each place type exist for the given population. Then, assignments to places indicate where people generally belong to. For example, each person gets assigned to exactly one household so that the household distribution by size and age of people is achieved [22]. Assignments to schools and workplaces tell where someone works or goes to school.

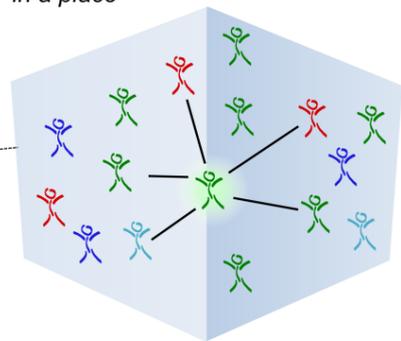
(2) Step two builds a system of daily routines. In a time step, each person visits one or more places. Parameterization of daily routines can be tough since it usually combines national data about places and social contact data. Time steps are considered to be atomic. This means that a person has only two options: Visiting a place within a time step or not. However, it makes sense to allow visiting multiple places during a

time step. For example, if a time step represents one day, then persons would visit all places of their whole daily routine. But then it is not possible any more to differentiate on a finer level when they go there or how much time they spend there. Desiring more detailed daily routines requires shorter time steps like hours or even minutes. The daily routine of a person depends on several factors: On personal attributes, if a person is feeling sick, on the day of the week. Assigned places assure that people always go to the same household, school or workplace where they meet the same colleagues and friends, and variations, especially for transport and leisure, provide a mixing of random people.

*daily routine of a person*



*contacts of a person in a place*

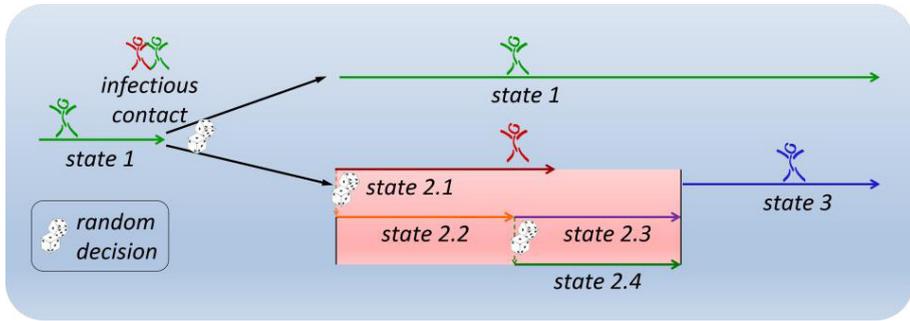


**Fig. 1.** Exemplary visualization of a person's daily routine and contacts within a place. A person visits several places each day, depending on many factors and influences. In places, contacts happen between present people; number and distribution of contacts are according to knowledge on social studies.

(3) This step finally creates contacts in every place between all present people within a time step. Social studies, often based on empirical research, serve as a basis for contact patterns [21]. This means to model contacts between random people in a place, so that contact numbers and age-distribution of studies are resembled. Depending on the place and the circumstances, contacts might be loose or close, short or long, physical or non-physical.

### 3.4 The Disease Module

The duty of the disease module is the handling of all aspects concerning the disease. An agent-based model involves information on an individual level, which can be split up into three parts: First, disease and health states of a person, second, state changes caused by contacts and third, state changes independent of contacts. Figure 2 shows an exemplary visualization of these tasks.



**Fig. 2.** Exemplary visualization of disease states and state changes. An infectious contact can cause an infection, resulting in an automatic disease process of different states until the person recovers.

The states are attributes of persons to describe their condition concerning the disease. In simple models there might be only a few states like healthy and sick, and every person is in either one or the other state. In more detailed models, states can consist of a combination of attributes like susceptibility, being infected with the pathogen, being infectious, symptoms severity, feeling pain or passed duration of an infection. The state also contains information about the immune system and about the patient's medical history. State changes upon contacts typically define transmissions of the disease but technically can trigger any desired action. Whenever a contact happens, the involved persons can change their state, depending on their own and their contact person's states, and also on the contact characteristics like duration or intensity. For complex diseases it is suggested to split this action into an attack phase, where the infected person transmits the disease with a probability, followed by a receive phase where the susceptible person either gets infected or defeats the disease. Often, transmission of a disease is not a deterministic process, but happens under special conditions with a certain probability. The other state changes might depend on the state of the person and their environment. They typically represent progression of a disease, recovery, or general changes of the personal health.

Disease states can also affect other modules. For example, a disease might affect the daily routine of persons so that they only stay at home and do not go to work or school. Disease might cause the death of a person which is handled by the population module.

