Tutorial Note on Agent-based Modeling and Simulation: Application to Diffusion Models

Alexis Drogoul and Benoit Gaudou

1 UMI 209 IRD-UPMC UMMISCO, Bondy, France
2 UMR 5505 CNRS IRIT, University of Toulouse, Toulouse, France

Abstract. This tutorial note aims at introducing agent-based paradigm for the modeling and simulation of complex systems. It will focus on its key concepts and highlight its specific features and benefits. A big part of the paper is dedicated to provide examples of applications taken in the diffusion model literature illustrating the versatility of agents and benefits it can bring to model in terms of heterogeneity (concerning agents or the environment).

1 Introduction

After years dominated by macroscopic approaches of the modeling, that describe with equations the behavior of the studied system or phenomenon only from a global point of view, modeling and simulation have undergone a deep revolution with the application of Multi-Agent Systems [1] to the problem of the modeling and simulation of complex systems. Agent Based Modeling and Simulation (ABMS) is a paradigm that allows modelers to reason and represent the phenomenon at the microscopic (individual) level, and to take into account heterogeneity and complexity both in the individual layer and the environment layer. ABMS has been successfully used in various research fields such as in Ecology [2] or Social Sciences [3].

We concentrate our presentation of examples on diffusion models and in particular to disease spreads, opinion dynamic, innovation diffusion and diffusion of culture models. We aim at illustrating benefits of the agent-based approach over the macroscopic one.

Words model and simulation in the sense we will use them along the paper are first defined (Section 2). Then we introduce the key concepts of the agent-based paradigm (Section 3) before presenting examples of models of diffusion phenomena (Section 4). Finally we conclude by presenting the current research trends and issues (Section 5).

2 Model, Simulation, Experiment

In the sequel we will refer a model in the sense of Minsky [4]: "To an observer B, an object A* is a model of an object A to the extent that B can use A* to answer questions that interest him about A". In our case, the object A will be named reference or target system. It is important to note that this definition highlights the fact that the model is developed with relation to a question on the system. One of the major flaw of a lot
of modeling and simulation projects is linked to the fact that the question is not well defined. Among all existing models, some (named static models) describe the structure of the reference system (the elements constituting the system and their relations). Inversely dynamic models address the question related to its evolution. The execution of a (implemented) dynamic model on a computer is called a \textit{simulation}.

The analysis of the model will need a lot of simulations to explore its behavior in function of various values of its parameters. The process will be quite similar to experimentations: controlled perturbations of the system (reference system or dynamic model) to answer a question.

Next section describes the few key concepts used to write agent-based models.

3 Agent-based Approach: A Small Set of Key Concepts

Figure 1 illustrates the key concepts of a Multi-Agent System. To be short, an agent-based model will be a kind of 1 to 1 matching between entities of the studying real system and agents living in a simulated environment. The agents and the environment can be viewed as a \textit{virtual micro-world}, that can be perturbed to be studied (with a freedom that we cannot have on studied systems).

![Fig. 1. Key concepts of an agent-based model.](image)

3.1 Key Concepts

\textbf{Agent.} Entities of the studied reference system will be represented by an agent. Following Wooldridge [5], an agent is a hardware or software entity with following properties: autonomy (he can act without the direct control of a human being), social ability (he can interact and communicate with other agents and even keep an image of its social environment), reactivity (he can perceive the environment and react to change the world) and pro-activeness (he can exhibit a goal-driven behavior).

Practically, this characterization of the term \textit{agent} induces that an agent is a \textit{self-contained} [6] entity with an \textit{internal state} (containing all the attributes that characterize an individual) and some \textit{behaviors} that will induce the dynamic of the simulation.
Environment. Agents live, evolve and interact in a simulated environment\(^3\). The environment can have various topologies: it can be continuous (in particular when modeler wants it to be created from GIS data) or discret (e.g. a grid or a network) \(^4\). It provides various services to agents: it gives the possibility to compute neighborhoods (depending on the environment topology). It also allows agents to interact and in particular to communicate with others.

