Adaptive Automata Community Detection and Clustering – A generic methodology –

Rawan Ghnemat(1), Cyrille Bertelle(1) and Gérard H.E. Duchamp (2) *

Abstract — **We present in this paper a generic methodology based on genetic automata for modelling community detection. With Communities, we deal with dynamic organizations which are self-organized from two aspects, the spatial one and the functional one. We propose in this paper a general methodology which extends cellular based modelling - like Schelling models - to more sophisticated approaches based on agent behavior modelling. These agent behaviors are modelled by automata with multiplicities. These automata-based models allow to define powerful operators like genetic operators and behavioral semi-distance. Mixing these operators, we can propose a complex computing, dealing with spatial self-orgnaization coupled with behavioral similarity for adaptive self-organized systems.**

Keywords: Agent-based modelling, Community Detection, Genetic Automata, Self-Organization, Segregation models.

Submitted to ICCIIS'07

1   Introduction

The current effort on Complexity Theory and its formalism, able to cover a wide area in many aspects of Science, allows today to make relevant links between social, biological and physical systems [5]. The complicated and reductionnist methodologies avoid since many years to light the natural connections between many aspects of the World representation. For example, ecosystems and urban systems cannot be modelled with disconnected theories. Even if the first ones deals with natural processes and the second with social ones, the both are composed with a wide range of interacting entities crossed by energetic fluxes. These fluxes, whatever they are: light, displacment or simply informational fluxes, act on systems in endless adaptive processes. These fluxes compose, rearrange, destroy and recompose dynamic structures or systems. Lighten by social description, we can describe some of these evolutive systems as *Communities*. Studying and modelling communities consists in the description of spatial self-organizations, controled by some global objectives.

This paper presents the new way of modelling communities and social systems based on complex systems approach. We will start our study by cellular automata which are widely used in geographical systems modelling. One of the most current studied communities is the result of segregation Schelling model [14] is the basis of such simulations. Section 2 describes the self-organization concept in social systems which can be considered as the most complex ones. We present computing concepts and tools capable to produce efficient emergent computing to implement such self-organization phenomena. Section 3 describes the segregation-based models as specific spatial self-organization models using cellular automata and grid decomposition where basic elements are fully reactive grid cases controlled by rules-based systems. Our purpose is to deal with more sophisticated behaviors like agent-based ones. So, section 4 proposes an algebraic formalism for agent behavior based on automata with multiplicities, characterized by powerful operators like behavior semi-distance and genetic operators. Section 5 presents a methodology for Community detection. The general processus proposed here, is based on the following properties which deal with the characteristics of these complex dynamic communities: (i) the process is based on an adaptive computation; (ii) self-organized behavioral systems are automatically computed using algebraic properties of the automata formalism; (iii) a spatial self-organization detection finally results as an emergent property of the overall process. Section 6 concludes and gives the perspectives of our work.

2   Modelling Self-Organized Social Systems Using Emergent Computing

The complexity modelling for social systems theory focuses nowadays on self-organization processes [7]. The initial self-organization description generally comes from physical or chemical, and it means of self-structuration of mater. The social systems can be considered as the most complex ones because we must be aware of reminding

---
*(1) LITIS - Le Havre University, 25 rue Philippe Lebon, 76058 Le Havre Cedex, France, rawan.ghnemat@gmail.com, cyrille.bertelle@gmail.com; (2) LIPN - Paris XIII University, 99 avenue Jean-Baptiste Clément, 93430 Villetaneuse, France, gheduchamp@gmail.com

that these systems include human behaviors which are active and self-conscious subjects, and so generate some self-creative consequences. The social systems theory deals with dialectic approaches between systems and individual human behaviors, under the fluxes of economy and politic: systems are the results of human interactions and these systems influence individual thinkings and actions on itself.

Today, Computer Sciences both on its conceptual improvment for modelization and on its technological development, allow to give experimental tools for social systems theory [6]. The emergent computing gives nowadays relevant simulation tools for complexity study and specifically social systems. It can be mainly classified in two categories:

- Cellular automata which generally deals with spatial self-organizations; Cellular automata [**?**] are tools for study some diffusion models of spatial organizations. They are based on grid case description as rules-based systems. We present this kind of models in section 3.