### 3.5 The Protocol Module

The protocol module keeps track of all information of interest to generate the desired results. There are two possible approaches. The first possibility is that the protocol permanently checks everything that happens and stores the information of interest. This approach can be extended to a so called VOMAS (Virtual Overlay Multi-Agent System), a well-known method where an overlay is added to the model containing another type of agents that watch and log [23]. The other approach is having a listening protocol module while it is the agents' duty to report all information of interest.

After the simulation the protocol module usually provides statistical calculations as results based on all collected individual data.

Concerning epidemics, two classes of data that are evaluated by the protocol module are of particular interest: states of agents and changes of states. For example, one might be interested in the number of people that are infected but also in the number of people that get infected in a time step. More detailed analyses are able to reveal where and under what circumstances transmissions happen and help to indicate population subgroups that are particularly responsible for spreading the disease.

### 3.6 Simulating the Model

The following paragraph describes a proposed simulation strategy that splits up into three phases and is also shown in Figure 3. In the initialization phase, all modules initialize. They create the persons, equip them with all required attributes so that they represent the desired population and also set up the environment (e.g., the places). Then the simulation phase starts. The model simulates over time using time steps, which are by definition atomic and cannot be split up further. This means that they consist of several actions that are not temporally ordered. Instead, a smart collision handling is required, which decides upon logical rules what to do. It is highly suggested to preserve personal states during a whole time step, store proposed changes

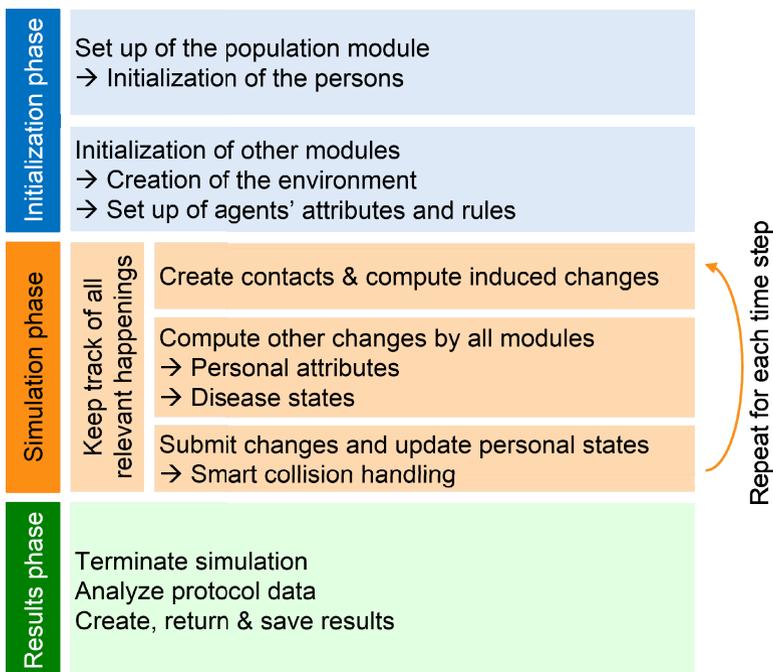


Fig. 3. Overview of phases and simulation of the agent-based epidemic framework

and perform changes only between time steps. At first, contacts are created and contact-based infections and changes are calculated. Then, other changes are created by the modules. Finally, every person has a set of proposed changes. The logical procedure of deciding the personal state for the next time step highly depends on the actual disease. Generally it makes sense to apply all attributes that do not affect each other, while competitive changes within one time step, for example new infections, recoveries, infections with different diseases or vaccinations, might be chosen case dependent. The only restriction is to find a well-defined state for every person for the next time step. The protocol module keeps track of the information of interest (e.g., personal states and proposed and performed state changes) and also of other relevant happenings (e.g., constructed contacts). This procedure is repeated for each time step and results in an evolving system. Finally, in the results phase the protocol module stores and represents the collected data in an appropriate format. Commonly this includes histograms and time series but, depending on the actual model, further statistical analyses and experimental visualizations might be possible and useful.

## 4 Parameter Settings and Interpretation of Results

Correct parameterization and interpretation of the model results are crucial aspects for modeling of epidemics. Most aspects are time related. The model itself does not provide any real timing like days, weeks or years. Instead, it calculates a number of time steps, which have to be interpreted as real time units. For example, for fast-evolving epidemics like ebola or influenza it makes sense to define a time step as a short interval like some hours or days. Slow-evolving epidemics such as HIV do not require a daily simulation, hence one time step might represent a week, months or even years. Following the good practice in modeling to make a model as simple as possible and as complex as required, it is recommended to set the time steps short enough to obtain all desired dynamics and long enough to prevent unnecessary overhead [24].