3.2 Agent-based Model Characteristics

The great power of agent-based models compared with other kinds of modeling paradigms is its huge expressive power. An agent is a very versatile object: it can represent any kind of entity of the real system, at any time or space scale, with any kind of formalism. A model can be heterogeneous in terms of the kinds of agents. This paradigm is thus very well-adapted to integrated models (for example of socio-economic and environmental systems) containing heterogeneous entities. For example, a model could contain agents representing farmers, watershed areas, fields or economic market (see Section 4.2 for a description of the MAELIA model). Each kind of agent has his own dynamic (i.e. behaviors) described in its own formalism.

It is important to note that agent-based models are generative models: they produce a result observed at the macroscopic level from lower levels dynamics, while equation-based models are analytical ones: they aim at characterizing equilibria. The modeler has in addition freedom on the level of both observations and simulations. Some indicators can be interesting at a meso-level (some agent aggregates). Similarly, it is sometimes interesting to aggregate some agents and give a dynamic to this aggregate as a whole (we name this kind of model a multi-level model).

3.3 How to Implement Agent-based Models

A lot of tools have been developed last two decades to help modelers to implement their conceptual models. The main interest of these tools is to provide features to write easily models (e.g. built-in primitives to create agents or display the simulation and its results...), to make them as expressive as modelers want them to be (e.g. integration of GIS, 3D, social network, database access or differential equation solvers) or to link simulations with additional external tools (e.g. for the analysis of the results)\(^5\).

We will introduce in the sequel three open-source tools. The two first ones are well-established tools with two opposite approaches. On the one hand, Netlogo \(^9\) is definitely the most used platform. It is dedicated to non-computer scientist: it provides a simple modeling language that can be used by any modeler to write its model. Nevertheless its language structure and its performances limits often its use to simple models

\(^3\) Note that this environment can be considered also as an agent containing all the other agents.

\(^4\) In addition to continuous/discret, Wooldridge gives additional features to environment such as: dynamic/static (the state of the environment changes or not during the simulation), deterministic/non-deterministic (if it is dynamic, the changes are deterministic or stochastic), accessible/inaccessible (can agents have access to all information of the environment) \(^1\).

\(^5\) Interested readers can have a look at \(^8\) for an overview.
or prototypes. On the other hand, Repast (Symphony) [10] provides an environment to develop agent-based simulators mainly in Java (and Groovy), it is thus dedicated to computer scientists that do not need a simple language (and even are much more comfortable with a traditional language) but look for additional powerful and dedicated tools (for example in terms of integration of GIS) and want to have simulations with a huge number of agents.

The GAMA platform [11] is an intermediate solution: it provides a modeling language and a powerful Integrated Development Environment to ease non-computer scientist to develop model with powerful features in terms of GIS integration or high-level decision-making algorithm to integrate into the agents. In addition, both the language and the software have been designed to allow the development of big models with a huge number of agents.

Of course any generic programming language can also be used to implement an agent-based simulator. Despite the powerful existing tools, lot of teams have chosen to implement from scratch their own ad hoc simulator for a dedicated project. The main reason is that the power of existing tools comes often with either some constraints in the manipulated concepts or a heaviness due to the genericity of the tool.

4 Applications

A huge number of models has now been developed in lots of research fields [12]. In the sequel, we illustrate agent-based modeling on two paradigmatic kinds of models: very simple (and often abstract) models to study in depth a single phenomenon (e.g. diffusion-related models [13]) and an example of a very complex integrated socio-environmental model (the MAELIA project [14]).

4.1 Simple Agent-based Models: Examples of Diffusion Models

Opinion Dynamics. First studies on opinion dynamics come from social psychology in order to understand group decision-making process [15]. An interesting phenomenon observed when a group is looking for a consensus is the emergence of extremist opinions whereas it would be expected that the group reaches a mean opinion between individual opinion. First models [16] were based on statistical physics and considered only binary opinions. They have been extended to take into account continuous opinions and conviction level.