- Agent-based modeling which generally deals with adaptive complex behavior. We present this kind of models in section 4.

3 Segregation-Based Models using Cellular Automata

We can observe that humans are more or less sorted into groups of similar people, according sometimes with racial and ethnic features but also in many others ways according to various characteristics. Thomas Schelling proposed in the 1970's a very simple rule-based model able to describe segregation phenomena. His description can easily be adapted to cellular automata models and it generates from random population, a spatial self-organization which is observed as patterns of clusters of similar persons. The cluster formation does not follow linear dependance from any parameter dealing with the evolution of the rule-based systems. More generally, Schelling model can be used as generic spatial self-organization concept for various kinds of systems: A. Singh and M. Haahr, for example, use a variation of this model to study the topology adaptation in P2P networks [16].

The original Schelling's model can be extended in concurrent version in order to obtain an exact definition of the implementation for each time step, not depending of the grid way [9].

In the following, we will make use of the (somewhat extended) notion of derangement. For the classical notion see [17]:

http://mathworld.wolfram.com/Derangement.html
http://en.wikipedia.org/wiki/Derangement

Here it will be called *generalized derangement*.

Definition 3.1 *Let* X, Y *be two sets* $X \subset Y$*. A generalized derangement from X to Y is an into mapping* $\alpha : X \mapsto Y$ *such that*

$$(\forall x \in X)(\alpha(x) \neq x) \tag{1}$$

To describe Thomas Schelling's concurrent model, we start with a cellular automata board which is a rectangle of $n \times m$ (n lines and m columns) points (each point will be located by its coordinates (x, y) with $1 \leq x \leq m$; $1 \leq y \leq n$). A *state* of the board will be simply a mapping $s : [1..m] \times [1..n] \mapsto \{0, A, B\}^1$ indicating whether a point at (x, y) has a value corresponding to

$$\begin{cases} \bullet \ \text{nothing } s(x, y) = 0 \\ \bullet \ \text{an element of type } A, \ s(x, y) = A \\ \bullet \ \text{an element of type } B, \ s(x, y) = B \end{cases} \tag{2}$$

the dynamics of the system will be described by a sequence of states $s_0, \ s_1, \cdots, \ s_n, \cdots$ generated by the following rules.

- one fixes a threshold (in percent) $0 \leq t \leq 1$

- the original state is s_0 (a distribution of A, B and empty cells along the board)

- at each step, for each (filled) cell of type X at (x, y), one counts the ratio $r(x, y)$ of neighbours of type X over the number of neighbours

- if $r \geq t$ the cell is marked s (stay), if not it is marked m (move)

- let M be the set of cells marked "move" and E the set of empty cells

- choose randomly (uniform distribution) α among the generalized derangements $M \mapsto M \cup E$

- then $s_{n+1}(x, y) = s_n(x, y)$ if the cell was marked s and $s_{n+1}(x, y) = \alpha(s_n(x, y))$ otherwise.

This algorithm is typically controled by grid diffusion processus using elementary rule-based systems like most of the cellular automata models do. The rules leading to each move is based on local computation of neighbourhood, the moves are global, holistic self-organizational. The following section proposes more sophisticated behavior modelled by agent systems.

---

[1] We can extend the model using greater alphabet cardinal

4 Agent-Based System Modelling

In this section, we give the basis of the conceptual tools that allow to extend the reactive and diffusive grid cases behavior to more sophisticated entities, using agent-based model. We propose to model the agent behavior with automata with multiplicities which are powerful algebraic stuctures.

4.1 Basic agent-based concepts and complex systems modelling

According to General System Theory [13], a complex system is composed of entities in mutual interaction and interacting with the outside environment. A system has some characteristic properties which confer its structural aspects, as schematically described in part (a) of Figure 1:

- The set elements or entities are in interactive dependance. The alteration of only one entity or one interaction reverberates on the whole system.