Once the time steps are defined, the other parameter values can be set. Number of contacts and many disease parameters are given time-dependent; hence they need to correlate to the timing of the model. For example, if the recovery time is known to be two weeks and the model simulates day by day, then the recovery time has to be set as 14 time steps. If the recovery time is known as a half day only then one might revise the timing and consider defining one time step as 12 hours.

Wrong parameter settings might not only distort results, they can even change the dynamic of the whole system. Sensitivity analysis is a good way to overcome this issue, to show the impact of uncertain parameters and to analyze the range of possible results. Like for statistical calculations, processing results of agent-based models always leaves room for interpretation. Hence, it is crucial to scrutinize the way results are generated from agents' data and discuss whether results correctly represent what they pretend.

## 5 Discussion and Outlook

The agent-based concept offers great benefits. Constructing a system that produces an epidemic in a similar way as it happens in reality produces reliable results and allows a wide range of testing assumptions. A real-looking structure leads to a better understanding of the model so that it is easier to communicate, gains a higher acceptance and allows non-modelers to comment on it, which helps improving it.

Consideration of vaccinations allows testing of various scenarios. In this context, a few interesting effects occur, which are represented accordingly in the agent-based model. Stochastic calculations and models are able to consider such effects only if they are known for every single parameter setting. Often, this is a severe issue because they cannot be measured for future scenarios and also cannot be extrapolated from available data. In the agent-based model, where the epidemic automatically emerges upon basic rules, herd immunity and serotype shift both occur and can even be examined without prior knowledge. This does not only make results more trustworthy, it even allows fundamental research on the spread of epidemics. Yet it is a matter of current research to find a definition and a standardized way for measurement of herd immunity and serotype shift [5]. Beyond that, observation and analysis of the basic reproduction rate  $R_0$ , which commonly describes the strength of an epidemic, is possible [25].

However, the method also has an issue that requires additional effort: If a relevant basic factor is not considered or included in a wrong way, then the model might still produce epidemics, but the propagation of the disease will happen incorrectly. Hence, the results will be wrong. Validation is a term containing a wide class of methods to overcome this issue by asking whether the model is being developed according to the model question [26–28]. Hence, it helps to identify problems and errors and increases the credibility of the model.

Another key benefit is the modular structure, especially for well-planned interfaces between the modules. First, the modular approach allows independent development of the modules, reusability of modules, combinations of existing modules. Thus, a pool of modules can be developed to create specific models just by assembling them. For example, various population modules representing different societies allow efficient transferability of the model to other countries. Development of new disease modules makes it possible to use an existing model for simulating other diseases. Second, it assures that changes only affect one specific module, which makes changes, adoptions and improvements easier. This highly supports an iterative modeling process. A third benefit is that separate validation of the single modules supports validating the whole model. Hence, validated modules might be reused for other projects, which decreases effort and increases accuracy.

## 6 Conclusions

Agent-based modeling is a promising way to simulate epidemics. It models individual people, their contacts and transmissions. The global spread of the disease is not

explicitly defined and happens automatically. Such a system has great benefits because it allows modeling complex dynamics based on simple rules. The structures, which finally lead to an epidemic, should correspond to reality. If these real-world structures incorporate all relevant basics then they result in a real-world behavior. This leads to direct calculation of different scenarios and high quality results and allows examination of widely disputed and unknown effects. Studying the infection and propagation process and its impact on the overall spread is also possible. From a technical point of view, it is generally simple to make changes for different assumptions in agent-based models, especially modular ones, because they only affect a small part of the model while other model parts remain unchanged.