First the bounded confidence model [17] uses continuous opinion value and acceptability threshold. When two agents (representing individuals moving in an abstract environment) meet each other they share their opinions. If they are not too far (distance below a threshold), opinions are altered in order to come closer. Depending on parameters (interaction frequency, initial opinion distribution, or even interaction network topology), various kinds of convergence can appear: either convergence to an intermediate consensus or to one or two extremist opinions.

This model has also be extended by introducing a second parameter related to the opinion: a confidence value (that will also represent a persuasion power value) [18]. This is the so-called relative agreement model. An opinion will thus be described by a
value and an incertitude interval around it. The smaller this interval is, the more confident in his opinion the agent is. When two agents meet each others, they will have an influence on each other opinion only if their interval intersects. In this case the opinions will tend to move towards the most confident agent’s opinion. This will also make the confident interval smaller because in some way the opinion has been confirmed by interaction. This model provides good results compared to observed data. In particular in an unconfident population the opinion of extremists will be an attractor for agents’ opinion, whereas they will not have such a big impact on a population without a strong uncertainty. This last model has been used as a basic element of another kind of diffusion phenomenon: the diffusion of innovations.

**Diffusion of Innovations.** Diffusion of innovations is the study of how and why an innovation (a new product for example) spreads in a population. Linked questions are for example: will an innovation percolate? how to improve this diffusion? how to predict the rate of adoption? Rogers [19] has laid the foundation of this research field. He proposed a typology of individuals with relation to their adoption time: Innovators, Early Adopters, Early Majority, Late Majority and Laggards. A successful diffusion of an innovation in a population follows most of the time a S-curve (see for example [20] about the diffusion of hybrid seed corn). This curve is the result of interactions (mainly communication and information sharing) between individuals and is due to their heterogeneity with regard to the innovation adoption.

The most well-known macroscopic model of innovation diffusion is Bass’ one [21]. He attempted to formalized Rogers’ observations by giving equations reproducing the S-curve. He considered that 2 phenomena take part to the diffusion of an innovation: innovation and imitation. He thus considered only two kinds of individuals: **Innovators** and **Imitators** (as an aggregation of Early Adopters, Early Majority, Late Majority and Laggards). First (the small number of) innovators will decide by themselves whether they adopt or not the innovation (which produces the slow begin in the adoption curve). Imitators are sensitive to social influence. Their probability to adopt an innovation will depend on and increase with the number of previous adopters (which produces the exponential part of the curve). The last stable part of the curve appears when only few imitators remain. This model has successfully been applied to several big U.S. companies to reproduce diffusion curve or to predict innovation diffusion. But due to its simplicity, lots of limitations have been highlighted and the model has been extended in particular to take into account missed innovation diffusion factors, such as price, advertisement or international market (interested readers can have a look at [22] for an overview).

In order to improve the descriptive power of the models by taking into account heterogeneity in the population (for example differences of culture) and the network effect that is observed in innovation diffusion, some microscopic models have been proposed and in particular agent-based models [23].

Examples of individual model are threshold model [24]. In this model, agents will adopt an innovation when the number of neighbor agents that have already adopted the innovation is higher than their threshold. By introducing a specific threshold for each agent, we can have an heterogenous population. The threshold value distribution will
have a huge impact on the diffusion of innovations, in particular when it is coupled with
some specific position in the interaction social network.

Previous model has been extended to take into account important adoption factors
linked to communication between agents and individual choice process. The IMAGES
project [25] has produced an interesting agent-based model aiming at studying the dif-
fusion of innovations in the case of new agro-environmental policies. This model is
deeply linked to actual world because it has been built thanks to farmer surveys about
their perception of the innovation consequences and their social network. Each agent
represents an agricultural unit (which is quite similar to a farmer). These agents are
based on the Relative agreement model [18], which is a major benefits comparing to
other innovation diffusion models.