- A global organization emerges from interacting constitutive elements. This organization can be identified and carries its own autonomous behavior while it is in relation and dependance with its environment. The emergent organization possesses new properties that its own constitutive entities don't have.

- The global organization retro-acts over its constitutive components.

The interacting entities network as described in part (b) of Figure 1 leads each entity to perceive informations or actions from other entities or from the whole system and to act itself.

A well-adapted modeling consists of using an agent-based representation which is composed of the entity called agent as an entity which perceives and acts on an environment, using an autonomous behaviour as described in part (c) of Figure 1.

To compute a simulation composed of such entities, we need to describe the behaviour of each agent. This one can be schematically described using internal states and transition processes between these states, as described in part (d) of Figure 1. So an automata with multiplicities as described in the following section is well-adapted for the agent behavior modelling.

4.2 Automata-based modelling for agent behavior

An automaton with multiplicities is based on the fact that the output data of the automata with output belongs to a specific algebraic structure, a semiring, including real, complex, probabilistic, non commutative semantic outputs (transducers) [11, 17]. In that way, we will be able to build effective operations on such automata, using the power of the algebraic structures of the output data. We are also able to describe automaton by means of a matrix representation with all the power of the new (i.e. with semirings) linear algebra.

Definition 4.1 (Automaton with multiplicities)
An automaton with multiplicities over an alphabet A and a semiring K is the 5-uple (A, Q, I, T, F) where

- $Q = \{S_1, S_2 \cdots S_n\}$ *is the finite set of state;*

- $I : Q \mapsto K$ *is a function over the set of states, which associates to each initial state a value of K, called entry cost, and to non- initial state a zero value ;*

- $F : Q \mapsto K$ *is a function over the set states, which associates to each final state a value of K, called final cost, and to non-final state a zero value;*

- T *is the transition function, that is $T : Q \times A \times Q \mapsto K$ which to a state S_i, a letter a and a state S_j associates a value z of K (the cost of the transition) if it exist a transition labelled with a from the state S_i to the state S_j and and zero otherwise.*

Remark 4.2 *We have not yet, on purpose, defined what a semiring is. Roughly it is the least structure which allows the matrix "calculus" with unit (one can think of a ring without the "minus" operation). The previous automata with multiplicities can be, equivalently, expressed by a matrix representation which is a triplet*

- $\lambda \in K^{1 \times Q}$ *which is a row-vector which coefficients are $\lambda_i = I(S_i)$,*

- $\gamma \in K^{Q \times 1}$ *is a column-vector which coefficients are $\gamma_i = F(S_i)$,*

- $\mu : A^* \mapsto K^{Q \times Q}$ *is a morphism of monoids (indeed $K^{Q \times Q}$ is endowed with the product of matrices) such that the coefficient on the q_ith row and q_jth column of $\mu(a)$ is $T(q_i, a, q_j)$*

Definition 4.3 (Automata-Based Agent Behavior)
We represent the agent behavior by automata with multiplicities (A, Q, I, T, F) over a semiring K:

Figure 1: Multi-scale complex system description: from global to individual models

- *The agent behavior is composed of a states set Q and of rule-based transitions between them. These transitions are represented by T; I and F represent the initial and final transitions;*

- *Alphabet A corresponds to the agent perceptions set;*

- *The semiring K is the set of agent actions, eventually associated to a probabilistic value which is the action realization probability (as defined in [8]).*

4.3   Agent Behavior Metric Space

The main advantage of automata-based agent modelling is their efficient operators. We deal is this paragraph on a innovative way to define behavioral semi-distance as the essential key of self-organization processus proposed later.

Definition 4.4 (Evaluation function for automata-based behavior)
Let x an agent whom behavior is defined by $\mathcal{A}$, an automaton with multiplicities over the semiring K, we define the evaluation function $e(x)$ by:

$$e(x) = V(\mathcal{A})$$

where $V(\mathcal{A})$ stands for the vector of all coefficients of (λ, μ, γ), the linear representation of $\mathcal{A}$, defined in remark 4.2.