## References

1. Casti, J.L.: *Would-be worlds: how simulation is changing the frontiers of science*. J. Wiley, New York (1997)
2. Okhmatovskaia, A., Verma, A.D., Barbeau, B., Carriere, A., Pasquet, R., Buckeridge, D.L.: A simulation model of waterborne gastro-intestinal disease outbreaks: description and initial evaluation. In: *AMIA. Annu. Symp. Proc.*, pp. 557–561 (2010)
3. Liccardo, A., Fierro, A.: A Lattice Model for Influenza Spreading. *PLoS One* 8, e63935 (2013)
4. Topley, W.W.C., Wilson, G.S.: The Spread of Bacterial Infection. The Problem of Herd-Immunity. *J. Hyg (Lond.)* 21, 243–249 (1923)
5. John, T.J., Reuben, S.: Herd immunity and herd effect: new insights and definitions. *Eur. J. Epidemiol.* 16, 601–606 (2000)
6. Lysenko, E.S., Lijek, R.S., Brown, S.P., Weiser, J.N.: Within-Host Competition Drives Selection for the Capsule Virulence Determinant of *Streptococcus Pneumoniae*. *Curr. Biol.* 20, 1222–1226 (2010)
7. Hsu, K.K., Shea, K.M., Stevenson, A.E., Pelton, S.I.: Changing Serotypes Causing Childhood Invasive Pneumococcal Disease. *Pediatr. Infect. Dis. J.* 29, 289–293 (2010)
8. WHO: Changing epidemiology of pneumococcal serotypes after introduction of conjugate vaccine: report. *Wkly. Epidemiol Rec.* 85, 434–436 (2010)
9. Kermack, W.O., McKendrick, A.G.: A Contribution to the Mathematical Theory of Epidemics. *Proc. R. Soc. Math. Phys. Eng. Sci.* 115, 700–721 (1927)
10. Dietz, K.: Epidemiologic interference of virus populations. *J. Math. Biol.* 8, 291–300 (1979)
11. Lipsitch, M.: Vaccination and Serotype Replacement. In: Dieckmann, U. (ed.) *Adaptive Dynamics of Infectious Diseases: in Pursuit of Virulence Management*, pp. 362–374. IASA, Cambridge University Press, Cambridge, New York (2002)
12. Matrajt, L., Longini, I.M.: Critical immune and vaccination thresholds for determining multiple influenza epidemic waves. *Epidemics* 4, 22–32 (2012)
13. Xue, Y., Kristiansen, I., de Blasio, B.: Dynamic modelling of costs and health consequences of school closure during an influenza pandemic. *BMC Public Health* 12, 962 (2012)
14. Eubank, S., Kumar, V.S.A., Marathe, M.V., Srinivasan, A., Wang, N.: Structure of Social Contact Networks and Their Impact on Epidemics. *AMS-DIMACS Spec. Vol. Epidemiol.* 70, 181–213 (2006)

15. Bauer, A., Pöll, C., Winterer, N., Miksch, F., Breitenecker, F.: Analysis and comparison of different modelling approaches based on an SIS epidemic. In: Proceedings of the International Workshop on Innovative Simulation for Health Care 2012, Vienna, pp. 115–120 (2012)
16. Wooldridge, M.: Agent-based software engineering. *IEE Proc. Softw. Eng.* 144, 26–37 (1997)
17. Jennings, N.: On agent-based software engineering. *Artif. Intell.* 117, 277–296 (2000)
18. Macal, C.M., North, M.J.: Tutorial on agent-based modelling and simulation. *J. Simul.* 4, 151–162 (2010)
19. Blower, S., Go, M.-H.: The importance of including dynamic social networks when modeling epidemics of airborne infections: does increasing complexity increase accuracy? *BMC Med.* 9, 88 (2011)
20. Parker, J.: A Flexible, Large-Scale, Distributed Agent Based Epidemic Model. In: Proceedings of the 2007 Winter Simulation Conference (2007)
21. Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G.S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., Edmunds, W.J.: Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Med.* 5, e74 (2008)
22. Miksch, F., Zauner, G., Popper, N., Breitenecker, F.: Agent-Based Population Models For Household Simulation. In: Snorek, M., Buk, Z., Cepek, M., Drchal, J. (eds.) Proceedings of the 7th EUROSIM Congress on Modelling and Simulation, Prague, Czech Republic, pp. 567–572 (2010)
23. Muaz, A.N., Hussain, A., Kolberg, M.: Verification & Validation of Agent Based Simulations using the VOMAS (Virtual Overlay Multi-agent System) approach. In: Proceedings of the Second Multi-Agent Logics, Languages, and Organisations Federated Workshops, Torino, Italy (2009)
24. Sánchez, P.C.: As simple as possible, but no simpler: a gentle introduction to simulation modeling. In: Proceedings of the 2006 Winter Simulation Conference, Monterey, CA, pp. 2–10 (2006)
25. Keeling, M.J., Grenfell, B.T.: Individual-based Perspectives on R0. *J. Theor. Biol.* 203, 51–61 (2000)
26. Balci, O.: Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Ann. Oper. Res.* 53, 121–173 (1994)
27. Sargent, R.: Verification and validation of simulation models. In: Proceedings of the 2010 Winter Simulation Conference, Baltimore, MD, pp. 166–183 (2010)
28. Klügl, F.: A validation methodology for agent-based simulations. In: Proceedings of the 2008 ACM Symposium on Applied Computing, pp. 39–43. ACM Press, New York (2008)