Spread of Diseases. Traditionally, models of disease spread are based on equation-
based models [26] such as the SIR archetype and its extensions SIRS, SEIR and so on.
Such models consider an homogeneous population in which any agent can interact with,
i.e. infect, any other one. This approach provides rigorous analytic results but remains
very simplistic.

The main improvement brought by the agent-based models of disease spread was to
release the strong hypothesis of possible uniform interactions between agents by intro-
ducing space: agents can only infect agents close to them. They also move from place
to place every day or for travel purpose and then can bring with them disease to an
uninfected place (city for example) [27]. Agent-based models also allow modelers to
introduce heterogeneity in the agent population and to explicitly represent social net-
works. Such models can thus represent much more realistic scenarios of disease spread.
Agent-based approach are thus well-adapted to tackle questions about the influence of
public information on the epidemic and its spread [28] or vaccination policy: they allow
to test various vaccination scenario in terms of who, when and in which quantity it is
more efficient to vaccinate.

Among the plenty of agent-based models of disease spread, we can present Episims
[29] and its sequels. Episims is a framework to develop agent-based models of epidemic
spread at various scales: it has been used from the scale of the city to the one of the
state. Agents represent individuals. Their social relationships, the places where they
used to go, their schedules (work, leisure or study) and their transportation network are
also represented. The aim of these models is to study various mitigation policies taking
into account contact social networks emerging from the transportation network and the
various schedules.

Dissemination of Culture. Axelrod proposed an agent-based model of the culture
dissemination in a population [30]. He aims at study the spread of culture from local
(inter-individual) interactions and the emergence of cultural stable areas (and in partic-
ular their number, i.e. the number of culture at the end).

The author represents a culture as a set of features. The model considers a set of
fixed agents with a culture (a set of values for each feature). At each simulation step,
one agent and one of his neighbor are chosen. These two agents will have a probability
of interaction that will be the ratio of similarity between their culture. If agents interact, the chosen agent will adopt the value of a random feature from his neighbor’s culture.

The first interesting result of the model is that an unique culture does not always appear. In addition the author shows that the number of different cultures decreases with the number of features, the interaction range and the growth of the geographic territory size (above a threshold) and increases with the number of traits per features.

**Conclusion.** Boundaries between these various diffusion models are not so clear. For example, epidemic models have been used to represent the diffusion of an information [31] (an information is then similar to a disease, an agent “infects” another agent by communicating him a piece of information). Similarly the threshold model [24] has been applied to various models of diffusion of innovation or dynamic of opinions [18].

In all these cases, the tends to use agent-based models is always driven by the need to introduce heterogeneity into the agent’s population and to use richer environment such as social networks. In addition this approach allows the modeler to access directly to the individual cognitive processes and to tune them and observe the interactions between this inner state (i.e. motivation) and the external social pressure.

4.2 Example of a Complex Model: The MAELIA Project [14]

The MAELIA project aims at developing an agent-based platform for the simulation of socio-environmental impacts of norms designing the management of the water resource. For the purpose of this project, a meta-model of socio-environmental system describing entities (actor and resource) and processes has been developed [32]. In order to evaluate impacts of norms on a territory where renewable natural resources are at the same time submitted to concurring exploitations and dependent on bio-geochemical and physical patterns, the platform will be endowed with a lot of kinds of agents: actors of the model (individual farmers, prefect, water police...) and resources (watershed areas, farms...).

These different kinds of agents are described using various formalisms. For example, the water flow in the watershed areas is described using the SWAT simulator equations [33], implemented into the GAMA platform [11]. A dedicated model of norms have also been implemented and a rule-based architecture developed for the prefect agent [34]. The farmer agents have the most complex cognitive architecture: a multi-criteria decision-making process to plan crop on their fields [35]. Expected extensions of the model includes the introduction of a model of innovation diffusions in the population of farmers, based on their neighbourhood and/or social networks.