Definition 4.5 (Behavioral semi-distance)
Let x and y two agents and $e(x)$ and $e(y)$ their respective evaluations as described in the previous definition 4.4. We define $d(x,y)$ a semi-distance or pseudometrics [2] between the two agents x and y as

$$d(x,y) = ||e(x) - e(y)||$$

a vector norm of the difference of their evaluations.

5   Community Detection using Genetic Spatial Automata Population

In this section, we give an operational definition of community in terms of functional concept dealing with complex and social modelling. To model such communities, we have to complete the concept of automata behavior with some spatial aspects and with some adaptive capabilities that genetic operators can allow to implement. With these concepts, we can model communities by evolutive population of these genetic spatial automata.

Definition 5.1 (Community operational definition)
A community is a system or an organization which is characterized by a spatial property, a behavior property and the interaction between the both.

Example 5.2 *In ecology, a community is a group of plants or animals living in a specific region and interacting with one another.*

The spatial patterns generated by Schelling models are some examples of communities and these spatial patterns are linked with the very specific behavioral rules implemented for each grid case. Our purpose here is to give a more generic processus which can be mixed with sophisticated agent behaviors. Using agent communication protocol, we can extend the diffusion process linked with cellular automata to more distant communications and interaction between spatial agents. In the following, we define this notion of spatial automata-based agents and then we develop the genetic operators allowing to transform automata with multiplicities to genetic automata. We will explain how the definition of adapted fitness will generate the detection processus.

Definition 5.3 (Spatial Automata-Based Agent)
A spatial automata-based agent is defined by its structural representation:

- *An automata with multiplicities corresponding to its behavior as a whole processus managing its perceptions and its actions over its environment. They include its communication capabilities and so its social behavior;*

- *A spatial location defined on some specific metric space.*

5.1   Genetic operators on automata population

We consider in the following, a population of automata with multiplicities which are each represented by a chromosome, following the genetic algorithm principles. We define the chromosome for each automata with multiplicities as the sequence of all the matrices associated to each letter from the (linearly ordered) alphabet. The chromosomes are composed with alleles which are here the lines of the matrix [3].

---

[2]see [4] ch IX

In the following, genetic algorithms are going to generate new automata containing possibly new transitions from the ones included in the initial automata.

The genetic algorithm over the population of automata with multiplicities follows a reproduction iteration broken up in three steps [12]:

- *Duplication*: where each automaton generates a clone of itself;

- *Crossing-over*: concerns a couple of automata. Over this couple, we consider a sequence of lines of each matrix for all. For each of these matrices, a permutation on the lines of the chosen sequence is made between the analogue matrices of this couple of automata;

- *Mutation*: where a line of each matrix is randomly chosen and a sequence of new values is given for this line.

Finally the whole genetic algorithm scheduling for a full process of reproduction over all the population of agents is the evolutionary algorithm:

1. For all couple of agents, two children are created by duplication, crossover and mutation mechanisms over their behavioral automata. The location of the children can be choosen from many ways: on the linear segment defined by the parents location or as the node of a square described by them and their parents (more details are given in [10];

2. The fitness for each automaton is computed;

3. For all 4-uple composed of parents and children, the performless agents, in term of fitness computed in previous step, are suppressed. The two agents, still living, result from the evolution of the two initial parents.

Remark 5.4 *The fitness is not defined at this level of abstract formulation, but it is defined corresponding to the context for which the automaton is a model, as we will do in the next section.*

5.2 Adaptive processus to implement community detection

The community detection is based on a genetic algorithm over a population of spatial automata-based agents. The formation of the community is the result of the population evolution crossing by a selection process computed with the fitness function defined in the following.

For this computation, we deal with two distances defined on agent set. The first is the spatial distance associated to the agent spatial location and the second is the behavioral semi-distance defined in the definition 4.5.

Definition 5.5 Community clustering and detection fitness
Let $\mathcal{V}_x$ a neighbourhood of the agent x, relatively to its spatial location. We define $f(x)$ the agent fitness of the agent x as :

$$f(x) = \begin{cases} \dfrac{card(\mathcal{V}_x)}{\displaystyle\sum_{y_i \in \mathcal{V}_x} d(x,y_i)^2} & if \displaystyle\sum_{y_i \in \mathcal{V}_x} d(x,y_i)^2 \neq 0 \\ \infty & otherwise \end{cases}$$

where $d(x,y)$ is the behavioral semi-distance between the two agents x and y.