5 Perspectives: Current Trends and Issues in ABMS

5.1 Multi-level Aspects

Multi-level models are a long-term research question in ABMS, but it becomes even more important with nowadays simulations including hundreds of thousand or millions of agents. As saw above, an agent is very versatile and can thus represent almost anything, even an aggregate of agents. A multi-level model is thus a model including agents
at various levels, i.e. individual agents and aggregates of agents. Each of these agents can have its own behavior. The point is to link behaviors of individual agents and ones of agents that aggregate them. [36] present one of the most advanced work in the domain. Authors propose an operational meta-model for handling multi-scale models. It is integrated in the GAMA platform [11]. It allows any agent to capture any other agent and to redefine its behavior. Captured agents evolve thus inside the body of the agent (that becomes their environment). This approach also allows to tackle the issue of the multi-environment simulations: an agent can be part of several environments, e.g. it can be member of several social networks or located on several grids. An example of application is when agents are in an environment moving relatively to any one (e.g. passengers in a moving train).

Despite Vo et al.’s work, a lot of related issues are still open, either on modeling aspects (e.g. definition of behavior for aggregate agents depending on individual agents) or in terms of visualisation and interaction of and with these aggregates.

5.2 Articulation between Paradigms: Issues Linked to Interdisciplinarity

As presented above, agent-based paradigm, thank to the versatility of agents, is well-adapted to integrate various heterogeneous models from various research fields, using often various formalisms: agent-based models are the ideal place for cross-fertilization between disciplines and formalisms. Nevertheless, it requiers, in addition to technical issue related to model coupling, new methodologies to guide this coupling of paradigms and points of view. The mete-model proposed in the MAELIA project is a first step on this way.

We can find a good example of articulation between formalisms in epidemiology where the spread can be represented either by an SIR equation system or by using agent-based formalism. We can imagine various ways of interactions between these two paradigms that can be complementary. Equations can directly be used to describe the behavior of some agents [37]: in a city network, each city epidemic state can be modeled using SIR equations. People moving between cities can be represented by agents with an epidemic state that will evolve thank to interactions with other agents they will meet during their trip. So equation-based approach will be well-dedicated to describe dynamic of an aggregation of a huge number of agents (in situation where it is useless or too much time-consuming to represent all individuals). Another way to articulate both paradigms is by using one to describe or improve the other one: e.g. equation-based models to calibrate the agent-based model or agent-based model to infer equation-based model.

5.3 Towards Virtual Laboratories...

As highlighted in the Section 2, simulations are to models what experimentations are to reference systems: a way to study them by perturbing them in a controlled environment. From this remark emerges the idea to build virtual laboratories: they would be computer softwares allowing to implement a model, to run simulations and to analyse the results. An experimental approach for multi-agent simulation should thus be developed.

Agent-based models have the specificity to have a lot of parameters on which the
experiment should play. In addition such models have most of the time a stochastic part. These two reasons induce the need to make a lot of simulations to explore the model. The virtual laboratory should thus allow modelers to automatise the execution of the simulations. This also imposes needs in terms of computational power and distributed computation on clusters or grids to fulfil the study of a model. In addition to purely hardware issues, new exploration algorithms should be developed to exploit better this computation power.

6 Conclusions

This tutorial note has introduced the field of agent-based modeling and simulation illustrated with some examples of diffusion models. These examples showed the versatility allowed by this kind of approach. We showed which benefits this approach has brought to traditional macroscopic models in terms of expressivity. It allows modelers to represent finely individual behaviors and allows them to perturb and experiment these models in some kinds of virtual laboratories.

The cost of this expressivity power is of course the complexity of the models with all the consequences this could have: difficulty to understand the model and how the macroscopic results are produced, to calibrate it, to experiment it deeply and even to communicate it to the community. Most of these issues are now active research fields.

References