The genetic evolution of the spatial automata-based agents leads to a self-organization which creates a clustering of the agents set in such way that each cluster contains agents of similar behavior. During the evaluation process, genetic algorithms can be turned such that individuals *outside communities* be attracted to them. The center of the clusters, the size of the clusters and the behavior of the agents in the center of each cluster are the result of the overall genetic processus which generates self-organization communities.

6 Conclusion and Perspectives

The goal of this paper is to present a methodology for communities detection based on genetic automata population. This methodology is generic and can be adapted to many kinds of applications. We presently, implement this methodology on Geographical Information Systems (GIS) using RePast and Agent Analyst tool [18]. Spatial automata-based agent model associated to GIS, allows more powerful tools both for spatial and behavior description than cellular automata do and in that way, we can extend the reactive segregation-based modelling to urban area sharing under the influence of both citizen behavior but also under the influence of more elaborated decision making. These applicative experiments are in progress.

References

[1] Benenson, I., Torrens, P.M. (2004) *Geosimulation - Automata-based modeling of urban phenomena*, Wiley.

[2] Berstel, J., Reutenauer, G. (1988) *Rational series and their languages*, EATCS.

[3] Bertelle, C., Flouret, M., Jay, V., Olivier, D., Ponty, J.-L. (2001) *Genetic algorithms on automata with multiplicities for adaptive agent behaviour in emergent organizations* In *SCI'2001*, Orlando, Florida, USA, 22-25th July 2001.

[4] Bourbaki, N. (1998) *Elements of Mathematics: General Topology*, Chapters 5-10, Springer-Verlag Telos.

[5] Capra, F. (2002) *Hidden Connections. Integrating the biological, cognitive and social dimensions of life into a science of sustainability*, Doubleday ed.

[6] Edmonds, B., Hales, D. (2005) *Computational Simulation as Theoretical Experiments*, Journal of Mathematical Sociology, Vol. 29(3):209-232.

[7] Fuchs, C. (2003) *Concepts of Social Self-Organization*, Research paper INTAS Project "Human Strategies in Complexity".

[8] Ghnemat, R., Oqeili, S., Bertelle, C., Duchamp, G.H.E. (2006) *Automata-Based Adaptive Behavior for Economic Modelling Using Game Theory*, in Aziz-Alaoui, M.A., Bertelle, C., *Emergent Properties in Natural and Artificial Dynamical Systems*, Springer.

[9] Ghnemat, R., Duchamp, G.H.E., Bertelle, C. (2007) *An exact resolution of Schelling's segregation model using generalized derangements*, (submitted).

[10] Ghnemat, R., Bertelle, C., Duchamp, G.H.E. (2007) *Self-organized communities modelling by genetic spatial automata*, (submitted to the workshop EP-NACS'07 within ECCS'2007, Dresden, Germany).

[11] Golan, J.S. (1999) *Power algebras over semirings*, Kluwer Academic Publishers.

[12] Goldberg, D.E. (1989) *Genetic Algorithms*, Addison-Wesley.

[13] Le Moigne, J.-L. (1999) *La modélisation des systèmes complexes*, Dunod.

[14] Schelling, T.C. (1971) *Dynamic Models of Segregation*, Journal of Mathematical Sociology, Vol. 1:143-186.

[15] Schutzenberger, M.P. (1961) *On the definition of a family of automata*, Information and Control, Vol. 4:245-270.

[16] Singh, A., Haahr, M. (2004) *Topology Adaptation in P2P Networks Using Schelling's Model*, workshop on Emergent Behavior and Distributed Computing, PPSN.

[17] Stanley, R.P. (1999) *Enumerative combinatorics*, Cambridge University Press.

[18] Strout, N. (2006) *What is Agent Analyst*, The 2006 ESRI International User Conference, San Diego.

[19] Weiss, G. (ed.) (1999) *Multiagent Systems*, MIT Press.